A Review of Statistical Research at the U.S. Bureau of Labor Statistics

Cathryn S. Dippo[1]

Abstract: Research projects currently being conducted by the statistical research units at the U.S. Bureau of Labor Statistics are discussed. Major efforts are under way to control, measure, and reduce nonsampling errors associated with questionnaires and data collection procedures. Theoretical research with supporting empirical studies is being conducted on current and alternative estimators of level, change, and variance.

Key words: Sample survey; questionnaire design; computer assisted interviewing; Laspeyres index; variance estimation; time series.

1. Introduction

The U.S. Bureau of Labor Statistics (BLS), established in 1884, is responsible for the analysis and publication of data series on employment and unemployment, prices and living conditions, wages and industrial relations, productivity and technology, occupational safety and health, and economic growth and employment projections. BLS is just one of the many agencies within the U.S. government which has a statistical function. For example the National Center for Health Statistics, the Energy Information Administration, and the National Agricultural Statistical Service are responsible for collecting, analyzing and disseminating information about the health status of the population, energy, and agricultural crops and livestock,

respectively. The Bureau of the Census is responsible for conducting the population, economic, and agricultural censuses, along with some demographic and economic surveys.

Each agency may maintain its own data collection operation or contract with another (or a private organization) for data collection. For example, among the BLS data series are some based on establishment surveys conducted by BLS or under a cooperative program with state governments, and others based on household surveys conducted by BLS or for BLS by the Bureau of the Census. In addition, BLS conducts wage and price surveys for other agencies, such as the Office of Personnel Management and the Department of Defense.

BLS is actively conducting both economic and statistical research. This review will only discuss statistical research conducted within the four statistical units within BLS. Statistical Methods Divisions exist in each of the three

[1] Senior Mathematical Statistician, Office of Mathematical Statistics, Bureau of Labor Statistics, 441 G Street, N.W., Washington, D.C. 20212, USA.

major program offices: Prices and Living Conditions, Employment and Unemployment Statistics, and Wages and Industrial Relations. The fourth is the Office of Mathematical Statistics, located in the Office of Research and Evaluation.

The Statistical Methods Divisions are responsible for planning, developing, managing, and evaluating all statistical aspects of the programs within their respective offices. The Office of Mathematical Statistics is responsible for planning and directing a continuing statistical research and evaluation program for improving the quality of current statistical output and for the long-term development of Bureau programs. Although it does not have any direct program responsibilities, it provides review and consulting services to the Statistical Methods Divisions. Statistical research is conducted by all of these groups, but the emphasis differs.

In this paper, I have grouped the research projects currently being conducted by these statistical units into three sections:

2. Questionnaire Design,

3. Data Collection Methodology, and

4. Sample Design and Estimation.

In selecting the research projects to be discussed, I have tried to highlight those for which a reference has been published in a journal or the proceedings of the American Statistical Association. Further information can be obtained from the cited authors at the Bureau of Labor Statistics, 441 G Street, N.W., Washington, D.C. 20212, U.S.A.

2. Questionnaire Design

Questionnaire research has been focused on response error in the Consumer Expenditure Survey (CE); however, an overall research plan has been developed for examining and revising the Current Population Survey (CPS) questionnaire. A Questionnaire Design Advisory Conference, sponsored by BLS in

January 1987, focused on obtaining comments on our current and planned research and on developing ideas for further or additional research with respect to the CE and CPS programs (see Dippo, et al. (1987)).

The continuing CE program, which began in 1980, consists of two surveys. In the Interview Survey 5 000 respondents are interviewed once a quarter for five consecutive quarters. The data collected in the first quarter is used for bounding only. In the independent Diary Survey, 5 000 respondents per year are asked to keep a diary of all purchases for two consecutive weeks.

Four projects designed to investigate response error in CE data have been recently undertaken by BLS staff. Since one of the major uses of the CE data is in the development of the commodity weights for the Consumer Price Index, there is particular concern over differential reporting by expenditure category, along with a general concern for the quality of the data for economic research.

