

Easley Hoy, Bureau of Labor Statistics

## I. Survey Requirements

The Bureau of Labor Statistics (BLS) initiated a quarterly sample survey of about 2000 establishments to measure timely national changes in the price of labor <sup>1/</sup> in the fall of 1975. The survey requirements included periodic publication of national Laspeyres type indexes for wages, benefits, and total compensation. In addition to the total index, subindexes for five major industries (Construction, Manufacturing, Communications/ Transportation, Trade, and Services) nine major occupation groups (Professionals, Administrative, Sales Worker, Clerical, Craftsmen, Operatives, Transport Operatives, Laborers, and Service Workers) four geographic regions, union/non-union and metropolitan/non-metropolitan areas were requested. Currently, the reference period of the data for the survey is the week containing the twelfth for months March, June, September, and December. The period of collection is about 1½ months so that publication of the data may be released near the end of months, May, August, November, and February.

Although the ECI is currently limited to the private nonfarm economy (excluding households) in the continental United States and changes in wage and salary rates, work is underway to expand this sector to include Alaska and Hawaii next year and the cost of fringe benefits <sup>2/</sup> and federal, state, and local government in 1980. In subsequent years full coverage will be achieved by including the farm economy and private household workers. Thus, when fully developed the ECI will be a comprehensive and timely measure of compensation change throughout all industries and occupations in the civilian economy.

The basic structure of the ECI survey design has maximum flexibility for achieving other goals, subject to the availability of resources for the desired improvements. Significant expansions in resources could provide greater occupational, industrial and geographic detail.

## II. Survey Planning Elements

### A. Universe

Since the ECI survey was to produce both major industry and major occupation subindexes, samples of units in both these groups would be necessary. For occupations, the 1970 Census of Population tabulations by occupation provided an acceptable universe. The Census tabulations were for 441 specific occupations within the nine major occupation

groups (MOG). Cross classification of these occupations were available at a two digit 1967 Standard Industrial Classification (SIC) level. For purposes of ECI sampling, consolidations were made in some cases so that the data used in sampling was for 62 industry groups. The availability of this data provided an objective resource, at exceedingly low cost, for both occupational sampling as well as a basis for index weighting later.

For industries, the 1971 BLS list of all reporting units from the Unemployment Insurance (UI) system (obtained through the cooperation of the individual state agencies) provided the basic universe of establishments in the sampling frame. For several industries in which the U.I. listing might have been incomplete, supplementation from more reliable sources was used to reinforce the sampling frame.

### B. Analysis of Available Information for the Development of Survey Design<sup>2/</sup>

As with all surveys, budgetary and resource constraints on the ECI survey became part of the survey design development. The speculations on unit cost were based on experience in the conduct of BLS Area Wage Surveys Industry Wage Surveys and other surveys in which similar activities were involved. Information on unit costs of (1) personal visit initiation of sample establishments for occupational wage surveys, (2) editing of data, and (3) mail survey operations with telephone and personal visit followup etc., were among those studied.

In addition to unit costs, rough approximations of various sampling variance components were derived. These approximations together with the estimates of unit costs were used in the consideration of alternative designs to help determine the initial optimum design. (When actual variances become available to evaluate the initial ECI design, some possible modifications may result.) Data from a few BLS Area Wage Surveys were used to approximate the various variance elements. The relative homogeneity of wage rate changes for detailed occupations within broad occupational groups within establishments were examined. The relative homogeneity across broad occupational groups was similarly considered. Some between primary sampling unit variances were also generated.

Within the available budget, it was decided that the overall national Employment Cost Index would have the greatest importance.

Thus, the subindexes would have only that precision as would be possible within the overall "best" national index. It was believed that the initial budget would permit an overall index with about a 5 percent relative error with a 95 percent confidence probability\*. At the same level of confidence, it was speculated that there would be about a 15 percent relative error for each of the regional subindexes and each of the five major industry subindexes and about a 20 percent (or less) relative error for each of the nine MOG subindexes.

