REDESIGN OF THE SURVEY OF CONSTRUCTION PERMIT SAMPLE ESTIMATION PROCEDURE

Katherine J. Thompson, Angela M. Rekowski, Lynn M. Imel, U.S. Census Bureau[1]
Lynn M. Imel, U.S. Census Bureau, Room 2136-4, Washington DC, 20233

Key Words: ratio estimation, complex sample survey, collapsing

1. Introduction

The Census Bureau conducts the Survey of Construction (SOC) to provide monthly estimates of residential building activity. These economic indicators include total number of housing units started and completed for single-family and multi-unit buildings at the regional and national levels. The SOC universe comes from two surveys, with the majority of the sample in the Survey of the Use of Permits (SUP). The SUP design is described in Section 2.1.

The unbiased SUP estimates are ratio adjusted to Building Permit Survey (BPS) estimates of permit authorized houses. Currently, the ratio adjustment uses 14 estimation cells (based on the number of housing units, region, and metropolitan status), separately repeating the estimation procedure for 60 months of permit authorizations. Authorization month estimates are then aggregated. Thus, each authorization month's estimate is a combined ratio estimate. Considering authorization months as strata, the aggregated estimates are separate ratio estimates. This ratio estimation procedure is described in Section 2.3.

Under the current procedure, some estimation cells contain a small number of observations, yielding potentially unstable estimates. In order to increase the number of observations per estimation cell, we evaluated several modifications to the procedure using four non-consecutive months of survey data. Results are presented in Sections 3 and 4, along with the recommended new ratio estimation methodology. Section 5 provides a comparison between the current and recommended procedure including estimates, standard errors, and coefficients of variation for six months of SUP data. Section 6 contains our conclusions.

2. Background

SOC estimates are the sum of estimates from two different surveys: the Survey of the Use of Permits (SUP) and the Nonpermit Survey (NP). Both surveys share the same first stage of selection. If a building permit is required for new construction in a land area within the PSU, then the land area within the PSU is included in the SUP frame of permit-issuing places. Otherwise, it is included in the NP frame. We did not consider changing the procedure used to calculate NP estimates, thus they are not discussed in the following sections of the paper.

2.1 SUP Sample Design

The SUP is a stratified three-stage cluster design. The first stage is a PPS subsample of PSUs, performed **once** in 1984. The second stage is a stratified systematic sample of permit issuing places within sample PSUs (also performed in 1984). The third stage is performed monthly: each month, field representatives select a sample of building permits from the permit offices in each sampled permit-issuing place. All sampled multi-unit and single family not-for-sale buildings are followed from permit authorization until completion. Single-family for-sale buildings are followed from permit authorization until completion or sale (whichever is last).

2.2 Characteristics of Housing Data

After a permit is selected for the SUP, its construction activity is monitored. Housing units on sampled permits are allowed up to 60 months to be started or completed. Any structure that contains two or more housing units is considered multi-unit for this paper.

The majority of houses are started within 3 months of permit authorization. The completion activity begins where the start activity ends, and most units are generally completed 9 months after their start.

---

[1] This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

2.3 Current Ratio Estimation Procedure

We use the following notation to describe the current and proposed ratio estimation procedures. Let j be the number of months since the permit was authorized, where $j = 1$ corresponds to the current survey month ($j = 1, 2, ..., 60$).

X_{ij} is the true number of permits authorized in estimation cell i in authorization month j. Y_{ij} is the true number of housing units that have been started/completed during the survey month in estimation cell i whose permit was authorized in month j. $\hat{X}_{kj}$ is the auxiliary estimate of total number of permits authorized in estimation cell k in authorization month j obtained from the BPS (k is not necessarily equal to i). These are control total estimates. $\hat{x}_{ij}$ is the unbiased SUP estimate of X_{ij} and $\hat{y}_{ij}$ is the unbiased SUP estimate of Y_{ij}.

The first step of the current **within-authorization-month** ratio estimation procedure is to calculate ratios of unbiased estimates ($\hat{y}_{ij}$ / $\hat{x}_{ij}$) for the 14 estimation cells displayed in Table 1 (for each authorization month j). Region names are abbreviated as follows: Northeast (NE), Midwest (MW), South (S), and West (W).

