



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

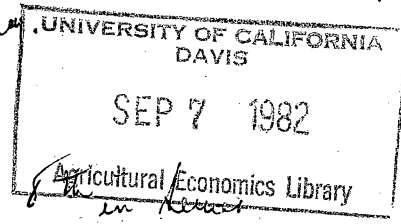
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Consumption

Dec '74 The Labor Review - For Better Jobs in America



1972-73 Consumer Expenditure Survey
Processing and Dissemination

Stephen Baer, ^{LUS} Bureau of Labor Statistics

1st in 12 years

The Bureau of Labor Statistics (BLS) over the past 18 months has been publishing tabulated results from the 1972-73 Consumer Expenditure Survey. This paper will briefly describe the Survey, report on the status of the dissemination of data in publications and tapes, and discuss some of the factors research analysts must consider as they plan projects which will make use of the data. A description of the survey is contained in an article by Michael Carlson, entitled "1972-73 Consumer Expenditure Survey" which appeared in the December 1974 Monthly Labor Review.

The 1972-73 Survey, the 8th in a series and the first in 12 years, is the only comprehensive source of detailed information on family expenditures and income related to socio-economic and demographic characteristics of U.S. families. These data are also a primary source of information to update the expenditure weights and pricing samples leading to a revision of the Consumer Price Index (CPI). The revised CPI is planned for publication in April of 1977. The July 1974 issue of the Monthly Labor Review contains an article on this subject by Julius Shiskin, Commissioner of BLS.

The Survey consisted of two components, each with its own collection vehicle and sample: A Diary or recordkeeping survey completed by respondents for two one-week periods, and a Quarterly Interview Panel Survey

For presentation to the American Agricultural Economics Association,
Atlantic City, New Jersey, September 16, 1976

AAEA 1976

in which families reported information to interviewers every 3 months over a 15 month period. The Diary Survey was conducted from July 1972 to June 1974. The Quarterly Interview Survey was conducted for the calendar years 1972 and 1973.

The Survey design was tailored so that information on larger expenditures was collected by interviewer assisted periodic recall methods and information on frequently purchased less expensive items was collected by diary recordkeeping techniques. Furthermore, the Interview design took account of the notion that some expenditures are easily recalled over long periods while others are more accurately recalled only over short periods.

Although the Diary was designed to obtain the detailed information on food purchases as well as housekeeping supplies, non-prescription drugs and medical supplies, and personal care products and services, respondents were requested to report all purchases made each week except those incurred while on trips.

The Quarterly Interview Survey obtained detailed information on family income, changes in assets and liabilities and major items of expense. Included was information on housing, utilities, housefurnishings, housing repairs and maintenance, clothing, major and minor equipment items, auto purchases, leasing, and repairs, recreation, personal insurance, medical care, education, reading, tobacco, and miscellaneous expenses. Global estimates of food purchases both for at home consumption and at restaurants were also obtained. A comprehensive set of family characteristic information for each family supplemented the expenditure data in each survey component.

The Diary sample was divided equally into 52 weekly subsamples for each of the 12 month collection periods. This permitted an accounting for the seasonal nature of expenditures over time. The Interview Survey samples were initiated in the first calendar quarter of each year. Data collection was conducted over 5 quarters.

The 1972-73 Survey was the first in which the Bureau of the Census took an active part in the collection of such data. The value of its participation is reflected in the cooperation rates attained. The first year diary survey experienced an 80% response rate. This grew to 90% for second year. Over 20,000 one-week diaries were completed each year. In both years of the Quarterly Interview Survey, cooperation rates exceeded 90%, resulting in 10,000 completed questionnaires in each year.

1972-73 Survey
In addition to the collection of data, the Census Bureau was responsible for field editing, quality control, diary coding, production of ^{population} weights from BLS specifications, data keypunching and certain computer edits including separation of consumption expenses from business expenses. Tapes of edited data by quarters were transmitted to the BLS at the conclusion of these activities.

The design of the Survey has influenced the processing activities, and plans for dissemination, and creates the need for users to be familiar with concepts and definitions before applying the information to analytic uses.

Data from the Diary Survey were coded by the Census Bureau field staff. Over 1700 separate item codes were used. Consistency checks were performed to assure that characteristic codes for information such as age, sex, race, and the like, were assigned.

At the conclusion of data coding and review, information was key-punched and subjected to computer edit. Computer files of the survey information were produced and transmitted to the BLS where the tapes were further screened for consistency and for valid codes.

