SAMPLING FOR EMPLOYMENT AT NEW ESTABLISHMENTS IN A MONTHLY BUSINESS SURVEY

George Stamas[1], Karen Goldenberg, Bureau of Labor Statistics; Kerry Levin, David Cantor, Westat
Karen L. Goldenberg, Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Room 4985, Washington, DC 20212

Key words: Sampling, Business Births, Establishment Survey

1. Introduction

The Current Employment Statistics (CES) Survey is conducted monthly by the U.S. Bureau of Labor Statistics (BLS). This survey samples business establishments to produce estimates of employment, hours and earnings, by industry, for the nation, states and metropolitan areas. The sampling frame for CES is a file of administrative records from state unemployment insurance (UI) programs. An updated file is available about 9 months after the end of each calendar quarter.

Like any survey attempting to estimate characteristics of a current population, CES misses the contribution of new units that did not appear on that frame. In the present context, this means that CES does not reflect employment from newly-opened businesses. New businesses have been estimated to contribute between 1 and 4 percent of U.S. employment (Subcommittee, 1994), and their absence has been the subject of critical review (e.g., American Statistical Association, 1994). The study described here considers the possibility of incorporating employment from new businesses into the CES by sampling from files of newly-established UI accounts to identify new businesses, or "business births," and to obtain employment data from them.

2. Births and birth frames

Defining births. The sampling unit for the CES survey is an employer's unemployment insurance (UI) account. An important feature of the UI program is that it functions separately within each state, and according to the UI laws of that state. Employers open accounts in each state where they operate business establishments. Most states require employers to register for an account within either 90 or 180 days of becoming liable for UI taxes in that state, although compliance with this requirement varies. While many new UI accounts go to new businesses, employers also obtain them for reasons such as changes in ownership, mergers, acquisitions, name changes, and incorporations.

There are at least two ways to define business births. Following a strictly administrative definition, any new UI account is a birth. Therefore, we could create a frame of newly-issued UI accounts and select a sample of those accounts to represent business births. This approach would require that we treat any employer in our sample with a *change* in UI account numbers as a business death. As noted above, however, not all new accounts go to new businesses. Also, every death resulting from a change in UI accounts requires sample turnover and recruitment of replacement respondents, with associated costs and response burden.

An alternative to the administrative definition is an economic definition of business births. Using this approach, a business birth is an establishment which formerly had no chance of being selected into the CES survey. The *reason for the new UI account* is the key to whether or not an employer's business is a birth, because it determines the zero or nonzero probability of prior selection. Accounts assigned to newly-established businesses had no prior probability of selection, and so clearly are births, as are accounts assigned to businesses hiring employees for the first time. In general, continuing economic ventures (such as ownership changes or corporate reorganizations) receiving new UI accounts had a prior probability of selection into CES, and are not business births. The exception is partial acquisitions of businesses. The portion of the business that remains in operation under the old ownership is not a birth, but that business will show a loss of employment as a result of the sale. Therefore the acquired portion, under a new account, must be treated as a birth.

Because of the costs and sample turnover associated with treating all changed UI accounts as deaths, we felt that the economic definition of births would serve us better than the administrative one. The pilot study reported here evaluates the feasibility of using the economic definition. This required that we interview recipients of new UI accounts to determine the reason for the new account. Because of the cost involved, we proposed interviewing a sample of these employers.

Birth frames. CES is conducted for BLS by and within the individual States. The CES sampling frame is taken from each state's Quarterly Unemployment Insurance (QUI) program administrative files, a compilation of employer Quarterly Contribution Reports (QCR), i.e., unemployment insurance tax reports. The files are due from the states to BLS in Washington five months after the calendar quarter, with additional time allowed for editing before businesses in this file are available for sampling. As a result, a business that opened in the first month of a quarter and paid UI taxes for that quarter would not be eligible for sampling for as much as nine months, at which point it is no longer truly a "new" business. The question we raised was whether similar

data on new employers was available in the state agency, and accessible for sampling at an earlier time.