Table 1: Current Estimation Cells For Unbiased Rates Estimates For a Given Authorization Month

<table>
<tr><th rowspan="2"></th><th colspan="2">1 HU</th><th colspan="2">2 HU</th><th colspan="2">3-4 HU</th><th colspan="2">5+ HU</th></tr>
<tr><th>MET</th><th>NON</th><th>MET</th><th>NON</th><th>MET</th><th>NON</th><th>MET</th><th>NON</th></tr>
<tr><td>NE</td><td>1</td><td>5</td><td colspan="2" rowspan="4">9</td><td colspan="2" rowspan="4">10</td><td colspan="2">11</td></tr>
<tr><td>MW</td><td>2</td><td>6</td><td colspan="2">12</td></tr>
<tr><td>S</td><td>3</td><td>7</td><td colspan="2">13</td></tr>
<tr><td>W</td><td>4</td><td>8</td><td colspan="2">14</td></tr>
</table>

These ratios are multiplied by the BPS control total estimates ($\hat{X}_{kj}$) to get combined ratio estimates (for each authorization month) of the characteristic for all 32 estimation cells.

Table 2: Ratio Estimation Procedure For a Given Authorization Month

	1 HU		2 HU		3-4 HU		5+ HU	
	MET	NON	MET	NON	MET	NON	MET	NON
NE	1	5	9	13	17	21	25	29
MW	2	6	10	14	18	22	26	30
S	3	7	11	15	19	23	27	31
W	4	8	12	16	20	24	28	32

This procedure is repeated 60 times for each authorization month j. The combined ratio estimates for a given cell k ($k = 1, 2, ..., 32$) are then summed over cells (k) and authorization months (j) to obtain totals.

Note that the single-family estimates are direct estimates, and the multi-unit estimates are synthetic. Even when the SUP sample was redesigned in 1984, the expected numbers of sampled permits in the multi-unit cells were small. Permits for 5+ unit structures are included with certainty, but smaller multi-unit structures are sampled, and the number of sampled permits in those estimation cells is very small, especially when disaggregated by authorization month. There are similar sample size issues with the single-family non-metropolitan estimation cells; approximately 20 percent of the SUP sample is located in non-metropolitan areas.

3. Evaluation of the Within-Authorization-Month Ratio Estimation Procedure: Changes to Ratio Estimation Cells.

We first examined the within-authorization-month ratio estimation procedure performed for each authorization month. Table 3 presents the outcome of this evaluation: a new within-authorization-month ratio estimation procedure. The asterisks distinguish new estimation cells from the old estimation cells.

Table 3: Revised Procedure's Estimation Cells for Within-Authorization-Month Ratio Estimates

	NE	MW	SO	WE
1 HU	(cell 1)*	(cell 2)*	(cell 3)*	(cell 4)*
2+ HU	(cell 5)*	(cell 6)*	(cell 7)*	(cell 8)*

Our primary goal was to simplify the SOC within-authorization month estimation procedure without sacrificing precision. We were interested in making the following changes: drop the metropolitan cross-classification in the ratio estimation procedure for single-family cells; drop the separate synthetic ratio estimation procedure for 2 and 3-4 unit cells. Instead, perform a multi-unit ratio adjustment (structures with 2+ units) by region; and use direct estimation instead of synthetic estimation for multi-unit structures.

3.1 Examination of Sample Sizes

Initially, sample size considerations motivated our investigation. We have two sample size considerations: the unweighted number of permits authorized in month j of estimation cell i, denoted n_{ij}; and the unweighted number of sampled permits with the characteristic of interest, denoted m_{ij}.

In the single family non-metropolitan cells and the multi-unit cells, the m_{ij}'s are very small (usually less than 15). An analysis based strictly on unweighted sample sizes can be misleading when using data from a sample survey, so we need to relate the sample size to the variance of the unbiased estimate. We do this using the "effective sample sizes" of y_{ij} and x_{ij} as defined in Rao and Scott (1992). Recall that the magnitude of CV($\bar{x}$) and CV($\bar{y}$) are directly related to the variance of the unbiased estimates, which in turn are functions of the efficiency of the sample survey design for the characteristics under consideration. Using design effects (Kish, 1965) measures the efficiency of the SUP sample relative to that of a SRS. The design effect of an estimated proportion (start or completion rate) in estimation cell i, authorization month j, is given by

$$D_{ij} = (\hat{V}_{pij}) / (\hat{P}_{ij}(1 - \hat{P}_{ij}) / n_{ij})$$

where $\hat{V}_{pij}$ is the modified half-sample (MHS) variance estimate of the proportion of characteristic y in estimation cell i, authorization month j. These are sample survey variance estimates. $\hat{P}_{ij} = \hat{y}_{ij} / \hat{x}_{ij}$ and n_{ij} is the number of sampled permits authorized in estimation cell i, authorization month j.

