Evaluating Regression Imputation Models: An Example from the Services Sectors Portion of the Economic Census

Quatracia Williams and Katherine J. Thompson

1. Introduction

The Economic Census uses a variety of statistical models for item imputation. Historically, the services sectors portion of the Economic Census has relied on industry-average-ratio imputation as its primary statistical imputation model. This ratio imputation method uses **weighted** least squares (WLS) estimates from no-intercept simple linear regression models as imputation parameters, specifically fitting the model $Y_{ij} = \beta_i X_{ij} + \xi_{ij}$, $\xi_{ij} \sim (0, X_{ij}\sigma^2)$ for each data item *Y*, where *j* indexes the establishments within industry *i* that satisfy the ratio edit, $l_i \le \frac{Y_{ij}}{X_{ij}} \le u_i$. The best linear unbiased estimator (B.L.U.E) of β_i (the industry-average ratio) of this model is $\sum_j Y_{ij} / \sum_j X_{ij}$ (Draper and Smith, 1981, p.111). The weighted regression method compensates for known heteroscedasticity (unequal error variances). Thompson and Williams (2003), Thompson and Sigman (1996), and Huang (1984) demonstrate the plausibility of this weighted regression model for the services sectors data.

The data items that we discuss are all subjected to ratio edits prior to imputation, and the industry-average-ratio imputation model follows logically from the form of the ratio edit: a no-intercept regression model where the edit's numerator is the dependent variable and the denominator is the independent variable. From a statistical perspective, however, this imputation model is limited. First, the actual regression line may not go through the origin, and including an intercept in the imputation model could improve prediction. Second, industry-average-ratio imputation assumes that the variability in the dependent variable can be best "explained" by a single covariate, whereas a multiple regression model may have more predictive power. In fact, other sectors of the Economic Census successfully employ multiple regression imputation models for this reason. With our Economic Census data, the available pool of covariates is highly correlated, and consequently, the use of multiple regression models for imputation introduces the model-fitting issue of multicollinearity. The presence of multicollinearity and heteroscedasticity renders the "traditional" regression evaluation diagnostics unsuitable and prevents the use of the standard automatic model selection procedures (e.g., stepwise regression).

In an earlier stage of this research project (described in Thompson and Williams, 2003), we compared methods of developing and evaluating multiple regression models using data from the 1997 Economic Census. Our research found evidence of improved predictions over industry-average-ratio predictions for a variety of data items, recommending a combination of weighted least squares and ridge regression methods for parameter estimation. With highly multicollinear data, it is inappropriate to use regression diagnostics based on sums-of-squares statistics for model selection, since these statistics are not unique. Moreover, with heteroscedastic data, distributional assumptions are violated and the parametric tests (e.g., F- and t-tests) that use sums-of-squares variance estimates are invalid. So, we developed a regression model selection procedure based on robust statistics such as mean absolute error (MAE) and mean absolute deviation (MAD; average residual analysis) that relies primarily on cross-validation (Neter, Wasserman, and Kutner, 1989, pp. 466-468) to measure model performance and uses delete-a-group jackknife estimation (Rao, 1993) to obtain t-statistics for individual parameter estimates.

In theory, even with multicollinear data and unequal error variances, **WLS** estimates from multiple regression models can be used with other data sets for prediction, provided that the prediction region does not change. Recall, however, that these WLS estimates are unbiased but may not have minimum variance because of the heteroscedasticity and consequently may not be the "best" estimators. This paper builds on our prior study, exploring whether WLS imputation models developed from the 1997 Economic Census data can be used for imputation in the 2002 Economic Census. All discussed comparisons extend the industry-average-ratio model to the following model: $Y_{ij} = \beta_{0i} + \beta_{1i} X_{1ij} + \beta_{2i} X_{2ij} + ... + \beta_{pi} X_{pij} + \xi_{ij}$, $\xi_{ij} \sim (0, X_{1ij}\sigma^2)$, where β_{0i} can be zero (no intercept), and *p* (the number of covariates) is greater than or equal to one. Thus, we evaluate **both** simple and multiple regression models. Note that we assume that the error increase is proportional to a single independent variable (X_{1ij}) and differs for each covariate set.