The first study provided a preliminary evaluation of the quality of data from the Diary Survey. Almost one fourth of the diaries examined were completed through either partial or total recall rather than strictly through daily diary entries. The percentage of diaries mentioning a particular item at least once was, in many cases, significantly lower than that observed during the 1972/73 Diary. Items listed as examples in the questionnaire were more likely to be reported than those not listed. Utilizing these and other results, the second study, a field experiment testing two alternate diary formats, was designed and carried out in 1985. These data are now in the final processing stage, and the results of the study will be presented at the 1988 annual meeting of the American Statistical Association.

The third study used data from a supplemental survey administered to diary respondents and interviewers in the second quarter

of 1984. This supplement was designed to examine the relationship between a respondent's diary keeping attitudes and his or her actual behavior. Tucker (1985, 1986) found that although in many instances attitudes and behavior coincided, a significant portion of the respondents expressed attitudes which were at odds with their behavior. A typology of respondent styles was developed from attitude and behavior measures constructed from the items in the supplemental survey. The relationships of the typology and the characteristics of the respondents in each category of the typology with a latent response error measure created from internal response pattern indicators were examined. Several relatively inexpensive procedural changes were recommended as a result of this analysis and work continues in this area with the development of a set of ordered data quality categories which combine a measure of response error with one measuring the effect of nonresponse (Tucker (1987)).

In the fourth study, data from the Interview Survey were used to examine conditioning and recall effects which are considered major types of response error in panel surveys. Using a set of households who responded for all four usable interviews, mean expenditures were developed by commodity category, interview wave, and recall month. Silberstein and Jacobs (1986) found only moderate time-in-sample effects with less than half of the categories exhibiting significant differences in means between the second and fifth interviews. On the other hand, significant recall effects were observed in all but two expenditure categories for which month of expenditure is collected. Means for the month furthest in time from the interview, i.e., three months earlier, were 15 to 39 percent lower than means for the closest month. It is doubtful that internal telescoping or erroneous reporting of expenditure month would account for these differences. Research continued in this area

(Silberstein (1987)) to identify those respondent and household characteristics associated with certain reporting patterns. Detailed analysis was conducted on two major expenditure classes where substantial differences were found by recall month.

With respect to the CPS, a joint BLS and Bureau of the Census questionnaire design task force has developed a plan for research to investigate: question wordings which are ambiguous, operational definitions of key labor force concepts which are inconsistent or vague, interviewer control over labor force classifications, the properties of measurement classification procedures, and questions which involve complex tasks of recall and estimation. The plan includes laboratory research to develop a better understanding of the measurement process and to develop alternative question wordings and measurement procedures, small scale field tests of alternative questionnaires and procedures, a major field test of one or two revised questionnaires and procedures, and a period of parallel collection of data using current and new questionnaires and procedures to provide measures of the effect on the published estimates. Some of the methodological issues proposed for testing are: extension of the current one week recall period to two weeks, interaction between the interviewer and respondent, use of prior responses to reduce the erroneous reporting of change, definition of an acceptable respondent, and coverage procedures for listing persons within households.

3. Data Collection Methodology

The primary purpose of research on the use of modern technology for data collection is to control measurement error while at the same time increase productivity. Currently, research

is being conducted on the use of computer assisted telephone interviewing (CATI), computer assisted personal interviewing (CAPI), touch-tone data entry by respondents, and random digit dialing (RDD).

CATI is being investigated for use in the Current Employment Survey (CES), the Current Population Survey (CPS), and the Consumer Price Index (CPI) housing survey. Each of these surveys presents different problems in the development of CATI.

Currently, the CES uses a form which is mailed from and returned to one of the 51 State Employment Security Agencies responsible for collecting data from approximately 280 000 establishments each month.

