Assessing the Effect of Government Frame Refinement on Collecting Establishment Data for the National Compensation Survey December 2007

Glenn Springer
Bureau of Labor Statistics

Abstract

Non-response error is a common but undesirable feature of a survey. Survey practitioners use various techniques to reduce non-response error. In the National Compensation Survey (NCS), the state and local government frame is developed using administrative files maintained by the states for unemployment insurance (UI). In this paper, we explore the extent to which a frame refinement process, used to reconfigure UI reporting units so that they conform to how payroll records are kept in an establishment, improved the ability to collect data for the establishment as sampled, thereby improving response as well as reducing respondent burden and the need for post-collection sample-weight adjustment. An overview of the refinement process, the system tool developed for refinement, and several approaches to measuring the impact of frame refinement on non-response will be explored.

Keywords: sampling frame, non-response error, establishment survey

1. Introduction

The National Compensation Survey (NCS) is an area-based, multi-stage, establishment survey conducted by the Bureau of Labor Statistics (BLS) within the Office of Compensation and Working Conditions (OCWC). The survey publishes industry-specific, occupational compensation estimates for the U.S. as well as for large metropolitan statistical areas, some smaller metropolitan areas, and some non-met counties. The NCS sample is selected using a three-stage stratified design with probability-proportionate-to-employment sampling at each stage. To summarize the three-stage, stratified, probability proportionate to employment size (PPS) sample design, the first stage involves selecting survey areas (PSU's) with probability proportionate to employment. Within each survey area, an establishment sample is selected, also with probability proportionate to employment, to complete the second stage. Finally, for the third stage of selection, occupations are selected with probability proportionate to employment from each sampled establishment. For a detailed description of the NCS design see Izsak et al. (2005). The focus of this study is on the second and third stage of selection, which is the establishment and occupation sampling processes.

1.1 Sampling Establishments in the NCS

The selection of the establishment sample within each survey area uses a stratified, PPS approach. Sampling cells are defined by industry, using the North American Industry Classification System (NAICS). There is also implicit stratification by establishment size because sampling within industry cells uses a PPS systematic approach in which each frame unit's measure of size is determined by its employment. The source for this information is the Bureau's Longitudinal Database (LDB). The LDB is maintained as part of a Federal/State Cooperative Program within the BLS. This Program, the Quarterly Census of Employment and Wages (QCEW), provides for the aggregation of state-maintained, unemployment insurance (UI) files. The resulting aggregated file serves as a sampling frame for many of BLS' establishment surveys.

Although a portion of the NCS private establishment sample is replenished each year, the government establishment sample is selected, in its entirety, approximately once every ten years. This schedule coincides with OMB's redefinition of metropolitan statistical areas, following the results of the decennial census. The NCS reselects survey areas (PSU's) to reflect OMB's redefined statistical areas and a government establishment sample is selected within the new areas. Currently the NCS program surveys 152 areas which correspond to the areas selected to reflect the OMB's 2003 redefined metropolitan statistical areas. In general, selecting the government establishment sample does not follow the annual replenishment schedule of the private sample because the government sector is considered more stable in terms of new establishments coming into existence or establishments going out of business. Also, response rates are higher within the government sector than within the private sector. For this reason, it is not necessary to issue replenishment sample groups to account for attrition due to respondents dropping out of the survey. Additionally, staffing levels are considered more stable within the government sector and the need

to capture changes in occupational mix over time is not as prevalent.

2. Assessing the Need for Refinement

As previously stated, the sample design at the establishment level is a stratified, PPS design. The sampling strata within each survey area are defined by industry using NAICS-coded information from the Bureau's LDB. These industry strata are defined according to planned estimation cells which coincide with NCS' industry publication plans. For the state and local government sector, these published industries are as follows:

Goods-Producing (Sectors 21, 23, 31-33);

Utilities (Sector 22);

Wholesale & Retail Trade (Sectors 42-45);

Transportation & Warehousing (Sectors 48-49);

Elementary and Secondary Schools
          (Industry Group 6111);

Junior Colleges (Industry Group 6112);

