EVALUATION OF UNIQUE ASPECTS OF THE SAMPLE DESIGN FOR THE NATIONAL COMPENSATION SURVEY

Lawrence R. Ernst, Christopher J. Guciardo, and Yoel Izsak
Ernst.Lawrence@bls.gov, Guciardo.Chris@bls.gov, Izsak.Yoel@bls.gov
Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3160, Washington, DC 20212-0001

KEY WORDS: Rotating panel design, PSUs, PPS sampling

1. Introduction

We describe in this paper two aspects of the National Compensation Survey (NCS) sampling design that are atypical and perhaps unique to this survey program, and evaluate them through empirical studies.

Three of the Bureau of Labor Statistics (BLS) compensation survey programs, the Employment Cost Index (ECI), the Employee Benefits Survey (EBS), and the locality wage surveys, were recently integrated, creating one comprehensive program, NCS. The ECI publishes national indexes that track quarterly and annual changes in employers' labor costs, both wages and benefits, and also cost level information, previously annually but now quarterly, on the cost per hour worked of each component of compensation. The cost level portion of this program is known as the Employer Costs for Employee Compensation (ECEC). The former EBS publishes annual incidence and detailed provisions of selected employee benefit plans and is now known as the Benefits Incidence and Provisions Product. The locality wage surveys program publishes locality and national occupational wage data. There are two samples used in the integrated NCS program, a larger wage sample, used for those survey products that only present wage data, and a smaller wage and benefits subsample used for products that include benefits data. Unless otherwise stated, the sample referred to in this paper is the wage sample.

The integrated NCS sample, at any point in time, consists of five rotating replacement sample panels. Each of the sample panels are in sample for five years before being replaced by a new panel selected annually from the most current frame. The NCS sample is selected using a three-stage stratified design with probability proportionate to employment sampling at each stage. The first stage of sample selection is a sample of areas (PSUs). The second stage is a sample of establishments within sampled areas, with the frame of establishments presently partitioned into 23 industry sampling strata. The third stage is a sample of occupations within sampled areas and establishments. Currently, government and aerospace manufacturing establishments, although included in the survey estimates, are not part of the rotating panel design and are not studied in this paper.

The following are the two aspects of the NCS design studied in this paper. First, although for most of the frame one fifth of the sample establishments are replaced each year, a set of units, which can be thought of as five-panel certainty establishments, are selected only once every five years using a smaller sampling interval than used to select the remainder of the sample. This is done in part to eliminate the contribution of these units to the between establishments, within PSUs component of the variance. The second issue studied is that many of the sample PSUs are clustered together and the sample establishments selected across the cluster as a whole, rather than in each PSU independently. This is done principally to avoid the inefficiencies that would result if the second stage of sampling was done independently in each sample area $\times$ industry stratum cell, since even if the minimum number of sample establishments was set to 1 for each such cell, a large proportion of the sample would be used up in meeting these minimums. This would be particularly true if the expected number of sample units in many of the cells is much closer to zero than to one based on the measure of size used. The paper discusses these two aspects of the design in further detail and evaluates their impact compared to alternative sampling methods through empirical studies.

In Section 2 we outline the approach to be taken in the empirical studies. The five-panel certainty issue is discussed and evaluated in Section 3, and the clustering of sample PSUs is considered in Section 4.

2. Methodology

The sampling frame for the NCS, like many of BLS surveys, is the Longitudinal Database (LDB), a quarterly universe of establishment data obtained from state unemployment insurance records. The data items on this frame for each establishment include, in addition to other items, an industry code, the total establishment quarterly wages, and the monthly employment for each of the three months of the quarter. This data allows us to compute mean quarterly wages per employee for any establishment domain.

In each of the topics considered in this paper, our empirical study consists of: selecting a number of samples directly from the LDB using two or more sampling procedures; computing estimates for each sample chosen for each procedure; and then calculating the root mean square error (RMSE) of the estimates across runs for each procedure, which should be an accurate estimate of the true standard error. The samples selected are restricted to the current 154 sample PSUs. More specifically, we calculate an RMSE by first computing the squared difference of each sample estimate and the area-weighted frame estimate (obtained from the frame totals for each of the 154 sample PSUs multiplied by the reciprocal of the probability of selecting the PSU), and then taking the square root of the mean of these squared differences over the 100 samples. The advantage of computing the RMSEs using samples selected directly from the frame is that the comparisons cannot be distorted by problems with the variance estimator, which could be the case when the comparisons are based on a single sample using our standard variance estimator. This is particularly true in the case of NCS, because we calculate variance estimates under the assumption that the sampling of establishments is done independently in each of the 154 sample PSUs, which, as discussed in Section 4, is not the case. This variance estimation issue as it relates to NCS is discussed in detail in Guciardo, Dorfman, Ernst, and Sverchkov (2004) and Sverchkov, Dorfman, Ernst, and Guciardo (2004).

