

INTRODUCING BIRTHS INTO THE EMPLOYMENT COST INDEX SURVEY

Mary K. Gessley, E. Thomas Kelly, and Chester H. Ponikowski

Mary K. Gessley, Bureau of Labor Statistics, Suite 3160

2 Massachusetts Ave., NE, Washington, DC 20212

KEY WORDS: design, collection, estimation

I. INTRODUCTION

In 1991, the Bureau of Labor Statistics developed a plan to represent births in the Employment Cost Index (ECI) survey in a more timely manner. Births are defined as units which did not have a prior chance of being selected to the sample of establishments under the existing ECI sampling process.

Virtually all births are new business formations, i.e., new economic units. Establishments that may have changed name or address, or been formed as the result of reorganization, yet include essentially the same occupations as before the change occurred are not considered births. Similarly, establishments that have changed ownership, but have continued or intend to continue the same economic enterprise, i.e., the same business activity, are not considered births.

The existing ECI sampling process replaces about one-fifth of the sample every year. Therefore, four-fifths of the sample represents a universe which is one or more years old. Without modifications to the sampling process, births which occur in the aged portion of the universe would only be represented when the sample is replaced.

Werking (1988) estimated that there may be about three-quarters of a million new business births each year in the United States, each with an average employment of four. Therefore, about three million employees a year may be employed by these businesses. The wages and benefits provided to employees in these new businesses may be different from those provided in well established (existing) businesses. Also, since wages and benefits in these units may change more rapidly, the ability to represent these units in the sample design as they are formed is a necessary and critical component for providing a comprehensive indicator of changes in labor costs.

This paper describes the ECI survey design (Section II), describes the methodology used to define, sample, and introduce birth units in the ECI survey (Section III), and presents results from the introduction of the first birth sample into the ECI (Section IV).

II. SURVEY DESIGN

The ECI survey is an establishment survey conducted quarterly by the Bureau of Labor Statistics. The goal of the survey is to produce estimates of the rate of change in employee compensation, which includes wages, salaries, and employers' cost for employee benefits. All State and local governments and private sector industries, except for farms and businesses conducted in private households, are covered in the survey. Businesses which employ only owners or immediate family members are outside of the scope of the survey.

The Universe Database (UDB) serves as the sampling frame for the ECI survey. The UDB is created from State Unemployment Insurance (UI) files of establishments which are obtained through the cooperation of the individual state agencies. UI accounts are assigned to all employers in the United States who are required to pay for unemployment insurance. The assigned accounts are updated and maintained on quarterly basis by employer tax reports.

The ECI sample is selected using a 2-stage stratified design with probability proportional to employment sampling at each stage. The first stage of sample selection is a probability sample of establishments and the second stage of sample selection is a probability sample of occupations within the sampled establishments.

The ECI sample of establishments is drawn by first stratifying the sampling frame by industry group and establishment employment. The industry groups usually consist of 2 or 3-digit Standard Industrial Classification (SIC) groups, as defined by the Office of Management and Budget, which are covered by the survey. The number of sample establishments allocated to each stratum is approximately proportional to the stratum employment. Some industries are sampled at a higher rate than other industries because of publication requirements. Two independent half-samples are selected within each stratum for variance estimation purposes. Each sampled establishment is

selected within a stratum with a probability proportional to its employment.

The sampled establishments stay in the sample for approximately 5 years. About 20 percent of sampled units are replaced each year. The sample replacement is done by industry groups. For a more detailed description of the ECI sample design, refer to the BLS Handbook of Methods (Bulletin 2414, September 1992).

III. METHODOLOGY

Under the current ECI sample design, the sample represents five subuniverses. Since approximately 20 percent of the sampled units are replaced each year, 80 percent of the survey sample represents subuniverses which are one to five years old. Births which occurred in these subuniverses were not represented until the sample was replaced. To represent the births in a more timely manner, BLS introduced a procedure for sampling these establishments annually.

