A COMPARISON OF ESTIMATORS FOR THE MEAN OF A FINITE POPULATION, BASED ON A SYSTEMATIC SAMPLE

Larry Huff, Diem-Tran Kratzke, Gordon Mikkelson, Sandra West
Diem-Tran Kratzke, 2 Massachusetts Ave, N.E. Washington, DC 20212

KEY WORDS: Response Analysis Survey, Systematic Sampling, Model Based Estimator

1. Introduction

In this paper we empirically compare several mean estimators for a finite population based on a systematic sample. This research began with a quality improvement project for two Bureau of Labor Statistics' establishment programs that collect monthly employment data: the Covered Employment and Wages (ES-202) program and the Current Employment Statistics (CES) survey. The ES-202 program is compiled using data from quarterly reports of business establishments that are covered under the Unemployment Insurance laws in the United States. The CES survey collects monthly employment from a voluntary sample of business firms and uses the ES-202 universe employment data to annually adjust its industry employment totals. A Response Analysis Survey (RAS) was conducted in order to determine the comparability and accuracy of employment data reported to these two programs. Each sample unit was asked several questions pertaining to their response practices for both of these programs. The samples were selected from among the CES reporters of ten participating states. The sample consisted of four panels that were selected approximately three months apart. The panel samples were selected with probability proportional to a measure of size based on size of firm and percent difference between reported employment to the CES and ES-202 programs.

Various characteristics of the population are estimated based on their responses to four specific RAS questions. Alternative mean estimators and their estimated standard errors are compared in order to determine the most appropriate estimation techniques. We compare the estimators that treat all four panels separately, as well as estimators that combine the four panels.

The general sampling design issues are discussed in Section 2. Background information about the RAS, including a description of the population and sample design, is given in Section 3. The estimation techniques and specific estimators tested are presented in Section 4. Section 5 provides the empirical results. The conclusions of the study are summarized in Section 6.

2. General Description

In this section, the problem that generated this research is presented in general terms. The population and sample design will be described in this section, and the motivation will be given in the next section.

Consider the following population design: Let r_i and c_j represent some measures of size, for $i = 1,\ldots,m$ and $j = 1,\ldots,n$. Let $N_{i,j}$ denote the number of population elements that are in the (i, j) cell, where $N = \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j}$.

An element in the (i, j) cell has a probability of being selected proportional to r_i times c_j. That is, the probability of the selection of element k, given it is in the (i, j) cell, is $z_{k,i,j} = r_i \cdot c_j \Big/ \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} \cdot r_i \cdot c_j$.

A sample of size n_t elements is drawn, where the subscript t denotes the time period the sample is drawn. Let $\pi_{k,i,j}$ denote the probability that the k^{th} element is in the sample. A systematic sample is drawn under a method that satisfies $\pi_{k,i,j} = n_t z_{k,i,j}$. Thus

$$\pi_{k,i,j} = n_t\, r_i \cdot c_j \Big/ \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} \cdot r_i \cdot c_j .$$

Let Y denote the variable of interest, which is a binomial variable that takes on the values 0 and 1, with unknown probabilities $1-p$ and p respectively. The objective of the study is to estimate p. Let $Y_{k,i,j}$ denote the value of Y for the k^{th} element in the (i, j) cell. Since there are N elements, the value of p is:

$$p = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{N_{i,j}} Y_{k,i,j} \Big/ N .$$

Since we only know the value of Y for the sample elements we need an estimator for p. One possibility is:

$$\hat{p} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ N, \qquad (1)$$

where $n_{i,j}$ is a random variable denoting the number of sample elements in the (i, j) cell, and $w_{k,i,j}$ denotes a weight, for which there are a number of possibilities. One possibility is the sampling weight; that is,

$$w_{k,i,j} = 1/\pi_{k,i,j} = \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} \cdot r_i \cdot c_j \Big/ [n_t \cdot r_i \cdot c_j],$$

resulting in the estimator:

$$\hat{p}^{s} = \frac{K}{N n_t}\sum_{j=1}^{n}\sum_{i=1}^{m}(1/r_i c_j)\sum_{k=1}^{n_{i,j}} y_{k,i,j}, \qquad (2)$$

where $K = \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} \cdot r_i \cdot c_j$.