The reliance on respondent recorded entries in the diary led to the need for disaggregation of combined reports in that survey. In some cases, respondents did not supply sufficient detail necessary to meet desired specifications. Item allocation routines were developed by BLS to assign proportions of the value reported for a combined entry among the components of that entry. For the most part, these assignments were made using proportions derived from specific reports in the other completed diaries. For example, non-specific reports of cheese purchases were allocated among soft or semi-soft cheese, solid or hard cheese, or cheese spreads.

Initial pre-tabulations of data were produced by expenditure code to disclose extreme values. These were cross-referenced with the schedule reports and updates were performed. Extreme values confirmed as correct were not adjusted. Because of the large volume of diary expenditure reports, only a limited number of tolerance updates could be carried out. These were sufficient to ensure the reasonableness of the resulting publication tables; however, users of final micro-data tapes should be careful to inspect the values because of the possible existence of undetected keypunch errors.

Test tables of expenditure data classified by important family characteristics were produced, and separate frequency distributions of individual item codes by characteristics were run. The test tables were reviewed for

reasonableness. The weighted frequency counts of socio-economic characteristics were checked against comparably defined data from the Current Population Series reports of the U.S. Department of Commerce. These evaluations led to further correction of keypunch errors in the income data and program correction in the assignment of family characteristic codes.

The pre-tabulations were further reviewed for the statistical reliability of mean expenditures. This review helped to determine the level of detail for publication. In the publication tables, the number of sample families in certain columns and cells is small, since the tables present finely classified breakdowns of families and purchases. Column totals based on fewer than 70 sample families or cells containing fewer than 5 sample observations are likely to have large sampling errors. When these circumstances exist, they are referenced by footnotes to the tables.

Income data were carefully inspected for completeness. While no imputation was conducted, incomplete income reports were identified and coded. The value of income reports from major sources such as wages and salaries, income from profession or unincorporated business or farm was cross-referenced with reports of weeks worked by each family member. Families were coded as incomplete reporters if the head reported weeks worked, but failed to report income from such employment. For cases in which the family head did not work, the procedure was applied to the spouse or other family members. A number of diaries contained members, all of whom did not work during the income reference period. These, most typically, were small families with retired heads. In such cases, incomplete income codes were

assigned if families reported receiving pension, annuity, or social security income, but did not provide dollar reports for these income items. Similarly, those families reporting no work experience but indicating receipt of nonemployment income such as welfare benefits or alimony payments were coded as incomplete income respondents if no values for such items were reported. Publication tables by income class contain a separate column of data for families defined as incomplete income reporters. These represent about 11% of the usable diaries. It appears that many respondents who provided poor or no income data also provided scanty expenditure data.

The collection design for the Quarterly Interview component specified differing recall periods for various classes of items. Data on income, assets and liabilities and housing expenditures were collected annually. Data on large expenditures such as major appliances were collected at six month intervals. Frequently purchased items were collected each quarter. Some information for food and transportation services was obtained by weekly or monthly expense pattern estimates in selected quarters. Data for utilities were collected by quarter from reports on specific bills together with billing periods of reference. These collection procedures were designed to assist the respondents in supplying accurate reports, but later introduced computational problems for data tabulations representing a uniform period.

The Census Bureau had the primary responsibility for initial processing of the Interview Survey data. The family characteristic components for each family record were screened for coding completeness. The expenditure

data were subjected to computer adjustments to separate the business portions of expense entries from the consumption portion. These were most typically found in the housing, transportation, and trips components. Also, computer edits were performed to derive missing information in the liability section of the questionnaire. Furthermore, the Census Bureau supplied special nonresponse codes in expenditure fields for which purchases were indicated, but no expenditure data reported. This was possible because most expenditure entries were preceded by conditional yes-no questions. For example, in the collection of clothing data, respondents were first asked whether or not they incurred an expense for an item before being asked for the expense itself. A special non-response code was entered if the respondent reported making a purchase, but failed to subsequently provide the cost.

Final tapes of the edited and coded data were then transmitted to the BLS. At this stage, special codes were created for data classification purposes. For example, family size codes were produced from the family member data. Programs were written to arrange the expenditure data into desired expenditure classes for publication. Tolerance listings were produced to expose keying errors or questionable reports of great magnitude. Because of the enormous volume of data, only limited tolerance error updates could be undertaken. These have helped to assure the reasonableness of published tabulations, but because not all extreme values were inspected, errors may exist on the micro data tapes. Ultimate users of the tapes should carefully review the data for unusual reports.

