AN IMPUTATION STUDY FOR THE MONTHLY RETAIL TRADE SURVEY

Elizabeth T. Huang, Bureau of the Census

I. Introduction

The Census Bureau conducts the monthly Retail Trade Survey of the business universe in order to provide timely estimates of the level and trend sales. The data for each establishment are subjected to a series of edit checks and to be imputed if they are missing. The problem of handling missing data for the Monthly Retail Trade Survey is examined in the paper. The Monthly Retail Trade Survey is composed of a list sample and an area sample, where the list sample contains 96% of the total sample size. The list sample consists of a fixed panels of certainty units (which report every month) and rotating panels of noncertainty sampling units (which report every three months). A stratified random sampling design was used (See (4)). The main variables collected in the rotating panel cases are the monthly retail sales for the current month and the previous month. For fixed panel cases, only current monthly sales are collected. These items are sometimes not reported or suppressed because of edit failure.

The current imputation procedure in the Monthly Retail Trade Survey takes advantage of the rotating nature of the sample panels and 'historical' data. The procedure operates by multiplying a nonresponding unit's 'historical' data by a measure of trend computed from those responding units whose size and kind of business characteristics are similar to the nonresponding unit's. This method assumes that trends in the nonresponse stratum are similar to those in the response stratum. The sample is partitioned into imputation cells defined by kind of business (KB), firm size (Group I and Group II) and size of sales. In each imputation cell, the trend is calculated from the reported items. If the 'current' month sales are missing, it is imputed based on the 'previous' month sales of the same unit. Let y_i be the current month sales and x_i be the previous month sales of the i^{th} unit that reported in the current month. Let z_i be the previous month sales reported 3 months ago by the i^{th} unit of the same panel. For the list sample of noncertainty units, the trends or the so-called ratios of identicals for each imputation cell are calculated by

$$\hat{R}_p = \sum w_i x_i / \sum w_i z_i \qquad (1.1)$$

$$\hat{R}_c = \sum w_i y_i / \sum w_i x_i \qquad (1.2)$$

where w_i denotes the sampling weight of the i^{th} responding unit. The summations in $\hat{R}_p$ are taken over all units in the imputation cell whose data x_i and z_i were reported. The ratio, $\hat{R}_p$, estimates the previous month to previous three months ago sales trend for each imputation cell in the domain of respondents. Similarly, the summations in $\hat{R}_c$ are taken over all units in the imputation cell whose data, y_i and x_i were reported. The ratio, $\hat{R}_c$, estimates the current month to previous month sales trend for each imputation cell in the domain of respondents.

After forming the ratio of identicals for each imputation cell, the next step is to test whether the ratio $\hat{R}_p$ satisfies the conditions $\hat{R}_p \in [m_1, M_1]$ and $N_1 > 15$, where N_1 denotes the number of units defining the ratio $\hat{R}_p$. The interval limits, m_1 and M_1, vary by KB and by month. If one or both of these conditions are not met in a given imputation cell, then the ratio $\hat{R}_p$ is recalculated over all reported x_i and z_i units within a collapsed cell which is defined by KB and firm size. In a similar manner, the ratio $\hat{R}_c$ is tested for each imputation cell for possible distortion and recalculated when necessary. If the ratio is accepted, the ratio will be used to impute the missing item. The ratio in (1.1) is used to impute the missing item in the case of previous month sales (x), and the ratio in (1.2) is used to impute the missing item in the case of current month sales (y).

Cassel, Särndal and Wretman (1979) outlined an approach that builds on an underlying linear regression model for estimation of the finite population mean when nonresponse has occurred. They developed two estimators; one estimator can be constructed to have built-in adjustment for varying response probabilities, and another estimator is simplified by leaving out such adjustment. The latter case takes the risk of design biased inferences when nonresponse occurs and the underlying model is false. They also extended the techniques to the case when only sample auxilary information is available instead of population auxilary information.

The current imputation procedure of Monthly Retail Trade Survey (as I view it) is a kind of latter case where the linear model going through the origin is assumed for each of the imputation cells of the sample. The missing item is imputed from the model using the sample auxiliary information.

For each KB, for each imputation cell ij (group size x sales size) i=1,...I, j=1, ...J, the current month sales (y) are assumed to have a linear relationship with the previous month sales (x),

$$y = R_{ij}\, x + \varepsilon, \ \varepsilon | x \sim N(0, x\sigma^2) \qquad (1.3)$$

where x is assumed to be known for every unit in the sample.