We consider model-selection and validation problems that originate with a specific collection set (the Economic Census), but our results can be extrapolated to other data sets. Section 2 provides background on our research data and briefly describes the imputation parameter development production environment. Section 3 addresses the model selection problem of determining which multiple regression models are “better” than simple regression models given issues of multicollinearity and heteroscedasticity. We explore whether the same **regression models** will perform well from one data collection period to another, given the same prediction (imputation) region, then provide some insight on model stability. For those recurring regression relationships (our recommended models), we assess whether we can use the parameters developed for a prior census for prediction in the subsequent census in Section 4. Finally, we provide some concluding remarks in Section 5.

2. Background

Every five years the U.S. Census Bureau conducts an Economic Census of business establishments.
This census is the principal source for measuring economic activity in nineteen business sectors, which comprise about 98% of the U.S. economy. The services sectors portion of the Economic Census collects data from fifteen of the nineteen sectors. The industries within a sectors are then assigned to one of five trade areas: Wholesale Trade, Finance Insurance and Real Estate (FIRE), Services Industries, Transportation Communication and Utilities Industries (Utilities), and Retail Trade. Although each sector collects several different items, they all collect a core set of basic data items: annual payroll, first quarter payroll, receipts (sales), and employment. In addition to these basic data items, tax-exempt Services Industries collect operating expenses and Wholesale Trade collect operating expenses, costs of purchases, beginning-year inventories and end-of-year inventories. All of the trade areas calculate two sets of imputation parameters for the Economic Census: the first set (known as the cold deck parameters) are developed from the **complete** prior census data; and the second set (known as the warm deck parameters) are developed from the current census data after sufficient cases have been processed. For our research we use data from two different collection periods, namely the 1997 and 2002 Economic Censuses. In a sense, the research that we describe in Sections 3 and 4 mimics the Economic Census imputation parameter-development environment, in that **complete** 1997 census data are available for research, but the 2002 data sets comprise only cases received by mid-February 2004.

Thompson and Williams (2003) recommended adding several new regression models to the services sectors available imputation suite. However, because this research was confined to one data set, 1997 Economic Census data, we were hesitant to recommend immediate implementation of the recommended models for several reasons. First, our recommendations included several multiple regression models whose performance we were unable to validate in more than one collection period. The 1997 Census was collected using the North American Industrial Classification System (NAICS), and the 1992 Census used the Standard Industrial Classification System (SIC), so industry-level comparisons would be confounded. And, in our previous study the 2002 data were not available for model validation. Second, employment data was collected differently in 2002 and 1997. The 1997 Census requested **total** employment (which included direct and leased employees), whereas the 2002 Economic Census collected payroll and employment data for direct and leased employees separately.

For the research presented in this paper, we selected one set of industries per trade area. These industries are not a representative sample of the trade area; instead, we chose them because they contained at least 40 companies in the 1997 census. Due to some edit-processing problems, our research excludes tax-exempt Service industries. We also exclude beginning-of-year and end-of-year inventories from all analyses because they are edited and imputed separately from the other basic data items with a software program that does not incorporate multiple regression imputation. Each research data set consists of full year reporting units whose basic data items satisfy all ratio edits. Because our 1997 and 2002 data are tested with exactly the same ratio edits, the prediction (imputation) regions for both data sets are the same.

3. Comparison of Regression Models Between Censuses

3.1. Model Selection Procedure

We developed a three-stage model selection procedure to evaluate regression models within each data set (1997 or 2002 Census). The first two stages of this procedure use cross-validation. To perform cross-validation for model selection, we randomly split the establishments in each industry into two equal-sized sets: a model-building set and a validation set. Then, we calculated the regression parameters for all-possible multiple regression models within a trade area from the model-building set data and used these parameters for prediction in the validation set. We used mean absolute error (MAE) to evaluate the predictions in the model validation set. The MAE is the average of the absolute difference between the predicted and the true values over all industries for each trade area (DeGroot, 1987, pp. 209-211) and is given by

$$\sum_{i}^{n}\left|\hat{Y}_i - Y_i\right| / n \qquad (3.1)$$

where i indexes the industry within trade area, $\hat{Y}_i$ is the tabulated value obtained from imputing data item Y in the validation data set using model-building data set's regression parameters, and Y_i is the tabulated value (from reported data in the validation data set) of the data item in industry i. The first stage of our model selection procedure was to sort all models for a given dependent variable (within trade area) by ascending MAE and drop all models that did not perform better (in terms of MAE) than the best performing industry-average ratio for each dependent variable.

