Adjusting Sampling and Weighting to Account for Births and Deaths

October 2009

Lawrence R. Ernst[1], Randall K. Powers[1], Andy Sadler[2], Dave Slack[2]
[1] Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 1950, Washington, DC 20212,
[2] Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3650, Washington, DC 20212

Abstract
In a single-stage sample where the units report on multiple occasions, it is typical for some units to drop out of the sampling frame over time (the deaths); for new units to be added to the frame (the births); while the remaining units stay in the original frame and, if selected, in sample. In this paper we discuss an approach to adding sample units from the frame of continuing and birth units to compensate for the deaths, and reweighting the modified sample so that it results in unbiased estimates in the sense that the expected weight of each unit in the modified frame over all possible samples is 1. A possible application to sample augmentation for the first-stage sample of the Producer Price Index, conducted by the Bureau of Labor Statistics, is discussed, as are alternative sampling procedures that focus on partially converting the augmented sampling process to the selection of a single PPS sample.

Key Words: Sampling frame, continuing units, modified frame, probability proportional to size sampling

1. Introduction

This paper discusses the following sampling problem. For a stratum in a sampling frame *I*, which is fixed, an initial sample of *n* first-stage units is selected with probability proportional to size without replacement. (As a simplification, we assume, without loss of generality, that *I* consists of a single stratum.) The specific sample selection method will be discussed later. It is intended that all *n* units remain in sample for a specified number of data collection periods. However, gradually over the data collection periods, more and more units, referred to here as deaths, go out of business or out of scope, or become nonrespondents. We consider such units to have left both the sample and the frame. (Note that units that were not selected in the initial sample that would have become nonrespondents if they had been selected are not treated as deaths since we do not know what their response status would have been if they had been selected in the initial sample. See the final paragraph of this section for further discussion of the treatment of nonrespondents.) Also, other units, designated as birth units, gradually join the frame.

At first there is no attempt to compensate for the deaths by adding to the original sample some combination of births and continuing nonselected units from *I*, where continuing units are units that were in *I* originally and do not leave *I* or the initial sample if selected in the initial sample. Eventually though, after enough data collection periods, the number of units that have left *I* may become large enough that the remaining units are deemed insufficient in number to continue to publish estimates for the survey. Then the sample would be augmented by adding additional units to the continuing initial sample units to increase the sample size, typically, although not necessarily, back to the original sample size of n units. The additional sample units are chosen from an updated frame U consisting of all units in the original frame plus all birth units, minus all units that were in the initial sample or dropped out of I. The augmented sample consists of all continuing initial sample units plus all additional sample units. Note that U, unlike I, is not fixed since U depends on the continuing initial sample units and the nonrespondents selected from I.

There are two general approaches to augmenting the sample that will be considered in this paper. One approach is to select the additional sample units from U using PPS systematic sampling with a new measure of size (MOS). This approach will be discussed in Section 2, with a simple illustrative example presented in Section 3. This approach is relatively straightforward, with the resulting augmented sample a union of the continuing units from the initial sample and the additional sample units chosen from U. We also illustrate this approach empirically in Section 4 by selecting 10000 augmented samples of establishments from one industry stratum using Producer Price Index (PPI) sampling frames; calculating for each sample the mean of the augmented sample weights averaged over all the units in the stratum eligible to be in an augmented sample; and then calculating the mean of these means averaged over the 10000 samples.

A second approach, discussed in Section 5, to augmenting the sample is to expand the original PPS sample by adding sample units in a manner that attempts to result in the combined initial sample and additional sample constituting a single sample selected PPS. (The first approach cannot do this since with that approach the continuing units have two chances of being selected, once from I and once from U, while the birth units have only one chance of being selected from U, resulting in a continuing unit generally having a higher probability of selection than a birth unit of the same new MOS.) As discussed in Section 5, this second approach is partially, but not completely, successful at attaining the goal of the augmented sample being selected PPS. Finally, we briefly mention in Section 6 some additional future research issues related to the subject of this paper.

