

THE NATIONAL COMPENSATION SURVEY: THE NEW BLS INTEGRATED COMPENSATION PROGRAM

Stephen H. Cohen

Bureau of Labor Statistics; 2 Massachusetts Ave., N.E.; Room 3160; Washington, D.C. 20212

Key Words: Compensation Surveys, Sample Design, Estimation

1. Overview

For many decades, the Bureau of Labor Statistics has conducted studies of wages by occupation and industry. The best known of the early studies stemmed from a Senate resolution of March 3, 1891, which instructed BLS to investigate the effects of tariff legislation on wages and prices.

Systematic collection of wage data by occupation and industry continued after the turn of the century, with changes in coverage dictated mainly by government needs.

Survey activity shifted in the early 1940's defense period to heavy industries essential to war production. Implementation of wage stabilization policy during the war required a large-scale program of occupational wage studies by industry and locality. The emphasis on data by locality continued after 1945 within the framework of industry studies generally designed to yield national and regional estimates.

Area wage surveys, initiated in the late 1940's, were designed to meet the growing demand for pay data related to office clerical and manual jobs that are common to a wide variety of manufacturing and nonmanufacturing industries within a metropolitan area. In 1960, the program was converted to a statistically selected group of areas which data could be estimated to represent all metropolitan areas of the 48 contiguous United States.

Also in 1960, the Bureau began conducting an annual nationwide survey of professional, administrative, technical and clerical jobs in a broad spectrum of private industries. In 1990, Congress passed legislation that provided for local variations in Federal white-collar pay scales. BLS combined its area wage program with the national white collar pay survey into the Occupational Compensation Survey Program (OCSP) to meet this administrative need along with providing compensation analysts with an enhanced local/national products.

In 1976, the Bureau began to publish the Employment Cost Index (ECI) which measures the rate of change in employee compensation, which includes wages, salaries, and employer's cost for employee

benefits. Several elements distinguish the ECI from other surveys of employee compensation. It is comprehensive in that it includes costs incurred by employers for employee benefits in addition to wages and salaries; and covers all establishments and occupations in both the private nonfarm and public sectors. It measures the change in cost of employing a fixed set of labor inputs, so it is not affected over time by changes in the occupational composition of the labor force. The survey is timely in that statistics are published quarterly. The ECI also enables users to compare rates of change in occupational group, industrial, geographic, union coverage, and public/private sub series.

In 1980 the Bureau began annual full-scale comprehensive surveys on the incidence and characteristics of employee benefits for full and part time employees: the Employee Benefits Survey (EBS).

Since the early 1980's discussion has ensued within the Bureau to integrate the occupational wage surveys, industrial wage surveys, ECI and EBS into a single integrated compensation program. One purpose of the goals of integration is to find ways to eliminate the duplication that exists through the current use of three separate surveys to study compensation. With the new joint occupation selection system, each establishment would only be contacted one time (in each sample cycle), rather than possibly being visited for more than one survey. Therefore, the burden on respondents will be reduced significantly, especially for those large establishments that are likely to be selected in multiple samples with many of the same data elements captured in multiple surveys. While the new sample size would be larger than that for an individual survey, it would be smaller than the sum of the samples for the current surveys. Therefore, fewer establishments of all sizes would be visited, so the overall burden on respondents would go down.

In October 1996, BLS began the process of integrating the above program into a single program: The National Compensation Survey (NCS). NCS allows for the continuation of data collected by OCSP, ECI and EBS. A review of some of the key concepts underlying the programs follows.

2. Concepts

Previous occupational pay surveys studied occupations that provided a representation of the range of duties and responsibilities associated with white collar jobs, skilled maintenance trades, and custodial and material movement jobs. In many instances, occupations have been divided into two or more work levels, based upon differences in incumbents' experience and responsibilities. Survey occupations are defined in advance with a uniform set of job descriptions. Survey occupations were judgements chosen. Weekly salaries reported for individuals in white collar jobs relate to regular straight-time salaries paid for standard workweeks. Earnings information for plant workers excludes late-shift differentials and premium pay for overtime.

Data on weekly work schedules; paid holiday and vacation practices; and health, insurance, and retirement plans were recorded separately for nonsupervisory office workers and for production and related workers.

The current industrial scope includes all establishments employing 50 or more employees in the private sector and all establishments for the nonfederal public workforce.

The ECI is designed as a Laspeyres, fixed-weight index at the occupational level, thus eliminating effects of employment shifts among occupational groups. The index weights are derived from the Occupational Employment Survey modified by the 1990 Census. The wage and salary component of the index is represented by average straight-time hourly earnings in an occupation. Straight-time earnings are defined as total earnings before deductions, excluding premium payments for overtime, weekend, and late-shift work. Earnings include production bonuses, commissions, and cost of living allowances but exclude nonproduction bonuses, payments in kind, and tips.

