Collecting All-Employee Earnings Data in the Current Employment Statistics Survey:
A Comparison of Two Concepts

Karen Goldenberg, Anthony Gomes, Marilyn Manser, Jay Stewart, Bureau of Labor Statistics
Jay Stewart, BLS, 2 Massachusetts Ave. NE, Rm. 4945, Washington, DC 20212-0001

Background

The Current Employment Statistics (CES) program produces monthly estimates of employment, earnings, and hours from a survey of approximately 400,000 business establishments. These estimates are fundamental inputs in economic decision processes of private enterprise, government, and organized labor. Respondents are currently asked to provide data on payroll and hours for production (in goods-producing industries) or nonsupervisory (in service-producing industries) workers for the pay period that includes the 12th of the month (PP12). A review of the CES program has indicated that users' needs would be better served if the CES collected earnings and hours for *all* employees using a more comprehensive concept. The Bureau of Labor Statistics (BLS) conducted a pilot study to examine the operational feasibility of meeting these needs.

We identified two major competing concepts to measure all employee earnings:

- Total earnings for the entire month. In addition to wages and salaries, total earnings includes bonuses and other nonwage payments (NWP). Total earnings is similar to the concept used by the ES-202 program.
- Gross total payroll for PP12, plus a measure of NWP. Gross total payroll includes wages and salaries, but excludes bonuses and other NWP that are not paid at least monthly. Gross total payroll is the earnings concept that is currently used by the CES (790) program to collect earnings for production and nonsupervisory workers.

The objective of the pilot study was to see if one concept would work better than the other in terms of respondent ability to supply the information, ease of reporting, and data accuracy. As part of the test, we also collected all-employee hours data for the time periods corresponding to the earnings reports.

In addition to the two concepts described above, we separately examined the reporting of NWP. NWP include bonuses, commissions and other lump-sum payments that are paid less frequently than monthly.

Study Design and Data Collection

The sample for the pilot study consisted of 1,033 establishments from 8 states in three major industries. All of the establishments were CES respondents who were making the transition from mail to Touchtone Data Entry. Normally, only establishments with fewer than 50 employees undergo this transition, so our sample was restricted to establishments in this size range. However, when we analyzed the data we found some larger establishments in the sample.

The study design had both advantages and disadvantages. The main advantage of sampling from respondents currently in CES was cost. There were not sufficient funds to carry out a separate data initiation and collection effort. By using CES respondents, BLS's Atlanta Data Collection Center (DCC) could conduct the interviews in conjunction with regular CES data collection. However, this sample design was less than ideal from a statistical point of view. It did not include medium- and large-sized establishments (except as noted above). In addition, because respondents already report payroll, the results could be biased toward finding that it is easier to report, even if no actual difference exists. A minor disadvantage of using the Atlanta DCC is that the respondents were drawn from eastern states, rather than the whole U.S.

We divided sample establishments into two groups: the 202 group and the 790 group. Respondents in the 202 group were asked to report *total earnings* for all employees for the entire previous month. These respondents were also asked to report total hours paid for all employees during the entire previous month. Respondents in the 790 group were asked to report *gross total payroll* for all employees for PP12. Respondents in the 790 group were also asked to report total hours

paid for all employees during PP12. Both groups were asked to report, separately, any NWP paid during the previous month.

For establishments that paid different employees at different frequencies, such as weekly and biweekly, we asked respondents in both groups to report these totals separately for each payroll frequency. We did this because adding up payroll amounts is not straightforward when the pay periods are of different lengths. In these cases, payroll amounts must be adjusted to a common time frame (usually a week) before they can be aggregated to an establishment total. We were concerned that asking respondents to add up the payroll amounts would introduce error and increase respondent burden.[1]

We collected data covering the fourth calendar quarter of 1997. Data were collected once a month for three months for the 202 group (November 1997 through January 1998) and for four months for the 790 group (October 1997 through January 1998). The extra month of data collection for the 790 group was necessary because we collected payroll for the current month, whereas we collected NWP for the entire previous month. By collecting data for an entire calendar quarter, we could compare the pilot study microdata to earnings reported to the ES-202 program. The fourth quarter was desirable because of the high incidence of NWP, especially bonuses, that are commonly paid during December.

The research design called for 25 percent of the sample to come from the manufacturing sector, 50 percent from services, and 25 percent from Finance, Insurance, and Real Estate (FIRE). Establishments were to be randomly assigned to the 790 and 202 groups. Because the 790 group's data collection began a month ahead of the 202 group's collection, some states ran out of respondents in the study industries for the second (202) group. As a result, the industry allocation was achieved for the 790 group, but was short on establishments in FIRE in the 202 group. More importantly, a larger fraction of establishments in the 202 group were delinquent reporters, and so would be less likely to participate

in the pilot study. These differences had to be addressed in the data analysis.