As mentioned earlier in this section, we consider units selected in the initial sample that become nonrespondents to have left the frame and to be treated as deaths, even though these units still exist. This is because these units have dropped out of the sample and when we add units to the original sample to compensate for the deaths, these nonrespondents will be given no chance of being added to the sample. We could of course give these units a chance of being an additional sample unit if we want to make an attempt to convert such units to respondents. However, this would complicate the weighting process, since these units might then be selected twice. This also might result in a large proportion of the additional sample units being nonrespondents. An alternative approach would be to allow attempts to convert the nonrespondents in the initial sample to respondents, but only before the additional sample units are drawn, and to consider the units for which this conversion was unsuccessful to be deaths and not eligible to be in an augmented sample. This is the approach that we will take in this paper.

2. The Two PPS Samples Approach to Sample Augmentation

One approach to the sample augmentation problem is as follows. Select an initial sample of n units from the original frame I using PPS systematic sampling, although actually any PPS without replacement sampling procedure can be used. Systematic sampling is just generally the simplest. Note that I is a fixed set.

At a later point in time it may be necessary to select additional sample units to compensate for the units that drop out of the frame, that is the deaths. These additional units are selected from an updated frame U consisting of all units in the original frame plus all birth units, minus all units that were in the initial sample or dropped out of I, generally due to going out of business or out of scope, or becoming nonrespondents that were not successfully converted to respondents. Note that while I is fixed, U *is* not fixed as explained in the Introduction. The MOS is updated for each unit in U. The process for selecting the additional units is also systematic PPS, although again any PPS without replacement sampling procedure can be used. The number of additional units selected from the updated frame is generally the number that have dropped out, although this is not a requirement. However, to insure that the estimates satisfy the requirement, that the expected value of an estimate of total over all possible samples is the population total, we require that any continuing unit not selected in the initial sample has a chance to be selected in the additional sample. One way to meet this requirement is by adding the constraint that if no units in the initial sample have dropped out, then at least 1 unit is selected from the updated frame PPS. It is also possible to add a larger or smaller number of additional sample units to form the augmented sample than the number of deaths if it is desired that the augmented sample have, respectively, a larger or smaller number of sample units than the initial sample. The initial sample units that remain in sample, that is the continuing initial sample units, together with the additional sample units, constitute the augmented sample.

Having described the sample selection process for this approach, we now explain the corresponding sample weighting process, which we consider the most interesting part of the procedure. (The weighting process to be described will be based only on the initial sample units selected from I in the case of the initial sample, and the additional units selected from U together with the continuing initial sample units in the case of the augmented sample. Units that go out of business or out of scope have their weights adjusted to 0. Nonresponse adjustment for sample units that become nonrespondents and are not converted back to respondents will not be discussed.)

For linear estimates that only incorporate the initial sample units and are for time periods when the initial sample units are the only ones to contribute to the estimates, the sample weight, w_{Ii}, for unit i in the initial sample is simply the reciprocal of the probability of selection of that unit in the initial sample, that is

$$w_{Ii} = 1 / p_{Ii} \tag{2.1}$$

where p_{Ii} is the probability of selection of this unit in the initial sample. The weight for each unit in I that is not selected in the initial sample is 0. These weights lead to unbiased estimates in the sense that the expected value of the sample weight for each initial sample unit over all initial samples is 1. See Ernst (1989) for an explanation of why the expected value of the weight of each sample unit being 1 can be considered a necessary and sufficient condition for obtaining unbiased estimates.

Other estimates are intended to incorporate data from the augmented sample, that is the continuing initial sample units together with the additional sample units selected from the updated frame. These estimates are intended for use for time periods after the additional sample units are drawn and data for these units are collected. The weighting is more complex, with four different types of computations, depending on the type of unit and the sample for which it was first selected, as will now be discussed.

Case 1. Sample Birth Units. The simplest of the first three cases is the set of sample birth units, which can only be selected as part of the sample of additional units. If U is the updated frame, B_i is the i-th birth unit in U, and B_i is selected in the birth sample conditional on U with probability p_{BiU}, then the sample weight for this unit, w_{BiU}, conditional on U, is

$$w_{BiU} = 1/p_{BiU} \tag{2.2}$$

if the unit is selected in the additional sample, while $w_{BiU} = 0$ if the unit is not selected in sample; hence $E(w_{BiU}) = 1$ for each birth unit and updated frame. Consequently, $E(w_{Bi}) = 1$, where $E(w_{Bi})$ is the expected weight for B_i over all updated frames.