MAE measures model effects for the entire trade area. Furthermore, since MAE is an average, it is highly sensitive to outliers and can be overly affected by models in one or two industries with unusually good or poor performance. The second stage of our model selection procedure specifically examines model performance on an industry-by-industry basis using mean absolute deviation (MAD), defined as

$$\sum_{j}^{n_i}\left|\hat{Y}_{ij} - Y_{ij}\right| / n_i \qquad (3.2)$$

(Nordholt, 1998) where j indexes the establishment within industry i and n_i is the number of establishments in industry i. The MAD is simply an average absolute residual for each industry.

For model selection, we wanted to see how often the (remaining) candidate models yielded more precise imputations than industry-averages by **industry** within trade area. After sorting the models within trade area by ascending MAD, we observed an interesting clustering pattern, as illustrated in Figure 1:

Model	MAD	Ratio	Classification
APR = β_0 + β_1RCP + β_2QPR	79.7	N/A	High Performing
APR = β_0 + β_1EMP + β_2RCP+ β_3QPR	79.9	1.0	High Performing
APR = β_1EMP + β_2RCP+ β_3QPR	80.0	1.0	High Performing
APR = β_1RCP + β_2QPR	80.7	1.0	High Performing
APR = β_1EMP + β_2QPR	82.3	1.0	High Performing
APR = β_0 + β_1EMP + β_2QPR	82.6	1.0	High Performing
APR = β_0 + β_1QPR	83.1	1.0	High Performing
APR = β_1QPR	85.5	1.0	High Performing
APR = β_1EMP + β_2RCP	273.6	3.2	Low Performing
APR = β_0 + β_1EMP + β_2RCP	274.3	1.0	Low Performing
APR = β_1EMP	304.0	1.1	Low Performing
APR = β_0 + β_1EMP	307.8	1.0	Low Performing
APR = β_1RCP	309.5	1.0	Low Performing
APR = β_0 + β_1RCP	321.8	1.0	Low Performing

Figure 1: Sorted MAD Statistics for a Fictional Industry (Predicting Payroll)

Notice that the first eight models (the "high performing models") in this fictional industry have approximately the same value of MAD, followed by another cluster of six models ("low performing models"), all having a MAD value approximately three to four times larger than those models in the first cluster. We observed similar phenomena in all trade areas, regardless of dependent variable. We used the ratio of ranked models to the prior model to automatically discriminate between the high and low performing models (illustrated in the third column of Figure 1); the first ratio of ranked MADs that was greater than 1.2 indicated the ranked model belonged to the "low-performing" cluster. [We obtained the cut-off value of 1.2 after studying several industries' data for a variety of dependent variables]. Then, we obtained summary counts of "high performing occurrences" for each model within trade area, where each count represented the model's performance within an industry. Finally, we dropped all models that were "high performing" in **less** than one-third of the trade area's test industries.

The third stage of our model selection procedure examined cumulative "significance" of regression model parameters for the **remaining** candidate models. Instead of using the WLS sums-of-squares standard errors, we computed delete-a-group jackknife standard errors of the parameter estimates for all remaining candidate models. For this, we randomly split the establishments in each industry into 16 groups and obtained each delete-a-group jackknife replicate estimate by dropping one group at a time, weighting the remaining observations by (16/15) and calculating the parameter estimates from the remaining 15 groups. Delete-a-group standard errors of each parameter estimate $\hat{\beta}_p$ are:

$$\hat{\sigma}(\hat{\beta}_p) = [15/16]\sum_{k=1}^{16}(\hat{\beta}_{pk} - \hat{\beta}_p)^2 \qquad (3.3)$$

where *p* indexes the regression parameter. Within each trade area, we computed the frequency of statistically significant parameters in each candidate model by industry (using t-statistics at the 90% confidence level), dropping all models that did not have significant parameters or intercepts in at least one-third of the test industries, thus obtaining our final reduced model set for a given dependent variable in each trade area.

