

Characterization, Evaluation and Management of Prospective Benefits, Costs and Risks in the Development of New Statistical Programs for Energy October 2009

John L. Eltinge and Polly Phipps

Office of Survey Methods Research, U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Ste. 1950 Washington, DC 20212

Abstract

Statistical agencies often need to begin new programs, or to undertake substantial changes in current programs, due to changes in: (a) the stakeholder needs for statistical information; (b) the resource base directly available to the statistical agency; or (c) the overall environment in which collection, analysis and dissemination of data takes place. This paper reviews some factors that may be important in the management of the resulting changes, and in the development and implementation of related methodology. These factors include primary stakeholders, their information needs, and translation of these needs into specific inferential goals; potential sources of applicable data, including both surveys and administrative records; feasible methods for collection and analysis of these data; investments required for the sound development and implementation of these methods; and evaluation of related prospective benefits, costs and risks arising from methodological choices.

Key words: Adoption and diffusion of innovations; Constrained optimization; Human capital and institutional capital; Incentives; Operational risk; Public goods; Satisficing; Stakeholder utility functions; Survey data quality; Survey design; Systemic risk; Total survey error.

1. Introduction

We thank the session organizers for the opportunity to speak here today. The first two speakers in the session have covered the topic of energy statistics in a considerable amount of depth. Consequently, we will focus our presentation primarily on some features of the process of innovation within statistical agencies, and illustrate some of those features with examples from energy statistics.

1.1. Energy Statistics

The Energy Information Administration (EIA) is the primary source for energy data in the United States. In existence since 1977, EIA is an independent statistical agency within the U.S. Department of Energy. EIA's data collection and dissemination efforts cover a wide range of topics, including energy production, reserves, consumption, distribution, imports, exports, and prices; and encompass numerous energy sources. EIA is responsible for data on traditional energy sources, including coal, petroleum, natural gas, electricity, and nuclear energy. These "core" data collection efforts require constant attention, care, and modification to keep up with changes in industries, markets, public policies, and new technologies. In addition, EIA is on the forefront of data development and collection for newer energy sources, including renewable and alternatives fuels, such as solar, wind, geothermal, biomass, and ethanol. New data development and collection

efforts require high levels of investment, including extensive design, testing, and analysis. The EIA also conducts energy analysis and forecasting; to do this the agency maintains energy databases and models, which require extensive development and continuous updating. To carry out its mission, EIA is funded at approximately 110.6 million dollars (in 2009); staffing includes 382 full-time equivalent positions, and 375 employees, of which 73 are statisticians.

1.2. Changes in Statistical Programs

Large-scale statistical programs generally involve a substantial amount of infrastructure, and substantial changes in statistical programs generally involve significant costs and risks. Consequently, statistical agencies usually change their production programs only when there are fairly strong reasons to do so. These reasons often involve one of three types of change. First, in some cases there are fundamental changes in the level of stakeholder interest in specific types of statistical information. For example, when the spot price for oil peaked at over \$140 per barrel last summer, stakeholders naturally had strong interest in the features of alternative fuels markets, e.g., ethanol and biodiesel. In the same time period, some stakeholders expressed substantial interest in “peak oil” concepts (e.g., Simmons, 2008).

Second, a statistical agency may need to change its programs due to substantial changes in its available resource base. This resource base naturally includes the level of funding for the agency, but may also include other components that are relatively hard to quantify. For instance, statistical programs tend to be rather capital intensive, with most of the capital in intangible forms, e.g., skills of available personnel or the methodology applicable to a given program. These forms of intangible capital may change over time due to, e.g., the cumulative effects of investment in personnel training or methodological research. In addition, the resource base for a statistical agency generally includes legal, regulatory, or market factors that affect its access to data. For instance, legal or regulatory changes may affect the structure, timeliness, accuracy and accessibility of some types of administrative-record data used in statistical programs for energy and economics.

Third, statistical programs may change in response to changes in the overall environment in which the agency carries out its work. For example, changes in the structure of energy markets over the past 30 years have led to changes both in the underlying substantive questions that are important for policymakers and researchers, and in the realistic options available for data collection. The dissemination of data is another notable example of change at the EIA and other statistical agencies over the past 30 years. In 1979, most data dissemination necessarily took place through tables of numbers printed on paper, while in 2009 most dissemination takes place through agency websites and other electronic forms.

1.3. Outline of Primary Ideas

In response to identified needs for change in statistical programs, a statistical organization will encounter a wide range of practical questions, including the following.

- (A) Identification of primary stakeholder needs for statistical information, and the impact that program changes may have on stakeholder utility functions.
- (B) Development and implementation of specific modifications in statistical programs.
- (C) Evaluation, and possible optimization, of the impact that modifications in (B) will have on the stakeholder utility functions identified in (A).