Case 2. Continuing Sample Units Selected from I. Cases 2 and 3 both apply to continuing units. These two cases are more complex because a continuing unit has two chances of being selected in the sample, either as a unit selected from I (Case 2) or as a unit that is not selected from I but is selected from U (Case 3). The sample weight w_{Ii} for a continuing unit i selected as an initial sample unit from I is the reciprocal of the probability p_{Ii} of this unit being selected in the initial sample from I multiplied by the weighting adjustment factor $1/(2-p_{Ii})$, that is

$$w_{Ii} = (1/p_{Ii})/(2-p_{Ii}) \tag{2.3}$$

The purpose of this weighting adjustment factor will be explained in detail later in this section, but the basic reason for using it is to compensate for the fact that continuing units generally have two chances to be selected in the augmented sample rather than one chance. Since p_{Ii} is the probability of this unit being selected in the initial sample, then by (2.3) the contribution to the expected weight of this unit arising from selection of the unit in the initial sample is

$$p_{Ii}(1/p_{Ii})/(2-p_{Ii}) = 1/(2-p_{Ii}) \tag{2.4}$$

Note again that I is a fixed set.

Case 3. Continuing Sample Units Selected from U. The sample weight w_{Ui} for continuing unit i if it is selected in sample from the updated frame U, that is the unit is an additional sample unit, is the reciprocal of the probability p_{Ui} of the unit being selected in the additional sample given U, multiplied by the weighting adjustment factor $1/(2-p_{Ii})$, that is

$$w_{Ui} = (1/p_{Ui})/(2 - p_{Ii}) \tag{2.5}$$

Note that continuing unit *i* in updated frame *U* matches with continuing unit *i* in the initial frame *I*. Also note that, as mentioned previously, *U*, unlike *I*, is a random set since it excludes the continuing initial sample units and the nonrespondents selected from *I*. The contribution to the expected weight of continuing unit *i* arising from the selection of this unit from *U*, conditional on *U*, is $p_{Ui}(1/p_{Ui})/(2 - p_{Ii}) = 1/(2 - p_{Ii})$ since p_{Ui} is the probability of this unit being selected in the additional sample, conditional on *U*. Finally, the contribution to the expected weight arising from the selection of continuing unit *i* from among all updated frames containing the unit is

$$(1/(2 - p_{Ii}))\sum_{U} p_U = (1 - p_{Ii})/(2 - p_{Ii}), \tag{2.6}$$

where p_U is the probability that *U* is the updated frame and the summation is over all updated frames which include unit *i*. The reason that (2.6) holds is that $1 - p_{Ii}$ is the probability that continuing unit *i* in *I* is not selected in the initial sample, which is the same as the probability that the updated frame includes continuing unit *i*, which is the summation in (2.6).

<u>Case 4. Nonselected Continuing Units</u>. There is one final case, that is units that are not selected as sample units from either *I* or *U*. The sample weight for each such unit is 0.

It remains only to show that the expected weight $E(w_{Ci})$ for each continuing unit C_i is 1 over all possible initial and updated samples combined, since we have already shown that by Case 1, $E(w_{Bi}) = 1$ for each birth unit B_i. However, $E(w_{Ci}) = 1$ follows immediately by simply combining (2.4), (2.6), and Case 4, since

$$1/(2 - p_{Ii}) + (1 - p_{Ii})/(2 - p_{Ii}) = 1 \tag{2.7}$$

It is (2.7) that explains why we have been using $1/(2 - p_{Ii})$ as a weighting adjustment factor throughout this section.

In summary, the first step in the procedure is to select an initial sample PPS. For estimates corresponding to the units selected in the initial sample, the sample weights are given by (2.1). Then the additional units are selected PPS from the updated frame *U* corresponding to the initial sample. For estimates that incorporate sample birth units, continuing units selected from the initial frame, and continuing units selected from the updated frame, the sample weights are given by (2.2), (2.3), and (2.5), respectively.