Colleges, Universities, & Professional Schools
          (Industry Group 6113);

Rest of Education (Sector 61 excl. 6111-6113);

Hospitals (Sub sector 622);

Nursing Homes (Sub sector 623);

Rest of Health & Social Services
          (Sector 62 excl. 622-623);

Public Administration
          (Sector 92 excluding sub-sector 928);

Rest of Service-Producing
          (Sectors 51-56, 71-81 excl. 814);

Because the government establishment sample had not been replenished for approximately ten years, a study was conducted to assess current reporting practices within the Bureau's LDB and the extent to which these reporting practices would provide an adequate sampling frame for NCS' state and local government survey. For detailed results from this study see Schildkraut et al. (2004). The study involved reviewing UI reporting practices for fifteen major cities, all of which represented current NCS survey areas. Results from this study indicated that UI reporting within these cities, for both state and local government, was characterized by large numbers of disaggregated reporting units. In many cases identifying information for single reporting units was inadequate for the purpose of the unit of collection, that is to say, reporting units differed only by changes in several identification numbers whereas, physical identification information, such as address or company name, was the same across reporting units. This was true particularly among units associated with state government university systems. Several state university systems reported hundreds of units for a given system within a survey city, where reporting practices were along the lines described above.

Another observation, provided by the frame analysis, was that in many cases city-reporting units for several large cities were very aggregated in their reporting practices. Several city-reporting units reported for over 10,000 employees on one UI record. As shown above, published industry breakouts would necessitate identifying separate frame units within these city government systems. For example, for a given city government system, employees working in the utilities sector would be represented by one frame unit, whereas employees working in the administrative sector would be represented by another frame unit, thereby allowing for frame units to be stratified by industry, according to the breakouts provided above.

Based on the results summarized above, the study concluded that frame refinement should be considered for the new government sample.

3. Frame Refinement Systems and Procedures

After assessing the reporting patterns within the state and local government sector and determining that frame refinement would provide a more collectible set of establishments which were coded according to NCS' planned sampling and estimation cells, a cross-functional team was formed to develop and initiate refinement procedures. Members included representatives from the statistical methods group as well as systems staff, program office representatives and regional office staff.

The three major tasks of the team were to develop a set of refinement procedures, to develop an on-line database system to capture and edit changes which were made to UI reporting units during refinement, and to conduct training on the procedures for regional office field economists. The final product would be a refined establishment frame file, for each NCS survey

area, which could be used by the statistical methods group to select the establishment sample within each survey area.

In developing a set of refinement procedures, the team recognized the need for an objective, clear establishment definition for the state and local government sector. Because this definition would serve as the basis for the field economists' decision-making during refinement and would guide the way in which UI reporting units may be reconfigured, it was considered a critical part of the procedures. The establishment definition along with applied refinement guidelines and procedures is provided below.

<div align="center">

Establishment Definition
State and Local Government Sector

</div>

A state or local government establishment is defined as an agency or entity such as a school district, college, university, hospital, nursing home, administrative body, court, police department, fire department, health or social service operation, highway maintenance operation, urban transit operation, or other governmental unit. It provides services under the authority of a specific state or local government organization within a defined geographic area or jurisdiction. Unlike in the private sector, physical location is not necessarily a key characteristic used in defining an establishment.

The UI file serves as the basis for the government sampling frame, although UI reporting practices are sometimes not consistent with the establishment definition. It may be necessary to reconfigure UI reporting units for collection considerations (see rule 3 below), or with regard to planned publication breakouts by NAICS (see industry breakouts below) or survey area/sub-area. Within the National Compensation Survey, there are 27 survey areas which correspond to core-based statistical areas, designated as Combined Statistical Areas by OMB (see OMB reference for defining Combined Statistical Areas). For these areas we would like to publish sub-area estimates, if we have sufficient data to do so. These sub-areas are defined according to the metropolitan or micropolitan area definitions for the CSA (Combined Statistical Area).

