# RECENT EXPERIENCES IN SURVEY COORDINATION AND SAMPLE ROTATION WITHIN MONTHLY BUSINESS ESTABLISHMENT SURVEYS

Mark Crankshaw, Laurie Kujawa, George Stamas
Laurie Kujawa, BLS, 2 Massachusetts Ave., NE Room 4985, Washington, DC 20212

**KEY WORDS: Sample Rotation, Sample Overlap, Permanent Random Numbers**

## I. Introduction

The U.S. Bureau of Labor Statistics (BLS) conducts several business establishment surveys that serve as indicators of current economic activity. The Current Employment Statistics (CES) survey is a monthly payroll survey designed to produce estimates of employment, hours, and earnings. The Job Opening and Labor Turnover Statistics (JOLTS) survey is a monthly survey designed to measure job openings, hires, and separations. Both surveys rely on the cooperation of sample members each month to provide accurate, time-sensitive data. In an effort to minimize the burden on respondents, CES and JOLTS have coordinated their sample processing procedures. Sample coordination minimizes the number of businesses that are selected as sample members of both surveys. In addition, there is an effort to rotate both samples thereby reducing the length of time a unit remains in the sample.

Both CES and JOLTS sample from a common source - the Longitudinal Database (LDB). The LDB is the universe that the Bureau of Labor Statistics uses to draw samples for establishment surveys. The LDB contains approximately 7.5 million U.S. business establishments, representing nearly all nonfarm elements of the U.S. economy. The primary source of data for the LDB is the quarterly contribution reports filed by employers for each Unemployment Insurance (UI) account. Before BLS receives the data, the State Employment Security Agencies (SESAs) review the data as part of the Covered Employment and Wages Program. The LDB contains employment and wage information from employers, as well as location information. The LDB is updated on a quarterly basis; however, the most recent data available is always six to nine months out of date.

The CES and JOLTS samples are coordinated through the use of Permanent Random Numbers (PRNs). Permanent Random Numbers are assigned to all records on the Longitudinal Database. Ohlsson provides an explanation for the use of PRNs in survey sampling. (Ohlsson, 1995) PRNs are uniformly distributed over the interval [0,1). Each PRN is carried out to 12 decimal places to avoid duplication of numbers. PRNs were initially assigned to all records on the 1995 second quarter sampling frame. As birth units are added each quarter, new PRNs are assigned independent of existing PRNs. A collocation technique is used to insure the birth units are evenly distributed over the interval [0, 1). Because collocation distributes the PRNs evenly on the interval [0,1), the collocated PRNs allow the proper representation of births in the sample (Butani, Robertson, and Mueller, 1998). Each record is also given a PRN date. This date corresponds to the year and quarter that the record first appeared in the database. Because the frame is linked from quarter to quarter, PRNs can be assigned for the life of the establishment. As a result, when the establishments are sorted by PRN they remain in essentially the same order. This minimizes the number of sample units that are cancelled as well as the number of new units that need to be solicited during sample selection.

This paper addresses issues that arise with survey coordination and sample rotation. First, a brief overview of the sample designs for CES and JOLTS is provided. Next, an explanation of the constraints facing both surveys is described. In section IV, proposed sample rotation plans are discussed. Section V summarizes the issues associated with coordinated samples.

## II. Overview of Sample Designs

### CES

The Current Employment Statistics survey is in the process of implementing a major redesign. (Werking, 1997) The survey is moving from a quota-based sample design to a probability-based design. Currently, all industries with the exception of Services are under the probability sample design. CES is a survey of more than 300,000 business establishments. The CES Redesign is a State-based, stratified simple

random sample. The Unemployment Insurance (UI) account is the basic sampling unit of the survey. For multi-establishment UI accounts, the sample unit includes the entire cluster of worksites within a UI account that appear on the frame at the time of sample selection. The sample is stratified by state, industry and employment size. With 11 industries and 8 size classes, there are 88 total allocation cells per State. (As CES moves from the Standard Industrial Classification (SIC) system to the North American Industry Classification System (NAICS), there will be 13 expanded supersectors.) Within each allocation cell, units are sorted by MSA. Because the sampling rate is uniform across the entire allocation cell, implicit stratification by MSA insures a proportional number of units are sampled from each MSA. The sampling rates for each stratum are determined through an optimum allocation process. The process distributes a fixed number of sample units over a set of strata in such a way as to minimize the variance of total nonfarm employment at the statewide level. As a result of the optimum allocation, more sample units are generally selected from cells that cost less to collect, have more units, and a larger variance.