Note that the sampling weight is approximately equal to $N_{i,j}/E(n_{i,j})$ which can be estimated by $N_{i,j}/n_{i,j}$. In this situation the estimator is:

$$\hat{p}^{p} = \left[\sum_{j=1}^{n}\sum_{i=1}^{m}(N_{i,j}/n_{i,j})\left(\sum_{k=1}^{n_{i,j}} y_{k,i,j}\right)\right]\Big/ N. \qquad (3)$$

Another possibility is a weight of 1, which results in:

$$\hat{p}^1 = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} y_{k,i,j} \Big/ n_t . \quad (4)$$

Next consider estimators from a model based approach. Let $Z^*_{k,i,j} = r_i c_j$ and consider the model:

$Y_{k,i,j} = \beta_{i,j} Z^*_{k,i,j} + \varepsilon_{k,i,j}$, where $\varepsilon_{k,i,j} \sim N(0,\sigma^2 (Z^*_{k,i,j})^\delta)$.

Note that if there is a relationship between $Z^*_{k,i,j}$ and Y (or some function of Y) then the population could be ordered on $Z^*_{k,i,j}$ prior to sample selection.

We have $\hat{Y}_{k,i,j} = \hat{\beta}_{i,j} Z^*_{k,i,j}$ as the fitted model. One way to estimate the total is by taking the sum of the responses from the sampled elements plus the sum of the responses from the non-sampled elements, which is estimated by $\hat{Y}_{k,i,j}$. Thus the mean estimator is:

$$\hat{p}^* = \left[\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\sum_{k\in S} y_{k,i,j} + \sum_{k\notin S}\hat{y}_{k,i,j}\right)\right] \Big/ N , \quad (5)$$

where S denotes the sample.

The choice of δ and the portion of the data the model is applied to will determine $\beta_{i,j}$. First consider δ =1 and fit the model separately for each (i, j) cell, then

$$\hat{\beta}_{i,j} = \frac{\sum_{k\in S} y_{k,i,j}}{\sum_{k\in S} Z^*_{k,i,j}} = \frac{\sum_{k\in S} y_{k,i,j}}{n_{i,j} r_i c_j},$$

and the inner sum becomes:

$$\left(\sum_{k\in S} y_{k,i,j} + \sum_{k\notin S}\hat{y}_{k,i,j}\right) = \sum_{k\in S} y_{k,i,j} + (N_{i,j} - n_{i,j})(\hat{\beta}_{i,j} r_i c_j)$$

$$= \frac{N_{i,j}}{n_{i,j}} \sum_{k\in S} y_{k,i,j}.$$

In this situation $\hat{p}^*$ becomes

$$\hat{p}^* = \left[\sum_{j=1}^{n}\sum_{i=1}^{m}(N_{i,j}/n_{i,j})\left(\sum_{k\in S} y_{k,i,j}\right)\right] \Big/ N , \quad (6)$$

which is the same as $\hat{p}^p$.

The same result, $\hat{p}^p$, is obtained under the model with any value for δ . Note that within an (i, j) cell, $Z^*_{k,i,j}$ is the same for all k within the cell.

Now consider the situation where all the terms are estimated; that is,

$$\hat{p}^* = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{N_{i,j}} \hat{y}_{k,i,j} \Big/ N ,$$

and use the model with $\delta = 2$ and $\beta_{i,j} = \beta$ for all (i, j).

Under this model:

$$\hat{\beta} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} \frac{y_{k,i,j}}{Z^*_{k,i,j}} \Big/ n_t ,$$

which results in

$$\hat{p}^* = \frac{1}{Nn_t}\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} (y_{k,i,j}/r_i c_j)\{\sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} r_i c_j\} \quad (7)$$

which is the same as $\hat{p}^s$.