The following rules should be used when reconfiguring UI reporting units:

1.) UI reporting units that collectively report employees for a specific government establishment should be combined (clustered) to form one unit according to the establishment definition. Clustering should be consistent with planned publication breakouts by NAICS (see industry breakouts provided in section 2 above) and survey area. An example of this is faculty departments on the same campus reporting separately for a university. The refined unit must be collectible and updatable.

2.) A UI reporting unit that reports employees across several industry breakouts (listed in section 2 above) or across several survey areas, including sub-areas, should be split to form units consistent with industry and area publication plans. An example of this is a city government UI reporting unit that combines social services with administrative offices. The refined units must be collectible and updatable.

3.) Decisions to reconfigure UI reporting units may also hinge on what constitutes a collectible, updatable unit. As a last resort it may be necessary to deviate from the establishment definition, as well as planned industry and survey area publication cells, in forming units for which payroll records exist. Here are two examples, both of which involve a State Government that has a number of departments (such as Departments of Conservation, Revenue, Social Services, Transportation, etc.) operating within an NCS area.

> Example 1: State Government operations are reported as one U.I. unit in an area. It is known that data <u>must</u> be collected separately by individual departments of the State, and a different respondent exists for each department. Since the units must be collectable and updatable, reconfigure as necessary.

> Example 2: State Government operations are reported as separate U.I. units for each department. It is known that all data are available from one central source. Although possible to collect data separately, it is highly inefficient to repeat entries of the same benefit data, company data, and documentation. Since the sample units are collectable and updatable at one central source, reconfigure as appropriate within NAICS and sub-area requirements.

As the team was developing an establishment definition along with applied refinement guidelines and procedures, we also developed a system tool to facilitate the refinement process, which primarily involved clustering or splitting UI reporting units. After these two tasks were complete, the team developed and delivered, using the Bureau's remote

video-conferencing facilities, training materials and exercises to several representatives from each of the Bureau's six regional offices.

4. Frame Refinement and Sampling Efficiencies

After the regional offices completed refinement we were able to assess the degree to which UI reporting units were reconfigured. Although the refinement procedures allowed for both combining and splitting UI reporting units according to the establishment definition and implied industry classifications, the overriding effect of the refinement was to reduce the number of units in the area frames. As shown in the table below, due to clustering UI reporting units, the frame unit count was substantially reduced for both state and local government. Review of the refined units confirmed that they followed guidelines provided by the establishment definition. The industry in which most of the clustering occurred was the educational services sector. It was through this clustering that the final establishment sampling frame represented a list of frame units that recognized a natural grouping within the target population (see Williams reference). Sampling from this frame would be substantially more efficient than from a frame where this natural grouping was not recognized.

In terms of the slight differences between the unrefined and refined employment for state and local government, this is a result of isolated cases in which the regions updated employment levels or deleted reporting units which were determined to be outside the scope of the survey.

Decrease in number of frame units and employment for the 152 NCS refined area frames.

	Unrefined Frame	Refined Frame
State Govt		
Units	30,684	7,659
Employment	2,839,917	2,863,514
Local Govt		
Units	71,037	36,036
Employment	9,164,037	9,156,107
Total		
Units	101,721	43,695
Employment	12,003,954	12,019,621

5. Collection Issues

Although refinement resulted in frame units that recognized natural clustering among units in the target population, and these refined units conformed to guidelines provided in the establishment definition, we still experienced challenges during collection. For example, in some cases where a UI reporting unit, representing a large city government, was disaggregated according to industry stratification, the field economist was unable to isolate the sampled disaggregated units for collection. In cases where UI reporting units were clustered there were isolated incidents in which the field economist could not collect the assigned unit because combined payroll records did not exist, thereby hindering the efficient collection of clustered units, or the field economist did have information regarding what constituted the assigned unit.

6. Post-Collection Sample Weight Adjustment

In our attempt to assess the effect of frame refinement on collecting establishment data for the NCS, we looked at several areas related to post-collection weight adjustment. Our analysis focused on the extent to which collected employment data, at the establishment level, differed from sampled/assigned establishment employment.