The "effective sample sizes" are obtained for y_{ij} and x_{ij} by dividing m_{ij} and n_{ij} by their respective D_{ij}. Note that when $D_{ij} < 1$, the survey sample is more efficient than a simple random sample.

For our analysis, we calculated **average** design effects for the 14 estimation cells from four months of SOC data. Our $\hat{V}_{pij}$ used a four-month **average** of MHS variance/covariance estimates, as recommended in Bell (1999), as do our SRS variance estimates.

Single unit cells' average design effects for houses both started and completed (using the current estimation cells as defined in Table 1) are generally larger than one, probably due to the clustering of permits within selected places. Consequently, our effective sample sizes are much smaller than the unweighted sizes (many of which were already problematic). Combining estimation cells by eliminating the metropolitan cross-classification is an obvious way to address some of the single-family sample size concerns.

In contrast, the multi-unit cells' average design effects for both characteristics are quite small, perhaps unreasonably so. The design effects in the 2 and 3-4 unit estimation cells are calculated from a small number of sampled permits and should be evaluated accordingly. The design effects for the 5+ unit cells are slightly more believable, since most of the survey sampling variability in those cells comes from the first and second stages of selection.

At any rate, the effective sample size issue isn't relevant to multi-units because of these small design effects. However, there are usually less than five unweighted observations per estimation cell in any given authorization month that have the characteristic (m_{ij}), resulting in highly variable estimates. Performing the ratio adjustment by region for all multi-unit structures – that is, omitting the ratio estimation procedure for 2 and 3-4 unit structures – is not expected to yield much improvement in the stability of the multi-unit estimates because of the small contribution to the total multi-unit sample from those cells.

3.2 Revisions to Single-Family Ratio Adjustment Procedure: Omitting Metropolitan Status Adjustment

The current single-family ratio estimation procedure is an eight-celled adjustment (region by metropolitan status). If the estimated stratum proportions ($\hat{P}_{ij}$) vary from stratum to stratum, then using a separate ratio estimator can result in large gains of precision relative to a simple expansion estimate (Cochran, 1977 and Wang and Takeuchi, 1995). Otherwise, a combined ratio estimator is preferable.

When the stratum proportions are constant, then there is no estimation benefit from a separate ratio estimator. That is, if estimation cells i and i' have the same proportion for a characteristic, then collapsing the two cells will neither change the mean of the ratio-adjusted estimator nor the variance (Wang and Takeuchi, 1995).

We performed Bonferonni adjusted z-tests within region cell and authorization month to test the hypothesis that $P_{region,\ MET,\ j} = P_{region,\ NONMET,\ j}$. In general, the cell proportions for single-family housing starts did not significantly differ by metropolitan status cell within region.

These results need to be interpreted cautiously. Both the unweighted and effective sample sizes for non-metropolitan single-family starts and completions are quite small, and the variances on those cell proportions consequently can be fairly high, affecting the magnitude of absolute difference that we can detect at the 95% confidence level. Nevertheless, we felt that these results provided sufficient evidence to justify eliminating the ratio adjustment by metropolitan status for single-family structures.

3.3 Revisions to the Multi-Unit Ratio Estimation Procedure

Our decision to eliminate the separate ratio estimation procedure for 2 and 3-4 unit structures was intuitive. First, it is impossible to perform any type of statistical test to evaluate equivalence of cell proportions with the current ratio estimation procedure. Second, the effective sample sizes in most cells were quite reasonable because the design effects for multi-unit estimation cells were so small. However, especially in the 2 and 3-4 unit cells, the average unweighted number of sampled permits (n_{ij}) in our data sets was quite small. For the starts data sets, there were about 25 permits per cell (aggregated over all authorization months) in the 2 and 3-4 unit cells. For the completions data sets, this number decreased to 20 per cell (again, aggregated over all authorization months).

Cochran (1977) states that the bias in a combined ratio estimator is negligible relative to the standard error provided that CV($\hat{x}_{ij}$) is below 0.10. We used this rule to evaluate whether to perform the 2 and 3-4 unit ratio estimations. Recall that the individual ratio estimates are accumulated over authorization months to obtain the estimate. The **aggregate** bias relative to the **aggregate** standard error can be much larger than the individual cell bias relative to the cell standard error.