If the model considered is :

$Y_{k,i,j} = \alpha + \varepsilon_{k,i,j}$, where $\varepsilon_{k,i,j} \sim N(0,\sigma^2)$,

then $\hat{\alpha} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} y_{k,i,j} \Big/ n_t$ and hence,

$$\hat{p}^* = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{N_{i,j}} \hat{y}_{k,i,j} \Big/ N = N\hat{\alpha}/N = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} y_{k,i,j} \Big/ n_t , \quad (8)$$

which is the same as $\hat{p}^1$.

All the above estimators of p assume that N is known. In practice this is not necessarily true.

Consider the following model where it is assumed that $\delta = 0$, and the independent variable is equal to one for all cases. That is,

$Y_{k,i,j} = \beta_{i,j} + \varepsilon_{k,i,j}$, where $\varepsilon_{k,i,j} \sim N(0,\sigma^2)$.

$\hat{Y}_{k,i,j} = \hat{\beta}_{i,j}$, where $\hat{\beta}_{i,j} = \sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ \sum_{k=1}^{n_{i,j}} w_{k,i,j}$.

The weights $w_{k,i,j}$ serve the purpose of making $\hat{\beta}_{i,j}$ a design consistent estimator of the population regression coefficient $\beta_{i,j}$.

When the total is estimated by:

$$\hat{T} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{N_{i,j}} \hat{y}_{k,i,j} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{N_{i,j}} \hat{\beta}_{i,j} = \sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j}\hat{\beta}_{i,j},$$

then the estimator of the mean becomes:

$$\hat{p}^* = \frac{\hat{T}}{N} = \frac{1}{N}\sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j}\hat{\beta}_{i,j} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j} \sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ \sum_{k=1}^{n_{i,j}} w_{k,i,j}}{N} \quad (9)$$

If the (i, j) cells are formed into h adjustment cells, the above formulae is the same as 2c and 3c in Section 4. Instead of modeling in each cell, as above, consider modeling over the entire data set. In this case $\beta_{i,j} = \beta$, which is estimated by:

$$\hat{\beta} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} w_{k,i,j} ,$$

which leads to the estimator

$$\hat{p}^* = \frac{1}{N}\sum_{j=1}^{n}\sum_{i=1}^{m} N_{i,j}\hat{\beta} = \hat{\beta} = \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{n_{i,j}} w_{k,i,j} \quad (10)$$

If the (i, j) cells are formed into h adjustment cells, the above formula is the same as 2a and 3a in Section 4. Note that the estimators in (9) and (10) will be about the same if $\sum_{k=1}^{n_{i,j}} w_{k,i,j} y_{k,i,j} \Big/ \sum_{k=1}^{n_{i,j}} w_{k,i,j}$ is about the same for each (i, j); that is, when $\hat{\beta}_{i,j}$ is approximately equal to $\hat{\beta}$.

3. Description of RAS

3.1 Background

Payroll employment for the United States are available from two major Bureau of Labor Statistics' (BLS) programs. The Employment and Wages (ES-202) program is compiled using Unemployment Insurance (UI) reports which virtually all businesses must file quarterly with their State. The Current Employment Statistics (CES) program is a monthly survey of nearly 400,000 employers. The CES employment estimates are available approximately one month after the reference month while the ES-202 universe employment data are available five months after the reference quarter. The ES-202 universe employment counts are used to adjust the CES employment estimates on an annual basis.

The employment data collected in these two programs are conceptually the same for all but a few employers which have employees who are exempt from the State UI laws. Many of the reported employment to the two programs are different, and some are substantially different. These reported employment differences were the motivation for a Response Analysis Survey (RAS) of about 8,000 CES reporters in ten cooperating states. The RAS was designed to determine the causes of reporting differences in the two programs and to evaluate the overall quality of collected employment..

3.2 Population and Sample Design

Conceptually, the population under study consisted of all sample units reporting to the CES survey, since they are all covered under the UI laws. Practically, the population excluded delinquent ES-202 reporters and other categories of CES reporters such as those with special reporting arrangements.