The goal of the procedure was to improve coverage of births in the current ECI sample. Due to a lack of additional data collection resources, the birth sample which would have arisen from sampling births in all aged universes could not be introduced at once. Therefore, we decided to introduce a procedure which samples births annually using the same probabilities of selection that were used to select the existing sample. This procedure allows us to represent births which occurred in a given year and to represent all births in the ECI survey design after four years. The implementation of the procedure began in 1991 with the selection of births which had occurred in 1990. The survey processing for this first birth sample was completed in mid-1993 and will be the main focus of this paper.

The sampling frame for births which occurred in 1990 was constructed by matching four consecutive quarters of units on the UDB, i.e., matching the fourth quarter of 1989 with the first quarter of 1990, the first quarter of 1990 with the second quarter of 1990, the second quarter of 1990 with the third quarter of 1990, and the third quarter of 1990 with the fourth quarter of 1990. Matching successive quarters was done to capture all potential births. Simply matching the fourth quarter of 1989 with the fourth quarter of 1990 could result in the exclusion of some birth units. For example, units which have failed to report their employment data for

four quarters are dropped from the database even though they could still be in business. Units that did not meet the BLS matching criteria which include identical UI number, SIC, ownership code, name and address, and some other predecessor and successor information, were considered units new to the UDB and were included on the sampling frame. Certain SICs were eliminated from all four quarters of the frame since they were scheduled to be sampled in replacement groups using a frame the same age as the birth frame. Other SICs were eliminated from either the first, the first and second, or the first, second, and third quarters of the frame if they were selected from a frame the same age as the birth frame. This was done because units in those SICs had a previous chance of selection. For example, units in SICs 63, 64, and 65 were eliminated from the first, second, and third quarters of the birth frame because these SICs were sampled in a regular replacement group using a third quarter 1990 UI frame. The birth sample was designed to represent only those units that have come into existence since the time that the industry was last sampled for the ECI, and not those units that would have a chance of selection from an upcoming sample drawn from the 1990 frame. Finally, the four quarters of birth data were merged, and the 1990 birth frame was reduced to include only those units with fewer than 100 employees. This was done due to lack of data collection resources at that time. Also, we assumed that most of the true births would have less than 100 employees.

In both a regular replacement sample and the birth sample, the sample of establishments is selected by first stratifying the sampling frame by industry group and by establishment employment. The units within each stratum are sorted by Census region, 2-digit SIC, and employment size, and systematic probability proportionate to size sampling is used to select the sample. The employment of a unit is used as its size measure. In a regular replacement group, sampling intervals are calculated by dividing the total stratum employment by the number of sample units to be selected from that stratum.

For each SIC in the birth sample, the mean sampling interval used in the selection of the regular sample of the SIC was determined. In a regular replacement group, a given SIC may have several strata with slightly different sampling intervals. Therefore an average of the intervals in all strata for a given SIC was used to determine the number of birth units to be selected. The number of units to be selected was obtained by dividing the total birth employment within

each SIC by the mean sampling interval used the last time that a given SIC was sampled in the ECI replacement group. This process yielded a total birth sample of 450 units.

Due to deficiencies in the birth sampling frame, not all of the 450 units were expected to be new economic units. Implementation of new UI rules in some states requiring single site reporting complicated the birth determination. Some "new" UI numbers were for units that had previously existed on the UDB under "old" numbers representing multiple work sites. This disaggregation of the UI file began in 1989 and was completed at the end of 1992.

A screening questionnaire was developed in order to aid in the identification of true business births. For each sample unit, the data collector requested information pertaining to starting date, reorganization, and ownership in an attempt to determine which sample units were true births. These questionnaires were also useful in characterizing the units that were determined to be established businesses (nonbirths).

Also, each record on the UDB frame file has a liability date associated with it. This liability date reflects the approximate time that a unit was required to start reporting to the state for Unemployment Insurance purposes. This date, coupled with the information gathered from the questionnaires, determined which units would be introduced into the ECI. If a sample unit was determined not to be a true birth, that unit was dropped from the sample.

A sample unit was determined to be a true birth (new economic unit) if it fell into one of the following categories:

- (1) Its starting date (from the questionnaire) was 1990 or later, and no reorganization or ownership change was indicated.
- (2) Its starting date (from the questionnaire) was 1989, no reorganization or ownership change was indicated, and its liability date was 1989 or later.
- (3) Its starting date was not available, no ownership change or reorganization was indicated, and its liability date was 1990 or later.