States identify new employers in several ways, any of which result in the employer being issued a UI account number. Every state requires a new employer to submit a status determination form (SDF) and so to register for UI as part of the procedure for becoming an employer. A business may neglect this paperwork but file the QCR and pay taxes at the end of the first quarter in business. A previously-unknown employer may come to the attention of the state agency because a laid-off employee filed a UI claim. Regardless of how the state learns about them, the state eventually records all new employers, along with all other UI-liable employers, on its Employer Master Information (EMI) file.

BLS staff conducted a series of discussions with state UI program staffs to identify other data sources. The EMI appeared to be the earliest source of machine readable data on new UI accounts (Subcommittee, 1994). This file includes a record for every employer issued a new UI account for any reason, and seemed to provide a suitable frame for business births. However, there is one important difference between this births frame and the CES sampling frame. The births frame includes employers that are *expected* to file QCRs but might not ever be legally required to do so, while the CES sampling frame consists only of employers that have UI liability and so have filed QCRs.

Incorporating employment information into CES: The primary reason for studying business births is to measure their employment and to include that employment in the CES estimates. For birth units, our plan was to add the weighted employment (selection weight times reported employment) to the total employment. We anticipated enrolling those units for monthly participation in the CES survey until they (and the units that they represented) had an opportunity to be selected from the ongoing frame. For various reasons, this phase of the research was never implemented.

3. The Business Births Pilot Study

The primary purpose of the Business Birth Pilot Study (BBPS) study was to determine whether we could effectively separate employers with new UI accounts into economic births and continuing economic units by means of a short telephone computer assisted telephone interview (CATI). We learned during the initial development of the questionnaire that concepts associated with defining business births were complex, even for the researchers defining them. As a result, we invested considerable effort into questionnaire design, to ensure that respondents could provide information about their reasons for new UI accounts. We began with cognitive interviews with employers who were issued new UI accounts; these resulted in a series of concept and questionnaire revisions (Goldenberg, 1996). Because the survey was to be conducted over the telephone, we then tested the questionnaire in a telephone environment.[2] We conducted three rounds of telephone pretests with employers from the birth frames, refining the questionnaire after each round (Goldenberg et al., 1997). Once the questionnaire was operational in CATI mode, some interviews were tape recorded and subjected to behavior coding, leading to still other changes in questions and procedures (Levin et al, 1996).

Frame construction and sampling. Because of the variations in practice across states, we wanted to receive frames from, and select samples across, as many states as possible. Seven states provided usable frames in time for the pilot survey. We sampled 250 UI accounts the first month and 500 UI accounts in each subsequent month. We took samples from 3 to 5 states per month. We stratified the frames by state and employment size, where one size class was unreported employment or fewer than 50 employees and the other size class was 50 or more. We randomly selected sample cases within each stratum.

Interviewing. We conducted interviews on current samples monthly from July through November, 1996, incorporating the experience of the previous month into the next field period. Initially, the data collection periods were one week. Because of the high noncontact rates, we extended the field period to two weeks, and added weekend and evening calls, in order to reach representatives of small businesses who were difficult to contact.

Data collection consisted of two phases of interviewing: telephone screening and regular interviews. During telephone screening, interviewers attempted to reach each selected sample unit, verified that the unit was in fact the sampled business, and identified the appropriate respondent for the pilot study (the person responsible for payroll activities). The interviewer also asked whether the establishment had a fax machine, and if so, obtained the fax number. As a result of the screening process, units that were out of business or otherwise out of scope (e.g., hired domestic worker) were not contacted for the regular interview.