Using current survey data, from our recent government collection experience, we compared survey results based on reconfigured establishment frame units from a refined establishment frame, to survey results based on establishment frame units, not reconfigured during frame refinement, from the same establishment frame.

Using less recent collection experience from government surveys in which no frame refinement was conducted, we compared survey results based on unrefined frames to results from our refined frames. In making these comparisons we were careful to isolate collection experience associated with the initial collection of sampled establishments. Because establishment employment tends to increase over time we did not want this type of change, observed in subsequent visits during update, to be associated with any inability to collect the assigned unit.

6.1 Comparing Results Using Current Survey Data from a Refined Government Frame

As described in Section 1.1, the government establishment sample is selected every ten years. We began the initiation of the new sample during the July-September quarter of 2006. Initiating the sample involves establishing contacts or respondents for each

sampled establishment. Prior to obtaining the necessary occupational compensation information, the third-stage selection of occupations must be completed. This involves establishing an employee list for the sampled establishment from which a PPS, systematic sample of occupations can be selected. The measure of size for this stage of selection is the number of employees within an occupation. More populated jobs within the establishment have a higher probability of selection. The sample weights at this stage of selection are the inverse of the occupation's selection probability. After the field-economist establishes an employee list and selects the occupations, occupational compensation data are then collected for the employees in the selected occupations. Introducing or initiating one-fourth of the government sample each quarter, the entire government sample will have been initiated by the end of the April-June quarter of 2007.

Using current survey data from the first three quarters, we can assess the degree to which field economists were able to correctly identify and collect the assigned government unit. Within the first three quarters we have initiated approximately 3,300 establishments out of the 4,400 establishment constituting the total government establishment sample. More than 1,100 of the 3,300 initiated units involved reconfigured frame units which were modified during the frame refinement process. The balance, approximately 2,100 establishments, corresponded to units which were not modified during refinement. That is to say their assigned employment corresponded to what was originally reported to the UI.

Using documentation recorded by the field economist regarding to what extent they were able to collect the assigned unit, we can compare their collection experience for the units reconfigured during refinement to that of the units which were not reconfigured during refinement and make inferences about the degree to which frame refinement improved our collection experience. Among the 2,148 establishments which were not reconfigured, we were able to identify only 37 cases, representing 1.7% of the non-reconfigured units, for which the collected employment represented more or less than the employment of the sampled unit. Among the 1,119 reconfigured units, we were able to identify only 59 cases, representing 5.3% of the reconfigured units, for which the collected employment represented more or less than the employment of the sampled unit. Although this represents a slightly higher percentage of units for which a post-collection weight adjustment would be made, we must keep in mind that all of the refined units represent units that were in some way

reconfigured (clustered or split) during refinement. That is to say that during refinement, UI reporting units were either combined or split to form a unit which could more likely be collected. Had these reporting units not been reconfigured, collection of the assigned units would have been less likely. As discussed in Section 2, the reporting practice among education units tended to be very disaggregated, with little identifying information available to distinguish the units for collection. As discussed in Section 4, the number of establishments on the refined frame was substantially reduced during refinement when disaggregated reporting units were clustered to form one frame unit.

In reviewing the 59 schedules, for which the assigned, reconfigured frame unit was not collected, we found many cases in which the field economist was not aware of how the original frame unit(s) had been reconfigured during refinement. In some of these cases the respondent provided data for the sampled location, corresponding to the way in which the unit originally reported to the UI, thereby accounting for only part of the assigned unit. Throughout the initiation of the government sample we have emphasized the need to reference changes made during refinement as they relate to the assigned establishment. As this source of information becomes more familiar among our collection staff we expect our ability to correctly identify the assigned unit to improve.

It bears mentioning that we did not experience a higher loss rate associated with refusals among the refined units. Among the unrefined establishments, 94% of the assigned units were collected. Among the refined, reconfigured establishments, 95% of the assigned units were collected.

6.2 Comparing Results Using Previous Unrefined Government Frames

After comparing our collection experience among units from our current refined frame, we looked at the degree to which assigned, sampled units from prior, unrefined frames were correctly identified and collected. We again used assigned employment and collected employment data for our analysis.