All earnings are computed on an hourly basis. Benefit costs are also computed on an hourly basis. The benefit data portion of the ECI includes paid leave benefits, supplemental pay, insurance, pension and savings, legally required benefits, and severance pay and supplemental unemployment benefit funds.

More detail on concepts can be found in the BLS Handbook of Methods (1997).

3. Integration Efforts

When NCS is fully implemented, most of the information currently produced by these three surveys will continue to be available; the difference is that the source will be a single sample. Integration of the three programs is necessary to minimize the burden to

respondents, to reduce the cost of collection and processing these critical data, and to link data elements of the three surveys together.

One aim of NCS is to produce data on both the local and national level. BLS will therefore sample establishments in 154 localities. The set of localities will include the 31 largest metropolitan areas in the U.S., 5 additional self-representing PSU's, and 118 randomly sampled localities. Data on compensation will be available for 34 of the self-representing largest metropolitan areas, a limited number of other areas, broad geographic regions, and the nation as a whole.

In each locality, BLS will take a sample of establishments. Within each establishment, BLS will take a sample of jobs. BLS will collect information on the work levels of selected jobs, compensation, hours worked, and other characteristics.

"Work level" is defined as the degree of knowledge, skill, judgment, scope, and working conditions associated with a job. The data on the work level of jobs will be collected by classifying each job according to a "generic" method that can be universally applied. Each job will receive some number of points based on a series of factors, allowing the job to be sorted into a level. This methodology, known as "generic leveling" is based upon OPM's Factor Evaluation System (FES) developed in the 70's. The compensation data will include the following information:

- Median and mean earnings by occupation;
- Earnings by job level (and aggregation of jobs by level to Major Occupational Group, total private, etc.), as defined by generic leveling;
- Local and national breakouts on earnings;
- Employer cost of benefits;
- Employee cost of benefits;
- Number of people participating in benefit plans;
- Provisions of benefit plans.

The types of benefits information that will be collected include:

- Health, life, and disability insurance;
- Retirement plan information;
- Leave information;
- Cost of mandated benefits, including Social Security, Workers' Compensation, and Unemployment Insurance;
- Overtime, shift, and bonus pay.

The benefit plan provision information to be collected includes:

- Availability of managed care in health insurance;
- Typical deductibles and copayments in health care plans;
- Other health care information, including availability of coverage of hospitalization, home health care, mental health, and substance abuse treatment;
- Dental, vision, and prescription drugs;
- Average levels of coverage for life and disability plans;
- Detailed pension plan provisions including data on eligibility, survivor options, and disability provisions;
- Detailed savings plans provisions, including match rates, investment choices, and loan options;
- Average number of vacation days, sick days, holidays, and other forms of leave.

4. Sample Design

The survey design for NCS is detailed below. The major features of the design are outlined with issues to be addressed by research to improve the sample design discussed. Other papers in these meetings: Black, Ernst and Tehonica (1997a, 1997b), and Paben and Ernst (1997), discuss some of this research in more detail.

4.1. Stratification and Sample Selection

The NCS sample is selected using a 3-stage stratified design with probability proportional to employment sampling at each stage. The first stage of sample selection is a probability sample of areas, the second stage is a probability sample of establishments within sampled areas, the third stage of sample selection is a probability sample of occupations within sampled areas and establishments.

The selection of sample areas is done by first dividing the entire area of the United States, consisting of counties and independent cities, into primary sampling units (PSU's). In most States, a PSU consists of a county or a number of contiguous counties. In

New England and Hawaii, minor civil divisions are used instead of counties. Metropolitan areas, as defined by Office of Management and Budget (OMB), are used as a basis for forming PSU's. Outside of metropolitan areas, each county defines a PSU.

The PSU's with similar average wages as measured by Unemployment Insurance reports wage are grouped into strata within each of the 9 Bureau of the Census' economic divisions. Then one PSU is selected from each stratum with the probability of selection proportional to the employment of the PSU. There are 36 PSU's that are self-representing, and these include the 18 Consolidated Metropolitan Statistical Areas (CMSA's) and 15 largest Metropolitan Statistical Areas (MSA's). The remaining strata are formed by combining PSU's that are MSA's and have similar average annual pay into 45 MSA strata and PSU's that are non-MSA's and have similar average annual pay into 73 non-MSA strata. The PSU's selected with probability proportionate to PSU employment from these strata, are non-self-representing because each one chosen represents the entire stratum. The NCS sample design was supplemented with 3 PSU's to meet requirements of Federal Employee Pay Comparability Act of 1990.

The sample of establishments is drawn by first stratifying the sampling frame for each PSU by industry and ownership.