The requirements for the RAS sample design included both reporters with reporting differences in the two programs and reporters with no reporting differences. Since most of the reported employment from the two programs are equal or nearly equal and most of the CES sample units are of small and moderate sizes, an equal probability selection design would result in selecting very few CES reporters with reporting differences and mostly small and moderately sized employers. One method of satisfying the requirements of oversampling large CES reporters and those with reported employment differences was to assign to each unit a measure of size which increases as the reporting size increases and also as the reported employment difference increases. Each population unit was assigned an employment size class (from 1 to 9) and an employment difference size class (from 1 to 6) based on the percentage difference between the reported CES and UI employment. This percentage difference was calculated by dividing the average absolute employment difference by the average CES employment over the months that have reported data for both programs. If the average CES employment is zero, then the percentage difference is not defined and the unit is put in size class 1. The sample was selected with probability proportional to the measure of size which is the product of these two size classes. The population units were ordered by this measure of size and selected systematically.

Since it takes each state about a year to interview 800 sample units, the sample was selected in four panels about three months apart to insure that we have the most recent data available at the time of interview. The measures of size were recalculated using updated information immediately prior to the selection of each panel.

4. Estimation

Due to the movement of establishments into and out of the CES survey, the population at each panel is not fixed. The inference population is established across all four panels by taking the union of all four panel populations. Estimation for the inference population could be done separately for each panel and averaged across panels. Another alternative is to combine the four panels and consider that all sample units were from one panel and base the estimation on the combined panel.

Our objective is to determine an estimator for the inference population mean, $\overline{Y}$ (which was referred to as p in section 2), whose population size is N.

4.1 Non Response Adjustment

The weighting approach, in which the sampling weights for responding units are inflated by dividing them by estimates of the probability of response is used to handle unit nonresponse. Every population unit is assumed to have a non-zero probability of responding if sampled. The simplest nonresponse model assumes that all population units have the same probability of responding and that data are missing at random throughout the population. Another approach is to classify sample units into adjustment cells. This approach assumes that the response rate is different from cell to cell and that data are missing at random within the adjustment cells.

We allow that large establishments and establishments with large differences between the reported employment in the two programs have a different response rate than smaller establishments and establishments with small differences. The adjustment cell will be percent difference (PD) by size where:

PD Class	Percent Diff.	Size Class	Employment
1	[0,5)	1	0 - 09
2	[5,20)	2	10 - 19
3	$[20,\infty)$	3	20 - 49
		4	50 - 99
		5	≥ 100

Let,

N be the inference population size,

N_h be the inference population size in adjustment cell h,

N_{ht} be the population size in adjustment cell h in panel t,

N_t be the population size in panel t,

n_h be the total sample size in all four panels in adjustment cell h,

n_t be the sample size in panel t,

$noob_t$ be the number of out of business units in panel t,

n_{ht} be the sample size in adjustment cell h in panel t,

π_{kht} be the probability of selection for unit k in adjustment cell h in panel t,

θ_{kht} be the response probability for unit k in adjustment cell h in panel t,

y_{kht} be the characteristic measured for unit k in adjustment cell h in panel t.

Since θ is not known, it will be estimated by $\hat{\theta}$, the observed response rate. In the adjustment cell model, the response rate is the same for each establishment in an adjustment cell, so the subscript k will be dropped. In the simple model, the response rate is computed for the whole population and both subscripts k and h will be dropped. When the summation is not over adjustment cells, the subscript h will be dropped from π and y.

4.2 Estimators

1. *Unweighted Estimator* (corresponding to $\hat{p}^1$)

a. Simple Model: At each panel t, the unweighted estimator for the simple model is

$$\hat{\bar{Y}}_t = \sum_k y_{kt} \Big/ (n_t - noob_t)\hat{\theta}_t ,$$

where the summation is over the respondents in panel t.

b. Adjustment Cell Model: At each panel t, the unweighted estimator for the adjustment cell model is

$$\hat{\bar{Y}}_t = \sum_h \sum_k (\frac{1}{\hat{\theta}_{ht}} \cdot y_{kht}) \Big/ (n_t - noob_t) ,$$

where the outside summation is over adjustment cells and the inside summation is over respondents in adjustment cell h in panel t.