When nonresponse y occurs, the $\hat{R}_{ij}$ is calculated from the response data of imputation cell ij by using (1.2) which is a least squares estimate of R_{ij} under model (1.3) and incorporating the sampling weights. If model is true, both least squares estimate ($\Sigma y_i/\Sigma x_i$) and $\hat{R}_{ij}$ are unbiased estimate of R_{ij}. $\hat{R}_{ij}$ is one of the estimators under model (1.3) discussed in Cassel, Särndal and Wretman (1979). The imputed value for the missing item y is $\hat{R}_{ij}$ x. The current imputation procedure puts further restrictions on the estimate $\hat{R}_{ij}$. If $\hat{R}_{ij}$ is not in the prior limits $[m_1, M_1]$ or the number of respondents in cell ij is less than 15, a collapsed cell is defined within group i. The following linear model is assumed in the collapsed cell i,

$$y = R_i\ x + \varepsilon \quad \varepsilon|x \sim N(0,\ x\sigma^2) \quad (1.4)$$

which assumes that the R differs by firm group size.

The same model assumption is used for the previous month sales (x) of the current month reporting unit, and the previous month sales (z) reported 3 months ago. All missing items of the previous month sales (x) are imputed before imputing the missing items of current month sales (y).

When nonresponse occurs, under the current stratified sample design and the current imputation procedure, the Horvitz-Thompson estimator of total sales y is a ratio type estimator. (See Huang (1984).)

II. Examining Current Monthly Retail Trade Survey Data - December 1982 Retail Trade Survey - SIC 562 (Women's Ready-to-Wear Stores)

Monthly retail sales reported data were examined to see whether the current model holds. In the current imputation procedure, for each imputation cell, the current month sales (y) and previous month sales (x) are assumed to have the following relationship:

$$y = \beta\ x + \varepsilon \quad ,\quad \varepsilon \sim N(0,\ x\ \sigma^2) \qquad (2.1)$$

The missing item y_i is currently estimated by $\hat{\beta}x_i$,

where $\hat{\beta} = (\sum_i w_i y_i)\ /\ (\sum_i w_i x_i)$,

w_i is the sampling weights corresponding to unit i, and the summation is taken over all reported x_i and y_i's.

Four alternative linear regression models are examined for each imputation cell:

$$y = \alpha + \beta\ x + \varepsilon\ ,\ \varepsilon \sim n(0,\ \sigma^2) \qquad (2.2)$$

$$y = \alpha + \beta\ x + \varepsilon\ ,\ \varepsilon \sim n(0,\ x\sigma^2) \qquad (2.3)$$

$$y = \alpha + \beta\ x + \varepsilon\ ,\ \varepsilon \sim n(0,\ x^2\sigma^2) \qquad (2.4)$$

$$\log y = \alpha + \beta \log x + \varepsilon\ ,\ \varepsilon \sim n(0,\ \sigma^2) \quad (2.5)$$

If the intercept α is not different from 0, the linear regression models in (2.2), (2.3), and (2.4) will reduce to the following models:

$$y = \beta\ x + \varepsilon \quad ,\quad \varepsilon \sim n(0,\ \sigma^2) \qquad (2.6)$$

$$y = \beta\ x + \varepsilon \quad ,\quad \varepsilon \sim n(0,\ x\sigma^2) \qquad (2.7)$$

$$y = \beta\ x + \varepsilon\ ,\ \varepsilon \sim n\ (0,\ x^2\sigma^2) \qquad (2.8)$$

The least squares estimates for β in models (2.6), (2.7), and (2.8) for a simple random sample of size n are

$$(\sum_{i=1}^{n} x_i y_i)/(\sum_{i=1}^{n} x^2{}_i),\ (\sum_{i=1}^{n} y_i)/(\sum_{i=1}^{n} x_i)\ ,$$

$$(1/n)\sum_{i=1}^{n} (y_i/x_i)$$

respectively.

The data used are from the December 1982 Monthly Retail Trade Survey for SIC 562. The total sample size is 2937.

The data used for model fitting are restricted to establishments in the list sample with reported nonzero sales for both current and previous months. The sample size for the reported data is 1448.