Another element in the survey planning was the method and frequency of survey data collection. Respondent burden (amount of data collected) and initiation costs were to be balanced. No information was available to BLS on the occupational structure within each of the "establishment" units on the sampling frame. However, the final ECI sampling unit had to be an occupation within an establishment. Therefore, there was a need to control the extent of respondent burden once an "establishment" unit was selected for the sample. Large respondent burden would lead to significant refusal rates in a voluntary survey. Yet, if only one or two quotes were reported per period, there would be waste in initiation cost as well as in subsequent data collection activities.

Additional questions on data collection methodology and frequency were answered in a pretest. The pretest sample of about 350 establishments tested three frequencies of reporting (monthly, quarterly, and semi-annually) by two styles of reporting (occupational average and sample employee micro data). Furthermore, where quarterly and semi-annual collection and compensation averages were requested, collection for two months at a time (current and prior month) was tested.

### C. Conclusion and Planned Survey Design<sup>3/</sup>

The interaction of all of the above considerations led to the following conclusions. The ECI survey design would consist of a stratified, single stage, double sample of both occupations and establishments. For Phase I of the double sample the detailed occupations would be stratified into certainty and noncertainty occupations within each of the 62 industry groups. Data suggested that within each industry group, 5 detailed occupations would be designated as "certainty" occupations. The remaining detailed occupations would be stratified into the 9 major occupation groups (MOGs) for each industry group. Eighteen occupations (2 from each of the 9 MOG's) would be selected on a probability basis for each industry group for a total of 23 occupations. Also for Phase I, the establishments would be stratified by the 62 industry groups.

In Phase I, the 5 detailed occupations with the largest employment in each of the 62 industry groups would be identified from the 1970 Census and designated as "certainty" occupations. The remainder of the detailed occupations in each of the 62 industry groups would then be sampled. The selection of the probability occupations in each MOG would be based on a systematic probability proportionate to size (PPS) approach (relative proportion of employees in detailed occupation to total employees in noncertainty occupations in the MOG in each industry group). Also some form of controlled selection<sup>3/</sup> would be used to provide a diversity of sample occupations within each major industry group.

The first phase of the double sample of establishments would consist of about 10,000 units selected across the 62 industry groups. The sampling interval would be determined and "certainty" establishments would be designated. The selection of the remaining sample units would be by a PPS (to total employment) procedure. These 10,000 units would be surveyed by mail in order to obtain occupational employment mix information concerning the respective 23 Phase I sample occupations in the appropriate one of the 62 industry groups described in the preceding paragraph.

The occupational mix data obtained in the first phase sample units would be used to establish measures of size for the second phase sample selection of both establishments and occupations. The second phase sample of establishments would be approximately 2,000 units. A two way controlled selection<sup>3/</sup> would be used in the second phase sampling to assure that the joint probability selection of occupations within each sample establishment would be within reasonable upper and lower respondent burden boundaries of reporting occupational compensation rates. In general, the phase two sampling method would use a PPS procedure and adopt a principle first suggested by Lahiri<sup>4/</sup>. First, the Phase I occupational employment data would be converted to relative measures of size within each Phase I occupation. Second, the maximum relative measures of size for each Phase I establishment would be identified. Third, the data on the maximum relative measures of size would be used as a basis for a systematic PPS selection of the Phase II sample establishments. Fourth, the data on the within occupational relative measures of size would be recast into relative quantities compared to the Phase II sample establishments' maximum relative measure of size. Fifth, these relative quantities would be used to establish the two way control selection sample patterns from which the final sample of establishments and occupations would be selected. Credit for much of this basic survey design belongs to Mr. Joseph Steinberg, formerly of the Bureau of Labor Statistics.