Over the 5 months, interviewers successfully reached and identified as eligible 1756 of 2248 cases. They were unable to locate 270 sample cases (12 percent) through contacts with the listed telephone numbers, calls to directory assistance, and other sources. Most of the "non-locatable" cases are probably not active businesses, but we had no way to verify this. Following screening, we faxed or mailed a letter describing the survey to the designated respondent. Interviewers began conducting telephone interviews with designated respondents within a week.[3]

Response. Response based on all eligible units identified during screening was 69 percent. Assuming that 90 percent of the non-locatable units and no answer units are really out of business, this rate rises to 86 percent. Eight-five percent of the successfully screened units completed the interview during the five-month period. Increasing the field period from one to two weeks made a big difference. The number of noncontacts because respondents were not available during the field period or did not answer the phone dropped from about 12 percent in July to 2 to 5 percent in the remaining months of the survey. Refusals were low, about 4 percent overall. Telephone screening was effective in identifying the correct respondent, as 92 percent of the respondents in the regular interviews were the people identified during screening (Goldenberg et al., 1997).

4. Estimates of population characteristics

Classification. Over the months from August to November, 73.5 percent of the cases were clearly identifiable as business births. These percentages range from a low of 65.8 in August to a high of 76.4 in October. The data are based on a different combination of states each month.

Among the units classified as births, more than half were new businesses. Another 25 to 31 percent were businesses that hired employees for the first time, and so became liable for taxes under UI laws. All other reasons accounted for 18 to 22 percent of births.

Birth dates. The onset of liability for UI taxes differs from state to state. Some states hold the employer liable when the first dollar has been paid in wages, while others defer liability until the federal minimum standard has been met ($1,500 paid during a calendar quarter). Based on this requirement, it seems reasonable to expect that most of the cases on the sample files have been in business for a relatively short time. In fact, it appears that about half of the units filed by the end of the sixth month. At the time of the interview, just over a fourth (26 percent) of the new businesses with employees had had those employees for 6 months or less. Another 10 percent each had employees for 7 or 8 months, with the median falling at the ninth month. Factoring in both the time for the states to transmit files to BLS and for BLS to process those files and select a sample, the 9 month median is probably just over 6 months of actual calendar time since onset of employment.[4]

On the other hand, it appears that there is a fair amount of noncompliance with state procedures, and as a result the states do not learn about many births as promptly as they should. Fewer than three-fourths of the employers in the sample files appeared on the EMI within a year of hiring employees, where we would have expected to see nearly all of them. An additional fourth of respondents' businesses were from 1 to 2 years old, while 3 percent had been employers more than two years prior to the date of the pilot survey. Apparently, even the EMI files miss a lot of business births.

5. Quality and timeliness of the birth frames

We have several dimensions on which to assess the quality of our birth frames apart from the pilot study. These include coverage of births, completeness of the information on the files, and timeliness—whether records on the birth frames reach us earlier than they would through normal CES procedures.

Coverage. We compared the month the businesses appeared on the birth frames with the initial appearance of those units on the CES sampling frame. From what we learned in the pilot test, more than half of the birth units from the July, August, and September files were in operation during the first quarter of the year, and so should appear on the CES frame for the first quarter. However, only 20 percent of them do so. Fifty-seven percent are present by the second quarter CES frame. Equally important, about 30 percent of the births on the CES frame for the states in question do not appear on the birth files for those states. This provides further evidence that EMI coverage is incomplete, again raising questions about its value as a sampling frame.

Completeness. Important data elements on the birth frames include company name, address, telephone, industry, and employment. As of April 1997, 45 states had transmitted at least one birth frame to BLS. A review of these files shows that name and address information are nearly complete. CES increasingly uses the telephone to enroll new sample units, and 73 percent of the records received had a phone number. However, 8 states provided files without phone numbers.

The primary purpose of the CES is to produce employment data by industry, which means accurate industry codes are essential. In addition, both industry and employment are used in CES sampling. Overall, 63 percent of the records had a useable industry code assigned. Industry codes were missing completely from 3 states' files, and 4 states coded fewer than 25 percent of the records. In terms of reported employment, 10 of the states did not report employment or reported it as zero. Another 5 states reported employment of 1 unless the employer indicated otherwise. Because of the importance of industry and employment in CES, these shortcomings again suggest serious limitations in using EMI files as a sampling frame.