3.2. Model Selection Results

To determine whether the same regression **models** could be used in consecutive census data sets, we performed the model selection procedure described above in Section 3.1.independently on our 1997 and 2002 census data sets and compared the resulting reduced model sets within trade area for each data item. Tables 1 and 2 summarize the number of models that exhibit better performance than corresponding industry-average-ratio models after our three-stage model-selection procedure[1], presenting results for multiple regression and simple linear regression models, respectively. For each basic data item, we first present all recommended models from each data year and then present the number of common models to both years. Note that Table 2 includes simple linear regression models **with intercepts**, but excludes industry-average-ratio models.

[1] i.e., have consistently lower MAE, are categorized as “high performing” based on MAD, and contain statistically significant parameters in most test industries.

Table 1: Results of Model Selection Procedure – Number of Recommended Multiple Regression Models

		Trade Area				
Data Item	Data Collection Period	FIRE	Retail	Services	Utilities	Wholesale
Annual	1997	1	6	3	4	6
	2002	3	4	1	1	4
	Both	1	4	1	1	2
Employment	1997	0	0	0	0	0
	2002	0	0	0	0	0
	Both	0	0	0	0	0
Receipts	1997	1	2	1	0	2
	2002	1	2	0	0	0
	Both	0	2	0	0	0
Operating Expenses	1997					4
	2002					2
	Both					2
Purchases	1997					5
	2002					3
	Both					3

Table 2: Results of Model Selection Procedure – Number of New Recommended Simple Regression Models

		Trade Area				
Data Item	Data Collection Period	FIRE	Retail	Services	Utilities	Wholesale
Annual Payroll	1997	0	1	1	0	0
	2002	1	1	0	1	0
	Both	0	1	0	0	0
Employment	1997	1	1	1	1	1
	2002	1	1	1	1	2
	Both	1	1	1	1	1
Receipts	1997	1	1	1	1	0
	2002	1	1	0	1	0
	Both	1	1	0	1	0
Operating Expenses	1997					1
	2002					0
	Both					0
Purchases	1997					1
	2002					1
	Both					1

Many of the selected models shown in Tables 1 and 2 do **not** consistently exhibit better performance than the industry-average ratio models in both years. For the simple regression models, it is not unrealistic to expect either a slope or intercept change in both census years, especially in our regression models containing employment. For the multiple regression models, we assumed a high degree of multicollinearity in our data sets. We hoped, however, to isolate explanatory factors that contributed to selecting the **same** regression models in both years.

For any type of regression model, a slope change could indicate a change in the distribution of association between covariates. To assess this, we computed the WLS model-R^2 for all possible covariate pairs (e.g., annual payroll and first quarter payroll; annual payroll and employment) using the industry-average-ratio model (simple linear regression with no intercept). The model-R^2 measures the proportion of variability in the dependent variable "explained" by the independent variable [multicollinearity is obviously **not** an issue with simple linear regression]. Often, the Pearson (sample) correlation coefficient is used to measure association between paired items. However, this statistic assumes equal error variances. Recall that our Economic Census data are heteroscedastic, requiring a weighted least squares regression procedure. Thompson et al (2001) considered using ranked correlation as a proxy for the appropriate model-R^2 ranking to determine imputation order (e.g., rank recommended industry-average ratio models by descending correlation). While they found that the relative rankings of the two statistics were the same for both measures of association (and thus both could be used for model **ordering**), the calculated values of the two statistics diverged greatly, with the sample correlation coefficient often greatly **overestimating** the strength of association.

Here, the distributions of model-R^2 were remarkably consistent from one census to the next in all trade areas. We were initially rather surprised by this, given the inter-censal definitional change in employment described in Section 2. For comparison purposes, we also computed sample pairwise correlations in all trade areas and examined their distributions. Again, the distributions were almost identical between censuses. Similar to Thompson et al (2001), sample correlations were generally much larger than their corresponding model-R^2 statistics (thus overestimating the level of association); however, the measure of spread (sample range) was very similar for each pair regardless of statistic (i.e., sample correlation or model-R^2).