The 4 current imputation cells and 2 collapsed imputation cells for SIC 562 are defined as follows:

Imputation cell

1. Sales* $\geq$ \$50,000, Group 2
2. Sales* < \$50,000, Group 2
3. Sales* $\geq$ \$50,000, Group 1
4. Sales* < \$50,000, Group 1

Collapsed imputation cell

1. Group 2 (firm size code = 6)
2. Group 1 (firm size code = 2,3,4)

*The sales size indicator depends on which panel the unit belongs. For fixed panel, the previous month sales are used. For rotating panels, the current month sales of 3 months ago are used.

The data of imputation cell 1 (group = 2, sales* $\geq$ \$50,000) are first used in fitting the different models. By looking at the plots of the residuals, it seems that model (2.2) with constant variance does not fit well. To find the approximate relationship of the variance of the current month sales y with the previous month sales x, units were first sorted by the previous month sales and then grouped with 20 units in each class. The variance or the standard error of y and the mean of x for each class were calculated. The relationship of the variance of y with the mean of x can be estimated by least squares method using the log transformation of the following

$$(s_y^2)_i = \lambda \ \bar{x}_i^{\rho} \ , \ i = 1,\ldots k.$$

The estimated λ, ρ for 4 imputation cells and 2 collapsed cells for the data of SIC 562 of December 1982 are tabulated in Table 2.1.

It seems that the error variance $x^2 \sigma^2$ is more appropriate than $x\sigma^2$ for each imputation cell. (The error variance for each imputation cell of other KB's and of SIC 562 of February 1983 was also investigated. See Huang (1984).) A linear model with error variance $x^2\sigma^2$ (equation (2.4)) was then used to fit each of the 4 imputation cells data to see whether the intercept is significantly different from zero.

In fitting each imputation cell data, the outliers were also examined and deleted in the analysis of residuals based on the Cook and Studentized statistics.

We can conclude from the model fitting for each imputation cell and collapsed cell of December 1982 for SIC 562:

(1) The error variance of the linear regression model for each imputation cell is approximately $x^2\sigma^2$. (Same conclusion for February 1983's data.)

(2) By fitting the linear regression model with error variance $x^2\sigma^2$ to each imputation cell, it showed that at 0.01 level the intercepts of all 4 imputation cells are not significantly different from zero; i.e., the ratio model (2.8) is more appropriate than the regression model (2.4) for the data. However, the intercepts of cells 1 and 2 are significantly different from zero with probability 0.0311 and 0.0396, respectively. (For SIC 562 of February 1983's data, the intercepts of cells 1 and 2 are not significantly different from zero. However, the intercepts of cells 3 and 4 are significantly different from zero with probability less than 0.01.

(3) The log scaled linear model (2.5) was also fitted to the data in each of the 4 imputation cells. The histograms of the standardized residuals and the scatter plots of the residuals showed no gross deviations from the assumptions of the model.

Since our data came from a stratified sample, the inclusion probability (or the sampling weight) for each sampling unit varies considerably for units in the different strata, especially between the certainty stratum and the noncertainty strata. The mean sampling weights for all reported data is 20.407, the range is from 1 to 512.080. The regression analysis described before is the standard test assuming that the data come from a simple random sample and all the model assumptions are met.

DuMouchel and Duncan (1983) proposed to use the difference between the weighted and unweighted estimates (where the weights are sampling weights) as an aid in choosing the appropriate model and hence the appropriate estimator.

The alternative way to write the ratio model (2.8) for all units in the 4 imputation cells (labled as Model A) is

$$R \equiv y/x = \sum_{j=0}^{3} \alpha_j z_j + e \ , \ e \sim n(0, \sigma^2) \quad (2.9)$$

where

$z_0 \equiv 1$

$z_i = 1$, if the unit is in the imputation cell i, $i = 1, 2, 3$
$= 0$, otherwise.

To test whether there is any difference of the weighted and unweighted regression coefficients in the model, $E\ (\beta_w - \beta) = 0$, we use method A of DuMouchel and Duncan (1983). The ordinary regressions were used to regress R on z, and R on z and zw (where zw is the variable z multiplied by the sampling weights w).

The test shows that there is a significant difference between $E(\beta_w)$ and $E(\beta)$ at 0.0157 significant level. Hence there is a difference in using sampling weights in estimating the parameters in Model A.

The alternative way to write the ratio model (2.7) (the current imputation model) for all units in the 4 imputation cells (labeled as Model B) is

$$R = y/x = \sum_{j=0}^{3} \alpha_j z_j + e, \ e \sim n(0, \sigma^2/x) \quad (2.10)$$

where z's are defined in (2.9).