Average CV($\hat{x}_{ij}$) used for the housing starts and completions ratio adjustment in these cells are all considerably larger than 0.10. Based on these results, we decided to combine the 2 and 3-4 unit structures with the structures containing 5+ units (a 2+ classification).

4. Determining the Number of Authorization Month "Strata" for the Separate Ratio Estimator

4.1 Preliminary Analysis

We next turned our attention to the separate ratio estimation procedure. The current procedure repeats the within-authorization-month ratio estimation separately for 60 authorization months. The purpose of ratio estimation is to increase the precision of the estimate by taking advantage of the high correlation between the characteristic (y) and the controlling variable (x). With the data, this correlation decreases to zero as j increases. Consequently, we decided to combine the units authorized after a certain number of authorization months (denoted as "s*" in equation 4.1) into a single estimation "stratum," yielding a separate ratio estimator of the form

$$\hat{Y}_i = \sum_{j=1}^{s^*-1} \hat{X}_{ij} \left(\frac{\hat{y}_{ij}}{\hat{x}_{ij}} \right) + \sum_{j=s^*}^{60} \hat{X}_{ij} \left(\frac{\sum_{j=s^*}^{60} \hat{y}_{ij}}{\sum_{j=s^*}^{60} \hat{x}_{ij}} \right) \qquad (4.1)$$

where i indexes the within-authorization-month estimation cells described in Section 3. The correlation in the final cell is expected to be non-zero.

We chose our s* as the union of our "best" s* for single-family housing starts and our "best" s* for single-family housing completions. We used the new single-family estimation cells for these analyses since it comprises the majority of the SUP sample.

To get a feel for candidate values of s*'s, we examined published estimates and plots of estimated proportions of total and single-family houses started and completed against number of months since authorization.

A separate ratio estimator works well when the n_{ij}'s are large and the number of strata (J) is small. Cochran (1977) notes that the ratio of the bias of the separate ratio estimator to the standard error is of order $\lambda = \sqrt{J} \times cv(\hat{x}_{ij})$. If this quantity is greater than 0.30 then this additional bias is not negligible. Using MHS replicate estimates of $cv(\hat{x}_{ij})$, we calculated λ for single-family starts and completions estimates (i = 1, 2, 3,4) for J = 60, 18, and 13. When J=60, the λ were considerably larger than 0.30. As expected for housing starts, the reduced number of strata did not reduce this statistic. However, for housing completions, in all cases reducing J to 13 reduced λ to less than 0.30, and in most cases reducing J to 18 also reduced λ to less than 0.30. This provided some evidence that collapsing the authorization strata cells would have beneficial results for the estimates.

Cochran's rule assumes single-stage simple random stratified sampling with a textbook variance estimator. Since the standard textbook rules did not strictly apply to our data sets, we used a modeling approach to determine a recommended value of s*.

4.2 Logistic Regression Analysis

We were interested in measuring the contribution of the individual authorization-stratum estimates to total proportion of houses started/completed in estimation cell i (regions 1 though 4). Since y_{ij} is a binary outcome, we used the following logistic regression model for each estimation cell i: $f(Y_i) = log(p_i/(1-p_i)) = X_i\beta_i$ where Y_i is a $n_i \times 1$ vector of 1's and 0's indicating whether a unit in estimation cell i was started/completed during the current survey month, X_i is an $n_i \times J$ design matrix of indicator variables containing J-1 indicator variables for number of months since authorized, and β_i represents the contribution from the estimate in authorization month stratum j to the probability of being started/completed during the given month.

This saturated model measures the contribution to the total estimate of cell i from each authorization month stratum estimate. It is **not** a predictive model. We are interested in the **significance** of the individual parameters. If an individual β_{ij} is not significant, then the authorization month stratum that it represents is a candidate for collapsing with an adjacent month.