The CPS has a rotating panel design of 60 000 housing units per month with a sample unit being interviewed for four consecutive months, out of sample for eight months, and then interviewed again for another four consecutive months. Currently, the field staff of the Bureau of the Census, which is the data collection agent for the CPS, collects most of the second through fourth and sixth through eighth interviews by telephone from their homes. The CPI housing survey consists of two interpenetrating samples: one of 20 000 owner housing units and another of 40 000 renter housing units. The owner units are split into 12 panels and interviewed annually; the renter units are split into six panels and interviewed semiannually. Vacant units are not excluded from the survey as in the CPS and most other housing unit surveys. Data for approximately one-half of the sample units are collected by telephone each month by the BLS data collection staff.

Because the CES did not collect any data by telephone, CATI research began in 1984 with a small two state feasibility test of telephone data collection (Werking (1984)). In 1985, two states were used to test the conversion of current CES mail reports to direct monthly CATI collection. The University of California at Berkeley micro-based CATI software was used. While the results from this CATI test were highly successful (see Werking et al. (1986)), the direct application of CATI to the full CES sample in 51 State agencies was not deemed economically feasible. This led to two paths of collection methodology research in 1986: CATI production tests and touch-tone telephone self-response tests. The CATI production tests were aimed at determining the most cost effective application of CATI in the CES. The touch-tone self- response tests were aimed at determining the feasibility of having the respondent directly key enter the survey data using a touch-tone phone. During 1988, a major effort will be devoted to improving the basic CATI system, including the expansion to a local area network environment and linking the touch-tone data entry collection as a subfunction of CATI.

CATI research for CPS began in 1985 with a feasibility operations test. A test sample of households were first interviewed in person and then by telephone from a central CATI facility operated by the Bureau of the Census for three succeeding months. The second phase of testing with approximately 1 500 cases per month in a sample of metropolitan areas began in late 1986. Since the major objective of this testing phase is to examine the effects of CATI on key estimates, the CATI system emulates to the extent possible the type of CATI system which would be used in the CPS on a production basis. Preliminary information indicates that important differences may exist between the current collection system which is a combination of personal visits and *decentralized* telephone data collection and the test collection system which is a combination of personal visits and *centralized* telephone data collection. Further analysis of the data is being conducted to examine the possible causes of the observed differences. A pilot test of the use of CAPI for CPS is tentatively scheduled for mid 1989.

The initial CATI test for CPI housing concentrated on evaluating the impact of CATI interviewing on response inconsistency. Results (Kosary and Sommers (1987)) indicate that CATI has the potential to increase data quality as measured by gross discrepancy rates between item responses obtained during an interview and later during a reinterview, simple response variances, and stability of survey estimates. Of special interest is the indication that CATI results in lower refusal rates for sensitive items, such as income. Due to its enforced consistency of procedures and question wording, CATI is an aid to obtaining more consistent results. Further CATI testing of the CPI housing schedule will take place in the fall of 1987. At the same time, research is beginning on developing structured questionnaires for the commodities and services portion of the CPI, so that CAPI and CATI testing can begin.

In 1988 BLS will sponsor a CATI feasibility test for the Continuing Point of Purchase Survey. Currently, sample households in this survey are asked the names and addresses of business establishments where they purchase approximately 140 categories of goods and services during a specified reference period which varies by category, e.g., one week for meat and poultry or six months for men's shirts. The outlets reported by the sample households become the frame for the sampling of outlets to be used in the Consumer Price Index. The point of purchase information is currently collected in a personal interview that lasts approximately 90 minutes. In the test, RDD and list telephone samples will be asked information on either 20 or 40 categories. The purpose of the study is to examine the effects of questionnaire length on outlet reporting, while at the same time testing the feasibility of obtaining names and addresses of outlets using telephone interviewing techniques.

4. Sample Design and Estimation

Since the majority of the BLS's major programs are ongoing, major sample design changes usually occur during major revisions of the entire survey program. For example, the CPI and CPS are revised approximately every ten years. The allocation of sample items and outlets for the latest revision of the Consumer Price Index, effective in January 1987, resulted from the development of a system of models relating design variables to data collection costs and the variance of average price change. Nonlinear programming techniques were used to solve for design variables that minimized the total variance of average price change, subject to data collection cost constraints (Leaver et al. (1987)).