Another important data element is a predecessor code. In continuing economic units, this is a designation that links the unit to another, now discontinued, UI account. Predecessor codes are assigned to some but not

all continuing economic units, with rules for supplying them based on state legal standards.

Thirteen percent of the birth file records showed some predecessor codes. However, 8 states reported no predecessors. This could be because they had not determined predecessor status at the time they sent the files, or because the states had eliminated records with predecessor codes despite our request to include them.

Sample interviews and the birth frames. The screening and pilot study interviews also offer insights into the quality of the birth frames. Do they represent active businesses? Over the course of the pilot study, we found 3 percent of the contacted sample units were out of business. Another 3 percent did not answer the phone or had a non-working phone. A smaller fraction were out of scope. Interviewers were unable to locate 12 percent of the sampled businesses using information provided on the files, even with the help of directory assistance. Since businesses generally need to be accessible in order to remain viable, it seems likely that most of these non-locatable cases are not active firms.

Of the cases still operating and eligible for the screening instrument, between 15 and 20 percent of the telephone numbers were no longer or had never been associated with the sample establishment. Ninety-nine percent of sample cases had business names supplied on the files, but according to respondents only half of those names were correct. Some names appear to have been truncated. Also, BLS originally requested one of two available names (legal or trade) and selected the legal name for sample files if both were available. It is possible that non-owner respondents are not familiar with their employer's legal name, and so reported that the business' correct name was incorrect. We addressed this problem later by requesting both names from all states. We were not able to determine whether the names were correct in a fifth of the cases, including non-locatable and noncontacted businesses.

As noted above, the number of records with predecessor codes was small. There should be very few predecessors among business births—and only for the case of partial acquisitions. Over the duration of the study, partial acquisitions ranged from 2.5 to 5.8 percent of business births, while the percentage of birth units with predecessor codes ranged from 1.8 to 11.3. Clearly, some birth units also had predecessor codes assigned.

A review of the October births with predecessor codes tells us more about the use of predecessor codes and about births in general. We found that some firms had UI accounts even though they are not liable for taxes, a situation that we did not anticipate. Apparently they reapplied for new accounts when they became liable. Two firms were spin-offs, a form of partial acquisition. Others were involved in corporate restructuring, which could also be viewed as partial acquisitions, although the questionnaire did not treat them that way. There were four cases that were classified as new businesses with no other explanation of why they might have a predecessor code. We believe the presence of predecessors may relate to the state becoming aware of the employer in more than one way and so assigning more than one UI account. Then, when a final determination is made on that business, one of the accounts might be assigned as a predecessor.

Timeliness. An important consideration in sampling for births in the manner proposed is whether using the EMI files as a sampling frame improves on coverage and especially on timeliness as compared to relying on the CES sampling frame. As an example, consider the birth frame files for new UI accounts assigned in August. Some states transmitted these files as early as the first week of September, but typically the files were not available to BLS until later in the month. Data from the August EMI files provided the sample for October BBPS interviews. In September, however, states were also sending BLS the first edition of the first quarter 1996 Quarterly Unemployment Insurance files, which becomes the CES sampling frame. The QUI files could include any of those employers that had established liability for January, February, or March, or even earlier. We know that more than half of the birth accounts on the August birth frame had hired employees and established liability prior to April, and so we would expect them to appear on the CES sampling frame for first quarter. The August birth frame also provides information on some of the new employers liable for taxes after the first quarter, as would all of the birth frames from April though July. We also know that only a small fraction of the August births are present on the August birth frame; most will appear later. Regardless of whether birth units appear on the sample frame, however, their existence will be reflected in the performance of the economy and in the UI benchmark source for CES for August.

Another way to view the birth frames and the first quarter CES frame is to consider what the January, February and March birth frames might have added to the CES frame, and when. We would expect half of the birth accounts on these frames to also be on the CES sampling frame for third quarter 1996, which was first available in April 1997. Another quarter of the records would appear three months later on the fourth quarter CES frame. Birth frames offer a limited head start on the CES sampling frame, but this is at a cost of quality in the frame, and still leaves the survey in need of an alternative method to determine a portion of employment.