A potential disadvantage of the approach used in the paper to calculate RMSEs is that it captures only the component of variance arising from the second of the three stages of sampling mentioned in the Introduction, that is the selection of the sample establishments. The component of variance arising from the final stage of sampling, the selection of occupations, cannot be captured in any study that only uses LDB data, since there is no occupational information on the LDB. The component arising from the first stage, the selection of areas, was not captured in our simulation study either, because each sample was selected from the same 154 areas and the squared differences involved the area-weighted frame estimates from these 154 areas. There are several ways that we could have captured the first-stage component, but we were not particularly interested in capturing either the between areas or within establishments component of variance because the methods compared in this study only differed in the way the establishments were selected and hence our key interest was in the impact on the between establishments, within areas component of variance. In fact, none of the methods studied impact the between areas component of variance. However, since they do affect the within areas, between establishments component of variance, they also indirectly affect the within establishments component of variance because any change in establishment selection probabilities affects this component.

3. Certainty Establishments

Typically in a rotating panel design with the sample in each panel selected PPS, the sample panels are selected independently with certainty units determined for each sampling cell using the sampling interval appropriate for the cell's sample size. In the case of NCS, where the design consists of five rotating panels with 1/5 of the sample replaced each year, the sampling interval for a cell would then be approximately 1/5 as large as it would be if instead of using a rotating panel design, a sample five times larger had been chosen with the entire sample replaced every five years. It was felt that the smaller certainty cutoffs that would occur in such a non-rotating sample design would lower the sample variance, or at least the between establishments, within areas component of this variance, in part because there would be a larger number of establishments that would not contribute to this component of variance since they would be certainty units.

As a result we decided to use a sample design that is a hybrid of a rotating and non-rotating design as follows. The total number of sample establishments is assumed fixed at n at any point in time. Every five years a sample of size n is selected, stratified PPS within each area cluster, where the strata are the industry sampling strata and the measure of size for each establishment is its area-weighted frame employment. Only the set of certainty units among them, which comprise say m establishments, are retained in the sample. These units are placed in that year's sample panel and the sample panels for the next four years. We consequently refer to these units as five-panel certainty units. Since each panel remains in sample for five years, each certainty unit supplies data to the estimates for at least nine years. The remaining sample units are selected annually from a sampling frame consisting of all the LDB units minus the five-panel certainty units. These units are referred to as single-panel units. There would be $(n-m)/5$ single-panel units selected in each sample panel. Consequently, each sample panel would have $(n+4m)/5$ units. However, there are at most n distinct units in sample from these five panels since the five-panel units are in each of these panels. (Actually, since generally at any point in time there are two sets of five-panel units in sample, there could be as many as $n+m$ distinct units in sample if each of these sets consisted of m units, although it is likely that there would be a great deal of overlap between these two sets of five-panel units.)

There are some drawbacks to this approach, in particular the differential time-in-sample effect due to the fact that the five-panel units are of different average age than the single-panel units since they remain in sample longer, nine years versus five years. As a result, unless the sample occupations are reselected during the nine year period, the sample occupations for the five-panel units would eventually become older on average than the sample occupations for the single-panel units and the five-panel units would consequently be slower to reflect changes in occupational distribution over time. However, this problem of differential age of sample units would occur under the current operational procedures even if all units were selected independently from sampling panel to sampling panel. This is because if the same unit is selected more than once in nearby sample panels, we would not choose new occupations because of respondent burden implications, but instead use the initial set of occupations in each sample panel for which the establishment is selected. Since it is more likely that larger establishments would be selected more than once in five years than smaller establishments, it is more likely that the sample occupations in larger sample establishments would have a greater average time in sample. This differential age problem is exasperated, however, by the five-panel certainty units aspect of the sample design, since it increases the expected number of units that are in sample in multiple panels. There are tentative plans for selecting a new sample of occupations for the five-panel certainty units at perhaps five year intervals to alleviate this problem.