The questionnaire design and CATI/CAPI research for CPS discussed earlier are just two components of an overall research plan developed by BLS with the Bureau of the Census for the next revision of CPS after the 1990 Decennial Census. Other components include: testing the feasibility of a two-sample design to produce both national and state estimates monthly, developing an overall cost-error model for sample allocation, and exploring alternatives for extending longitudinal analysis capabilities.

Similarly, sample design research for the CES is being conducted as a part of an overall modernization which includes the incorporation of new data collection methodologies and estimation techniques described elsewhere in this paper. One aspect of the modernization is the development of timely birth frames for sampling and estimation of new business employment. Tupek and Grzesiak (1986) studied the assignment of new unemployment insurance accounts to businesses to determine whether the new accounts actually represent new businesses and the length of time from

when a new business first hires employees until it is captured in the Unemployment Insurance system. A survey to determine characteristics of new businesses and the feasibility and accuracy of estimating birth employment using both mail and CATI began in 1986. The percentage of new accounts which are really new businesses has been estimated from the survey as well as the initial employment of the new businesses. The collection of this information considerably reduces the lagtime for introducing birth employment into CES estimates. Future research is planned to determine the feasibility of national implementation and to evaluate usage of the birth frame in sample solicitation.

In the area of wages and industrial relations, BLS has been requested by the U.S. Congress to develop an establishment survey to measure white collar wages. Current plans call for this survey to be built around several existing wage and benefit surveys, thus requiring both sample design and estimation research.

In conjunction with the testing of RDD for the Continuing Point of Purchase Survey mentioned in the previous section, BLS will be investigating a dual frame sample design referred to as a "list assisted" telephone interviewing design as suggested by the Survey Research Center at the University of Michigan. In addition to the households contacted via RDD, a sample of telephone numbers selected from a telephone directory or commercial list are sent a letter of introduction prior to the actual telephone interview. This research builds on the dual frame work of Sirken and Casady (1987).

Estimation research at BLS is currently concentrated in five areas: estimation of employment level and change together with their corresponding variance estimators for the Current Employment Survey, modified Laspeyres index estimators, measures of central tendency for censored earnings data,

variance estimators for complex surveys, and model-based seasonal adjustment.

The current estimator for CES is a stratified link relative estimator. Alternative estimators currently under consideration include modified link relative, separate link relative, and regression estimators. West (1983) presented results from a theoretical investigation of regression and ratio type estimators of the population total and month to month change supported by an empirical study using a real population under different sampling plans, response rates, and estimation procedures. A modified link relative estimator tended to do better than the others under most conditions. Grzesiak and Copeland (1987) have extended this research to longer time periods (24 months) and more industries, while comparing theoretical and survey results. An investigation (West (1984)) of variance estimators for the leading estimators of level indicated that the robust type estimators suggested by prediction theory were better than the jackknife, bootstrap, and Bayesian estimators studied.

The statistical research on indices is concentrated on statistical properties of estimators of fixed base Laspeyres indices as opposed to research on whether a different form of index such as chain Laspeyres or one of the superlative indices would more appropriately measure desired economic concepts. BLS conducts three price index programs – the CPI, the Producer Price Index (PPI), and the International Price Program (IPP). Each of the price index programs is based on probability samples of establishments and items. In concept the indices estimated by each program are fixed base Laspeyres although for a variety of practical reasons, mainly related to continually changing universes of items, the indexes depart from that concept. The estimators of long-term price change from the base period to the current period used in the CPI, PPI, and IPP are calculated by multi-

plying together successive estimators of one-period price changes. The one-period changes from one time to the next are estimated based on sample items which are priced in both periods. Estimators of short-term change, e.g., one-month or 12-month, are calculated as ratios of appropriate long-term changes. Alternative estimators are being studied which employ only the current period's sample items or which differentially weight the one-period changes in the estimator now being used for publication. Initial results from this research are reported in Valliant (1987a) with more detailed findings in Valliant (1987b). Follow-up work on alternative estimators is now being conducted.