2. *Weighted Estimator by Separate Panel*

a. Simple Model (corresponding to $\hat{p}^*$ equation (10)): At each panel t, the ratio mean estimator for the simple model is

$$\hat{\bar{Y}}_t = \sum_k (y_{kt}/\pi_{kt}) \Big/ \sum_k (1/\pi_{kt}) ,$$

where the summation is over the respondents in panel t.

b. Adjustment Cell Model: At each panel t, the ratio mean estimator for the adjustment cell model is:

$$\hat{\bar{Y}}_t = \sum_h \sum_k (\frac{1}{\hat{\theta}_{ht}} \cdot \frac{y_{kht}}{\pi_{kht}}) \Big/ \sum_h \sum_k (\frac{1}{\hat{\theta}_{ht}} \cdot \frac{1}{\pi_{kht}}) ,$$

where the outside summation is over the adjustment cells and the inside summation is over the respondents within adjustment cell h in panel t.

c. Post-stratified: (corresponding to $\hat{p}^*$ equation (9)): At each panel t, the post-stratified estimator for the simple model is

$$\hat{\bar{Y}}_t = \frac{1}{N_t} \cdot \sum_h [N_{ht} \cdot \frac{\sum_k (y_{kht}/\pi_{kht})}{\sum_k (1/\pi_{kht})}] ,$$

where the outside summation is over adjustment cells and the inside summations are over the respondents within adjustment cell h in panel t. Again the response rate is canceled out.

3. *Weighted Estimator by Combined Panel*

Alternatively, the four panels of sample units are combined in one single panel and the expansion estimator is applied to this single panel. When the units are put together in one panel, the inverse of the probability of selection is no longer appropriate. The expansion N_h/n_h will be used. The subscript t will be dropped for the mean in the following formulas.

a. Simple Model (corresponding to $\hat{p}^*$ equation (10)):: The combined-panel estimator for the simple model is

$$\hat{\bar{Y}} = \sum_h \sum_k (\frac{N_h}{n_h} \cdot y_{kht}) \Big/ \sum_h \sum_k (\frac{N_h}{n_h}) ,$$

where the outside summation is over adjustment cells and the inside summation is over respondents within adjustment cell h in panel t.

b. Adjustment Cell Model: The combined-panel estimator for the adjustment cell model is

$$\hat{\bar{Y}} = \sum_h \sum_k (\frac{N_h}{n_h} \cdot \frac{1}{\hat{\theta}_h} \cdot y_{kht}) \Big/ \sum_h \sum_k (\frac{N_h}{n_h} \cdot \frac{1}{\hat{\theta}_h}) ,$$

where the outside summation is over adjustment cells and the inside summation is over respondents within adjustment cell h in the combined panel.

c. Post-stratified: (corresponding to $\hat{p}^*$ equation (9)): The post-stratified estimator for the combined panel is:

$$\hat{\bar{Y}} = \frac{1}{N} \cdot \sum_h N_h \frac{\sum_k \frac{N_h}{n_h} \cdot y_{kht}}{\sum_k \frac{N_h}{n_h}} ,$$

where the outside summation is over the adjustment cells and the inside summation is over respondents within adjustment cell h in the combined panel.

4.3 Variance Estimator

The data analysis software, SUDAAN, was used to facilitate the estimation of variances which were calculated via Taylor Series Linearization method. Finite population correction factors were ignored..