In this section we compare variances of estimates corresponding to three different methods of selection of the sample panels. To simplify the work, we used only a single frame for Methods 3A and 3B below, as opposed to five frames, 12 months apart, used in production, and for these two methods we selected five independent samples from this single panel instead of selecting one sample from each of five panels. The three methods compared are:

3A. Select a sample of size n, stratified PPS. Retain only the certainty units among these units, say m in number. Then select 5 independent samples of size $(n-m)/5$, stratified PPS, from the remaining units on the frame.

3B. Select 5 independent samples of size $n/5$, stratified PPS, from the frame.

3C. All units are selected as a single large sample of size n.

Method 3A is the approach we are currently using. Method 3B is the more customary approach of sample selection for a rotating panel design. Method 3C is appropriate if the design is not a rotating panel design, that is if instead a single sample is used for a fixed length of time and is then completely replaced. Note that in addition to the methods differing in the way they determine certainty establishments, Method 3C selects establishments completely without replacement over a five year period, which may tend to produce lower variances than the other two methods for which the same establishment can be selected up to five times during the same period.

To evaluate the three approaches, 100 samples were selected for each approach. The target n is 37,286, which was the production sample size for the private sector excluding aerospace manufacturing prior to a recent sample cut. As noted in the Introduction, this is what we are restricting our frame to in this work. The sample was allocated among 54 area clusters $\times$ 23 industry strata using the method described in Ernst, Guciardo, Ponikowski, and Tehonica (2002). Some aspects of the allocation, in particular the use of 54 area clusters for sampling rather than an independent selection in all 154 areas, are discussed in the next section of this paper. Roughly, in allocating the sample among these 54 clusters, 36 of these clusters corresponded to individual certainty metropolitan areas or combined statistical areas, with most of these areas oversampled compared to the other areas. This serves to meet the needs of the Pay Agent, a key NCS customer who uses NCS data to recommend locality pay increases for federal workers. For each industry stratum, the remaining sample was allocated among the remaining clusters proportional to the area-weighted frame employment of all establishments in the cluster on the LDB. (In actuality, NCS uses modified frame employment in production instead of frame employment in calculating the measure of size. Modified frame employment differs from frame employment in only one way. The modified frame employment is one for establishments with frame employment of zero, in order to give a chance of selection to establishments that did not have any employees at the time of the frame formation but that do have employees at the time of the initial data collection. Since all the data in the study comes only from the frame, we used the unmodified frame employment.)

The frame used in this study is a more recent frame than the one used to select the current production set of five-panel units and consequently the set of five-panel units selected in Method 3A is not identical to the production set.

For each method, empirical RMSEs were obtained by selecting 100 samples and computing the RMSEs of employment and mean quarterly establishment wages for the following domains: each of the 23 industry sampling strata, each of the 18 geographic domains defined by the nine census divisions $\times$ {metropolitan, non-metropolitan classification}, and each of seven size classes. In the first through third numerical columns of Table 1, the relative RMSEs (RRMSEs) are presented for each of the three methods for total employment for the geographic domains only. For the same columns of Table 2, the RRMSEs are presented for the same domains for mean quarterly wages. The other domains are omitted from these tables partly due to space limitations and partly due to confidentiality issues. (The estimated RRMSEs were calculated as follows. First RMSEs were calculated from the 100 samples as described in Section 2. These RMSEs were then divided by the area-weighted frame estimate, with the quotient expressed as a percent.)

Note that total employment is needed as the measure of size variable used in selection of sample establishments and also as an estimation variable. For the former use, we take the frame employment in the third month of the quarter, and for the latter use, the mean of the three monthly employment figures. If we had used the same employment figures for both purposes, then the contribution to the estimated RMSE of total employment from any area cluster $\times$ sampling industry for which the cluster only consisted of one PSU would be zero. As explained in the next section, this applies to 52 of the 54 area clusters. Even though the two different employment figures used in this study are not identical, they are much closer than the corresponding figures in NCS production, partly because the employment used in estimation in production, obtained from the respondent, is often for a much later point in time than the employment used in sample selection and partly because the employment used in estimation excludes certain workers included in the frame employment.