Category (2) was included to encompass those units which would not have had a chance of selection in a previous sample due to potential lags in reporting for UI purposes. Some units which started business earlier than 1989, but for some reason were not included previously in the regular sampling frame were excluded by the above criteria. Future representation of these units is currently being considered.

Once the data was collected for all of the true births in the sample, the ECI cost level estimates with and without births were produced for the first quarter of 1993. The survey estimation procedure reflects the two-stage sampling design, that is, the separate probability proportionate to employment size sampling of establishments and occupations within establishments. Two weight adjustment factors are applied to the unit's data. The first factor is introduced to account for the establishment nonresponse. Nonresponse adjustment was performed using a "nearest neighbor" approach using only units determined to be "true births". The second factor, a post-stratification factor, is introduced to adjust the estimated employment totals to actual counts of the employment by industry for the survey reference date. These actual employment figures were obtained from the State Unemployment Insurance reports that correspond most closely to the reference date of the survey.

IV. RESULTS

Of the 450 sample units, 299 were in business and in the scope of the ECI survey and 151 were either out of business or outside of the scope of the ECI survey. Most businesses that were out of the scope of the survey were businesses in which there were no employees other than the owner or an immediate family member. Of the 299 units that were in business and in the scope of the ECI survey, 245 responded and 54 refused. Of the remaining 151 units, 130 were out of business, 15 were out of scope, and 6 could not be located or otherwise did not have response information. Not all of the 299 responding or refusing units were business births, however. The questionnaires indicated that many units had started prior to 1989, but had recently undergone some sort of organizational change, such as new ownership. Out of the 299 responding or refusing units, 110 were determined to be births. Of those 110, 85 (or approximately 77 percent) were respondents, and 25 (23 percent) were refusals.

As expected, many new UI numbers actually belonged to units that had been in existence for years. An existing unit undergoing an ownership change or reorganization would not be considered a birth, as it would have had a chance of selection under the previous UI number. Approximately 11 percent of the 450 units indicated that they had undergone an ownership change in 1989 or 1990, while approximately 10 percent of the 450 had reorganized during the same time period. Name or address changes might also account for an established business appearing on the birth frame.

After nonresponse adjustment for the 25 refusing units was complete, the "true birth" units were used in ECI estimation system to produce estimates for the first quarter of 1993. A comparison of indices with and without births was not made at that time, since there was no prior quarter rate including births from which to measure change. A selection of the ECI cost level estimates for the entire sample both with and without birth estimates, and for birth units alone, is provided in tables A through D.

The inclusion of the birth units seemed to have a moderate downward effect on the ECI estimates. Table A shows that the total compensation per work hour is \$16.66 when birth units are included, \$16.70 when birth units are excluded, and \$14.80 for birth units alone. The wages and benefit costs are both lower when birth units are included. This may indicate that more modest wages and benefits are offered to employees in the birth units. The differences in the cost estimates for births and nonbirths are not statistically significant. However, for all the industry and occupational groups as well as for overall estimates, any differences found indicated lower costs when birth units were included. This pattern was found to be significant at the $\alpha = .05$ level of significance using the paired sample sign test.

Tables B through D provide a selection of the costs per work hour with and without birth units, and costs for birth units alone for major industry divisions and occupational groups which had birth units sampled. In wholesale trade there are no differences in costs per work hour including and excluding births; the birth units alone showed an increase in compensation. This increase was not significant at the $\alpha = .05$ level of significance. In the construction industry, the total compensation per work hour is \$19.71 when birth units are excluded and \$19.61 when birth units are included. For birth units alone, total compensation is \$15.46, however, this decrease was also not significant at the α

$=.05$ level. Half of the difference in total compensation is due to lower wages and a half is due to lower benefit costs for birth units. Table D shows a cost decrease in total compensation of \$0.10 occurred in the managerial and administrative group between estimates including and excluding births. The difference between birth units and nonbirth units is approximately \$6.00. Other major occupational groups showed little or no change in costs with and without birth units, and differences ranging from a few cents to over \$5.00 between birth and nonbirth units. Again, all differences in costs between births and nonbirths for the major occupational groups and industry divisions were well within 2σ limits for the estimates. Although the differences in costs were not significant, if the current pattern continues, the differences may become significant when the process of phasing in the complete sample of births is completed.