To address these issues, the remainder of this paper is organized as follows. Section 2 outlines an overall framework for evaluation of data quality and the corresponding utility of statistical data for stakeholders. Section 3 reviews some elements of the general literature on adoption and diffusion of innovations, and suggests some ways in which that this literature may be useful for the management of innovation processes within statistical organizations. Section 4 discusses areas in which statistical agencies and other stakeholders may especially benefit from sharing of information. Section 5 explores further some areas in which other statistical agencies may especially benefit from the experiences of the EIA.

2. Data Quality and Changes in Statistical Programs

2.1. Multiple Dimensions of Data Quality

To explore the issues raised in Section 1, consider the following schematic representation for the quality of data produced by a statistical program. We emphasize that all models and figures presented in this paper are intended primarily to develop and explore general trade-offs among benefits, costs, and risks related to statistical programs and their stakeholders. In some cases, one could consider formal development of methods to collect objective data applicable to these models; to estimate the corresponding parameters; and to evaluate model goodness-of-fit. In other cases, these models should be interpreted only in a broad qualitative context.

To develop some notation, define a set of published statistical products $\hat{\theta}_k$, $k = 1, \dots, K$ (e.g., specified population means, totals, ratios, indexes, or other estimates) based on an underlying population Y . For product k , the program has identified J distinct quality measures Q_{jk} that are considered to be of practical importance. For example, Q_{jk} could be a standard summary of one of the six dimensions of data quality outlined in Brackstone (1999, 2001): accuracy, timeliness, relevance, accessibility, interpretability and coherence. For example, within the overall quality measure “accuracy,” one may be interested in the distinct components associated with a standard total survey error model as considered in Andersen et al. (1979), Groves (1989) and Weisberg (2005). These may include the effects of superpopulation models, frames, sampling, nonresponse, and measurement error.

In addition, suppose that the quality measures Q may be associated with three classes of predictor or control variables:

- X_1 : Variables that are directly under the control of the survey designer, e.g., sample size, initial selection probabilities or the collection instrument.
- X_2 : Variables that are observable during data collection, but not necessarily subject to direct control, e.g., interviewer turnover rates.
- X_3 : Other variables that may be important for Q , but are neither directly controllable nor observable in real time during data collection. Examples of variables X_3 would include some undetected failures in programming of the collection instrument, and changes in the underlying survey environment.

The program management may then consider the working model

$$Q = f(Y, X, \beta) + e \quad (2.1)$$

where $f(\cdot)$ is a function of known form; β is a vector of unknown parameters, e.g., from a superpopulation model or variance-function model; and e is a general disturbance term.

In a general sense, much standard methodological work can be viewed as attempt to optimize expression (2.1) conditional on certain constraints, e.g., a data-collection budget or limitations on respondent burden. A simple example is Neyman allocation (Cochran, 1977, Chapter 5), i.e., allocation of a fixed total sample size across strata in order to minimize the variance of a weighted mean estimator under stratified random sampling.

2.2. Linkage Between Standard Measures of Data Quality and Perceived Value for Stakeholders

Use of a standard data-quality model (2.1) is complicated by the fact that most statistical programs use a standard set of procedures to produce a large number of estimates, and these estimates are of interest to a wide range of stakeholders. In addition, the utility of a given set of estimates may vary substantially across stakeholder groups. To explore these issues, consider L distinct sets of primary stakeholders, and define an $L \times K$ dimensional matrix Z representing the utility of the K distinct published estimates as viewed by the L stakeholder groups. In addition, define the schematic model

$$Z = g(Q, E, \gamma) + d \quad (2.2)$$

where d represents a general disturbance term; E represents a set of general environmental factors (e.g., the changing degree of interest in biofuels); and γ is a vector of model parameters reflecting underlying perceived information needs of the stakeholders.

Figure 1 provides a schematic representation of three possible relationships between the methodological measures of utility Q (presented on the horizontal axis) and the stakeholder perceptions of utility Z (presented on the vertical axis). To simplify the representations, both J and L are restricted to equal 1, and both measures of utility are scaled to range from 0 to 1. For purposes of discussion, we focus attention on the case represented by the vertical red line, with methodological utility scaled to equal 0.6. The dashed line (with an intercept of 0 and a slope of 1) corresponds to the case in which stakeholder utility is coherent with methodological utility. For this case, improvements in standard measures of methodological quality lead to a corresponding improvement in utility for some primary stakeholders.

The solid curve at the top of Figure 1 represents a case in which methodological changes may be of relatively little interest to the stakeholder group, and may in fact be considered problematic. Specifically, increases in methodological utility above 0.6 lead to relatively small changes in perceived stakeholder utility, while decreases in methodological utility may lead to substantial decreases in perceived stakeholder utility. For this case, the stakeholder may be justifiably reluctant to consider methodological changes, while a methodologist – with a notably different utility function – may view that reluctance as unreasonable risk aversion.

Conversely, for the curve formed by open circles methodological utility scaled to equal 0.6 corresponds to a relatively low level of perceived stakeholder utility. Moderate increases in methodological utility above 0.6 lead to substantial increases in perceived stakeholder utility, while corresponding decreases in methodological utility lead to relatively small decreases in stakeholder utility. For such cases, the primary stakeholder may consider aggressive exploration of methodological alternatives to be warranted. Such cases may arise, e.g., when contracts provide substantial incentives related to certain performance thresholds like precision or response rates.