**JOLTS**

The JOLTS survey uses a probability-based sampling design that surveys approximately 16,000 non-farm business establishments to produce National and U.S. Census Region level estimates. (Crankshaw and Stamas, 2000) Unlike the CES survey, the basic sample unit of JOLTS is the individual business establishment. The JOLTS sample is a stratified simple random sample stratified by U.S. Census Region, establishment size, and industry division. The JOLTS sample was selected on a Standard Industrial Classification (SIC) industry division basis. The JOLTS survey, however, is designed to be able to produce both SIC industry division and NAICS Supersector based estimates. The JOLTS sample is stratified by four U.S. Census Regions and six size classes—with the largest size class being sampled with virtual certainty. There are eleven SIC industry divisions and fifteen NAICS Supersectors. (Federal government and state and local government are being counted here as divisions or Supersectors.) The JOLTS sample was allocated (by SIC) into 264 allocation cells using an optimal allocation process similar to that of CES. Data gathered by a JOLTS pilot study found that the standard deviation of the JOLTS variable (job openings, hires, quits, etc.) levels are proportional to the mean frame across establishments within stratum. Therefore the allocation process optimally allocated the JOLTS sample taking into account the frame employment variance within a given allocation cell. This ensured that, like CES, more sample units were generally selected from cells that had more units and had a larger variance in the variables of interest.

The JOLTS sample was divided into 18 non-certainty panels and a single near-certainty panel (sample units from the largest size class as well as cells with few units and other cells designated by the allocation process). Each non-certainty panel was designed to ensure that each panel could be representative of the entire non-certainty population. As a result each non-certainty panel is the same size and comprised of a similar allocation. This was done both to ease the enrollment burden of the new survey as well as to aid in sample rotation.

## III. Constraints on Resources

Operational issues have placed constraints on both CES and JOLTS. Limited resources coupled with increasing costs led to the development of procedures to further maximize overlap across sample redraws. Maintaining acceptable response rates for a probability sample required more resources than initially anticipated. As a result, annual sample overlap within surveys has currently been increased to a level that both CES and JOLTS can support.

During the first quarter of each year, the CES sample is redrawn. The sample update process provides the most recent information on the industry, size, and location of a unit. Annual sample selection helps keep the CES survey current with respect to employment for business births and deaths. Under the CES probability design, a simple random sample is selected from the frame where strata are defined by state, industry, and size class.

In addition to the annual reselection, the CES sample is also updated on a semi-annual basis. The semi-annual update provides an opportunity to sample births that were not on the frame during the first quarter annual reselection. Births make a significant contribution to the economy. Therefore, the semi-annual update attempts to capture a representative portion for the sample.

After births are identified on the frame, a simple random sample of units are selected. It also allows establishment deaths to be removed. The semi-annual update also tracks the sample that was selected during the first quarter and provides updated information on employment and location.

The original CES redesign methodology called for a quarterly update for business births, along with a first quarter annual redraw of the sample. Even with the maximum sample overlap provided by PRNs, the solicitation burden was unmanageable. A fourth quarter update was also problematic in terms of solicitation. Many births that appeared on the fourth quarter birth update would be dropped from the sample during the first quarter reselection. Operationally, this would prove to be an overwhelming burden on data collectors. As a result, CES moved to a semi-annual birth update.

Despite the reduction in birth updates, there was still an overwhelming number of new sample units. This led to the introduction of a swapping procedure. Swapping was introduced during the 1998 annual update. The procedure essentially replaces new sample members with existing sample members that would otherwise rotate out of the sample. The procedure is implemented after a random sample has been selected during the annual update process. All new sample members, i.e., sample members that were not included in the sample selected during the prior year, are identified. These are the units that are available for swapping. All units selected during the previous sample year and not selected during the current year are also identified. If a new unit is to be swapped with a previous sample unit, it must be within the same state, industry and size class and have the same PRN date. Newly selected units are replaced until all suitable replacements are exhausted. Units are generally available for swapping due to changes in MSA, industry, or size class. Swapping increases the sample overlap from year to year, thereby reducing the number of new units that need solicited. Establishing contact with respondents requires significant time and energy. In addition, the initial explanation of the requested data elements requires additional time. As a result of the procedure, resources could be redirected at improving response rates.

Table 1. Results of the swapping procedure

| Frame Year | New Units Selected | Units Available for Swapping |
|---|---|---|
| 1998 | 35,058 | 14,631 |
| 1999 | 40,234 | 14,814 |
| 2000 | 38,624 | 16,957 |
| 2001 | 37,726 | 11,592 |

Table 1 illustrates the results of swapping based on a fully implemented probability sample. In production, the swapping technique was only performed in industries that had switched to probability for at least one year. No swapping was performed within an industry on probability during the initial year of implementation. Because all industries have not been initiated under the probability sample, the number of units swapped in production was less than the numbers displayed in the table.