Overall, the pilot study identified several limitations of the EMI as a sampling frame for business births.

There are problems of completeness and especially of timing. Many of the new accounts go to businesses that have had employees for months or even years before coming to the state's attention, and so are not truly "births" at the time we find them. In addition, our experience with predecessor codes demonstrates that we cannot assume such codes serve as an adequate indicator of the business as a continuing economic unit. Finally, it is necessary to differentiate between economic births and other changes in UI accounts, because these data showed that over a quarter of new UI accounts went to continuing economic units.

6. Comparing the birth frames with additions to the CES sampling frame

We compared birth frames for April, May and June 1996 to the UI records that make up the second quarter and first quarter 1996 sampling frame. Eighty-six percent of the records matched second quarter 1996 UI files. These records matched on many key characteristics: 74 percent on industry code; 71 percent on UI liability date; 91 percent on ownership; 53 percent exact match on name; 81 percent of records with predecessor UI accounts matched; and 38 percent on employment. (These matches exclude cases with missing values on the birth frames.) Match rates may appear low but the birth frames are often incomplete, whereas the UI records from the sampling frame are based on tax reports and have benefitted from additional research required to correctly establish most of these data.

Comparing the birth frame records to longitudinal records from the QUI, we find 54 percent of the second quarter birth frame records can be found in the first quarter of 1996 and another 25 percent in the second through fourth quarter of 1995. These administrative files indicate that births occurred months earlier than our birth frames suggest. There are several reasons for this. First, not all records on the birth frames are economic births. The BBPS showed that as many as a third of the records are continuing enterprises. Second, as the survey results demonstrate, employers register with their states, on average, about 6 to 7 months after starting their businesses. The files that constitute the CES sample frame are not closed until 6 to 9 months following the reference month of the birth frames, allowing the states to add these birth records and their employment to earlier reference months. The birth frames tell us the firms exist earlier than the CES sampling frame can, but still do not tells us of the births as they occur. As a result, sample based estimates from the birth frame do not synchronize with the CES sampling frame and the final historic CES employment series. And, of course, there is the possibility that the frames do not reflect all births.

If we take the CES sampling frame for first and second quarter and match UI accounts, the unmatched accounts could be defined as births. We cannot find about 30 percent of these births when comparing these accounts with the birth frames from April through November of 1996. One might suppose that some of these records would have been in birth frames that were not collected for January through March of 1996 or even earlier. This still leaves the possibility that the birth files are incomplete. To better understand the coverage problems of these files we must continue collecting them and compare them to CES sampling frames. Communication with the states that compile the files is also critical if we are to refine a process to create a useful birth sampling frame.

7. Collection Costs

Costs of conducting an on-going survey of new UI accounts can be separated into telephone costs for prescreening and for the main interview, and personnel costs for prescreening, interview preparation and conducting the main interview. There are additional fax and postage charges for written prenotification about the survey. There are also the costs of creating and maintaining the birth frames and selecting samples.

By October, data collection had stabilized sufficiently for us to approximate costs in a production environment. During this period, we made about 1255 phone calls for initial screening, with 57 percent of the cases requiring only one call and another 18 percent requiring 2 calls. On average, each of the 500 screened cases required 2.25 calls. Conducting the main interview took another 1500 calls, averaging 4 calls to each of the 375 successfully screened sample units. Each of the original 500 cases required an average of 5.25 calls, or 8.5 calls for each of the 327 completed interviews. Assuming 5 minutes for each completed screening interview, 7 minutes for each completed main interview, and a 1 minute charge for each of the other calls, at 10 cents a minute each of the 500 selected cases averaged $1.38 in phone charges. Using the same assumptions, each of the completed interviews averaged $2.10.