Our primary interest is in how the RMSEs for Methods 3A and 3B compare. As listed in Table 3, Method 3B yielded lower RMSEs for total employment for 7 of the 23 industry domains, 4 of the 18 geographic domains and none of the 7 size class domains. So in each of these three sets of domains, Method 3A yielded lower RMSEs for the majority of the domains, as anticipated. Based on a two-sided sign test, the *p*-values for these three comparisons are 0.093, 0.031, and 0.016, respectively, as listed in Table 4. However for mean wages, the RMSEs for Method 3B were lower than for Method 3A, for 12 of the industry domains, 5 of the geographic domains and 1 of the size class domains, with higher *p*-values of 1.000, 0.096, and 0.125, respectively. Comparing Methods 3A and 3C, where we anticipated lower RMSES for Method 3C, this method did yield lower RMSEs for total employment for 13 industry, 17 geographic, and 7 size class domains, while for mean wage Method 3C yielded lower RMSEs for 12 industry, 14 geographic, and 5 size class domains. The corresponding *p*-values are 0.678, 0.000, and 0.016 for total employment, and 1.000, 0.031, and 0.453 for mean wages.

In general, the results appear to be more consistent with what we anticipated for total employment than for mean wages, particularly for size class estimates. The fact that the RMSEs of total employment were lower for Method 3C than for Method 3A is not surprising. As mentioned previously, lower RMSEs were anticipated for Method 3C than for Method 3A because an establishment can be selected once over five years for Method 3C and up to five times for Method 3A. In addition, the selection of establishments within an area cluster × industry sampling cell is systematic PPS, with frame employment as one of the sort variables. This sorting serves to produce some control over the allocation within each cell by size class, a control which is more effective in reducing variances for estimates of total employment for Method 3C than for Method 3A because of the larger sample size per sample panel for Method 3C. For example, if the sample size in a cell in a sample panel is one unit, which occurs much more often with Method 3A than Method 3C, there is no control at all in the allocation by size class.

4. Sampling Across Strata

As noted previously, there are 154 sample areas × 23 industry sampling strata in this study. The total number of sampling cells would then be 154×23=3,542 if we sampled independently in each sample area × industry sampling stratum cell. If we sampled independently in each of these cells with a minimum of one unit in each nonempty cell, then the total number of sample units to meet the minimums would be, provided there are few empty cells, more than 1/2 of the total sample of 6,876 units in each single-panel wage sample selected using Method 3A. We did not study the wage and benefits sample in this paper. However, the current single-panel wage and benefits sample size of 2,876 in production is actually less than the number of sample area × industry stratum cells. Thus, sampling independently in each of these cells, which we designate as **option 1**, results in an inefficient design for the wage sample, since for this option there is a relatively little sample remaining to allocate to the large cells after meeting the minimum allocation for each cell. Also it is an impossible design for the wage and benefits sample if there are more than 2,876 nonempty cells, unless the size of this sample is increased. The following are three alternatives options:

2. Collapse industry strata together for sampling purposes to reduce the number of cells.

3. Collapse area PSUs together for sampling purposes to reduce the number of cells.

4. Use a two-dimensional controlled selection procedure (Causey, Cox, and Ernst 1985), which guarantees that:

> The number of sample units in each sample area, in each industry stratum, and in each sample area × industry stratum cell is within one of the desired number for every possible sample. (4.1)
>
> The expected number of sample units in each of the domains listed in (4.1) over all possible samples is the desired number. (4.2)

By the “desired number” for each of these domains, we mean the sum of the area-weighted frame employments over all units in the domain divided by the same sum over all units in all sampling cells, with this quotient then multiplied by the total sample size, that is a proportional to size allocation.

Currently we use what might be considered a hybrid of options 1 and 3. The larger PSUs are sampled as in option 1, while the smaller areas are clustered together. For the wage and benefits sample, the 11 largest areas are currently sampled as in option 1, with the remaining 140 PSUs collapsed into a single cluster. For the wage sample, 52 areas are sampled individually; the three non-metropolitan areas from Alaska and Hawaii are collapsed to form one cluster; and the remaining 99 areas are collapsed into a second, much larger cluster. The reason for the two separate clusters for the wage sample is that the NCS wage estimates are used in determining locality pay for federal workers in the 48 contiguous states and the District of Columbia only, which led us to sample Alaska and Hawaii separately from the remainder of the nation. For each industry stratum in each of these two clusters, a systematic PPS sample is chosen among all the establishments in the cluster, where the measure of size for each establishment is the product of its frame employment and the reciprocal of the probability of selecting the sample area in which it is located. The establishments in each cluster are sorted first by sample area.