The estimate of the mean rate (or trend) for each imputation cell under Models A and B with or without sampling weights (designated as w/wt, wo/wt, respectively) is given in Table 2.2.

In the current imputation procedure, the estimated mean rate from each imputation cell is checked to see whether it is in the prior limits of 1.443879 and 1.718664. If any of the estimates do not fall in the range, it will be recalculated using the appropriate collapsed cell.

It can be seen in Table 2.2 that the estimates of all 4 cells from Model A wo/wt are in the desired range, while some others indicated by '*' are outside of the range. If the present range is a good prior, we'll conclude that Model A (wo/wt) is a good model for the data.

In the next section, a simulation study is conducted. The objective of the study is to evaluate the different imputation procedures.

III. Simulation Study

One way to compare different imputation procedures is to do a simulation study. The simulation study described below uses only full reported survey data as a complete data set, and simulates the missing values from the complete data set. Different imputation procedures are then applied on the simulated data set, the imputed values are then compared with the original values. Ford, Kleweno and Tortora (1980) did a simulation study using agriculture survey data, and Kalton (1981) did a simulation study using ISDP data.

The simulation study was conducted using the Monthly Retail Trade Survey data. The complete data set is the reported list sample of December 1982 retail sales of SIC 562, where the reported and imputed codes for both current and previous months are 1. There are 1445 units. A random mechanism is used to designate the missing values from the complete data set. (i.e., It is assumed that the data are missing at random.)

For each imputation cell, the establishment's current month sales are designated missing randomly according to the current nonresponse rate for the cell. Five sets of missing data are generated.

From the previous study of the complete data set, it seems that for each imputation cell, the ratio model is a reasonable model and the model error variance is proportional to the square of the previous month sales. The current imputation procedure assumed a ratio model with model error variance proportional to the previous month sales. Models A and B defined in (2.9), (2.10) are used for the imputation comparisons.

Models A and B are defined for 4 imputation cells. It also is defined for the collapsed cells. Recall that the current imputation procedure will collapse within the group if the mean rate of any cell does not fall within the prior limits. Hence 6 imputation procedures are applied to the 5 simulated data.

1. Model A (4 cells) without using sampling weights in the estimation procedure [Model A (4 cells) wo/wt].

2. Model A (4 cells) incorporating the sampling weights in the estimation procedure [Model A (4 cells) w/wt].

3. Model B (4 cells) without using sampling weights in the estimation procedure [Model B (4 cells) wo/wt].

4. Model B (4 cells) incorporating the sampling weights in the estimation procedure [Model B (4 cells) w/wt].

5. Model B (2 cells) incorporating the sampling weights in the estimation procedure [Model B (2 cells) w/wt].

6. Current procedure: Use Model B (4 cells) w/wt to estimate the mean rate, if any of these rates falls off the range, the rates will be recalculated using the collapsed cells.

The criteria used to evaluate different imputation procedures are the following:

A. The mean deviation defined as $\sum (\hat{y}_i - y_i)/m$, where $\hat{y}_i$ is the imputed value and y_i is the actual value for unit i, ($i = 1,2,\ldots m$), m is the number of missing values.

B. The mean absolute deviation defined as $\sum |\hat{y}_i - y_i|/m$.

C. The root mean square deviation defined as $\{\sum (\hat{y}_i - y_i)^2/m\}^{1/2}$.

D. The bias of the estimated totals due to imputation,

$$\sum_{i=1}^{m} w_i(\hat{y}_i - y_i) \text{ , where } w_i \text{ is the}$$

sampling weight for the i^{th} unit.

E. The relative bias of the estimated totals due to imputation

$$\left(\sum_{i=1}^{m} w_i(\hat{y}_i - y_i)/\sum_{i=1}^{n} w_i y_i\right) \times 100, \text{ where}$$

$$\sum_{i=1}^{n} w_i y_i \text{ is the estimated total}$$

from the complete data set.

The "errors" due to imputations of the above five types are calculated for each simulated data set, and the average of the five data sets is tabulated in Table 3.1.

The mean deviation measures the bias in the imputed values. On the average, Model A over imputes the actual value, Model B slightly under estimates the actual value. The average over imputed value using Model A (4 cells) wo/wt is $16,529; when incorporating sampling weights in the estimation, it is $15,954. The average under imputed value using Model B without and with sampling weights are $1,356 and $4,501, respectively. The average under imputed value under current procedure is $2,390.