The adjustment of index numbers to account for particular biases has been an area of interest within BLS. One example is a proposed adjustment of the CPI housing index to account for aging bias using linear regression. The effect of measurement error in the auxiliary variables used to make adjustments has been a concern. Miller (1986) has been involved in ongoing research into estimating linear relationships when errors exist in the auxiliary variables.

Alternative measures of central tendency for censored earnings data have been investigated by West (1987). Currently, BLS publishes median usual weekly-earnings of full-time wage and salary workers by age, race, and sex based upon a subset of the CPS monthly sample. Because of the problems of comparing medians and functions of medians over time, West (1986) developed a method for computing means using the Pareto distribution as a model for the tail of the earnings distribution with a modified maximum likelihood estimator for its parameter. This method resulted in a fairly robust estimator of the overall mean that has good statistical properties.

The problem of variance estimation from complex surveys has been examined by Dippo and Wolter (1984) and Valliant (1987c, 1987d, 1987e). Dippo and Wolter studied the stratified sampling properties of Taylor series, random group, and balanced half-sample variance estimators as applied to estimators which are functions of stratum means. We evaluated theoretical approximations to the variance estimators as applied to ratio-mean estimators, i.e., means estimated as the ratio of two estimated totals, using Consumer Expenditure Survey data. Relatively small reported sampling biases were found, but confidence interval coverage properties were generally poor in the simulation study we conducted. Follow-up empirical work on the same variance estimators plus the jackknife is now being conducted with additional expenditure data.

Theory and empirical work on generalized variance functions (GVFs) has been done by Valliant (1987c). Using a model-based approach, he gave some theoretical justification for certain types of GVFs in two-stage cluster sampling. He also reported the results of a simulation study using CPS data which compared several alternative forms of GVFs in terms of bias, precision, confidence interval coverage, and confidence interval length. He has also studied model-based, conditional properties of stratified ratio and regression estimators and accompanying variance estimators (Valliant (1987d, 1987e)). Valliant's papers illustrate some of the considerable differences that may exist in conditional and unconditional inferences drawn from stratified samples. In two earlier papers Valliant (1985, 1986) provided some superpopulation theory and empirical results on finite population estimation under a general nonlinear model with particular application to estimation of totals of binary variables.

Model-based seasonal adjustment is being tested on employment and unemployment series from the CPS and CES using signal extraction from seasonal ARIMA models.

Outlier treatment is incorporated into the modeling to prevent distortions in seasonal adjustment due to unusual, nonrepetitive behavior in the series. This contrasts with the automatic elimination of large irregulars in X-11 and X-11-ARIMA based on a standard error calculation. The series are checked for additive outliers, level shifts, and seasonal changes. As well as comparing model-based and X-11-ARIMA seasonal adjustment, Scott (1986, 1987) seeks to determine whether the model-based approach can improve the implementation of X-11-ARIMA on BLS series.

An earlier study (Scott (1984)) compared direct and indirect adjustment of CES services employment, and found indirect adjustment to perform as well as direct, as well as providing consistency between components and the aggregate. A more complicated nonhierarchical set of aggregate series occurs in the PPI program. A review and evaluation of the patterns of direct and indirect adjustment there is being considered.

5. References

Dippo, C., Palmisano, M., and Bienias, J. (1987): Report on the Bureau of Labor Statistics' Questionnaire Design Advisory Conference. Bureau of Labor Statistics, Washington, D.C.

Dippo, C.S. and Wolter, K.M. (1984): A Comparison of Variance Estimators Using the Taylor Series Approximation. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 113–121.

Grzesiak, T.J. and Copeland, K.R. (1987): Evaluation of Estimators for Employment Level and Change in Establishment Surveys. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.

Kosary, C.L. and Sommers, J.P. (1987): The Effect of Computer Assisted Telephone Interviewing on Response Inconsistency and Burden. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.