In evaluating these options, we compared three methods of allocating the sample of establishments. All three methods use the same approach as Method 3A in selecting certainty units and are assumed to produce the same set of certainty establishments. The methods differ in how they determine the single-panel allocations. The first method, designated Method 4A, allocates the sample among the 54 area clusters using the current sampling method, as outlined in the previous paragraph and described in detail in Ernst et al. (2002). Note that Method 4A is identical to Method 3A. The second method, designated Method 4B, differs from Method 4A only in how the total sample in the cluster of 99 PSUs is allocated among these 99 PSUs. Method 4B allocates the total single-panel sample among the 99 individual PSUs $\times$ 23 industry sampling strata using controlled selection, as described in option 4. In a sense a comparison of Methods 4A and 4B is akin to a comparison of options 3 and 4, at least restricted to the set of 99 PSUs. Controlled selection is not used for the cluster of three PSUs from Alaska and Hawaii since the strata represented by these PSUs comprise a very small proportion of the national employment.

In the third approach, designated Method 4C, each of the 154 area PSUs is given a fixed positive single-panel allocation through an iterative process. To do this, we initially assigned to each of the 36 certainty areas its current single-panel allocation, which for these areas combined is 4,574 sample units. The remaining number of sample units, 2,302, is allocated among the 118 remaining sample areas proportional to area-weighted frame employment of all units on the frame except the NCS wage five-panel certainty units. For any area for which the allocation is less than the minimum for the area, which is defined as the number of industry strata for which the frame (excluding five-panel units) is nonempty, the allocation is raised to the area minimum. Since this increases the sample size to 7,658, which is above the required 6,876, the allocations are reduced for those areas with allocations above the area minimum, including the certainty areas, to compensate. The allocations are reduced by multiplying the initial allocation in each of these areas, by a reduction factor (0.873) obtained by taking the number of units added to meet the minimums (782), dividing this number by the initial combined allocation in all areas above the minimum (6,153), and subtracting this quotient from 1. After multiplying by this factor, the reduced allocations for these areas need to be rounded to integer values. This is done by rounding up exactly enough of the allocations to the next integer and rounding down the remaining so that the total of 6,876 units is preserved. In doing this rounding, the priority in rounding up goes to the areas with reduced unrounded allocations with the highest fractional remainder.

After this reduction is completed, it is possible (although it did not occur in our work) that some areas may have their allocations reduced enough so that they are below the minimum for the area, in which case the allocation process would have to be iterated until we reach the point where the allocation to each of the 154 areas is at least equal to the minimum, and the sum of the allocations to all 154 areas equals 6,876. At that point we have the final allocation for each of the 154 areas.

Finally, for each of the 154 areas the total allocation to the area is allocated among the 23 industries proportional to size with the constraint that there is a minimum of one sample establishment in any cell for which the frame is nonempty. Method 4C corresponds to option 1.

Methods 4B and 4C represent two different ways of modifying the current clustering of areas. Method 4B, the controlled selection method, retains the cluster of 99 PSUs, but restores more control to the sample allocation in each of the individual PSUs that comprise this cluster than does Method 4A. Method 4C, the 154 areas $\times$ 23 cells method, restores control to all of the individual sampling cells, but at the expense of inefficiencies in the allocation due to the requirement of a minimum of one in each of a large number of cells.

Note that we have not considered option 2, the collapsing of industry strata, in any of the comparisons. This is because the industry sampling strata were determined by the publication plans by industry and it is felt that it is preferable, therefore, not to collapse by industry. Of course, it is always possible to produce estimates at a finer detail than a sampling cell, so collapsing of industry strata is a possible topic for future research.