The overwhelming number of new sample units also led to the reduction in the size of the CES sample. The number of sample units drawn for the probability sample was set approximately equal to the number of units in the CES quota sample. With the implementation of the first industry, it was evident the survey could not adequately support a probability sample of equal size. As a result, the sample was cut approximately 20 percent. The sample size cut will provide additional resources that will be devoted to data collection and follow-up techniques essential in a probability sample.

Table 2. CES Sample Size

| Frame Year | # of Sample Units |
|---|---|
| 1998 | 263,294 |
| 1999 | 250,118 |
| 2000 | 201,171 |
| 2001 | 207,140 |

Table 2 illustrates the sample size before and after the reduction. As the probability sample was being implemented, it was clear that it would require more resources than initially planned. With the 2000 sample, the number of units was reduced by approximately 20 percent.

## IV. Sample Rotation

CES has not currently implemented a sample rotation scheme. CES resources have been redirected due to the implementation of the probability sample and NAICS conversion. When the probability sample is fully

implemented under NAICS in June 2003, additional resources will be available. These resources will provide CES an opportunity to reduce (or eliminate) swapping and rotate the sample. Several techniques utilizing PRNs have been proposed; however, a decision regarding sample rotation has not been reached. Methodological as well as operational aspects associated with each technique are currently being reviewed.

The first rotation technique reviewed was the constant shift method outlined by Ohlsson in the paper titled "The System for Co-ordination of Samples from the Business Register at Statistics Sweden." The technique utilizes PRNs to shift the sample selection starting point a specified distance. The distance is calculated based on the amount of rotation desired. After an initial review, there are definite concerns with this technique. Operationally, this technique could prove to be problematic for CES. Because the sampling frame is updated quarterly, units shift selection cells due to changes in size and industry. As a result, a unit may rotate out of the sample for a year and back into the sample the next. In addition, birth units may rotate out of the sample after only a year. These issues would have to be addressed before this technique could be utilized.

The second rotation scheme that was reviewed was the random rotation group method. This method is also outlined in Ohlsson's paper (Ohlsson, 1992). In the random rotation group method, each unit is permanently assigned to a rotation group. During the first year, the PRN assigned to all units in the first rotation group is shifted to the left by an amount dictated by the amount of rotation desired. The PRN of units in the second group is shifted to the left by the same amount during the next year. This procedure continues each year for the subsequent groups. Like the constant shift method, there are concerns with this technique. Units can be rotated into the sample for a single year. Sample units can also move out of the sample one year and in again the following year. There is no guarantee the amount of time a unit will be out of the sample. This method of rotation prevents units from staying in the sample a specified number of years. It does not guarantee a minimum time for units to remain out of the sample. With limited resources, this places a significant strain on data collection.

The JOLTS program has not yet begun to rotate sample either. Before a decision is made, coordination of sample rotation techniques will need to be reviewed. If CES and JOLTS do not implement similar techniques, sample coordination will be lost. As a result, both the respondents and estimates may suffer.

JOLTS is currently looking at the effects of two different approaches for rotating sample. This exercise is further complicated with a change in the industry basis of the design. Also complicating matters is that the original sample was selected from a frame that is now two years older than the last frame. There have been many establishment births and deaths over that period.

A new JOLTS sample was scheduled for implementation after the last of the 18 non-certainty JOLTS panels had been enrolled into the JOLTS survey. The new sample will utilize the latest LDB establishment data and the PRNs on the LDB. The PRN dates can be used to identify births on the LDB frame. The PRN provides a position relative to some starting point in each allocation cell from which to select new sample units. The new JOLTS sample will be a reflection of new businesses entering the frame and other businesses leaving the frame, as well as establishment moving between size and to a lesser extent, industry strata.

Both sampling rotation schemes under consideration for the JOLTS survey involve reselecting a fresh 30-panel JOLTS sample from the original JOLTS PRN starting point. Panels 1-19 provide for 12 additional months of sample. The main difference between the two schemes is in the panel assignment for surviving and birth units. Surviving units are units in the original sample that remain on the latest LDB frame. Birth units are establishments that have been added to the LDB frame since the original sample was drawn.

The first scheme makes no attempt to fix the original panel of surviving units. Essentially, under this scheme, the frame is defined and ordered within strata by PRN. Sampling for panels 1 through 30 begins with the same starting PRN as the original sample. Because establishment births and deaths on the LDB frame are randomly assigned PRNs, the units surviving from the original panel can be shifted from their original panel. At the appropriate time the initial panel (panel 1) can be retired and be

replaced by panel 19. At regular intervals the original panels (2-18) can be sequentially retired and replaced by the new sample panels (20-30) until it is necessary for a new sample to be redrawn. The benefit of this scheme is that it is that selection is easier to implement. Also, because of the reliance on the PRN each panel should not have a bias in its composition of births and surviving establishments. On the negative side, firms growing in employment, as well as those decreasing in size, or going out of business, tend to push survivors into higher and higher numbered panel. This would require these establishments to participate in the survey longer. It would also require solicitation for replacement units for lower numbered panels. This scheme is similar to the constant shift method under consideration by CES.