Summarizing data from an earlier unrefined state and local government sample for the Employment Cost Index (ECI), there were 221 schedules in which the reported employment differed from the assigned employment by 100 or more and the difference represented at least 50% of the assigned employment. Because this comparison was made during the initiation phase of the survey, we attribute such

substantial differences to problems in identifying and collecting the assigned unit. During the initiation phase, differences of this magnitude, between assigned and reported employment, would not be attributable to expected changes in employment over time due to expansion or contraction. Additionally, employment changes over time, for the government sector, tend to be less pronounced. These 221 schedules represented over 25% of the total 797 government index schedules that were analyzed. Among the 1,039 schedules that comprised the Occupational Compensation Survey Program, a forerunner to the integrated NCS Program, there were 112 schedules for which the reported employment differed from the assigned employment by 100 or more and this difference represented 50% or more of the assigned employment. Combining the two sources we have 333 out of 1,836 schedules, or 18%, for which there were substantial differences between assigned and reported employment.

For comparison, we summarized the collection experience for sample units from our current, refined frame. From nearly twice as many total schedules we found just over one-half as many schedules with substantial employment differences, as described above. Among the 3,267 sample units selected from the refined frame there were only 189 (6%) for which a substantial employment difference was reported. We also want to note that this comparison assumes that implied design changes with subsequent samples have a negligible effect on collectibility.

7. Future Research

7.1 NCS' Rotating Panel Design

Because the NCS uses a rotating panel design, it takes approximately one year to initiate or introduce a new establishment sample into the program. The most recent government sample was introduced in the summer of 2006. The phase-in of the sample will not be complete until the end of June 2007. For this reason, results of our frame refinement experience will be somewhat preliminary until we are able to analyze all of our collection experience, associated with the latest government sample, sometime later in the year.

7.2 BLS Program Integration

Research is currently underway to explore the possibility of coordinating the Bureau's NCS Program with another BLS program, namely the Occupational Employment Statistics Program (OES), which publishes detailed occupational wage and employment data. Because the OES samples directly from the LDB, with little establishment frame refinement, further work will be needed to determine if a coordinated program could benefit from NCS' frame refinement approach for the government sector.

8. Conclusions

Results of the frame refinement for the state and local government surveys indicate that sample unit identification improved as result of refinement. In general, substantial differences between employment for assigned sample units and reported units were less predominant in the refined frames compared to the unrefined frames. Although the frequency of post-collection sample weight adjustment was similar among the refined and unrefined units, there is some indication that improvements in communication could reduce weight adjustment among the refined units. In many cases in which the field economist was unaware of what constituted the refined unit, communication with the refining analyst would have improved the field economist's ability to collect the assigned unit. In general, the refinement was completed in a reasonable timeframe and information gained during refinement, particularly among the large government systems, improved field staff's familiarity with the reporting units.

Admittedly, any decision to refine frames would have to weigh the benefit of improved estimates against the cost, in staff time, required to do the refinement. Given the reporting patterns associated with the state and local government sector, frame analysis should probably be considered for future government samples. Using experience gained from past refinement, future government frames will be more accurate as well as more focused to the needs of the NCS Program.

References

Izsak, Y., Ernst, L.R., McNulty, E., Paben, S.P., Ponikowski, C.H., Springer, G., Tehonica, J. (2005). Update on the Redesign of the National Compensation Survey. 2005 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

Schildkraut, J., Sunday, K. (2004). State and Local Government Frame Analysis. Internal BLS Working Document.

Williams, B. (1978). A Sampler on Sampling. John Wiley & Sons, Inc. New York.

For details on the 1990 census area definitions, see Federal Register (55 FR 12154 – 12160, March 30, 1990); or OMB Bulletin No. 95-04

The criteria for defining Metropolitan, Micropolitan, and Combined Statistical Areas are published in the Federal Register (65 FR 82228 – 82238, December 27, 2000

The criteria for defining Core Based Statistical Areas are published in the Federal Register (65 FR 82228 – 82238, December 27, 2000)