Procedures, similar to those performed for the Diary, were carried out for the coding of incomplete income reports in the Interview. Only

about 5% of the usable Interview sample of families were identified as being incomplete income reporters. The difference between the 11% incidence of incomplete income reports in the Diary and the 5% in the Interview is most probably accounted for by the fact that the latter survey benefitted from interviewer rapport built up during the collection of less sensitive data over the four interview quarters prior to collection of income information. The Diary, having a minimum of interviewer participation and a short collection period, was not so advantaged.

Data imputation routines were developed by BLS to account for missing expenditure entries using the special Census non-response codes. Regression testing yielded information on those characteristics most likely to explain expenditure behavior. The schedule returns were arrayed on a matrix defined by family income level, tenure arrangements, a family size/^{household size & number} type variable and a geographic variable. Schedules with missing expenditure entries were then subjected to a computer routine in which imputation values were obtained from like classified survey families reporting complete information. As might be expected, item non-response differed according to the sensitiveness of the queries. The procedure was useful, but surprisingly few imputations were necessary.

The last major processing step for the Interview Survey was to convert the data obtained by the various recall collection periods to annual estimates. At the same time, estimates of sales taxes were produced for the expenditure items. This was an extensive and complicated program. At

the conclusion of the data review and processing operations, a set of tabulations was produced which were reviewed for reasonableness and statistical reliability.

To date, the Bureau has made data available in publications through the Consumer Expenditure Survey Series of Reports. Data from the Diary component of the survey have been published for each twelve month survey period in selected one-way classifications and cross-classifications of family characteristics. Broad classes of expenditures are shown in these tables. By early 1977 data from both years of the Diary survey will be combined and published by family characteristics for the U.S. and each of four U.S. regions. These tables will present detail within broad expenditure classes. We also plan to publish expenditure variances, which can be associated with the magnitude of average expenditures for major consumption components and the number of families represented by the expenditure components.

Work is under way to produce a public-use tape of the detailed Diary income and expenditure data in micro form. Included will be wide range of socio-economic characteristics for each Diary record and those expenditures for which the diary was designed. Separate tapes covering each twelve month survey-period will be available for sale by the end of 1976. Expenditure components and detail on the tapes will be limited to that which will be published in the combined year reports.

Limited data from the Quarterly component of the survey have appeared in two reports thus far. Information on vehicle purchase and repair expenses by selected family characteristics for 1972 and 1973 were published in November of 1975.

Additional selected data on food, tobacco, utilities, housefurnishings, and equipment, clothing, transportation, reading, education, and gifts and contributions classified by selected family characteristics were published for 1972 and 1973. Expenditures are presented as weighted averages of all sample families represented by the table column characteristics, regardless of whether or not each family reported a particular expense. The mean expenditures of those families actually making purchases can be derived by dividing the weighted sample mean by the percent of families reporting the purchase. For example, BLS Report 455-1 indicates that U.S. families, on average, spent \$323.70 on the purchase of new domestically produced automobiles in 1972. However, only 9.16% of U.S. families made such purchases. They spent on average \$3,533.84. Work is continuing to supplement this report with tabulations to appear soon, which will include expenditure data on the above components plus housing, home repairs and maintenance, house hold operations, alcoholic beverages, health care, selected personal care expenses recreation and insurance. Thus, all the components covered by the Interview Survey will be available soon, but not all the detail for each component. These reports contain what we call summary tables and will cover the periods 1972 and 1973. More detailed tables are planned for next year.

Micro data public-use tapes similar to those for the Diary, but containing the summary information from the Interview component, are planned for release in early 1977. These tapes combine selected characteristics

with all expenditures reported in the Interview survey aggregated to the items shown in the tables.

To recapitulate: neither the Diary nor the Interview covers the total range of consumption expenditures. The Interview covers 90% of them, including global data on food purchases, alcoholic beverages, and tobacco. Three expense components: housekeeping supplies, personal care products and non-prescription drugs are not asked at all. A few other expense components in the Interview have small pieces missing, for instance, the telephone component contains no data for coin telephone calls.