The mean absolute deviations and the root mean squares deviations measure the "closeness" of the imputed value ($\hat{y}_i$) with the true value y_i. On the average, Model A (4 cells) wo/wt has $2,255 larger mean absolute deviation than Model B (4 cells) wo/wt, and $2,610 larger mean absolute deviation than the current procedure. The current procedure has the smallest mean absolute deviation $56,975. Model B (4 cells) w/wt gives the smallest mean square deviation $186,324.

Since our sample is a stratified random sample, sampling units from different strata have different inclusion probabilities. To estimate total sales, the bias due to imputation is of most interest. The average bias of these five data sets is \$2281 x 10^3 when using the current imputation procedure. The smallest bias is \$1,969 x 10^3 by using Model B (4 cells) with sampling weights. Model A (4 cells) without sampling weights has the largest bias \$31,659 x 10^3. Note that under the same model and the same number of cells, the bias of the estimated total is smaller by using sampling weights than not using sampling weights. This occurred for both Models A and B. In comparing Model B (w/wt) with 2 imputation cells and 4 imputation cells, the bias of the estimated total of 4 imputation cells is 40% less than 2 imputation cells.

The relative bias is 0.1394% for the current procedure, and 0.1203% for Model B (4 cells) with sampling weights, and 1.9344% for Model A (4 cells) without sampling weights.

For the current imputation procedure, all ratios of identicals of the five data sets for cells 3 and 4 exceeded the prior limits. Beside data set 1, the recalculated ratios of identicals from the other four data set for the collapsed cell of Group 1 still exceeded the prior limits.

IV. Summary and Recommendation

We have reviewed the imputation procedure of the Monthly Retail Trade Survey. The data of December 1982 retail sales were examined. We summarize the results as follows:

1. The current imputation procedure is a fairly simple procedure which assumes a ratio model (1.3) for the reported data in each imputation cell. By examining the reported retail sales data of December 1982, the error variance of the model for each imputation cell for most selected SIC's is proportional to x^2 instead of x (where x is the previous month sales). It seems that the current definitions of the imputation cells need to be modified so that the data will conform with the assumed model.

2. The simulation study using SIC 562 of December 1982's data (it is assumed that the data are missing at random) shows that the current imputation procedure gives lesser bias than the other imputation procedures studied in estimation of the population total. It also shows that 4 imputation cells give less bias than 2 collapsed cells; using sampling weights in the estimation gives less bias than not using sampling weights. It is suggested that finer imputation cells may be needed for each KB.

3. Field follow-up on the nonresponse data is necessary so we can better understand the nonresponse characteristics. This would show, for example, whether the distribution of nonresponse is the same as the distribution of response. Or whether the nonresponse rate systematically increases or decreases with sales size.

4. Revising or incorporating the prior limits used in the imputation procedure. For the data we examined (December 1982 - SIC 562), two ratios of identicals are outside of the prior limits (see Table 2.2 last column). The ratio will then be recalculated within a bigger cell and it will be used in the imputation whether the new ratio is in the prior limits or not. If the prior limits are good, it should be used in the imputation procedure when the ratio of identicals is outside of the limits. For example, using the closest bound of the limits to replace the ratio that is out of range. If the prior is out of date, it seems that it should be revised more often by using the existing ratios that have been calculated through the years.

5. The current imputation procedure is a mean imputation one (see Sedransk and Titterington (1980)), i.e., to impute for missing sales using a mean of the predictive distribution conditional on the known predictors. The mean imputation usually gives less variance of the total than the random imputation (where some error has been added to each predicted value). Since the objective of the Monthly Retail Trade Survey is to publish the total of the monthly sales, the mean imputation is used in the current imputation procedure. If furnishing the public use tape is also needed monthly, then in order to preserve the distribution of the monthly sales data the random imputation should be used, i.e., some 'error' should be added to the predicted value. These errors can take the form of random normal deviates defined in the model or randomly selected residuals from the model.

REFERENCES

1. Cassel, C.M., Särndal, C.E., and Wretman, J. H. (1979). 'Some Uses of Statistical Models in Connection with the Nonresponse Problem.' Symposium on Incomplete Data: Preliminary Proceedings, pgs 188-215.

2. DuMouchel, W. H. and Duncan, G. J. (1983). 'Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples. JASA 78, pgs. 535-543.