There are well-documented modifications to the standard logistic regression estimation techniques that take the sample survey design into account. Roberts, Rao, and Kumar (1987) showed that the correct estimated asymptotic covariance matrix of $\hat{\beta}_i$ is given by

$$\hat{V}_{\beta_i} = (X_i'\hat{\Delta}_i X_i)^{-1}\{X_i' D_i(w)\hat{V}_{pi} D_i(w) X_i\}(X_i'\hat{\Delta}_i X_i)^{-1}$$

where X_i is the $J{\times}J$ design matrix and $\hat{V}_{pi}$ is our MHS replicate variance-covariance matrix of the sample survey proportions in cell i;

$$D_i(w) = diag\left(\frac{\hat{x}_{i1}}{\hat{x}_i}, \ldots, \frac{\hat{x}_{iJ}}{\hat{x}_i}\right) = diag\left(W_{i1}, \ldots, W_{iJ}\right);$$

and

$$\hat{\Delta}_i = diag\left(W_{i1}\hat{f}_{i1}(1-\hat{f}_{i1}), \ldots, W_{iJ}\hat{f}_{iJ}(1-\hat{f}_{iJ})\right),$$

where $\hat{f}_{ij}$ is the predicted estimate i in authorization month stratum j.

We did not fit logistic regression models for housing starts. Regression parameters did not converge for authorization month stratum greater than four because very few houses are completed within four months of permit authorization.

We fit regional models for single-family completed houses, using data from authorization months three through 60. Our first saturated model fit in each region i was:

$$f(Y_i) = \beta_{i0} + \beta_{i4}X_{i4} + \beta_{i5}X_{i5} + \ldots + \beta_{i14}X_{i14} + \beta_{i15+}X_{i15+}.$$

β_{i0} is our baseline measurement (units completed during the current survey month within three months of their authorization). We did not include authorization months one and two: those months did not contain enough unweighted observations for the parameter estimation to converge. Similarly, we arrived at 15-or-more months as our final cell by sequentially fitting single-family region models and using the smallest number of cells with which all 16 models' parameter estimates would converge (starting with J=18 as determined from 4.1.).

We performed Wald tests to test the significance of the individual parameters (given all other covariates) in the saturated models. To determine whether combinations of parameters could be dropped, we used the Wald statistic for nested hypothesis provided in Rao, Kumar, and Roberts (1989). The usual caveats about the instability of the Wald statistics calculated from sample survey data apply here, namely that the test statistics' validity are highly dependent on the consistency of the estimated variance-covariance matrix (Fay, 1985). This assumption may be questionable even using an averaged variance-covariance matrix in our calculations.

In most of our data sets, the parameter estimates from authorization months 12 and 14 were not significantly different from zero at $\alpha = 0.05$. In 9 of 16 models, we failed to reject H_0: $\beta_{14} = 0$, and in 10 of sixteen models, we failed to reject H_0: $\beta_{12} = 0$. Moreover, in 14 of our 16 data sets, the parameter estimates from authorization months 15 or greater (the final "catch-all" strata) were highly significant. Thus, collapsing authorization months at the "tail end" of the model (j = 12, 13,14) was virtually guaranteed to yield significant results for the final estimation cell in a "reduced" saturated model.

Although both the 12 and 14 month estimation cell parameters were not significant, given the other covariates, we did reject the hypothesis that $\beta_{12} = \beta_{14} = 0$ in most of our data sets. However, the **unweighted** sample sizes (both authorizations and completions) in these cells were generally quite small, which we believed affected the reliability of our tests. Consequently, we decided to collapse authorization month strata from 12-or-more months.

Further testing showed that when $\hat{\beta}_{11,i}$ was dropped from the saturated model, the other parameters contributed significantly, so no further testing was necessary. Thus, our "best" authorization month stratum for single-family completions begins with authorization month three and ends with a collapsed authorization month stratum at 11 or more months. Combining this model's s^* with the start*, we get a cut-off value of 11 months for collapsing.

This decision is based on an analysis of **single-family** structures. Multi-unit structures generally take longer to complete. Indeed, the SOC analysts did not endorse any cut-off s^* below one year after authorization for this reason. Consequently, the revised SOC estimator for housing starts and completions is

$$\hat{Y}_i = \sum_{j=1}^{11} \hat{X}_{ij}\left(\frac{\hat{y}_{ij}}{\hat{x}_{ij}}\right) + \sum_{j=12}^{60} \hat{X}_{ij}\left(\frac{\sum_{j=12}^{60} \hat{y}_{ij}}{\sum_{j=12}^{60} \hat{x}_{ij}}\right).$$

5. Comparison of Old and New Estimation Procedure on 6 Months of Data

Tables 4 and 5 compare averaged SUP estimates for six consecutive months.