In order to provide a complete account of all expenditures, it is necessary to integrate data from the two survey components. The Bureau plans to do this in publications of data to become available early next year, but does not plan to integrate at the micro-level because of the lack of a generally acceptable methodology to accomplish this. Thus, there will be no integrated tapes. Users will have to form their own decisions about data matching and merging techniques. Therefore, we want to call attention to the types of problems that have to be addressed before undertaking integration.

Plans originally called for the Diary to be initiated concurrently with the Interview Survey. However, because of the lateness in receiving results of Dairy pretests and the initial burden of Interview Survey data collection on the Census interviewing staff, the Diary was not fielded until late June of 1972.

The major problem for integration is this difference in the time period coverage between the two survey components, the Interview covering annual expenditures for the calendar year periods 1972 and 1973, the Diary providing weekly expenditure data for the twenty four month period July 1972 through June 1974.

There is one time period which holds potential for integration that of calendar 1973. Appropriate weeks from each year of the Diary can be converted to estimates for calendar 1973. While the collection periods for the two components overlap for the 18 month span July 1972 through December 1973, because of the variable length recall periods in the Interview Survey, integration over this period is not possible.

Those wishing to integrate on a calendar 1973 basis must address and solve a number of conceptual problems. Income estimates from the second-year Interview Survey refer specifically to calendar 1973 period, but this is not the case with respect to the Diary. Diary panel families sampled in calendar 1973 reported their income for the 12 month period prior to the week paneled. For example, a Diary respondent paneled in January 1973 provided an income estimate for the period January 1972 to January 1973. Thus, while the expenditures reported in the Diary can be related to the 12 month calendar period 1973, the income estimates for those families theoretically reflect an approximate mid-point period of December 1972 January 1973.

*Basic
Stage out*

Another integration problem is accounting for the difference in the income level between the two survey components. We have already noted differences with regard to the completeness of reports as collected. In addition, because the Quarterly Interview questionnaire collected income in great detail, these data may generate higher family income estimates than those from the Diary.

Finally, users should be aware that tapes of Quarterly Interview data will contain estimates of both family income before taxes and income after taxes. The Diary tapes will contain only the former. Estimates for personal income taxes were not collected in the Diary, thus after tax income is not available.

One difference in the concept of expenditures on the tapes of the Diary and Interview Surveys is in the treatment of gifts. Purchases of items given as gifts to persons outside the family are normally excluded from current consumption estimates. Expenditure data on the Interview Survey tape do not include such purchases. They are identified separately in a summary classification. However, they are not so separated on the Diary tape. This difference is trivial since the incidence of gift purchases for the expenditure classes on the Diary tape is known to be small.

An important issue in determining the form and detail level of data for the tapes is that of confidentiality. Appropriate controls on the level of characteristics and expenditure detail in the published tables

ensure that no family or individual respondent can be identified.

No such controls can be put on the uses to which data from the public-use tapes will be applied. The Bureau, in complying with its own standards as well as those of the Bureau of the Census, has chosen to handle the problem by limiting the level of detail on the tapes. This has been done in three specific ways.

First, the expenditure data and the item detail of expenditures was carefully inspected for unusual entries. The item classification levels were defined such that a maximum of expenditure detail could be provided but above a level permitting identity of the respondent.

Second, family and family member incomes were carefully screened. Reports less than \$2,000 and those exceeding \$30,000 were presented as coded variables rather than as direct observations.

Third, the geographic information available at the micro-level was limited to include only a region code, a code identifying 3 size classes of Standard Metropolitan Statistical Areas (SMSA's), each further broken down by inside and outside central city, and one code for non-SMSA's further defined by a urban-rural breakdown. No individual areas will be identified for the time being.

Tapes for both survey components contain a set of important characteristics about the family and family members. The tapes of Diary data include references to week of collection so that users may assess the frequency and dollar volume of purchases on individual items over time.

The Interview tapes will contain pointed characteristics on housing and housing services such as number of rooms, and type of utilities used.

The Bureau has learned a great deal about the collection, processing and dissemination of consumer expenditure and income data from the conduct of this survey. It would be unfortunate if another 12 years were to elapse before the survey could be conducted again, particularly in view of the need for current expenditure information.

The Bureau has proposed to conduct a survey of this type on a recurring basis. The merits of such a program are obvious. Limited planning funds have been provided by the Congress and the planning activities are already under way. Final, larger funding has not yet been approved. We are hopeful that the administration and the congress will provide this project with the priority we think it deserves.