As a result of the pretest and methodological research, the decision was made that the ECI data collection would begin as a quarterly collection vehicle. The quarterly survey would collect data for a single reference pay period with preference for collecting average compensation for each selected sample occupation. If companies preferred, data collection for a sample of employees would be accepted. In fact, probability subsampling of more detailed occupations would be permitted provided that the subindexes precision would be unaffected. In addition, the initiation of sample establishments would be by personal visit and the sample occupations matched. Also on the initiation visit, compensation information, occupational employment, appropriate establishment/occupation identifiers and response status codes would be collected. For subsequent (ongoing) quarters, the sample establishments would first be contacted via mail shuttle forms by the regional office. Some sample establishments might require an additional telephone follow-up and data (quotes) are collected.

The form of the ECI indexes were to be Laspeyres type indexes with fixed weights (or cost weights) as of the base period. A Laspeyres index uses the fixed weights to calculate a weighted average of change. According to ECI survey design, the process of selection of the occupations and establishments (especially the Phase I occupations) provides the fixed weights which are necessary to calculate Laspeyres indexes. A more specific discussion of the ECI estimation procedures is provided in an accompanying paper in this session<sup>2/</sup>.

### III. Implementation of Survey Design

#### A. Initial Sample

Many of the ECI survey design plans were implemented as anticipated; however, some modifications did occur. The Census list of 441 detailed occupations contained some miscellaneous categories which could not be matched in actual establishments. Therefore, these occupations were collapsed into more identifiable occupations before Phase I occupation selection occurred.

For the Phase I establishment selection, time and cost considerations lead to the decision to select this sample as a subsample from a much larger national probability sample. This larger probability sample of about 200,000 units was selected from a basic frame (1970-71) of about 3,000,000 U.I. reporting units. The selection of the ECI Phase I 10,000 sample units was a PPS procedure within a 3 digit SIC by employment size of the original larger sample.

When the ECI Phase I survey for occupational mix data was initiated, some ECI sample units were the same as those in the 1971 Occupational Employment Survey (OES). When this occurred, the OES data was modified and

used in order to reduce respondent burden and cost while adding to the variance slightly. The response rate for the ECI Phase I survey was about 70 percent. When a sample unit was nonresponse, occupational employment data were imputed using average data from a comparable group of reporting units. Then the occupational employment data for all eligible ECI Phase I sample units were compiled to provide a basis for the second phase sampling of establishments and occupations.

During the selection of the Phase II sample of establishments, some establishments were hit multiple times. From a weighting standpoint, two alternatives could have been implemented. One would be to place the establishment in the control selection process only once and weight the expected quotes by the multiple hit factor. The other alternative would be to place the establishment in the control selection process as many times as it was hit and weight it normally. Timing in delivering and launching the sample became a significant factor, and it was operationally easier to implement the first option. Therefore, the first option was selected, although, the affect on the precision of the final estimate will remain unclear until an evaluation is done.

During a review of distribution tabulations before the initiation of the control selection process, it was determined that in certain industry groups one control selection pattern would not provide a sufficient number of designated occupational quotes. Therefore, in these industry groups more than one control selection pattern was selected. This modification also affected the establishment/occupation weights to be used in the estimation of the ECI.

Once the sample had been designated and the BLS field representatives began survey initiation, the response rates and levels were monitored in BLS Washington. Overall oversampling had been provided in the designated sample to account for expected out of business, out of scope, and refusals. For the base period, overall out of business/out of scope and refusal rates were not overwhelmingly high at 15 percent respectively. However, certain industry groups and one occupation group were definitely short of necessary sample. The Construction Industry was the most critical industry group and the Salesworker Occupation Group was the most critical major occupation group.