Note that instead of using the adjustment factor $1/(2 - p_{Ii})$ in Cases 2 and 3, it would be possible to use the factors $\alpha/(2 - p_{Ii})$ in Case 2 and $(2 - \alpha - p_{Ii})/((2 - p_{Ii})(1 - p_{Ii}))$ in Case 3, where α is any constant between 0 and $2 - p_{Ii}$. With these alternative factors, $E(w_{Ci})$ remains 1 for any continuing unit *i* over all possible initial and augmented samples. Varying α allows for adjustments in the estimates that impact on their variances. Note that, in particular, when $\alpha = 0$, the factor corresponding to Case 2 units is 0, and consequently it is only the continuing sample units selected from the updated frame that contribute to the estimates. Analogously, when $\alpha = 2 - p_{Ii}$, and hence α

varies with the unit, it is only the sample units selected from the initial frame that contribute to the estimates. Finally, when $\alpha = 1$, the adjustment factor reduces back to $1/(2 - p_{Ii})$ for both Case 2 and Case 3 units. Both $\alpha = 0$ and $\alpha = 2 - p_{Ii}$ may lead to high variances, since in both situations some of the continuing sample units may not be used in the estimates.

Still another possibility would be to use the weighting adjustment factor 1 instead of $1/(2 - p_{Ii})$ in Cases 2 and 3, which is equivalent to there being no adjustment factor. However, in general $E(w_{Ci}) \neq 1$ for this approach, resulting in biased estimates and consequently this approach was not considered.

A final possibility would be to modify the sampling by including the continuing initial sample units in the updated frame, which would allow the initial sample units to be selected twice, that is in both the initial sample and the set of additional sample units. However, the staff responsible for the sampling for the Producer Price index (PPI), which is the application we will focus on, was concerned that such a "with replacement" type approach might have an undesirable impact on variances and consequently this approach was not considered either.

3. Illustrative Example for the Procedure of Section 2

In this section we present a simple example to illustrate the procedure described in Section 2. Before doing this we note that it is not really necessary to go through all the steps presented here in order to select the initial and augmented samples and to calculate the sample weights for these two samples. To obtain the initial sample it is only necessary to choose a sample from *I* of the desired number of units using PPS systematic sampling and then obtain the corresponding sample weights using (2.1) for the estimates that use the initial sample only for time periods prior to sample augmentation. (We ignore the fact that prior to sample augmentation some of the initial sample units may have already dropped out.) To obtain the additional sample units, simply select a sample from the updated sampling frame *U* also using PPS systematic sampling. The weights for time periods for which the entire augmented sample is used in the estimates are given by (2.2) for birth units conditional on *U* being the updated frame, (2.3) for continuing sample units selected from *I*, and (2.5) for continuing units selected from *U*.

In the example considered in this section, the initial frame has 5 units A, B, C, D, and E, with the original MOS for these units given in row 2. The initial sample consists of two units selected using PPS systematic sampling from the initial frame A-E sorted in alphabetical order. At the time of sample augmentation, unit A has dropped from the frame and unit F has become a birth, with the new MOS for units B-F given in row 3. Our ultimate goal is to show that the expected value of the augmented sample weight is 1 for each of units B-F, that is all units except the death unit A. The augmented sample consists of three units, which means that one additional unit is selected if A was not selected in the initial sample and two additional units are selected if A was in the initial sample, since A is a death unit. The probability of selection of each of the units A-E in the initial sample is given in row 4, with the corresponding initial sample weight without the adjustment factor, that is the reciprocal of probability of selection of each of these units in the initial sample, given in row 5. The corresponding initial sample weight with the adjustment factor is given in row 6, and is obtained from (2.3). The contribution of each of the continuing units in the initial frame, that is B-E, to the overall expected augmented sample weight arising from the selection in the initial sample, is given in row 7 and is obtained from (2.4). Note that there are no entries for unit A for rows 3, 6, and 7 because unit A is a death and no entries for unit F for row 2 and rows 4-7 because unit F is a birth, and hence entries in the indicated cells for these two units would not be used.

The possible initial samples are given in row 8 and the corresponding updated frames for selection of additional units are in row 9. Each of these possible updated frames is obtained from the corresponding initial sample by subtracting the death (unit A) from the initial frame A-E, and also subtracting the initial sample units, and adding the birth (unit F). For each frame in row 9, the corresponding cell in row 10 is the probability of that set of units being the updated frame. (Note that in this example there is a one to one correspondence between initial samples in row 9 and updated frames in row 10. This is not always the case. Initial samples with the same set of continuing units but different death units correspond to the same updated frame, which affects the probability of the possible updated frames. However, this does not complicate the calculation of the sample weights if one follows the weighting approach described in the first paragraph of this section.)