The pilot study was conducted as part of the regular CES CATI interview. To minimize interviewer effects, establishments from the two groups were randomly assigned to interviewers.

Analysis and Results

The analysis is based on establishments that reported all-employee counts to CES for the two months prior to the start of the pilot study. This limitation was needed because we found that some sample cases were delinquent reporters or former respondents who had previously dropped out of the CES survey. The resulting sample consisted of 866 establishments (423 in the 202 group and 443 in the 790 group).

Reporting. For both groups, a higher fraction of respondents reported all-employee earnings than reported production/nonsupervisory worker (P/NS) payroll. Overall, 67 percent of establishments provided all-employee earnings data for all months, whereas 59 percent provided P/NS payroll data for all months. These results do not change when broken down by group: these percentages are 63 percent and 53 percent for the 202 group, and 71 and 65 percent for the 790 group. Hence, respondents appear to be more likely to report all-employee earnings than to report P/NS payroll.

Response rates for earnings and hours appear to be higher for the 790 group than for the 202 group (see Table 1).[2] The fraction of respondents who reported all-employee earnings in all months of the pilot study was higher in the 790 group than in the 202 group (71 percent versus 63 percent). In addition, the fraction that reported no all-employee earnings data during the pilot study was lower in the 790 group (16 percent vs. 32 percent for the 202 group). However, the probability of reporting all-employee earnings is strongly related to the establishment's propensity to report P/NS payroll, and the two groups differed on this dimension.

Using the data in Table 1, we computed a "counter-factual" distribution of responses for the 790 group. We used the response rates from Table 1 for each level of CES P/NS payroll reporting, but assumed that the distribution of CES P/NS payroll

---

[1] Note that it was not necessary to collect data for each payroll separately for the 202 group, because adding up monthly earnings numbers is straightforward. However, collecting the information separately reduces respondent burden in some cases. Respondents in both groups were given the option of reporting all payrolls as a single number if it was easier. For respondents in the 790 group this would be the case only if the payrolls were paid at the same frequency.

[2] The response rates in Table 1 are for total earnings in the 202 group and gross total payroll in the 790 group. For establishments with more than one payroll, we tabulated only the primary payroll.

reporting was the same as for the 202 group.[3] That is, we reweighted the percentages for the 790 group using 202 group "weights." The resulting counter-factual distribution is closer to the 202 distribution than the actual distribution for the 790 group: 64.3 percent would have reported all-employee payroll in all months, 13.0 percent would have reported in some months, and 22.7 percent would not have reported payroll at all. Hence, much of the difference in reporting shown in Table 1 can be attributed to different underlying propensities to report earnings/payroll.

Our econometric analysis confirms this result (see Goldenberg et al. 1999). Binary and ordered probit equations show that, after controlling for differential propensities to report P/NS payroll, for interviewer effects, and for industry effects, there is no difference in response rates between groups.

The response rates for hours are very similar to those for total earnings and gross payroll, so we do not report them here. For more information, see Goldenberg et al. (1999).

It is somewhat more difficult to collect monthly data for NWP than for earnings or hours, as evidenced by lower item response rates (see Table 2).[4] A smaller fraction of establishments in the 202 group reported NWP for all months than reported gross payroll or total earnings, while for the 790 group the response rates were fairly similar for both variables. A larger fraction of establishments in the 790 group reported NWP for all months (73 percent vs. 56 percent in the 202 group), and a smaller fraction did not report NWP for any months (25 percent vs. 42 percent).

About half of the difference in the response rates for NWP can be accounted for by the propensity to report P/NS payroll. Computing the counter-factual response rates for the 790 group reveals that if the two groups' propensity to report P/NS payroll been the same, 67 percent of the 790 group would have responded in all months, while 32 percent would not have responded at all. Hence, sample composition accounts for about half of the difference in response rates between the two groups.

We examined whether other factors, such as interviewer effects and industry composition, could have caused the difference. Even after controlling for all of these factors simultaneously in a probit analysis, a large difference still remained.

Respondent Burden. To measure respondent burden, we asked respondents how long it took them to compile the supplemental data (payroll or earnings, hours, and NWP) in the first and last months of data collection. We found that respondents in the 202 group, on average, took 3.8 minutes longer than 790 respondents to compile the supplemental data items in the first month (12.6 minutes vs. 8.8 minutes), and 3.2 minutes longer in the last month.[5] Both of these differences are significant at the one-percent level.