#### B. Supplementation

A small supplementation helped the other industry groups, but despite a very heavy supplementation in the Construction Industry, the sample size was marginal with respect to the desired reliability. Also the passage of time would drive the sample size

lower (or the variance higher) than the survey specifications. A special study of the Construction Industry sample was launched and some interesting conclusions were drawn. For the ECI, the sample "establishment" was the U.I. reporting unit. In the construction industry the U.I. reporting unit was generally a project site or sites within a county. Since most construction projects last only two or three years and the construction company may leave the county, the time span between the date of the universe and the date of data collection became critical and was correlated with the out of business rates. Therefore, the older the universe data became, the higher the out of business rate which was the problem. One alternative would be to follow the actual construction company from county to county; however, difficult weighting and coverage problems would arise. It was decided, instead, to select an independent sample of Construction units with the most recent universe date and rotate them into the sample. It is now recognized that for the Construction Industry, sample rotation must occur fairly frequently.

Another conclusion of the study was that for the construction industry, the detailed Census occupation was too fine. The average number of construction industry job matches using the detailed Census occupation was definitely below the average job matches of other industries. The study also revealed that certain detailed occupations (eg. foremen) existed at all construction sites; however, because of their small Census numerical number, they had less probability of being selected. In addition, many construction contractors are so specialized (eg. electrical, air conditioning) with very specific occupations that a high average job match rate with the Phase I selected sample occupations for construction establishments would be difficult if not impossible to attain. Therefore, it was decided to attempt an occupational sample disaggregation approach. Instructions could be provided to field representatives to match all detail occupations within a sample establishment into the appropriate major occupation groups (or sub occupation groups). Then using the sample establishments' detail occupation employments as measures of size, the field representative were to select PPS sample occupations for which quotes would be obtained within each major occupation group (or sub occupation group). Unfortunately, the existing data processing system and potential sources of high nonsampling error prevented the selection of more than one sample occupation per MOG (or sub MOG). This new construction sample has been implemented and has resulted in 25% more job matches (or quotes) per sample establishment.

The Salesworker quote shortage required a different approach. A study of the number of Salesworker quotes collected by the ECI survey indicated that a supplemental sample

of establishments in the Trade industry group would provide sufficient quotes for the Salesworker MOG subindex. First, the Phase I sample of Trade establishments were stratified by whether they reported having salesworkers or not. Then a PPS selection with the maximum probabilities as measure of size was performed in each strata. Naturally, the stratum with the reported salesworkers was sampled much more heavily than the other stratum. In either case, if a supplemental hit overlapped an initial sample establishment, the original sample establishment and occupations were retained and the supplemental hit was in effect discarded. However, for all non overlapped supplemental hits, only certain Phase I sample occupations which needed supplementation were designated for job matching and subsequent data collection. This sample supplementation and subsequent weighting problem, increased the complexity of ECI estimation<sup>5/</sup>.

This fall, the ECI survey plans to supplement the Finance, Insurance, and Real Estate Industry sample in order to publish that Industry's subindex. This supplemental sample is similar to the Salesworker supplementation just described except that no preliminary stratification of Phase I establishments was necessary. The supplemental occupations will also be a subset of the Phase I occupation sample.

#### C. Review of Initial Sample Design

An initial review of the overall ECI survey design has indicated two general directions. The 441 Census detailed occupations appear to be too broad for large establishments and too narrow for small establishments. A use of broad entry level occupations with disaggregation sampling techniques would provide more opportunity for a job match and hence, possibly would lower the sampling error. Disaggregation at the establishment level could also provide data that is more shift free from employment at an establishment/occupation level. In addition, the two way controlled selection appears to accomplish its goal of balancing respondent burden against survey needs. Therefore, plans are to use both the disaggregation techniques and the controlled selection again in the ECI expansion into government. The actual calculation of variance<sup>6/</sup> components will reveal the result of this strategy. Also in government expansion, half samples are being designated in order to improve variance estimation.