Review of Figure 1 leads to three general ideas. First, it is important for statistical agencies to have a solid understanding of the general utility functions of their primary stakeholders. The EIA invests substantial time and energy in this area; see, e.g., Energy Information Agency (2008, 2009a, 2009b), Gruensprecht (2008), Ellerman (2008) and Cooper (2008). Section 3 below explores utility-function ideas in additional detail. Second, it is beneficial to focus a statistical program on production of published estimates that meet the information needs of stakeholders with a relatively cohesive set of utility functions. Conversely, distinct sets of stakeholders with very different utility functions (e.g., differing needs regarding timeliness, accuracy or degree of publication detail) may lead to establishment of distinct statistical programs, even if those programs are nominally covering the same substantive topics. Third, investments in methodological improvements generally should be focused on cases in which at least some primary stakeholder groups have utility functions roughly matching the middle or bottom curves in Figure 1.

3. Social Processes Related to the Adoption and Diffusion of Innovations

3.1. General Literature on Innovations

In considering issues related to stakeholders, their perceived utility functions, and changes in available products and processes, it is useful to explore the general literature on “adoption and diffusion of innovations.” This literature has developed since the turn of the previous century, and includes a large number of publications in the past two decades, primarily in application areas related to consumer electronics and other forms of consumer goods. For some general discussion of this field, see Rodgers (1995), Katz et al. (2004), von Hippel (1988, 2005) and references cited therein. One should use a considerable level of caution in review of this literature. For example, much of this literature focuses on case studies, and the generalizability of specific results may be uncertain. However, it would be of value to consider three of the primary themes from this literature, and potential applicability of those themes to exploration of stakeholder needs in statistical programs.

First, in the exploration of stakeholder needs, the literature emphasizes that the distribution of stakeholder needs may involve one or more of the following.

- (i) Direct response to the perceived added utility of an additional or “improved” product. For statistical programs, this could involve, e.g., publication of more detailed estimates, improvement of the quality of currently published estimates, or improved dissemination of current estimates.
- (ii) The degree of standardization or customization of the “product” designed to provide (i). One EIA example of this arose at the 2009 EIA Annual Conference. Industry analysts indicated they would like more standardized and better data to evaluate energy policies,

practices, programs, and goals across states (state data are variable in quality); these evaluations would likely contribute to regional and national-level discussions. An example provided was the lack of standardized data for electrical billing. However, there are likely different stakeholder expectations about (i), (iii) and (iv). For example, some companies may be less interested in standardization, since it is hard to develop a standard metric that fits all cases. Also, states could be less interested, as states are different, state utility commissions are important stakeholders, and their interests are likely to be different. Analysts also note that it takes federal initiatives to produce comparable state data.

- (iii) Aggregate resources available to invest to obtain the products defined in (i) and (ii). These resources may include money; scarce skill sets that are in high demand for other projects; and perceived control, institutional credibility and other intangibles. In some cases, these resources may be very time-dependent, and future availability of these resources may be unpredictable.
- (iv) Risk profiles related to the investments in (iii). The risk factors of interest may include termination of a development project; delays in timelines for development, testing and implementation of the production system; perceived subrogation in one or more dimensions of data quality; and failure to communicate accurately the prospective benefits, costs and risks inherent in the proposed statistical program.

Second, the literature often notes a distribution of behaviors of internal and external stakeholders, e.g., the customary partition into innovators, early adopters, early majority, late majority and others. Note that within the context defined by Section 2 and items (i)-(iv) above, each of these behaviors may be entirely rational. As part of development and implementation of a new (or revised) statistical program, statistical agencies can enhance their effectiveness through development of a solid consensus on identification of: the primary stakeholders; their expectations; and their likely responses to prospective innovations in statistical programs.

Third, the literature also directs a large amount of attention to design of processes for development, adoption and diffusion of innovations. Four critical factors are as follows.

- a. Development of the innovation (methodological, technological or managerial) itself. This development almost always requires multiple iterations; a substantial resource base for the initial development; and requires a nuanced understanding of time lags, feedback loops, and evaluation of prospective benefits, costs and risks. Also, in keeping with the three curves in Figure 1, stakeholders often have different views of the incremental benefit (through improvement of prospective benefits, costs and risks) of an additional piece of information (e.g., one more experiment). Dillman (1996) explores this issue within the context of improvements in survey methodology.
- b. This skills and information base of stakeholders at different points of the development/adoption/diffusion distribution.