Leaver, S.G., Weber, W.L., Cohen, M.P., and Archer, K.P. (1987): Item-Outlet Sample Redesign for the 1987 U.S. Consumer Price Index Revision. International Statistical Institute, Bulletin of the International Association of Survey Statisticians, in press.

Miller, S.M. (1986): Tests for Normality in Errors-in-variables Regression Models. American Statistical Association, Proceedings of the Section for Business and Economic Statistics, pp. 567–571.

Scott, S. (1984): A Case Study of Direct and Indirect Seasonal Adjustment of Aggregate Series. American Statistical Association, Proceedings of the Business and Economics Section, pp. 686–691.

Scott, S. (1986): A Comparison of Model-based and X-11-ARIMA Seasonal Adjustment. American Statistical Association, Proceedings of the Business and Economic Statistics Section, pp. 385–390.

Scott, S. (1987): The Impact of Outliers on Seasonal Adjustment. American Statistical Association, Proceedings of the Business and Economic Statistics Section, in press.

Silberstein, A.R. and Jacobs, C.A. (1986): Symptoms of Repeated Interview Effects in the Consumer Expenditure Interview Survey. American Statistical Association, Proceedings of the Symposium on Panel Surveys, in press.

Silberstein, A.R. (1987): Respondent Characteristics and Recall Bias in the Consumer Expenditure Interview Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.

Sirken, M.G. and Casady, R.J. (1987): Sampling Errors and Nonresponse Rates in Dual Frame Surveys. To appear in the

Edited Monograph of the International Conference on Telephone Survey Methodology, Charlotte, NC, U.S.A.

Tucker, C. (1985): Identifying Measurement Error in Consumer Expenditure Reports. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 476–481.

Tucker, C. (1986): An Analysis of the Dynamics in the CE Diary Survey. American Statistical Association, Proceedings of the Social Statistics Section, pp. 18–27.

Tucker, C. (1987): Toward the Development of an Ordered Set of Data Quality Categories. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.

Tupek, A.R. and Grzesiak, T.J. (1986): Measuring Employment of New Businesses in the Current Employment Statistics Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 505–510.

Valliant, R. (1985): Nonlinear Prediction Theory and the Estimation of Proportions in a Finite Population. Journal of the American Statistical Association, 80, pp. 631–641.

Valliant, R. (1986): Mean Squared Error Estimation in Finite Populations under Nonlinear Models. Communications in Statistics, Theory and Methods, 15, pp. 1975–1993.

Valliant, R. (1987a): Estimation of Laspeyres Price Indexes Using the Prediction Approach for Finite Population Sampling. Journal of Business and Economic Statistics, 5, to appear.

Valliant, R. (1987b): Laspeyres Price Index Estimation under an Autoregressive Model. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.

Valliant, R. (1987c): Generalized Variance Functions in Stratified Two-stage Sampling.

Journal of the American Statistical Association, 82, pp. 499–508.

Valliant, R. (1987d): Conditional Properties of Some Estimators in Stratified Sampling. Journal of the American Statistical Association, 82, pp. 509–519.

Valliant, R. (1987e): Some Prediction Properties of Balanced Half-sample Variance Estimators in Stratified Single-stage Sampling. Journal of the Royal Statistical Society, Series B, 49, in press.

Werking, G. and Tupek, A., (1984): Controlling Response Error in the Current Employment Statistics Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 658–663.

Werking, G., Tupek, A., Ponikowski, C., and Rosen, R. (1986): A CATI Feasibility Study for a Monthly Establishment Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 639–644.

West, S.A. (1983): A Comparison of Different Ratio and Regression Type Estimators for the Total of a Finite Population. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 388–393.

West, S.A. (1984): A Comparison of Estimators for the Variance of Regression-type Estimators in a Finite Population. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 170–175.

West, S.A. (1986): Estimation of the Mean from Censored Income Data. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 665–670.

West, S.A. (1987): Measures of Central Tendency for Censored Earnings Data from the Current Population Survey. American Statistical Association, Proceedings of the Section on Survey Research Methods, in press.