Table 4: Ratio of New Starts and Completions Estimates to Old Estimates

	6 Month Average	
	Starts	Completions
US Total	1.00	0.94
NE Total	0.99	0.91
MW Total	0.98	0.94
S Total	1.01	0.95
W Total	1.01	0.95
NE, 1 HU	0.97	0.93
MW, 1 HU	0.97	0.92
S, 1 HU	0.99	0.96
W, 1 HU	0.99	0.96
1 HU	0.98	0.95
2-4 HU	1.00	0.91
5+ HU	1.07	0.94
2+ HU	1.06	0.94

Starts ratios show little difference between methods. The ratios of new to old estimates presented in Table 4 are usually near 1. This negligible difference between estimates was expected since most starts occur within 6 months of their authorization date.

The new completion estimates however, are consistently smaller than estimates using the old method by 5 to 10 percent as shown in Table 5. The differences in single unit housing estimates are much smaller than the multi-unit estimates.

Table 5: Comparison of Starts and Completions CV's with Old and New Procedures

	6 Month Average			
	Starts		Completions	
	Old CV	New CV	Old CV	New CV
US Total	0.03	0.03	0.03	0.03
NE Total	0.08	0.09	0.09	0.09
MW Total	0.06	0.07	0.08	0.09
S Total	0.04	0.04	0.04	0.05
W Total	0.04	0.04	0.05	0.06
NE, 1 HU	0.08	0.08	0.07	0.08
MW, 1 HU	0.06	0.06	0.08	0.08
S, 1 HU	0.04	0.04	0.05	0.05
W, 1 HU	0.04	0.04	0.06	0.06
1 HU	0.03	0.02	0.03	0.03
2-4 HU	0.15	0.21	0.31	0.34
5+ HU	0.09	0.09	0.10	0.11
2+ HU	0.08	0.08	0.10	0.11

CV's for starts and completions for both methods appear to be similar. As with starts, the 2-4 HU contain more variation between methods, but all other CV's are very close.

6. Conclusions

This paper presents a revised ratio estimation procedure for the SUP estimates. We developed this procedure after a careful analysis of the current estimation procedure. Over time, the conditions that originally made the current estimation procedure viable have changed. Furthermore, the average length of time between authorization and housing completion has decreased to less than a year, making the separate estimation procedure for 60 months of authorization questionable.

Applying both the old and new methods to six consecutive months of data revealed that the new estimator has little or no effect on the starts estimates. However, the completions estimates are consistently reduced by 5 to 10 percent across the board. The estimates most affected by the new estimation procedure are the 2-4 HU, and to a lesser extent, the other multi-unit estimates. Consistently smaller estimates and comparable CV's are evidence that the new procedure improves the estimator by reducing the positive bias.

Acknowledgments

The authors would like to thank Richard S. Sigman, Douglas Bond, Aref Dajani and Catherine Hood for their helpful comments on earlier versions of this manuscript. We would also like to thank George Train for his assistance.

References

Bell, W. (1999). "Recommendations on Variances of Seasonally Adjusted Change Estimates for the Survey of Construction (SOC)." Internal Census Bureau memorandum.

Cochran, W.G. (1977) *Sampling Techniques, 3rd Edition.* New York: John Wiley and Sons.

Fay, R. (1985). A Jackknifed Chi-Squared Test for Complex Samples. *Journal of the American Statistical Association,* **80**, pp. 148-157.

Kish, L. (1965). *Survey Sampling.* New York: John Wiley and Sons.

Rao, J.N.K., Kumar, S. and Roberts, G. (1989). Analysis of Sample Survey Data Involving Categorical Response Variables: Methods and Software. *Survey Methodology,* **15**, pp. 161-186.

Rao, J.N.K and Scott, A.J. (1992). A Simple Method for the Analysis of Clustered Binary Data. *Biometrics*, **48**, pp. 577-585.

Roberts, G., Rao, J.N.K., and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. *Biometrika*, **74**, pp. 1-12.

Thompson, K., Rekowski, A., and Imel, L. (1999), "Redesign of the Survey of Construction Permit Sample Estimation Procedure," MCD Working Paper Number: Census/MCD/WP-99/xx, U.S. Census Bureau.

Wang, Y. and Takeuchi, D. (1995). Post-Stratification for Proportion Estimation with Use of a Selection Weight. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 917 - 922.