- c. Distinctions between specific innovations, and an anticipated stream of innovations (some incremental, some potentially disruptive).
- d. Governance of the process for development, adoption and diffusion of innovations. This governance process includes:
 - Allocation of resources for each step in the process of development, adoption, diffusion and maintenance, and expectations regarding the pace and hurdle rates for these steps.
 - Evaluation of the performance of the development/adoption/diffusion/maintenance process; and initiation, intermediate-stage, and termination decisions at each stage of the process.
 - Internal and external processes for governance, e.g., regulatory (formal OMB or industry standards) market-driven, formal internal competition, implicit internal decisions embedded in larger internal processes for resource allocation.
 - Incentive structures for all participants in the governance process as such, and in the full process of development, adoption and diffusion.

3.2 Special Issues for Government Statistical Agencies

As noted above, much of the literature on adoption and diffusion of innovations has focused on examples from consumer products and related marketing areas. For government statistical agencies, there are several additional complicating factors. For example, for most statistical agencies, estimates are published without direct charge to many or all of the stakeholders in question, and thus are treated as “public goods.” Consequently, the funding streams and institutional incentive structures are somewhat different for statistical agencies relative to the “producers” considered in much of the standard innovation literature. For some general background on “public goods” phenomena, see Wilson (1989) and references cited therein.

In addition, the statistical profession places a very high value on transparency in the development, presentation and evaluation of methodology. Thus, in many cases, much or all of a given methodological development is essentially placed in the public domain and thus also becomes a “public good” fairly early in the process of development, adoption and diffusion. This in turn has an additional effect on institutional and individual incentive structures.

Also, in comparison with consumer-product companies, governmental statistical agencies often operate with a more pronounced set of constraints, including requirements to meet a wider range of stakeholder needs; to comply with general regulatory requirements regarding personnel, contracting and the use of technology; and to match expenditures with constraints imposed through the Congressional appropriations process. For this reason, decision processes within statistical agencies often resemble minimax or satisficing processes (Simon, 1979, 1982) rather than the relatively simple optimization exercise often envisioned in standard methodological work.

On a related note, some of the literature in Section 3.1 places strong emphasis on specific internal management structures, and related incentives, for management of the process of development, adoption and implementation of innovations. For example, this literature often highlights the importance of an “internal champion” for a given set of prospective innovations. This internal champion often has strong positive incentives, and authority, to ensure that innovations with good prospects for success receive robust support in funding and other critical resources. Conversely, the same internal champion

has broad authority to terminate development projects that do not show sufficient progress. In the abovementioned literature, individual and institutional incentives often are tied closely to standard profit motives, and to the prospect for use of a portion of the profits in subsequent investment in additional innovation projects. Standard budgetary structures for government agencies generally do not have the same mechanisms for identification and re-investment of profits.

Finally, in standard consumer-product areas, product performance is directly observable by a wide range of stakeholders; the quality of reception for a cell telephone would be a simple example. In contrast with this, for many statistical products, there are limited direct feedback loops that link methodological improvements (e.g., reduction of nonresponse bias or reduction in sampling error variance) with perceived stakeholder needs. Instead, many stakeholders must obtain their information on statistical quality through third-party evaluations, e.g., reviews by objective expert panels like the ASA Committee on Energy Statistics.

In summary, government statistical agencies work in an environment that is not identical to the private-sector environment commonly studied in the literature on adoption and diffusion of innovations. Thus, in future work it will be of interest to study in additional detail the extent to which standard insights from this literature will apply to government statistical agencies. It will also be important to explore specific practical ways in which those standard approaches may require adaptation to account for features of the governmental environment.

4. Prospective Benefits and Costs of Communication and Collaboration Across Statistical Organizations

As noted above, the amount of common ground shared by statistical agencies is substantial, but is by no means comprehensive. Consequently, it is useful to identify some areas in which interagency communication and collaboration may be especially valuable.

First, there are several different types of interagency work. Some of that work centers on standards and other forms of regulation. Examples include the FCSM subcommittee on race-ethnic classification; periodic interagency committees on metropolitan area classification and industrial classification; the FCSM/OMB standards for performance of government statistical programs (revision of Statistical Directives 1 and 2, issued in final form, September, 2006), and standards for communication of information on sample design, instruments and paradata. Second, there are several forms of general information sharing. Examples include seminars; proceedings from conferences like the FCSM Research Conferences, the Joint Statistical Meetings, the series of International Conferences on Establishment Surveys; and edited books, e.g., on telephone surveys, measurement error, and nonresponse. A third category of information sharing involves consultation with, and formal review by, stakeholders and technical experts. Examples include reviews by the Committee on National Statistics (CNSTAT); the ASA Committee on Energy Statistics; external program reviews; the Federal Economic Statistics Advisory Committee (FESAC); the Census Bureau Advisory Committees and studies carried out under the ASA/NSF/agency fellow program. Fourth, there are specific focused efforts in the shared development and use of methodology, e.g., Q-Bank and the FCSM subcommittees on nonresponse bias, administrative records.