The number of sample establishments allocated to each stratum within a PSU is approximately proportional to the stratum employment. Each sampled establishment is selected within a stratum with a probability proportional to its employment. The industry strata were chosen because of a desire to produce estimates of major industry division as well as selected individual industries which are traditionally produced for the ECI and locality products. After the first round of surveys is completed, the allocation will be reviewed using the first round variance estimates to develop an optimal allocation that will efficiently produce the desired publication lines.

After the sample of establishments is drawn, occupations are selected in each sampled establishment. The number of occupations (4, 8, 10, 12, 16 or 20) selected in an establishment depends on the total number of employees in the establishment. The probability of an occupation being selected is proportionate to its employment within the establishment.

Approximately 1/3 of the total NCS sample of 36,000 establishments will be designated core sample for ECI quarterly wage and benefit collection, while for the remaining 2/3 of the sample wage only data will be

collected annually . In order to reduce respondent burden in these core units due to the increased collection time associated with benefits, the occupational selection rate will be reduced. The effect on variances on occupational wage data was found to be a reasonable tradeoff to avoid potentially significant additional nonresponse in these core units that require quarterly collection of both wage and benefit information.

There are a number of data collection issues that require special selection or weighting schemes to address data collection issues. The easiest conceptually is schedules that result from collecting data from units other than assigned. This problem occurs because the respondent can not supply data only for the frame unit because of the way the records are maintained in the establishment. The field documents the issue and the schedule is reweighted. The most difficult process that this effects is the modification of the variance formulae.

Another special handling situation is occupational selection of multi-national corporations. As discussed above each probability selected occupation requires extensive data collection to determine the appropriate skill level of the occupation. When selection is done by establishment, for companies with hundreds of locations and occupations, continued respondent cooperation becomes problematical. These are usually establishments that can significantly effect the estimates in some markets. For those reporters we developed a centralized occupational selection scheme. We select the occupations PPS from a master nationwide list stratified by major operating units of the corporation. Then for each establishment we survey only those occupations that are on the nationwide list. We then poststratify the weights in each establishment based upon the employment in the occupations in the establishment and its total employment.

4.2. Sample Replacement Scheme

The initial plan of NCS was to replace the OCSF samples and outputs using a fixed job list with generic leveling outputs that represent the entire market. The initial sample will be restricted to establishments employing 50 or more workers with expansion to occur over time. Beginning in April 1998, the ECI sample will be gradually replaced by the core NCS units.

When fully implemented the NCS sample will consist of five panels. Each panel will contain one-fifth of cross-area and cross-industry sample, i.e., it will be a representative sub-sample of the total sample

with one-third of the sample members in each panel designated for quarterly collection for the ECI. Each panel will be in the survey for 5 years and then it will be replaced. Under this scheme the entire sample is completely replaced after 5 years. Since there will not be sufficient resources to completely replace the ECI at one time, we will need to develop a process to blend the current ECI sample with the NCS sample to derive the index. Over time the current 23 benefit package studied will be modernized and here too a method to blend the series together will have to be developed.

The primary objective of the replacement scheme is to reduce reporting burden of individual establishments by rotating units out of the sample, and to insure that the establishment sample is representative of the universe it is designed to cover.

For each panel an additional questionnaire is proposed to probe establishments that pass birth screening criteria as applied to the BLS frame to ensure these units are economic birth establishments to the universe. This will enable BLS to weight all five panels to properly reflect the effect of births in our estimates.

4.3. Estimation Procedure

The survey will produce level estimates, such as average wage of professional workers at the entry level along with quartiles, first and last decile, and indexes. The estimation procedures for these two type of estimates are described below. Index procedure includes also seasonal adjustment. Note that both of these procedures involve weighting the data from each employee in the sampled occupation by the final weight, i.e., the inverse of the probability of the employee being in the sample.

The final weights for each occupational quote include the initial sample weights, adjustments to the initial sample weights, and three types of adjustments for non-response. The initial sample weight for an occupation in a particular establishment and PSU reflects the probability of selecting a particular PSU, probability of selecting a particular establishment within the PSU, and probability of selecting a particular occupation within the selected establishment and PSU. Adjustments to the initial weights are done when data are collected for a more or less than the sampled establishment. This may be due to establishments mergers, or splits, or inability of respondents to provide requested data for the sampled establishment. The three types of adjustments for non-response include adjustment for establishment refusal to participate in the survey, adjustment for respondent refusal to provide data for a particular occupation, and

adjustment for respondent unwillingness to provide occupation level information.

The estimation procedure for a level estimate, such as mean weekly earnings, involves multiplying data for each employee in the sample occupation by the final weight and dividing by the sum of the final weights. To generate the proportion or percentage that is desired, such as percent of employees participating in a particular benefit, appropriate employee counts of plan participants and establishment employment totals are used in the estimation procedure.

Level estimates are raked (benchmarked) against the Bureau's employment series to account for growth in the industry since the development of the frame. There is a 24 to 30 month delay from the reference date of the frame until the first estimates are produced from that frame.