As has been mentioned, the five-panel units are assumed, in order to simplify the comparisons, to be the same for all three methods considered in this section, even though it is only Method 4A that mimics our current allocation approach. Consequently, this is an oversimplification for the other two methods because we somewhat inconsistently assume that the five-panel units are selected using Method 4A, but then the single-panel units are selected using an alternative method. Actually, for controlled selection, Method 4B, the simplification does not change the set of five-panel units, since from (4.2) it follows that the number of sample units in each Method 4A sampling cell is the same for Methods 4A and 4B and, consequently, that these two methods yield the same set of five-panel certainty units. However, if the five-panel units were selected using the 154 areas $\times$ 23 industries cells method with a minimum allocation of one unit in each cell, that is method 4C, we would obtain a different set of five-panel units. For example, if that minimum was required in the selection of the five-panel units, then in those cells for which the PSU is part of the cluster of 99 PSUs and for which the number of frame units is 1, that unit must be a five-panel certainty unit, while that would generally not be the case for the current method.

We further explain the allocation for each of the 100 samples to each of the 99 PSUs $\times$ 23 industry strata cells that comprise the cluster of 99 PSUs for controlled selection, Method 4B. First, the total allocation of this cluster to each of the 23 industry strata is allocated among the 99 individual PSUs proportional to the product of the area weight and the frame employment, excluding five panel units, of the PSU × industry stratum cell. The results in a real-valued, not integer-valued, allocation to each of the 2,267 cells determined by the 99 PSUs × 23 industry strata. For each of the 100 samples an integer-valued allocation must be determined for each of these cells. To do this, following the methodology in Causey et al. (1985), a tabular array, that is a two way additive table, is constructed. The internal elements of this array have dimensions 99 × 23 and their values are the real-valued allocations to the corresponding cells just described. This tabular array constitutes the controlled selection problem, which we designate as $\mathbf{A} = (a_{ij})$. A solution to the controlled selection problem $\mathbf{A}$ consists of a set of integer-valued tabular arrays $\mathbf{M}_1 = (m_{ij1}),...,\mathbf{M}_\ell = (m_{ij\ell})$, of the same dimensions as **A,** and associated probabilities $p_1,..., p_\ell$, summing to 1, and satisfying the following additional conditions:

$$\left| m_{ijk} - a_{ij} \right| < 1 \text{ for all } ijk \tag{4.3}$$

$$\sum_{k=1}^{\ell} p_k m_{ijk} = a_{ij} \quad \text{for all } ij \tag{4.4}$$

where (4.3), (4.4) corresponds to (4.1), (4.2), respectively. The procedure for obtaining such a solution is described in Causey et al. (1985) and involves solving a sequence of transportation problems.

Once a solution to the controlled selection problem is obtained, then corresponding to each of the 100 samples, one of the $\mathbf{M}_k$, $k = 1,..., \ell,$ is selected with the associated probabilities. The internal elements of the selected array specify the allocation of the sample to the 2,267 cells that comprise the cluster of 99 PSUs.

As we did in Section 3, 100 samples were selected here for each of the Methods 4A, 4B, and 4C. For each of the Methods 4A, 4B, and 4C, RRMSEs were then calculated for total employment and mean quarterly wages for the same domains as in Section 3, with the RRMSEs calculated as described in that section. In addition, because Methods 4A and 4B differ only in how they handle the cluster of 99 PSUs, RRMSEs were also calculated for each of the Methods 4A, 4B, and 4C for the same domains restricted to the 99 strata containing these PSUs. The RRMSEs are presented in the indicated columns of Tables 1 and 2. Note that there is no column in either table for Method 4A for national estimates, since this column would be identical to that for Method 3A.

We also have rows in Tables 3 and 4 for the methods of this section that are analogous to rows in these tables for the methods of Section 3. However, we are restricted, for reasons of confidentiality, on the amount of data we can provide in Tables 3 and 4 for the Methods 4B and 4C rows. That is why many cells in Table 3 are blank and the corresponding *p*-values in Table 4 are simply presented as > 0.05. However, in the case of the *p*-values this is an indication that none of the suppressed data is particularly exciting.

In the comparisons of Methods 4A and 4B, the most striking result that we can discuss is that for estimates of total employment, Method 4B produced lower RMSEs than Method 4A for 15 of the 18 geographic domains ($p = 0.008$) for the full set of 154 strata and for 16 domains ($p = 0.001$) in the case when the estimates are restricted to the cluster of 99 strata. The only other set of domains and estimates for which the results were significant at the 0.05 level for the comparisons of these two methods was mean wages for all 154 strata for the geographic domains, where the RMSEs were lower for Method 4B for only 4 of the 18 domains ($p = 0.031$). We do not completely know the reasons for the dramatic differences in the comparisons between Methods 4A and 4B for total employment and mean wages for the geographic domains, except that, as in Section 3, employment comparisons appear to produce results closer to what we anticipated than wage comparisons, which we assume is due to the fact that estimates of employment are more closely tied to the sampling than estimates of wages.