Prospective benefits of interagency collaboration include efficiency; increased use of common technical language and shared understanding of the primary methodological and operational issues; and combination of resources that may be unique or complementary. One example of the latter would be the use of BLS and Census data as input for some BEA national-account estimates. However, there are also prospective

costs and risks associated with interagency work. For example, such work can become excessive, especially if the central issues involve divergent stakeholder utility functions that cannot reasonably be reconciled. Thus, it would be interest to explore the degree to which one could try to incorporate these factors into cost-benefit models for the development and implementation of statistical methodology within government agencies.

5. What Other Statistical Organizations Can Learn from Successes at the EIA

5.1 Review of Longstanding Substantive Issues in Energy Statistics

During the 30 plus years of EIA history, the agency has focused its efforts on many issues that are central to its responsibilities as an independent statistical agency and its mission to collect, evaluate, assemble, analyze, and disseminate energy data and information. Kent (2003) provides an outstanding examination of longstanding issues through his content analysis of topics that EIA brought to the Committee on Energy Statistics of the American Statistical Association between 1979 and 2001. The role of this federal advisory committee has been to review and provide advice to EIA on energy data collection and analysis, and technological and methodological issues.

Kent identifies a number of substantive issues that have been fairly constant over time and continue to this day. First, EIA's important consumption surveys, including the Residential Energy Consumption Survey (RECS) Commercial Building Energy Consumption Survey (CBECS), and the Manufacturing Energy Consumption Survey (MECS) are integral to the long standing issue of the measurement of demand. Specific concerns include survey scope and content, sampling frames and precision, collection methodologies, and timeliness (including length between collection cycles). Second, supply side measurement of all energy sources is another long-standing issue, with a particular focus on data series consistency, data quality issues (nonresponse, missing data), and estimation (including annual, monthly, weekly consistency). Third, forecasting and modeling energy markets is a longstanding substantive issue for EIA, including the thorny issues of how to balance policy analyses with other data needs, data quality, and the content and adequacy of models, including the National Energy Modeling System (NEMS), the Regional Short-Term Energy Model (RSTEM), and the System for Analysis of Global Energy (SAGE).

5.2. The ASA Committee on Energy Statistics

Demand, supply, and forecasting issues were all agenda topics between 1998 to 2003, when Phipps was a member of the ASA Committee on Energy Statistics. The specific issues under study by EIA staff covered all dimensions of data quality. Several areas of EIA data quality work and research stand out as best practice examples and we describe them below.

5.2.1 Integrating Behavioral Science Methods into EIA Survey Design and Data Quality Studies

In the mid-1980s, statistical agencies began adopting behavioral science methods and tools to explore survey data quality, primarily with household surveys. While EIA was not the first statistical agency to utilize behavioral science methods, the agency was probably the most successful in adapting these methods to fit agency needs. EIA developed a set of tools and a process to evaluate establishment surveys; the process focused on data quality issues such as accuracy and timeliness. The EIA approach was developed over time, discussed at numerous energy statistics committee meetings, and shared at professional associations meetings (e.g., Freedman and Rutchik, 2002,

Goldenberg et al. 2002). It included the use of focus groups, subject matter expert review, and the mapping of survey concepts and industry players; these tools were used to assess design issues and data requirements. At the actual survey design stage, pre-survey visits to companies were used to evaluate data availability and sensitivity, and mapping was used to display relevant findings and processes. Pre-survey visits were also used as a tool to explore modifications to on-going surveys. Finally, cognitive testing with employers was used to test the survey instruments. EIA's use of behavioral science methods was all the more successful in that it appears to have become institutionalized, i.e., used extensively and improved over time to this day.

5.2.2 Electronic Data Dissemination

As early as 1984, Energy Statistics Committee documents show that EIA began to focus on the role of information technology in the data dissemination process. In the early 1990s, EIA recognized the power of the internet for data dissemination activities, and later in the decade made a decision to put a lower priority on the preservation of traditional publication formats and hard copy reports and maximize the information provided on the web (Kent, 2003). After the establishment of the EIA website in 1995, EIA initiated research on web-based issues, with a focus on the data quality dimension of accessibility. This included customer satisfaction surveys beginning in 1994, and the adoption of the behavioral science tool of cognitive/usability testing to test website design in 1999. In addition, EIA recognized the importance of graphics in data dissemination, initiating an internal graphics competition in 1994 judged by committee members. And in 2000, EIA began cognitive testing of EIA graphics as part of web design and development. As with establishment survey design, EIA studies on dissemination and web-site design were discussed many times with the committee and shared with other agencies at professional association meetings.

Again, similar to behavioral science research on establishment survey design and development, research on website content and design is institutionalized and continues on an ongoing basis. In 2006, EIA moved from a decentralized to a centralized corporate approach to web development and operations, transferring program office web content and development into their National Energy Information Center (Pearson, 2009). Continuing a tradition of research-based development, the current NEIC focus is on research-based web design and development which encourages staff to concentrate on groups of users as the center of the development process: congressional staff / policy analysts, public citizens, journalists, energy producers, commercial energy consumers, and a commercial software application known as the data hound robot which automatically collects data on energy supply and inventory data (http://tonto.eia.doe.gov/abouteia/eia_explained.cfm). In addition to these new user personas, and the customer satisfaction surveys, cognitive interviews, and usability testing, EIA's web design research methods have expanded to include: card sort testing and analysis to inform information architecture choices, formal audience analysis, and web metrics and search log analysis (Pearson, 2009).