3. Ford, B. L., Kleweno, D. G., and Tortora, R. D. (1980). 'The effects of procedures

which impute for missing items: a simulation study using an agricultural survey.' Proceedings of the Survey Research Methods Section of ASA, pgs. 251-256.

4. Isaki, C. T., Wolter, K. M., Sturdevant, T. R. and Monsour, N. J. (1976). 'Sample Redesign of the Census Bureau's Monthly Business Survey,' American Statistical Meeting, Proceedings of Business and Economic Statistics Section, pgs. 90-98.

5. Kalton, G. (1981). 'Compensating for Missing Survey Data' Institute for Social Research, The University of Michigan.

6. Little, R.J.A. and Rubin, D.B. (1982). 'Missing Data in Large Data Sets.' Paper for small conference on the Improvement of the Quality of Data Collected by Data Collection System, Tennessee.

7. Little, R. J. A. (1982). 'Models for Nonresponse in Sample Surveys.' Journal of the American Statistical Association, 77, pgs. 237-250.

8. Little, R. J. A. and Samuhel, M. E. (1983). 'Alternative Models for CPS Income Imputation.' American Statistical Meeting, Proceedings of Survey Research Methods Section, pgs. 415-420.

9. Sedransk, J. and Titterington, D.M. (1980). 'Non-response in Sample Surveys.' Technical report, Bureau of the Census.

10. Wolter, K. M., Isaki, C. T., Sturdevant, T. R., and Monsour, N. J. (1976). 'Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys.' American Statistical Meeting, Proceedings of Business and Economic Statistics Section, pgs. 99-109.

11. Huang, E. T. (1984). 'An Imputation Study for the Monthly Retail Trade Survey.' Statistical Research Division Report Series, Bureau of the Census. CENSUS/SRD/RR - 84/13.

Table 2.1 The Estimated λ , ρ for each Imputation Cell

	December 1982					February 1983		
	GP	Sales	n	λ	ρ	n	λ	ρ
SIC 562	(Women's Ready-to-Wear Stores)							
Imputation Cells								
1	2	> $ 50,000	411	0.0513	2.0709	402	0.141970	1.87255
2	2	< $ 50,000	249	8.4806	1.6594	265	0.002676	2.22700
3	1	> $ 50,000	354	0.0468	2.1115	276	0.003741	2.21092
4	1	< $ 50,000	431	1.8031	1.7804	612	0.140595	1.60352
Collapsed Cells								
1	2	$(0, \infty)$	660	1.12407	1.7873	667	0.038760	1.99410
2	1	$(0, \infty)$	785	0.52022	1.8271	888	0.152564	1.93086
Total			1,445					

Table 2.2 Cell Mean Rate (Trend) for Each Imputation Cell Under Different Models

Imputation Cell	Model A (ratio model with $V(\varepsilon) = x^2\sigma^2$)		Model B (ratio model with $V(\varepsilon) = x\sigma^2$)	
	wo/wt	w/wt	wo/wt	w/wt
1 (GP2, Sales $\geq$ $50,000)	1.61598	1.61117	1.53371	1.50170
2 (GP2, Sales < $50,000)	1.68204	1.60483	1.66189	1.63338
3 (GP1, Sales $\geq$ $50,000)	1.48462	1.47382	1.48338	1.41526*
4 (GP1, Sales < $50,000)	1.49881	1.37942*	1.43179*	1.41607*

Table 3.1 Summary of the Results of the Model Comparisons from the Average of the Five Data Sets

	Model A (4 cells)		Model B (4 cells)		Model B (2 cells)	Current Procedure
	wo/wt	w/wt	wo/wt	w/wt	w/wt	
	$	$	$	$	$	$
Mean deviation	16.529×10^3	15.954×10^3	-1.356×10^3	-4.501×10^3	0.995×10^3	-2.390×10^3
Mean absolute deviation	59.585×10^3	59.892×10^3	57.330×10^3	57.347×10^3	56.994×10^3	56.975×10^3
Root mean square deviation	205.612×10^3	208.105×10^3	190.503×10^3	186.324×10^3	188.514×10^3	187.197×10^3
Bias of the estimated total	$31,659 \times 10^3$	$22,002 \times 10^3$	$6,316 \times 10^3$	$1,969 \times 10^3$	$3,247 \times 10^3$	$2,281 \times 10^3$
Relative bias of the estimated total (%)*	1.9344	1.3443	0.3860	0.1203	0.1984	0.1394

* The estimated total from the complete data is $1,636,659 x 10^3.