There were several characteristics that distinguished the regression models that performed well in consecutive collection periods from those that did not. When both conditions did not hold, model performance was inconsistent. First, for those that did, there was a strong degree of association between the independent and dependent variable(s), with average model-R^2 greater than 0.75. Moreover, the range of these model-R^2 statistics was quite small within trade area (generally less than 0.30). Second, in the case of multiple regression models, the model-R^2 between covariates (explanatory variables) was generally low (less than 0.50) or was **quite** variable between industry (with ranges up to 0.80). Thus, the multicollinearity in these models is not as severe as would be indicated by the sample correlation.

This conclusion is reinforced by the second stage of our model selection procedure. In Section 3.1, we described our model selection procedure for obtaining reduced sets of models using the MAD (average residual) statistic. Recall that our procedure retains only those multiple regression models that perform well consistently by industry. With highly multicollinear models, the MAD rank is quite variable by industry, so that these models are dropped (from either data set) by the selection procedure.

Having found a set of explanatory conditions that were common to **all** of our recommended regression models, we then turned to the question of whether the same model **parameters** could (or should) be used for prediction in consecutive time periods.

4. Parameter Validation: Comparison of 1997 and 2002 parameters

In Section 3, we found sets of regression models that performed well in consecutive collection periods and identified some explanatory characteristics for their performance. The previous analysis was concerned with the **form** of the models (i.e., covariate sets). Now, we turn to the consistency of parameters (within model) over time.

In Section 2, we described the imputation parameter development environment for the services sectors trade areas. Recall that all trade areas develop two sets of **industry-average-ratio imputation** parameters: "cold deck," developed from the complete prior census data and used for the first phase of census data processing; and "warm deck," developed from the current census data available after a certain percentage of establishments in each trade area have been processed. Thus, the question of parameter stability has strong implications for our programs. For example, finding evidence of frequent slope changes (industry-average-ratios) provides justification for a mass re-imputation with warm deck parameters when available.

We used cross-validation for model parameter evaluation. Now, however, we use the 1997 data as model building data set and the 2002 as validation data set. As with our model selection procedure, we rely primarily on trade-area level MAE to assess performance. We also calculated a mean ratio of predicted to true value (MRPT) to measure the practical impact of the MAE for a given data item within a trade area, where MRPT is given by

$$\sum_{i}^{n} \left| \hat{Y}_i / Y_i \right| / n \qquad (4.1)$$

For example, the MAE may be a large (quantitative) number, but on the average, the predicted and true values may be within a percent of each other.

In both the FIRE and Wholesale Trade industries, the multiple regression models exhibited consistently smaller MAEs than the corresponding industry-average ratio models for the same dependent variable, and with a few exceptions, the MRPTs from the multiple regression models were generally closer to one. This pattern was quite different for the other trade areas, where the multiple regression models had consistently higher MAEs than corresponding industry-average ratios (with one exception in the Retail Trade area). In all trade areas but Wholesale, the simple linear regression models that included intercepts had much higher MAEs and MRPTs than their corresponding industry-average ratios, although often, the MRPTs of the industry-average ratios were more than 10-percent above or below one (indicative of a slope change between censuses).

We hypothesized that the models for which 1997 parameters yielded poor 2002 predictions had significantly different parameters from one census period to another. To test this, we conducted paired t-tests of parameters (within model), again using delete-a-group jackknife standard errors, and examined the frequency of significantly different parameters for each model in the test industries. Since we did not have estimates of between-census correlation, we conducted two sets of tests per model, one assuming no correlation between censuses (extremely conservative) and the other assuming a high degree of correlation between censuses (ρ=0.90, extremely anti-conservative). In the majority of trade areas, the two sets of test results were identical.

We decided to reject the hypothesis of no significant difference in parameters when **at least one** parameter per model was significantly different in at least one third of our test industries (e.g., for 30 test industries, if one parameter was significantly different in 10 or more industries, we concluded that both sets of parameters were significantly different). In this case, we would not recommend using the 1997 parameters for 2002 data prediction. The converse is true when we failed to reject the hypothesis of no difference. Because of our concerns about the stability of multiple regression model parameters given the multicollinearity in our data, this criterion is fairly conservative.

Table 3 presents the results of these comparisons for our **recommended** models in each trade area. We include simple linear regression (industry-average-ratio) models to show in many cases there was a slope change.