Sample allocations for the PSU's except the 34 self-representing PSU's which are published areas are in proportion to the total stratum size that the selected PSU represents. This means that the small PSU's (as measured by total nonagricultural employment in the PSU) have a larger proportion of the universe in the sample. Unfortunately this implies that the larger areas which are economically more interesting have smaller proportions of their universe in the sample. The Bureau is beginning to embark on a research effort to design a small area estimation program to model estimates of these larger areas that do not have adequate sample for a designed based estimate.

The index computation involves the standard formula for Laspeyres fixed-employment-weighted index, modified by the special statistical conditions that apply to the NCS. An index for a benefit derived from the NCS data is simply a weighted average of the cumulative average benefit costs changes within each estimation cell, with base-period benefit bills as the fixed weights for each cell.

The estimation cell is defined on the basis of private or public sector/industry/major occupation group. For the private sector, 72 Pseudo-Standard Industrial Classification (PSIC) industries have been identified, most at the 2 digit SIC level. For the public sector, separate cells are identified for State and for local governments. Industries as broad as "public administration" and as narrow as "colleges and universities" are treated as separate estimation cell industries. For example, one estimation cell is identified as State government/public administration/clerical workers.

The index computations for the occupation and industry groups follow the same procedures as those for all overall indexes except for the summation. The bills

for the occupational groups are summed across industries for each group; the bills for the industry divisions are summed across occupational groups for each industry division. When NCS is fully implemented, indices are planned for the nine Economic Census Divisions and the larger CMSA's and MSA's.

4.4. Non-response

There are three types of non-response: permanent non-response, temporary non-response, and partial non-response. The non-responses can occur at the establishment level, occupation level, or benefit item level. The assumption for all non-response adjustments is that non-respondents are similar to respondents.

To adjust for permanent establishment or occupation non-response, weights of responding units or occupations which are deemed to be similar are adjusted appropriately. Establishments are considered similar if they are in the same 2-digit SIC by broad size class groupings. For occupations, each schedule must have a correct major occupational grouping (MOG) code or census occupation for the schedule to be considered complete. Missing wage data for occupations is reweighted within occupational level if known, otherwise within the same occupation code or MOG. If there are no sufficient data at this level, then a broader level of aggregation is considered for establishment and occupational non-response with suppression of detail below the non-response aggregation level.

For temporary and partial non-response, a replacement value is imputed based on information provided by establishments with similar characteristics. Imputation for benefit non-response is done separately for each benefit both in the base period and quarterly. If it is unknown whether or not a benefit practice exists for a non-respondent, the average cost is imputed based on data from all responding establishments (including those with no plans and plans with zero costs).

There is a continuous effort to maximize response rate. We are developing and providing respondents with new and useful products. We are looking into providing alternative methods for respondents to report their data.

4.5. Measuring the Quality of the Estimates

The two basic sources of error in the estimates are bias and variance. Bias is the amount by which estimates systematically do not reflect the

characteristics of the entire population. Many of the components of bias can be categorized as either response or non-response bias.

Response bias occurs when respondents' answers systematically differ from the correct values. Response bias can be measured by using a re-interview survey. Properly designed and implemented, this can also indicate where improvements are needed and how to make these improvements. Another aim is quick feedback to the field economist to avoid systemic problems from continuing to use incorrect procedures. This program is designed as a process improvement effort rather than data scrubbing.

Each field economist is also subject to Targeted Data Analysis and Regional Office Observational Visits. In these efforts a senior field economist reviews the schedule or observes the interview respectively. After the review, the schedule is discussed with the field economist to improve his/her techniques immediately.

Non-response bias is the amount by which estimates obtained do not properly reflect the characteristics of non-respondents. This bias occurs when non-responding establishments have, for example, benefits movements which are different from those of responding establishments. Non-response bias is being addressed by continuous efforts to reduce the amount of non-response.

Another source of error in the estimates is sampling variance. Industries and occupations are identified that contribute substantial portions of the sampling variance. Allocating more sample to these areas would improve the efficiency of the sample.

These variances will be considered in allocation and selection of the future replacement samples.

References

Black, S. R., Ermst, L. R., and Tehonica, J. (1997a), "Sample Design and Estimation for the National Compensation Survey," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.

Black, S. R., Ermst, L. R., and Tehonica, J. (1997b), "Statistical Problems in BLS Compensation Surveys when the Collected Establishment Data Differs from the Assigned Data," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.

Ernst, L. R. and Paben, S. (1997), Adjusting Establishment Selection Probabilities and Number of Occupational Selections to Reduce Variance in BLS Compensation Surveys, in *Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.

BLS Handbook of Methods, U.S. Department of Labor, Bureau of Labor Statistics, April 1997, Bulletin 2490.

Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.