The second scheme keeps surviving units in panels 1-18 in their original panel. We will refer to this as controlled rotation. In addition, this scheme identifies birth units by panel under the original design and places each into the appropriate stratum and panel under the new design. Additional sample units, which could be frame births or survivors, supplement the births and survivors to reach allocated totals for each stratum and panel. When the allocation called for fewer units in any stratum/panel than were available from survivors and births, survivors and births were dropped randomly using the PRN. Panels 19-30 are comprised entirely of new sample units. The panels can then be rotated out in a manner similar to the first scheme. The benefit of this approach is that it allows non-certainty units that have been in the survey for at least 18 months to depart the survey as originally scheduled. However for panels 1-18 to be representative until such time as they are replaced, they would have to be supplemented with births and other sample units to bring strata units up to the allocated levels.

Table 3 shows the workload under the two sampling schemes. Because the monthly collection under both methods is the same, we only look at births and other new units added to the sample. Under the original plan, data collectors would be required to contact and enroll 334 more sample cases in panels 1 through 18. This results from sample survivors shifting to higher numbered panels. With this shift in panel assignments, workload for panels 19 through 30 is lighter under the original plan by 1163 units. Of course this is at the cost of breaking a commitment of keeping an establishment for a fixed number of months, letting some establishment go before their time is up and the possibility of having to let an establishment go and then re-enroll it in a later panel. (An alternative is to have the employer continue to report and re-weight or ignore the response.)

Table 3 also suggests that controlling selection to keep survivors in their original panel might cause a bias by under-representing birth units. The original plan selects 10 percent more birth units than the controlled approach.

Table 3. Workload differences between sampling schemes

| Panels | Births | | Additional Sample Cases | | Total Additional Workload | |
|---|---|---|---|---|---|---|
| | Original | Controlled | Original | Controlled | Original | Controlled |
| 1-18 | 1,549 | 1,317 | 2,617 | 2,515 | 4,166 | 3,832 |
| 19-30 | 1,057 | 993 | 7,980 | 9,207 | 9,037 | 10,200 |
| 1-30 | 2,606 | 2,310 | 10,597 | 11,722 | 13,203 | 14,032 |

| Table 4. Comparison of birth sample and estimates | | | | | | |
|---|---|---|---|---|---|---|
| Panel | Original | | | Controlled | | |
| | Sample Birth Units | Birth Establishments Estimates | Birth Employment Estimates | Sample Birth Units | Birth Establishment Estimates | Birth Employment Estimates |
| 0-30 | 2,649 | | | 2,353 | | |
| 0-18 | 1,592 | 1,135,638 | 8,752,721 | 1,360 | 794,596 | 7,763,062 |
| Standard Deviation 1-18 | | 188,446 | 1,347,045 | | 127,487 | 1,340,156 |
| Population | | 1,175,473 | 9,031,628 | | 1,175,473 | 9,031,628 |

Table 4 shows the number of birth units in the original panels contrasted with the number of births in all panels under the two different approaches. Based on the estimates pooled across the first 18 panels, it is evident that the original method provides a sample that is much more representative of the population than the controlled rotation method. This is true both in terms of the estimated number of new records on the frame and the number of employees associated with those records. With the controlled rotation, every panel underestimates the number of new records and only two panels overestimate the number of employees. This suggests the controlled rotation along with the postratification of births and survivors provides a sample that is biased against new employer records.

## V. Summary

The practical business of collecting data from business establishments often puts constraints on sample design. The use of PRNs for sample selection does not limit our ability to deal with those constraints. This paper has shown how the CES has adjusted sample selection to meet the requirements imposed by data collection, and how JOLTS may deal with these issues.

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

## References


Butani, Shail, KennethW. Robertson, and Kirk Mueller, "Assigning Permanent Random numbers to the Bureau of Labor Statistics Longitudinal (Universe) Data Base," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 1998.

Crankshaw, Mark and George Stamas, "Sample Design in the Job Openings and Labor Turnover Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 2000.

Ohlsson, E. (1992), "SAMU - The System for Co-ordination of Samples from the Business Register at Statistics Sweden," R&D Report 1992:18, Stockholm: Statistics Sweden.

Ohlsson, E. (1995), "Coordination of Samples Using Permanent Random Numbers," in B.G. Cox et al., Business Survey Methods, New York: Wiley, pp. 153-183.

Werking, George S. "Overview of the CES Redesign," *Proceedings of the Section on Survey Research Methods, American Statistical Association, 1997.*