5.2.3. Other EIA Successes

EIA has successfully addressed many other survey issues. As discussed by Kent (2003), EIA spent considerable time on survey definitions, reporting to the energy statistics committee and seeking input. In the late 1990s the agency tackled the issue of coherence with a project to develop common data definitions across the agency programs that would be clear to a broad range of data users. The complex project involved integrating and rewriting survey definitions and related terms (e.g., finished motor gasoline, gasoline grades, gasohol), and collecting and integrating comments across EIA.

5.3 New Challenges: Substantive Issues in Energy Statistics That Are Relatively New, or Have Recently Encountered Increased Interest

As with other statistical agencies, EIA grapples with changing public policy and changing market and industry structures. These changes pose difficulties for longstanding substantive issues and data collections. They also generate new substantive issues that pose new challenges for statistical agencies. Currently, global energy issues and public policy debates and concerns have put forth multiple issues that EIA is exploring or may encounter in the near future; we provide several examples below. These issues have many methodological dimensions.

Collecting energy consumption data is a critical component of EIA's mission, and in the current energy climate, even more emphasis is likely to be placed on consumption data and data quality. At the 2008 EIA conference, the consumption surveys were a main topic on the agenda, including tracking change in end use, the use of consumption data to analyze policies, energy efficiency indicators, and tracking global energy consumption. In 2009, energy consumption data were discussed again at the EIA conference in a session on data needs. As discussed by Kent (2003), the residential, commercial, and manufacturing surveys have presented numerous data quality challenges in the past. The challenges persist, particularly in the area of accuracy and timeliness. Stakeholders would like greater sample detail and precision, new and refined questions included in survey instruments, and a smaller lag time between survey cycles.

Renewable energy and emerging technologies and industries, a major focus of the 2009 EIA conference, is high on the energy agenda for policymakers and analysts. Many agencies face new demands for new products, but perhaps none more than EIA at the current time. Emerging technologies are likely to require new data collections or changes in existing collections, with substantial lead time for survey development and design. Since industries and companies are emerging, sampling frames may be lacking and small sample size is an issue. New methods to identify and monitor emerging trends seem in order for statistical agencies, as decisions to undertake new data collections are particularly hard, given agencies must grapple with the risk of investing in new products with unpredictable shelf lives.

Market sensitive data is another issue that EIA is facing; with continued energy market and sector volatility, this issue is probably unlikely to disappear from EIA's agenda. Data accuracy and timeliness are again major data quality concerns, including outliers, revisions, and confidentiality. Other agencies such as BLS have grappled with market sensitivity with some regularity; cross-agency dialogue may provide a useful forum for this shared problem.

6. Discussion

In summary, this paper suggests that methodologists can contribute to innovations in statistical agencies in at least two areas. The first area is the traditional area of work by methodologists in evaluation and reduction of one or more components of total survey error. In that traditional area, it will be useful for methodologists to expand standard approaches to consider more systematic evaluation of the homogeneity of predictors and goodness-of-fit across multiple surveys. Also, issues of operational risk and systemic components of risk warrant further study.

Second, methodologists can contribute through a broader framework for assessment of prospective benefits, costs and risks associated with changes in statistical programs. These include components of data quality beyond accuracy; coherence of

stakeholder utility functions with standard methodological criteria; and systematic empirical assessment of fixed and variable components of survey cost structures.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the Bureau of Labor Statistics. The authors thank Tony Barkume, Edward Blair, John Bosley, Stan Freedman, Bill Mockovak and Bill Weinig for helpful discussions of some of the issues covered in this paper; and thank Randall Powers for preparing Figures 1 and 2.