The next five rows, 11-16, are concerned with selection of unit B from the updated frame. (Since unit A is a death, it is not in any possible updated frame.) In row 11, the first number is the probability of selecting B in the additional sample given that the initial sample is AC and hence the updated frame is BDEF. The corresponding weight without the factor for selection of B in the additional sample given that the updated frame is BDEF is in the same column of row 12 and is the reciprocal of the entry of in row 11 of that column. The weight with the factor given the same updated frame is in the same column of row 13 and is the weight without the factor multiplied by the factor. The second, and only other set of numbers in rows 11-13, apply when the initial sample is AD and consequently the updated frame is BCEF, with these numbers computed analogously to the first set of numbers in these rows, except that B is a certainty additional sample unit conditional on this updated frame. There are two sets of numbers in rows 11-13 since there are two possible updated frames that include unit B.

Next, the single number in row 14 is the contribution of unit B in the additional sample to the expected augmented weight of unit B and is calculated as follows. The contribution arising from the case when the updated frame is BDEF is the probability that BDEF is the updated frame multiplied by the probability that B is selected in the additional sample given that BDEF is the updated frame, multiplied by the weight for B in the additional sample with the factor given that the updated frame is BDEF. The contribution from the only other updated frame, BCEF, that includes unit B, is computed similarly and the two contributions summed to obtain the single number in row 14. Finally in the cell in row 15 immediately below is the expected augmented weight of unit B over all possible samples of initial and additional units, that is the sum of the values in rows 7 and 14 of that column. This sum is 1 as desired.

The analogous entries for unit C are in rows 16-20, and for units D-F in rows 21-25, 26-30, and 31-35, respectively. Note, in particular, that the numerical entries in rows 32 and 33 are identical since unit F is a birth unit and hence the weighting adjustment factor for it is 1.

4. Illustrating the Weighting Procedure of Section 2 Using PPI Data

In order for the sampling and weighting procedure of Section 2 to yield unbiased estimates it is necessary to establish that the expected value of the weight of each sample unit is 1 for both the initial and augmented samples. In the case of initial samples, this follows immediately from (2.1). In the case of augmented samples, this follows analytically from (2.2)–(2.7).

In this section we compare empirically for PPI for one industry, the overall estimated expected mean of the augmented sample weights over all units eligible to be in an augmented sample, with and without weighting adjustment factors. (That is, to simplify the work in this section, we will not be concerned, as we were in the previous section, with the expected augmented sample weight of each unit eligible to be in an augmented sample, but instead the expected mean of these expected weights over all such units. For the industry under study, 10000 augmented samples, each consisting of 69 establishments, were selected from the initial frame and the updated frame corresponding to the initial sample, with the sample weight for each establishment in an augmented sample given by (2.2), (2.3), (2.5), or 0, depending on whether the establishment for the particular sample is a sample birth unit, a continuing initial sample unit, a continuing sample unit selected from the updated frame, or a unit not selected to be in sample from either the initial or updated frames.

For this example, there are 334 units in the original frame of which 123 constitute the deaths. There are also 93 birth units leaving a net of 304 units (334-123+93) eligible to be in an augmented sample. (Note that in this example we assume that all of the 123 deaths are due to units that went out of business or out of scope prior to the formation of the updated frame and hence that none of these 123 units were eligible to be in an augmented sample. If some of these 123 units were refusals instead, these refusal units would only be counted as deaths, and hence excluded from the updated frame, if they were selected in the initial sample.)

To obtain each of the 10000 augmented samples, first select 69 initial sample units using PPS systematic sampling from among the 334 units in the initial frame and let d be the number of units among these 69 initial sample units that become deaths. Then select d additional units from among the $304-(69-d)=235-d$ units in the corresponding updated frame using PPS systematic sampling, with the $69-d$ initial sample units among the 304 units eligible to in an augmented sample excluded from being selected as additional sample units for the corresponding sample, since they have already been selected as initial sample units. For each augmented sample, the weight of each of the 304 units eligible to be in an augmented sample is given by (2.2), (2.3), (2.5), or 0 as explained above.