We also examined the distribution of time spent compiling the supplemental data. In the first month, 56 percent of 790 respondents compiled the data in 5 minutes or less, and 77 percent did so in 10 minutes or less. In contrast, 33 percent of establishments in the 202 group took 5 minutes or less to compile the data and 65 percent took 10 minutes or less.[6]

As with the item response rates, it is important to control for the propensity to report P/NS payroll data and interviewer effects mentioned above. We used robust regression procedures to estimate the difference in time to compile the supplemental data. The multivariate analysis confirms that it took longer for respondents in the 202 group to compile the data, with the difference being about 3.4 minutes in the first month and 3.3 minutes in the last month. When we allowed interviewer effects to vary across the two groups (through interaction terms in the regression), we found that they accounted for much of the longer reported compilation time for the 202 concept. This suggests that more extensive interviewer training could have a large impact on respondent burden.

Unfortunately, there does not appear to be much "learning" over the collection period. The average time to compile the data decreased only slightly (by 1.3 minutes for the 202 group and 0.7 minutes for the 790 group). Further, the fraction of respondents

---

[3] In the 202 group, 52.7 percent reported P/NS payroll in all months, 12.5 percent reported for some months, and 34.8 percent did not report P/NS payroll in any months. The corresponding percentages are 65.0, 14.2, and 20.8 for the 790 group.

[4] One possible reason for the lower response rate is that some establishments that did not pay any NWP may have shown up as missing in our data.

[5] Note these means were computed omitting two respondents in the 202 group that took longer than 60 minutes.

[6] These data were missing for many establishments. For these tabulations the sample size was 270 for the 202 group, and 347 for the 790 group. Note also that, because were looking at respondents only, we did not impose the scope restrictions.

in both groups who took 10 or fewer minutes to compile the data increased by only 4 percentage points,[7] suggesting that respondents followed the same procedure each month.[8]

Data Quality. As noted earlier, the motivation for collecting data for a complete calendar quarter was to compare the pilot study data with ES-202 data, which are typically regarded as "truth,"[9] from the same establishments

Our first step was to compute monthly and quarterly earnings for each establishment in the pilot study. Because establishments in the 202 group reported monthly earnings, quarterly earnings simply equal the sum of reported earnings for October, November, and December of 1997. For the 790 group, computing total monthly earnings was a little more complicated. First, we inflated the payroll amount from each payroll into a monthly figure using factors that depend on the month and length of the pay period. Then we summed the monthly earnings from each payroll and added nonwage payments for the month.[10] As before, we computed quarterly earnings by summing total monthly earnings over the three months of the quarter.

The resulting dataset consisted of one observation per establishment, and included data on quarterly earnings, hours and employment from the

CES and the pilot study, and quarterly earnings and employment from the ES-202. Only establishments that provided complete data for all months in the calendar quarter were included in this analysis. This left us with a sample of 541 establishments that provided earnings data for all three months of the quarter. Due to data collection problems, length of pay period was missing for 72 establishments in the 790 group. There were 468 establishments in the final sample, 245 in the 202 group and 223 in the 790 group.

For each establishment, we computed the difference between quarterly earnings from the pilot study and quarterly earnings from the ES-202 for each establishment, and expressed that difference as a percentage of aggregate earnings from the ES-202.

Table 3 shows the average (percentage) difference for the two groups. For the 202 group, the average difference is two percent, compared with 29 percent for the 790 group. However, this difference is not statistically different from zero because the of the large standard error for the 790 group.

The data contain two outliers, one in each group, so we recomputed the means and standard errors without them. The average difference for the 202 group fell slightly to less than one percent, while the average difference for the 790 group fell to 6 percent. Again, the difference between the two groups is not statistically significant at the 10-percent level.

The average differences can be misleading because positive errors offset negative errors. For this reason, we also computed the average of the absolute value of the percentage differences (which we refer to as absolute deviations) (see Table 3). When all observations are included, the mean is 27 percent for the 202 group and 51 percent for the 790 group, though they are not statistically different from each other because of the large standard error for the 790 group. Omitting the two outliers reduces the means to 25 percent for the 202 group and 29 percent for the 790 group. These means are quite close and are not statistically different from each other at the 10-percent level of significance.

Our regression results (not shown here, see Goldenberg et al. 1999) confirm that there is no significant difference in earnings discrepancies between the 202 and 790 groups when we control for other factors (previous reporting of earnings, and industry and interviewer effects). We estimated the equations over all observations using robust

---

[7] The results are qualitatively the same if the sample is restricted only to establishments that had valid responses to the time question in both the first and last months.

[8] During the cognitive interviews that were conducted at an earlier stage of our research (see Goldenberg and Stewart [1999]), a few respondents indicated that they could program their information systems to give us whatever data we requested on an ongoing basis. However, it is unlikely that respondents programmed their systems to provide data for the 3-4 months of the pilot study. More respondents will bear this cost once all-employee earnings and hours become regular data items. This will reduce response burden, but we have no way of knowing whether there will be a differential effect by earnings concept.

[9] Under the ES-202 program, employers submit to their state unemployment insurance (UI) offices the total earnings for all UI-covered employees during the preceding calendar quarter. The dollar total excludes earnings by a small number of non-covered employees, such as proprietors of unincorporated businesses.