V. SUMMARY

The problem of an aged universe is not unique to the Employment Cost Index survey, but a common one among most panel surveys. In implementing a procedure to sample births annually, BLS has improved the ECI's representation of economic births and its ability to provide a comprehensive indicator of changes in labor costs. By phasing in births gradually, over a four year period, the birth units will be represented in the ECI survey without the need for additional data collection resources.

The initial sample, representing the first of the four years of the phase-in process, yielded 110 "true birth" units. In general, when these units were included in the ECI estimates, compensation costs per hour worked seemed slightly lower than costs without the birth units included. There may be some indication that wages and benefits offered in birth units are lower than those offered in the existing sample establishments. As subsequent birth samples are introduced in accordance with the four year process, the additional units may give the inclusion of business births a greater impact on the overall estimates. Regardless of the overall effect on the estimates, the introduction of business births into the Employment Cost Index survey has met the initial goal of improving the survey's representation of the private and public sector economy and hence its estimates of the cost of compensation.

As mentioned earlier, there is some indication from the initial results that the characteristics of the birth units differ from those of the existing, more established units. The birth units may, for example, grow in

employment at a different rate or offer different combinations of benefits. It will be interesting to discover at what point the birth establishments become more similar to the existing establishments. We plan to study how the characteristics of the birth units change over time as more sample data becomes available, and hope to report on these findings in future papers.

REFERENCES

BLS Handbook of Methods (Bulletin 2414, September 1992), Washington, D.C.: Bureau of Labor Statistics, 56-66.

Cochran, W. G. (1977). *Sampling Techniques*, 2nd edition. New York: John Wiley & Sons, Inc.

Horvitz, D. G., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association*, 47, 663-685.

Working, George S. and Tupek, Alan R. (1988), *Modernizing the Current Employment Statistics Program*, ASA Proceedings of the Section on Survey Research Methods.

Table A:

Private Sector, All Industries, All Occupation Groups

	Cost per work hour without birth units (std error in parens)	Cost per work hour with birth units (std error in parens)	Cost per work hour birth units alone (std error in parens)
	<hr/>	<hr/>	<hr/>
Total Comp	16.70 (.133)	16.66 (.131)	14.80(1.324)
*Wages	11.90 (.097)	11.88 (.096)	10.77(.894)
*Benefits	4.80 (.043)	4.78 (.042)	4.03(.521)

Table B:

Wholesale Trade , All Occupation Groups

	Cost per work hour without birth units (std error in parens)	Cost per work hour with birth units (std error in parens)	Cost per work hour birth units alone (std error in parens)
	<hr/>	<hr/>	<hr/>
Total Comp	18.12 (.569)	18.12 (.554)	21.58(3.474)
*Wages	12.92 (.397)	12.92 (.382)	13.73(1.649)
*Benefits	5.20 (.205)	5.20 (.203)	7.86(2.612)

Table C:

Construction, All Occupation Groups

	Cost per work hour without birth units (std error in parens)	Cost per work hour with birth units (std error in parens)	Cost per work hour birth units alone (std error in parens)
	<hr/>	<hr/>	<hr/>
Total Comp	19.71 (.625)	19.61 (.576)	15.46(1.427)
*Wages	13.64 (.365)	13.59 (.333)	11.29(1.087)
*Benefits	6.07 (.280)	6.02 (.262)	4.16(.506)

Table D:

Occupation Group Managerial & Administrative, All Industries

	Cost per work hour without birth units (std error in parens)	Cost per work hour with birth units (std error in parens)	Cost per work hour birth units alone (std error in parens)
	<hr/>	<hr/>	<hr/>
Total Comp	30.45 (.834)	30.35 (.818)	24.53(3.641)
*Wages	22.14 (.599)	22.06 (.588)	17.12(2.035)
*Benefits	8.31 (.270)	8.29 (.267)	7.41(2.311)