In the comparisons of Methods 4A and 4C, Method 4C produced lower RMSEs than Method 4A for estimates of total employment for 16 of the 18 geographic domains ($p = 0.001$) for the 154 strata case and for 17 domains ($p = 0.000$) for the 99 strata case. For wage estimates for the geographic domains, the results at first glance seem less interesting, with lower RMSEs for Method 4C than for Method 4A for 11 of the 18 geographic domains for the 154 strata case and 12 of the 18 geographic domains for the 99 strata case. However, what is striking is that for both the 154 and 99 strata cases for non-metropolitan estimates, Method 4C yielded lower RMSEs than Method 4A for all 9 census divisions for both estimates of employment and wages, a result which can be ascertained from Tables 1 and 2. This is easily explained, since the requirement of a minimum of one for each of the 154 PSUS × 23 industry cells in Method 4C raises the allocations in the cluster of 99 strata, which includes all non-metropolitan areas.

Another interesting result, which cannot be ascertained from any of the tables, concerns the comparisons for industry domains for total employment for the cluster of 99 PSUs. Roughly, the RMSEs are lower for Method 4C than for Method 4A for all the industries in the bottom half of the industries in terms of sample size, where the size is based on the single-panel sample for the cluster of 99 PSUs for Method 4A. We have the following explanation for this result. In Method 4A our allocations to the 54 clusters with the requirement of a minimum of one in each cluster × industry, resulted in the cluster of 99 PSUs receiving disproportionately low allocations in the single-panel sample for those industries with small total sample sizes. The minimum of one for each of the 154 PSUS × 23 industry cells for Method 4C raises the allocations to the cluster of 99 PSUs in comparison with Method 4A, but in particular for those industries with disproportionately low allocations to this cluster for Method 4A.

In the actual NCS production sample, we realized the problem of the low allocations to the cluster of 99 PSUs produced by our allocation program in some industries and manually increased the allocations to this cluster in those industries.

5. References

Causey, B. D., Cox, L. H., and Ernst L. R. (1985). Applications of Transportation Theory to Statistical Problems. Journal of the American Statistical Association. 80, 903-909.

Ernst, L. R., Guciardo, C. J., Ponikowski, C. H., and Tehonica, J. (2002). Sample Allocation and Selection for the National Compensation Survey. 2002 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

Guciardo, C. J., Dorfman, A. H., Ernst, L. R., and Sverchkov, M. (2004). Variance Estimation for the National Compensation Survey When PSUs are Clustered Prior to the Second Phase of Sampling. 2004 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

Sverchkov, M., Dorfman, A. H., Ernst, L. R., and Guciardo, C. J. (2004). Zero-Calibrated Variance Estimators. 2004 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

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

Table 1. Percent Relative Mean Square Errors for Employment

Domain	National					99 Strata		
	Meth. 3A	Meth. 3B	Meth. 3C	Meth. 4B	Meth. 4C	Meth. 4A	Meth. 4B	Meth. 4C
New England Met	0.559	0.572	0.376	0.461	0.813	5.578	3.541	8.222
New England Nmet	3.325	3.890	2.172	2.838	1.690	3.325	2.838	1.690
Mid Atlantic Met	0.366	0.381	0.554	0.357	0.370	1.779	1.769	0.618
Mid Atlantic Nonmet	2.715	2.964	1.335	2.061	0.907	2.715	2.061	0.907
E-North Central Met	0.287	0.363	0.153	0.295	0.172	1.080	1.063	0.396
E-North Centrl Nmet	1.305	1.644	0.448	1.259	0.462	1.305	1.259	0.462
W-North Centrl Met	0.423	0.495	0.259	0.384	0.252	1.035	0.836	0.469
W-North Cent Nmet	1.129	1.363	0.432	1.167	0.307	1.129	1.167	0.307
South Atlantic Met	0.286	0.234	0.182	0.260	0.217	0.971	0.863	0.358
South Atlantic Nmet	1.308	1.486	0.373	1.063	0.321	1.308	1.063	0.321
E-South Central Met	0.745	0.787	0.476	0.614	0.410	1.639	1.346	0.689
E-South Centrl Nmet	1.121	1.046	0.347	0.920	0.286	1.121	0.920	0.286
W-South Centrl Met	0.434	0.444	0.331	0.387	0.256	1.697	1.390	0.579
W-South Centr Nmet	2.859	2.604	0.729	2.392	0.472	2.859	2.392	0.472
Mountain Met	0.565	0.576	0.360	0.633	0.359	0.957	1.112	0.599
Mountain Nonmet	2.808	3.096	0.936	2.294	1.318	2.808	2.294	1.318
Pacific Met	0.316	0.283	0.203	0.294	0.260	1.849	1.367	0.699
Pacific Nonmet	2.075	2.177	0.934	1.543	0.810	2.647	1.935	1.003