References

- Andersen, R., J. Kasper and M.R. Frankel (1979). *Total Survey Error*. San Francisco: Jossey-Bass.
- Brackstone, Gordon (1999). Managing data quality in a statistical agency. *Survey Methodology* **25**, 139-149.
- Brackstone, Gordon (2001). How important in accuracy? Proceedings of Statistics Canada Symposium 2001 – Achieving Data Quality in a Statistical Agency: A Methodological Perspective.
- Brewer, K.R.W. (1995). Ken Foreman Memorial Lecture, Australian Bureau of Statistics.
- Cochran, W.G. (1977). *Sampling Techniques, Third Edition*. New York: Wiley.
- Cooper, K.B. (2008). EIA's Challenges in Achieving Its Mission: EIA Study Team Reflections and Recommendations. Presentation to the Energy Information Administration 2008 Energy Conference, April 7, 2008.
- Dillman, D.A. (1996). Why innovation is difficult in government surveys (with discussion). *Journal of Official Statistics* **12**, 113-197.
- Ellerman, A.D. (2008). EIA's External Study Team: Introduction and Evaluation. Presentation to the Energy Information Administration 2008 Energy Conference, April 7, 2008.
- Energy Information Agency (2008). EIA – Energy Conferences and Presentations: http://www.eia.doe.gov/eia_conference_2008.html
- Energy Information Agency (2009a). The EIA website for the American Statistical Association Committee on Energy Statistics: http://www.eia.doe.gov/calendar/asa_overview.htm
- Energy Information Agency (2009b). State energy data needs assessment, January, 2009.
- Freedman, Stanley and Robert Rutchik (2002). Information collection challenges in electric power and natural gas. *Proceedings of the Annual Meeting of the American Statistical Association*, pp 52-61.
- Goldenberg, Karen, Amy E. Anderson, Diane K. Willimack, Stanley R. Freedman, Robert H. Rutchik, Luann M. Moy (2002). Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies. <http://www.jpsm.umd.edu/qdet/qdet-set.html>
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R. M. and Heeringa, S. (2006) Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 169, 439-457
- Gruensprecht, H. (2008). EIA's Mission and Budget: Priority Setting in the Short Run. Presentation to the Energy Information Administration 2008 Energy Conference, April 7, 2008.

- Katz, R., ed. (2004). *The Human Side of Managing Technological Innovation, Second Edition*. New York: Oxford University Press.
- Kent, Calvin A. (2003). The Role of a Federal Advisory Committee: A Case Study of the American Statistical Association (ASA) Committee on Energy Statistics. Paper presented to the ASA Committee on Energy Statistics, April 3, 2003.
- National Research Council (2009). *Principles and Practices for a Federal Statistical Agency, Fourth Edition*. Committee on National Statistics. Constance F. Citro, Margaret E. Martin and Miron L. Straf, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Pearce, Gina (2009). Developing and Implementing Web Management and Governance for the Energy Information Administration. United Nations Statistical Commission and Economic Commission for Europe, Warsaw, Poland.
- Rogers, E.M. (1995). *Diffusion of Innovations, Fifth Edition*. New York: Free Press.
- Simmons, M.R. (2008). The Peak Oil Debate as the EIA Turns 30. Presentation to the EIA 2008 Energy Conference, April 7, 2008.
- Simon, H. (1979). *Models of Thought, Vols. 1 and 2*. New Haven, Connecticut: Yale University Press.
- Simon, H. (1982). *Models of Bounded Rationality, Vols. 1 and 2*. Cambridge, Massachusetts: MIT Press.
- Von Hippel, E. (1988). *The Sources of Innovation*. New York: Oxford University Press.
- Von Hippel, E. (2005). *Democratizing Innovation*. Cambridge, Massachusetts: MIT Press.
- Weisberg, H.F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.
- Wilson, J.Q. (1989). *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