The overall estimate of the expected weight averaged over the 304 units eligible to be in an augmented sample is obtained by taking the mean over the 10000 samples of the mean of the weights of the 304 units and is 1.0020 if the weighting adjustment factors are used in the computation. In contrast, the corresponding overall estimate of the expected weight is 1.5681 if it is calculated without using the weighting adjustment factors. Thus, with the weighting adjustment factors included, their inclusion does result in estimated expected weights averaged over the 304 units numerically close to 1 and much closer to 1 than the corresponding estimate with these factors excluded. In fact, with a two-sided test of the null hypothesis that the expected mean weight is 1 against the alternative hypothesis that the expected mean weight is not 1, the corresponding *p*-value is 23.3 and hence the null hypothesis cannot by rejected at any typical level of significance despite the very large number of samples selected.

5. Approach to Sample Augmentation that Focuses on Obtaining a Single PPS Sample

As noted in the Introduction, a drawback to the procedure described in Section 2 is that the selection probabilities for the units in the augmented sample are not proportional to size. This mostly arises because each continuing unit has two chances of being selected in the augmented sample, that is either as an initial sample unit or as an additional sample unit, while each birth unit has only one chance of being selected in the augmented sample, that is as an additional sample unit. This results in a continuing unit generally having a higher probability of selection in the augmented sample than a birth unit with the same MOS.

One way of overcoming this problem would require the replacement of some of the continuing initial sample units by birth units. We will not consider this approach because of the extra expense involved in replacing units, but instead describe an approach to partially overcoming this problem without dropping any continuing initial sample units, using the following steps:

1. Select an initial sample of n units from the initial frame exactly as would be done if following the approach in Section 2. Then create an updated frame consisting of all births and all continuing units, including, unlike in Section 2, all continuing units that are initial sample units. (The continuing initial sample units would not be given a chance of being selected for a second time from the updated frame. They would simply be used in determining the allocation between the birth units and the continuing units selected from the updated frame based on a new MOS.) The target sample size for the augmented sample, consisting of the continuing initial sample units and the additional units selected from the updated frame, is generally n, that is the same number of units as in the initial sample.
2. Determine the set of certainty units in the updated frame created in step 1 and subtract the number of these units from n for the case when a sample of n units is selected PPS from the updated frame. Then allocate the remaining number of sample units to select from the updated frame between noncertainty continuing and noncertainty births units, proportional to size, based on the new MOS. These allocations generally are not integers.
3. To obtain an integer allocation for each of the two allocations obtained in step 2, that is the allocation of noncertainty continuing and noncertainty birth units, round each of these step 2 allocations to an integer. There are several ways to do this. One way, which we will use throughout this section, is to round the number of births and continuing units to the nearest integer, with a minimum of one unit of each type even if this increases the total sample size. Add back separately to these two rounded allocations the number of certainty births and the number of certainty continuing units calculated in step 2 to obtain a revised target numbers of total births and total continuing units to be selected from the updated frame. (An alternative rounding approach, which will not be considered here, is to use controlled rounding (Cox and Ernst 1982). For example, with controlled rounding, if the unrounded value of the

number of births and continuing units are, respectively, 1.3 and 4.7, then the two allocations are rounded to 2 and 4, respectively, with probability 0.3, and to 1 and 5, respectively, with probability 0.7, instead of always rounding the number of births to 1 and continuing units to 5.)

4. If the number of continuing initial sample units exceeds the number of additional continuing units allocated from the updated frame, as described in steps 2 and 3, then: retain all continuing initial sample units in the augmented sample; select no additional continuing sample units from the updated frame to be in the augmented sample; and reduce the allocated number of births by the difference between the number of continuing initial sample units and the number of allocated continuing units from the updated frame computed in step 3, except that the number of allocated births should be at least 1. The changes described in this step reduce the allocation of births in comparison with continuing units in the augmented sample, but this appears to be the best that can be done in terms of a proportional allocation between births and continuing units without dropping any initial continuing sample units. (A minimum allocation greater than 1 for births may be necessary, however, if an allocation of 1 results in large weights and large variances. This would increase the total sample size for the augmented sample.)
5. If, unlike in step 4, the number of continuing initial sample units is less than or equal to the number of allocated continuing units from the updated frame, then the final number of continuing units to select from the updated frame is the difference between the number of allocated continuing units in step 3 and the number of continuing initial sample units. The number of allocated births units is then the number calculated in steps 2 and 3. In this case, the allocation between births and continuing units is proportional to size if we ignore rounding.
6. With the number of additional continuing units and number of sample birth units allocated as described in the previous steps, the set of sample units for each of the corresponding two components is selected separately using PPS systematic sampling. These two samples plus the sample of initial continuing units constitute the augmented sample.
7. The sample weights for the sampling approach described in steps 1-6 for computing estimates that incorporate data from both the initial sample units and the additional sample units and that are intended for use for time periods after data from the additional sample units are collected, are computed as described in Section 2, that is the weights are calculated using (2.2) for birth units, (2.3) for continuing units selected in the initial sample, and (2.5) for continuing units selected in the updated sample.