5. Empirical Results

The RAS study involved ten states with approximately 800 sample units being interviewed in each state. Not all of the interviews were completed by the time this study began. We decided to use data from the three states that had the highest number of completed interviews from the first two panels: Florida with 334 interviews, New York with 316, and Oregon with 307. The RAS questionnaire consists of more than 30 questions for each of the CES and ES-202 respondents. Responses to four of the questions were used in the empirical comparisons. We selected the questions so that we would have at least one question for each of the three components that comprise correct reporting: timing, method, and content. These questions were also asked of all respondents. (Some questions were asked of only some respondents that correspond to skip patterns.) Timing refers to the time period for which employment is reported. The correct time period is the pay period that includes the 12th of the month. The correct method of counting employment involves an unduplicated count of individuals who worked or received a check or other form of payment during the pay period. The content component refers to the kind of employees to be included in the employment count. For the ES-202 program, it is based on who is covered by the state Unemployment Insurance laws. The content component may vary slightly between the CES and ES-202 programs. The following four questions were used in the empirical comparisons:

- Q10: Do you use the same pay period for all your employees? (timing)
- Q14: Is the employment figure you use for the monthly BLS report obtained from your payroll system? (method)
- Q29: Can [the employment count] include a person more than once? (content)
- Q31: What time period does this employment count represent? (timing)

Estimates for Q10 and Q14 represent the proportions of firms that answer 'yes' to these two questions. Estimates for Q29 represent the proportions of firms that answer 'no' to this question. Estimates for Q31 represent the proportions of firms that report employment for the pay period that includes the 12th of the month.

Results from the three states are similar. Only results from Florida will be presented. The results for Florida are shown in Tables 1-3 in the Appendix. The estimates and their standard errors are given in Table 1. The five types of estimates by panel are computed by taking the average of the two panel estimates. The associated standard error is the standard error of the average.

For each question and each estimation method, there was not much difference in whether the nonresponse correction was done globally or by strata. That is, for each question, the estimates by adjustment cells or by simple non response methods are similar. Thus, the global nonresponse correction is acceptable.

The differences between the CES and ES-202 estimates are shown in Table 2. For Q10, Q29, and Q31, all estimation methods produced positive differences. Similarly, all estimation methods produced negative differences for Q14. The unweighted estimates produced the largest differences for all four questions. The weighted estimates by combined panel produced the smallest differences for three of the four questions. These differences are similar across estimation methods.

The maximum and minimum values from among the eight estimates for each question and program are given in Table 3. The minimum values were produced by the unweighted estimates for six of the eight groups. The weighted estimates by panel produced the maximum estimates in five of the eight groups. The other three maximum estimates were produced by the combined panel estimates. The post-stratified estimate by panel and the combined panel adjustment cell model each produced three of the maximum estimates. The only estimation procedure that did not produce any maximum or minimum estimate was the combined panel simple ratio method. The difference between the maximum and minimum values for each question is relatively small.

6. Conclusion

All methods produced similar standard errors. In general, the unweighted methods produced smaller mean estimates than the other methods. However, due to the potential bias inherent in these methods, we do not recommend them. The remaining methods produced very similar results, whether or not we used the global nonresponse correction and whether or not we computed estimates by panel. As Table 2 and Table 3 show, the differences between the program responses are similar across methods, and the difference between the maximum and the minimum estimates within a response is relatively small. Based on the simplicity in implementation, the simple ratio method by panel is recommended. The simplicity comes from the fact that the probability of selection is readily available and the nonresponse adjustment factor need not be calculated. In addition, the underlying model for this estimator is intuitive. The model and resulting estimator, which is given in Section 2 equation (10), is very simple. It basically assumes that the value of the dependent variable for a particular unit is equal to the mean over the entire data set plus a random noise. By modeling over the entire data set, an outlier in any one cell will not have as much influence on the estimate. The collected data showed no evidence that units in different cells have different response rates.

References

Cochran, W.G. (1977), *Sampling Technique*, John Wiley & Sons, Inc.

Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J., and Kott, P.S. (1995), *Business Survey Methods*, John Wiley & Sons, Inc.

Deville, J.C., Sarndal, C.E., and Sautory, O. (1993), "Generalized Raking Procedures in Survey Sampling," *Journal of the American Statistical Association*, vol. 88, no. 423, Theory and Methods.

Kalton, G., and Maligalig, D.S. (1991), "A Comparison of Methods of Weighting Adjustment for Nonresponse," *Proceedings of the Bureau of Census Annual Research Conference*.