[10] Recall that payroll data were collected for the current month, and NWP data were collected for the previous month. This means that payroll data collected in October were combined with NWP data collected in November to obtain October earnings.

regression procedures, and found that, controlling for other factors, the average absolute difference between pilot study earnings and ES-202 earnings is slightly smaller for the 202 group than for the 790 group. However, this difference is not statistically significant at the 10-percent level.

We looked at whether establishments that generated monthly earnings and hours summaries for internal purposes provided better data, and found that respondents with monthly summaries were half again as likely to report earnings that were within one percent of their corresponding ES-202 earnings.

Our regression results also provided some evidence that NWP may be double counted for the 790 group. The coefficients on the industry dummy variables indicate that the discrepancy between the pilot study earnings and the ES-202 earnings tends to be larger for FIRE than for other industries. In earlier cognitive interviews (see Goldenberg and Stewart [1999]), we found that some CES respondents included NWP in their payroll reports. When this occurs NWP may be counted twice. Since NWP are more common in FIRE than in other industries, the positive coefficient on the FIRE dummy variable could be the result of this double counting.

To examine this hypothesis further, we recomputed quarterly earnings for the 790 group *excluding* NWP. After reestimating the equation, we found that the coefficient on the FIRE dummy fell by nearly 40 percent and was no longer significant. When we recomputed the means in Table 3 excluding NWP for the 790 group, we found that the average difference for the 790 group fell to 2 percent (not statistically different from zero). There was no change in the average absolute deviation. These results suggest that extra care needs to be taken when collecting earnings data in establishments that pay NWP.

Conclusions

We found no significant difference in the collectibility of all-employee earnings using either gross total payroll or total earnings, or in the collectibility of all-employee hours for the period corresponding to the earnings. NWP were more difficult to obtain than either hours or earnings. There was also no significant difference between total earnings and gross total payroll plus NWP in terms of the accuracy of earnings data when compared to ES-202. Additional research that includes larger establishments, and that addresses the collectibility and double counting of NWP, would be useful.

References

Goldenberg, Karen, Anthony Gomes, Marilyn Manser, Jay Stewart (1999) "The CES Concepts Pilot Study: Design and Analysis of Responses." Unpublished BLS report.

Goldenberg, Karen and Jay Stewart (1999) "Earnings Concepts and Data Availability for the Current Employment Statistics Survey: Findings from Cognitive Interviews," in the *Proceedings of the Section on Survey Research Methods, American Statistical Association.*

Table 1: Response rates for total earnings and gross total payroll for all employees, by group and P/NS payroll reporting

Reported all-employee earnings or payroll for payroll #1 in:	202 Group (Total Earnings)				790 Group (Gross Total Payroll)			
	Reported P/NS payroll in:				Reported P/NS payroll in:			
	No Months n = 147	Some Months n = 53	All Months n = 223	Total n = 423	No Months n = 92	Some Months n = 63	All Months n = 288	Total n = 443
No Months	54.4	41.5	14.4	31.7	52.2	22.2	3.5	16.3
Some Months	6.1	7.6	3.6	5.0	13.0	38.1	6.9	12.6
All Months	39.5	50.9	82.1	63.4	34.8	39.7	89.6	71.1
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 2: Response rates for NWP, by group and P/NS payroll reporting

Reported NWP in:	202 Group				790 Group			
	Reported P/NS payroll in:				Reported P/NS payroll in:			
	No Months n = 144	Some Months n = 52	All Months n = 223	Total n = 419	No Months n = 92	Some Months n = 63	All Months n = 288	Total n = 443
No Months	66.0	51.9	24.2	42.0	64.13	41.27	8.7	24.8
Some Months	0.7	0.0	3.6	2.1	0.0	6.35	1.4	1.8
All Months	33.3	48.1	72.2	55.9	35.87	52.38	89.9	73.4
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3: Average difference between pilot study earnings and ES-202 earnings

All Observations	Difference between pilot study earnings and ES-202 earnings (as a fraction of ES-202 earnings)		Absolute value of the difference between pilot study earnings and ES-202 earnings (as a fraction of ES-202 earnings)	
	Mean	Standard Error	Mean	Standard Error
202 group (n = 245)	0.0192	(0.0492)	0.2724	(0.0461)
790 group (n = 223)	0.2855	(0.2262)	0.5137	(0.2244)
Difference	-0.2663	(0.2315)	-0.2413	(0.2291)
Omitting Outliers	Mean	Standard Error	Mean	Standard Error
202 group (n = 244)	-0.0064	(0.0422)	0.2478	(0.0391)
790 group (n = 222)	0.0631	(0.0415)	0.2922	(0.0368)
Difference	-0.0695	(0.0592)	-0.0445	(0.0537)