#### D. Expansion of ECI and Survey Activities

Currently, the ECI survey is expanding to include the collection of benefit data along with the wage data. Also, the extension of ECI coverage to Alaska and Hawaii is underway. The Alaska and Hawaii sample (excluding Construction Industry) is a single phase PPS sample of establishments with total employment as a measure of size. The sample occupations were a subsample of the original

Phase I sample occupations. The certainty occupations were retained, and a one in two subsample of the probability occupations in each major occupation group was selected. For the ECI expansion into government, two main issues affect the planned survey design. According to 1972 Standard Industrial Classification manual, government ownership is split by industry into Public Administration and the other industries already mentioned in the private sector. In addition, based on preliminary contacts, it is believed that smaller decentralized governmental units, in general, would be more able and willing to provide ECI data in a timely fashion than larger centralized governmental units. Also, another BLS survey (Service Contract Act) is pretesting expansion into the government sector, and early indications are that job matching is better decentrally.

Therefore, for the Federal government, about 2000 submitting offices have been stratified into three major industry divisions; Public Administration, Services, and other. About 230 Phase I type submitting offices (including 21 certainty units) have been selected with PPS and plans are to request Phase I type occupational employment data from Civil Service by Major Occupation Group. Next, a controlled selection procedure would be applied to obtain a Phase II type sample of about 160 submitting offices with a expected average sample of about 5 Major Occupation Groups per sample. Then, within each sample SO/MOG, there would be a PPS selection of a sample establishment/ occupation.

For state and local government, the planned survey design has additional sampling. For the states, five states have been selected with certainty with respect to their state employment. Then, the remaining states have been stratified by geographic area and state employment size, and eleven probability states have been chosen by PPS. For the local government, the Census of Governments provided the sampling frame. Five municipalities and one county have been selected with certainty with respect to their government employment. The remaining local government units (municipalities, counties, townships, and special districts) were stratified by category and employment size class. Forty large and thirty small local governments have been selected by PPS. Therefore, a total of 16 states, and 76 local governments have been selected. A refinement survey has been launched to obtain a list of governmental units similar to the Federal Government submitting offices. Plans are then to select the final establishment/occupations for state and large local governments in a similar fashion (stratified, single stage, double sample) as that used for the private sector. For the thirty small governments, only a single stage, single phase sample is planned. The occupations will be selected by the sampling techniques

described in the construction supplement. For public schools, HEW provided the list of school districts as the sampling frame. About 640 Phase I type school districts have been selected by PPS and Phase I type occupational data has been received. Current plans are to select about 160 Phase II type establishments and the final establishment/ occupations for the schools in a similar fashion as that used for the private sector.

For public hospitals, HEW again provided the sampling frame. About 100 Phase I type hospitals will be selected by PPS. However, no Phase I type occupational data is planned to be collected because of potential response problems. Instead, a single stage, single phase sample of establishments followed by a direct disaggregation within each sample establishment and Major Occupation Group is planned.

#### IV. Future Plans

Much work has been done on the ECI; however, much work still remains. Future plans include the calculation of actual variances<sup>6/</sup> for both publication and survey redesign purposes. Also included are plans for a Quality Measurement Survey. The frequency of reported no benefit policy changes versus known legally required benefit changes has initiated some concern. Also planned is a sample rotation scheme as some sample members would have been in sample for 5 years by the time rotation is planned to be initiated.

#### REFERENCE

- 1/ Employment Cost Index: a measure of change in the "price of labor," Victor J. Sheifer, Monthly Labor Review, July 1975, pp 3-12.
  - 2/ How benefits will be incorporated into the Employment Cost Index, Victor J. Sheifer, Monthly Labor Review, January 1978, pp 18-26.
  - 3/ Sampling Aspects of the Employment Cost Index, Joseph Steinberg, unpublished paper within the Bureau of Labor Statistics.
  - 4/ Lahiri D.B., Bulletin of International Statistical Institute, 33, pt II pp 133-140.
  - 5/ ECI Estimation Procedures, Douglas Wright and Steven Kaufman, 1978 ASA Proceedings, San Diego, California.
  - 6/ Estimates of Variance of the ECI, David Prevert and Harry Marks, 1978 ASA Proceedings, San Diego, California.
- \* Assuming a monthly change of 0.5%