Lessler, J.T. and Kalsbeek, W.D. (1992), *Nonsampling Error in Surveys*, John Wiley & Sons, Inc.

Little, R.J. (1984), "Survey Nonresponse Adjustments," *The American Statistical Association Proceedings*.

Sarndal, C.E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag.

Stukel, D.M., and Boyer, R. (1992), "Calibration Estimation: An Application to the Canadian Labour Force Survey," *Social Survey Methods Division, Statistics Canada, Working Paper* SSMD # 009E.

Appendix - Empirical Results Using Florida Data

Table 1. Estimates And Their Standard Errors (in parentheses)

	Timing		Method		Content		Timing	
	Q10		Q14		Q29		Q31	
	CES	ES-202	CES	ES-202	CES	ES-202	CES	ES-202
By Panel -- (Average of Panels 1 and 2)								
Unweighted - Simple	82.84 (2.15)	79.52 (2.29)	89.77 (1.74)	95.69 (1.18)	90.02 (1.69)	68.21 (2.64)	85.01 (2.02)	60.52 (2.80)
Unweighted - Adj. Cell	82.92 (2.15)	79.46 (2.30)	89.69 (1.76)	95.60 (1.21)	90.21 (1.67)	68.30 (2.67)	84.90 (2.03)	60.21 (2.84)
Ratio - Simple	84.66 (2.57)	82.01 (2.67)	90.45 (1.83)	95.69 (1.34)	92.41 (1.57)	75.64 (2.61)	86.40 (2.23)	66.17 (3.38)
Ratio - Adj. Cell	84.70 (2.53)	81.95 (2.65)	90.35 (1.85)	95.61 (1.36)	92.49 (1.57)	75.42 (2.65)	86.33 (2.24)	65.82 (3.42)
Post- Stratified	84.59 (2.45)	81.79 (2.57)	90.94 (1.71)	95.74 (1.30)	92.49 (1.57)	75.18 (2.68)	86.70 (2.25)	65.22 (3.51)
Combined Panel								
Ratio - Simple	85.30 (2.38)	82.49 (2.53)	89.91 (2.03)	95.59 (1.27)	90.93 (1.86)	73.44 (2.86)	82.45 (2.77)	63.79 (3.30)
Ratio - Adj. Cell	85.47 (2.39)	82.67 (2.55)	89.75 (2.06)	95.54 (1.29)	90.85 (1.90)	73.40 (2.90)	82.19 (2.82)	63.78 (3.34)
Post-Stratified	85.31 (2.44)	82.57 (2.58)	89.81 (2.07)	95.65 (1.26)	90.73 (1.94)	73.29 (2.92)	82.12 (.286)	63.62 (3.37)

Table 2. Differences Between Program Responses (CES and ES-202)

	Timing	Method	Content	Timing
	Q10	Q14	Q29	Q31
By Panel -- (Average of Panels 1 and 2)				
Unweighted - Simple	3.32	-5.92	21.81	24.49
Unweighted - Adj. Cell	3.46	-5.91	21.91	24.69
Ratio - Simple	2.65	-5.24	16.77	20.24
Ratio - Adj. Cell	2.76	-5.26	17.07	20.51
Post-Stratified	2.81	-4.80	17.32	21.48
Combined Panel				
Ratio - Simple	2.81	-5.68	17.49	18.66
Ratio - Adj. Cell	2.80	-5.79	17.45	18.41
Post-Stratified	2.74	-5.84	17.44	18.50

Table 3. Differences Between Minimum And Maximum Estimates Across All 8 Estimation Techniques

	Q10		Q14		Q29		Q31	
	CES	ES-202	CES	ES-202	CES	ES-202	CES	ES-202
Maximum	85.47	82.67	90.94	95.74	92.49	75.64	86.70	66.17
Minimum	82.84	79.46	89.69	95.54	90.02	68.21	82.12	60.21
Difference	2.63	3.21	1.25	0.20	2.47	7.43	4.57	5.96