Table 2. Percent Relative Mean Square Errors for Mean Earnings

Domain	National					99 Strata		
	Meth. 3A	Meth. 3B	Meth. 3C	Meth. 4B	Meth. 4C	Meth. 4A	Meth. 4B	Meth. 4C
New England Met	1.787	2.108	1.696	1.800	1.875	5.442	6.950	6.966
New England Nmet	3.842	3.342	3.544	3.573	3.221	3.842	3.573	3.221
Mid Atlantic Met	1.455	1.692	1.675	1.474	2.090	2.889	3.439	3.051
Mid Atlantic Nonmet	3.695	3.374	3.243	4.047	3.262	3.695	4.047	3.262
E-North Central Met	0.936	0.921	0.919	0.905	1.012	1.654	2.171	1.763
E-North Centrl Nmet	1.879	1.975	1.402	1.883	1.443	1.879	1.883	1.443
W-North Centrl Met	1.553	1.455	1.300	1.564	1.512	2.911	3.000	2.946
W-North Cent Nmet	2.392	2.470	1.596	2.653	1.440	2.392	2.653	1.440
South Atlantic Met	1.166	1.141	1.069	1.093	1.120	2.441	2.383	2.385
South Atlantic Nmet	2.523	3.232	2.222	2.300	1.952	2.523	2.300	1.952
E-South Central Met	2.232	3.209	2.107	2.504	2.524	2.956	3.584	3.029
E-South Centrl Nmet	1.410	1.594	1.115	1.508	1.215	1.410	1.508	1.215
W-South Centrl Met	1.258	1.385	1.396	1.476	1.569	3.686	3.528	4.691
W-South Centr Nmet	3.982	4.529	1.841	4.084	1.820	3.982	4.084	1.820
Mountain Met	1.821	1.911	1.890	2.074	2.332	3.157	3.797	3.102
Mountain Nonmet	3.055	3.683	2.816	3.479	2.315	3.055	3.479	2.315
Pacific Met	1.361	1.362	1.543	1.397	1.509	4.527	3.068	3.352
Pacific Nonmet	2.785	3.156	2.521	3.019	2.760	2.753	3.264	2.665

Table 3. Number of Domains for Which RMSE of Indicated Method < RMSE of Method 3A (4A)

	Employment			Mean Wages		
	Ind.	Geog.	Size	Ind.	Geog.	Size
Total number of domains	23	18	7	23	18	7
Meth. 3B, all 154 strata	7	4	0	12	5	1
Meth. 3C, all 154 strata	13	17	7	12	14	5
Meth. 4B, all 154 strata		15			4	
Meth. 4C, all 154 strata	10	16	1	9	11	1
Meth. 4B, Rem. 99 strata		16			5	
Meth. 4C, Rem. 99 strata		17			12	

Table 4. p-Values for Two Sided Sign Test Corresponding to Table 3 Results

	Employment			Mean Wages		
	Ind.	Geog.	Size	Ind.	Geog.	Size
Meth. 3B, all 154 strata	0.093	0.031	0.016	1.000	0.096	0.125
Meth. 3C, all 154 strata	0.678	0.000	0.016	1.000	0.031	0.453
Meth. 4B, all 154 strata	>0.05	0.008	>0.05	>0.05	0.031	>0.05
Meth. 4C, all 154 strata	0.678	0.001	0.125	0.405	0.481	0.125
Meth. 4B, Rem. 99 strata	>0.05	0.001	>0.05	>0.05	0.096	>0.05
Meth. 4C, Rem. 99 strata	>0.05	0.000	>0.05	>0.05	0.238	>0.05