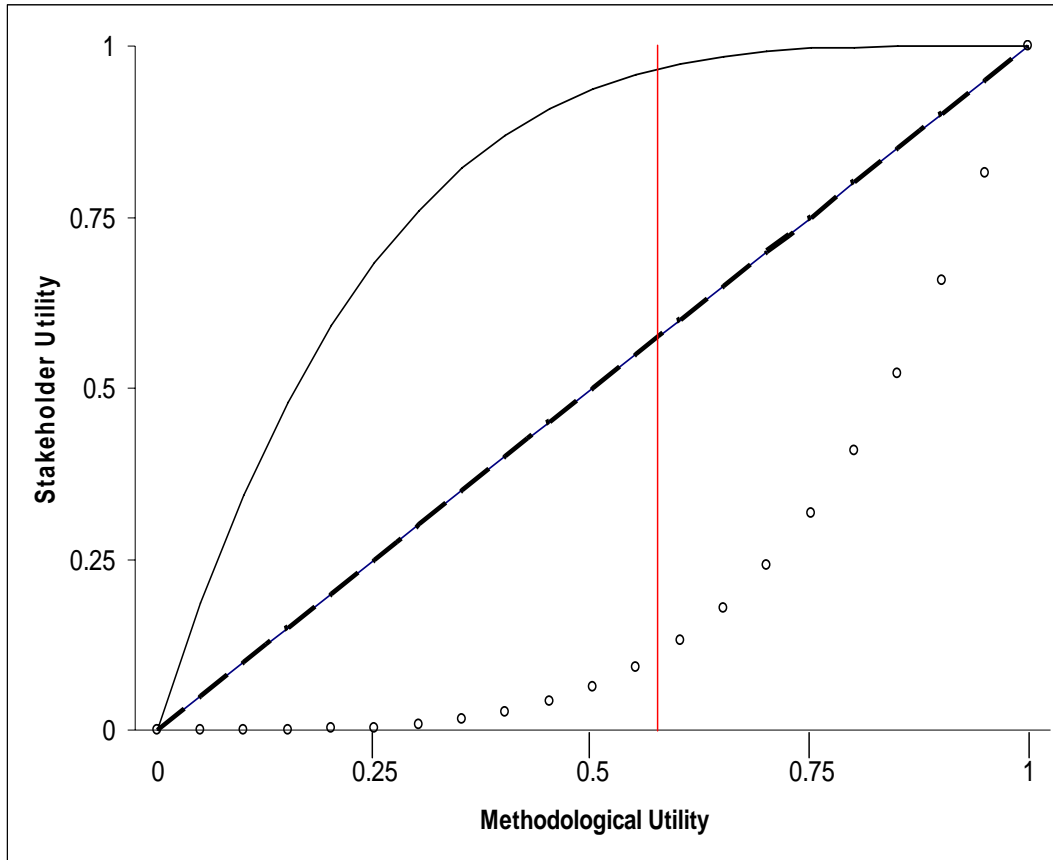


Figure 1: Schematic Representation of Three Possible Relationships Between Methodological Utility (Horizontal Axis) and Stakeholder Utility (Vertical Axis).

For purposes of discussion, attention centers on the case represented by the vertical red line, with methodological utility scaled to equal 0.6.

Dashed line: Stakeholder utility is coherent with methodological utility. Improvements in standard measures of methodological quality lead to a corresponding improvement in utility for some primary stakeholders.

Solid curve: Risk aversion may appear to be justified when the scaled methodological utility equals 0.6. Increases in methodological utility above 0.6 lead to relatively small changes in perceived stakeholder utility, while decreases in methodological utility may lead to substantial decreases in perceived stakeholder utility.

Curve formed by open circles: Aggressive exploration of methodological alternatives may be warranted. For this case, methodological utility scaled to equal 0.6 corresponds to a relatively low level of perceived stakeholder utility. Moderate increases in methodological utility above 0.6 lead to substantial increases in perceived stakeholder utility, while corresponding decreases in methodological utility lead to relatively small decreases in stakeholder utility.

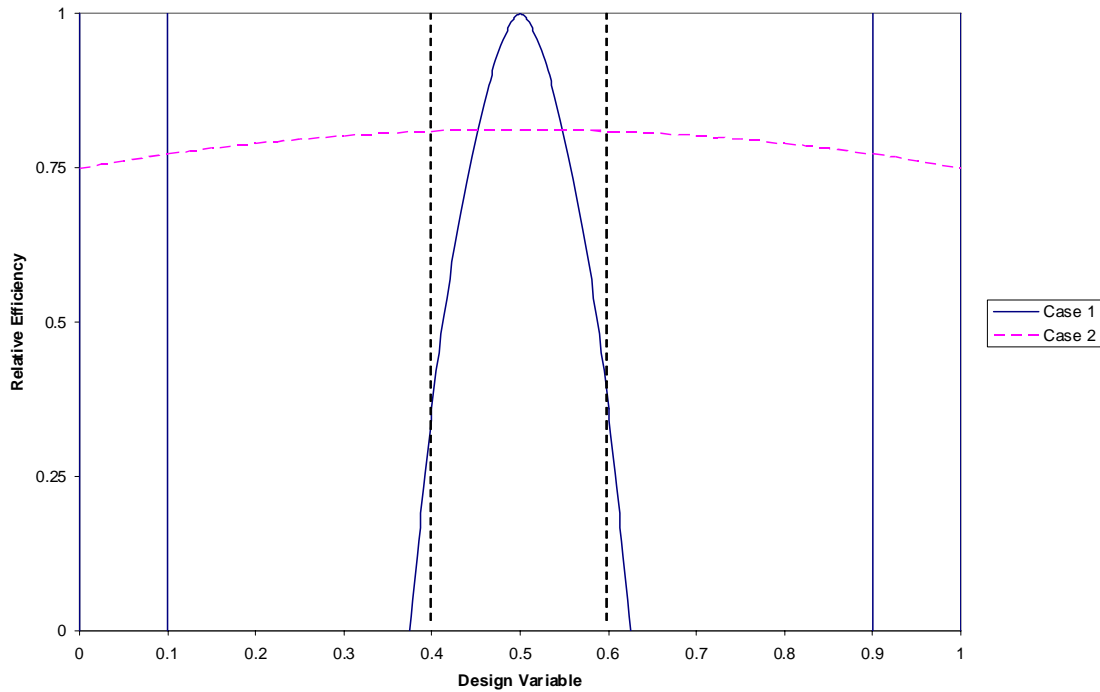


Figure 2: Schematic Representation of the Relationship Between a Design Variable (Horizontal Axis) and the Relative Efficiency of a Statistical Procedure (Vertical Axis).

For purposes of illustration, the design space (generally high dimensional) is represented as one-dimensional. For both case 1 and case 2, the design variable is scaled to have its optimal value equal to 0.5.

Case 1: A procedure that is fully efficient if the procedure can be consistently executed at the optimal design point of 0.5, but is fragile in the sense that moderate deviations from the optimal design point lead to large losses in efficiency.

Case 2: A procedure that is not fully efficient, relative to the idealized case 1, but is relatively robust against moderate deviations from the optimal design point of 0.5.