6. Possible Additional Research Issues

We briefly mention in this section some additional research issues related to the subject of this paper.

In Section 2 we considered the selection of two samples, an initial sample and an augmented sample, with the latter sample replacing the deaths among the initial sample units by additional units. It would be interesting to determine if it is possible to generalize this work to the case of three or more samples, where for each sample selection any new deaths are replaced by units that are in a newly updated sampling frame and that have not been selected previously.

In the end of the Introduction, we discussed different approaches to handling nonresponse with respect to selection of the additional sample units. This may be worth investigating further.

In Section 2, generalized weighting adjustment factors incorporating an α term are described. What value of α results in the lowest variances of the estimates? In particular, does $\alpha = 1$ result in the lowest variances?

Acknowledgements

The authors would like to thank John Eltinge for reviewing the paper and providing a number of useful comments.

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

References

Cox, L. H., and Ernst, L. R., (1982). Controlled Rounding. INFOR, 20, 423-432.
Ernst, L. R. (1989). Weighting Issues for Longitudinal Household and Family Estimates. Panel Surveys, 139-159. New York, John Wiley.

Table 1: Example of Procedure of Section 2

Row	Row Description	Unit					
1	Units	A	B	C	D	E	F
2	Original employment	100	80	60	50	10	
3	Employment at time of sample augmentation		120	30	100	25	30
4	Prob unit in initial sample	0.6667	0.5333	0.4	0.3333	0.0667	
5	Initial sample weight without adjustment factor	1.5	1.875	2.5	3	15	
6	Initial sample weight with adjustment factor		1.2784	1.5625	1.8	7.7586	
7	Contribution to expected augmented weight of unit from initial sample		0.6818	0.625	0.6	0.5172	
8	Initial sample	AB	AC	AD	BD	BE	
9	Updated frame	CDEF	BDEF	BCEF	CEF	CDF	
10	Prob updated frame	0.2	0.4	0.0667	0.2667	0.0667	
11	Prob B selected in additional sample given updated frame		0.8727	1			
12	Weight for B in additional sample given updated frame without factor		1.1458	1			
13	Weight for B in additional sample given updated frame with factor		0.7813	0.6818			
14	Contribution of B in additional sample to overall expected weight		0.3182				
15	Overall expected augmented weight of B		1				
16	Prob C selected in additional sample given updated frame	0.3243		0.2927	0.3529	0.1875	
17	Weight for C in additional sample given updated frame without factor	3.0833		0.3417	2.833	5.3333	
18	Weight for C in additional sample given updated frame with factor	1.9271		2.1354	1.7708	5.3333	
19	Contribution of C in additional sample to overall expected weight			0.375			
20	Overall expected augmented weight of C			1			
21	Prob D selected in additional sample given updated frame	1	0.7273			0.625	
22	Weight for D in additional sample given updated frame without factor	1	1.375			1.6	
23	Weight for D in additional sample given updated frame with factor	0.6	0.825			0.96	
24	Contribution of D in additional sample to overall expected weight				0.4		
25	Overall expected augmented weight of D				1		
26	Prob E selected in additional sample given updated frame	0.2703	0.1818	0.2439	0.2491		
27	Weight for E in additional sample given updated frame without factor	3.7	5.5	4.1	3.4		
28	Weight for E in additional sample given updated frame with factor	1.9138	2.8448	2.1207	1.7586		
29	Contribution of E in additional sample to overall expected weight					0.4828	
30	Overall expected augmented weight of E					1	
31	Prob F selected in additional sample given updated frame	0.3243	0.2181	0.2927	0.3529	0.1875	
32	Weight for F in additional sample given updated frame without factor	3.0833	4.5833	3.417	2.8333	5.3333	
33	Weight for F in additional sample given updated frame with factor	3.0833	4.5833	3.417	2.8333	5.3333	
34	Contribution of F in additional sample to overall expected weight						1
35	Overall expected augmented weight of F						1