Staff time required to conduct the survey accounted for a much larger part of the survey costs: 131 hours to prescreen the 500 cases; 65 hours to prepare the 375 main interview cases; and 155 hours to conduct the main interview. It took a total of 351 staff hours to screen and collect data during the October field period. Time required to close each of the 500 cases averaged 42 minutes. A completed sample interview (327) averaged 68 minutes, factoring in the cost of screening the other 173 cases that did not result in completed interviews. At $35 per hour of staff time including benefits and overhead for a mix of staff, costs run $25.88 per

selected case or $41.76 per completed interview including the telephone expense.

A recent quarterly update of the CES sample frame from first quarter 1996 to second quarter 1996 yielded 65,000 new accounts each month. This figure is smaller than the actual number of new UI accounts that we would expect to see, because frame records are linked to the extent possible. The average number of new records per month from 45 states submitting birth frames sums to 74,000. If we were to sample at a rate of 1 in 40, the annual bill would come to a conservatively estimated $434,000. This figure does not include costs of on-going monthly collection for those units identified as births. Nor does it include the cost of states providing the frames, of BLS file maintenance, or of sample selection. State design simulations for CES have shown relative standard errors in birth estimates selected at these rates range from about 5 to 20 percent (BLS, 1997). Before setting up an infrastructure around these frames, we will investigate less costly alternatives to estimating birth employment that could provide data that are equally (or more) reliable.

8. Discussion

We began with a question about the feasibility of using an economic definition and a short telephone interview to identify business births. Pilot test results show that, in fact, the procedure is feasible. Indeed, the employer population seems receptive to BLS's need to collect data on new employers, demonstrated by an overall 85 percent response rate. The BBPS also provided us with a look at the cost of interviewing employers to determine their status.

Assuming reasonable costs for conducting an ongoing survey of new employers, will employment data on new businesses be earlier, more complete, or otherwise better than what we how have with the CES sampling frame? We learned from the pilot test that the answer is "probably not." The birth frames offer less of an early start in identifying births than we anticipated. The quality issues of the frames add to survey costs. Given the incomplete nature of the birth frames and other problems associated with either frame, both approaches require supplementing employment estimates. This leaves the value of collecting birth frames and sampling from them to compute a birth employment estimate in doubt. As an alternative, we are pursuing statistical modeling as a less expensive and equally reliable solution to the problem of obtaining employment from new businesses.

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

ACKNOWLEDGMENTS

Thanks to Jurgen Kropf and Jon Schomaker for matching birth frames and birth sample files with the UI files that define the CES sampling frame.

REFERENCES

American Statistical Association. 1994. *A Research Agenda to Guide and Improve the Current Employment Statistics Survey.* Panel for the Bureau of Labor Statistics; Current Employment Statistics Survey.

Bureau of Labor Statistics, Statistical Methods Division. 1997."Five Year Simulations of the Sample Design" for various states. BLS internal reports, January.

Goldenberg, K.. 1996. "Using Cognitive Testing in the Design of a Business Survey Questionnaire." *Proceedings of the Section on Survey Research Methods.* American Statistical Association.

Goldenberg, K., K. Levin, T. Hagerty, T. Shen, and D. Cantor. 1997, forthcoming, "Procedures for reducing measurement errors in establishment surveys." *Proceedings of the Section on Survey Research Methods.* American Statistical Association.

Levin, K., T. Shen, S. Heltemes, T. Hagerty, D. Cantor, and K. Goldenberg. 1996. *Business Births: Initial Evaluation of CATI Survey.* Report submitted to the Bureau of Labor Statistics, August. Rockville, MD: Westat. (Also other WESTAT reports for each month of the Business Birth Pilot Study.)

Subcommittee on Business Births and Deaths Research. 1994. *Report of the Subcommittee on Business Births and Deaths Research.* Bureau of Labor Statistics, December.

NOTES

[1] Now with the Federal Aviation Administration.

[2] Westat conducted all telephone interviewing for this survey under contract to BLS. All but the pretest interviews were made using Westat's in-house Computer Assisted Telephone Interview (CATI) system.

[3] See Goldenberg et al., 1997, for a more detailed discussion of telephone screening and respondent contact procedures.

[4] As an example, the August files arrived at BLS by the end of September. We produced a sample for data collection early in October.