Table 3: Results of Significance Tests Comparing Parameters of Recommended Models

Trade Area											
		FIRE		Retail		Services		Utilities		Wholesale	
Item	Model Type	Total	Different Parameters by Census	Total	Different Parameters by Census	Total	Different Parameters by Census	Total	Different Parameters by Census	Total	Different Parameters by Census
Annual Payroll	MR	1	1	4	3	1	1	1	1	2	2
	SLI	0	0	1	1	0	0	0	0	0	0
	SL	3	3	3	3	3	2	3	1	4	3
Employment	MR	0	0	0	0	0	0	0	0	0	0
	SLI	1	1	1	1	1	1	1	1	1	1
	SL	2	2	2	2	2	2	2	0	1	1
Receipts	MR	0	0	2	2	0	0	0	0	0	0
	SLI	1	1	1	1	0	0	1	0	0	0
	SL	2	1	2	2	2	1	2	1	3	2
Operating Expenses	MR									2	2
	SLI									0	0
	SL									2	2
Purchases	MR									3	3
	SLI									1	1
	SL									1	1

MR = Multiple Regression; SLI = Simple Linear Regression with Intercepts; SL = Simple Linear Regression without Intercept (Industry-Average Ratio)

Table 3 shows that the majority of regression models contain at least one parameter that is significantly different from census to census, indicating slope changes, intercept changes, or both. This validates the currently used "cold deck/warm deck" parameter development procedures for the services sectors trade areas. Indeed, it is sensible to anticipate inter-censal slope or intercept changes with linear models: if within-industry relationships between data items are not changing (e.g., average wage per employee), then there would be no need to conduct a census.

Recall in Wholesale Trade and FIRE Industries, the multiple regression parameters constructed from prior period data performed quite well with the current data, although most models contain at least one statistically different parameter. In these two trade areas, we suspect that the **combined** effect of covariates in multiple regression models is not changing from census to census. Moreover, in the Wholesale Trade industries, the differences between the computed regression parameters for all recommended models are negligible. Since significance tests simply measure differences, we do not use them as a metric for predictive performance. Ultimately, the cross-validation results are a more appropriate measure.

5. Conclusion

We have presented a model selection and model validation problem for data sets that exhibit both multicollinearity and heteroscedasticity. Our model selection and validation problems are compounded by the time span between Economic Censuses (five years). We first address the model selection problem, presenting procedures that do not rely on sums-of-squares statistics. By independently conducting the model selection procedure in two Economic Census data sets, we identified models (sets of covariates) that performed well in both collection periods. Two conditions were always present for those models: there is a strong association between the independent variable and each of the covariates; and there is a variable degree of association among the covariates themselves.
This association is measured by model-R^2 (appropriate for WLS) instead of the inappropriate Pearson sample correlation coefficient. Although the multicollinearity in the data items is less severe than would be indicated by sample correlation, it is still present, and hence the model selection **order** is not expected to be consistent over time (even though both sets of data are restricted to the same prediction region).

Having determined which models could be used in consecutive censuses, we then examined whether we could use the same parameters in both data sets. For most of our recommended regression models, parameters were statistically different (also confirmed by cross validation), so we advise continuing the current procedure of developing new parameters for imputation from the most recent data whenever possible.

For all analyses, we used robust measures such as MAE and MAD. We chose these statistics because of their invariance to the problems of multicollinearity and heteroscedasticity. Besides being methodologically appropriate, these measures are quite understandable by the subject-matter experts because they are the same measures they use to compare alternative imputation models. Because the analysts understood our methods, they understood our results and agreed to adopt many of our recommended new regression models for imputation. Unfortunately, all basic data items, except for purchases, must be published by July 2004. As we said in Section 2, sufficient research data was not available until mid-February of 2004. So, when we presented our results to the subject-matter experts at the end of March, there was not enough time to add the new regression models into production for the 2002 Census. However, these models will be used in the 2007 Economic Census. Fortunately, there is enough time to implement our recommended new models for Purchases since the 2002 tabulations for this item are not published by the Wholesale Trade sector until December 2005.

References

DeGroot, M. (1987). Probability and Statistics. Reading, MA: Addison Wesley.

Draper, N.R. and Smith, H. (1981). Applied Regression Analysis. New York: John Wiley and Sons, Inc.

Huang, E.T. (1984). An Imputation Study for the Monthly Retail Trade Survey. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 610 – 615.

Neter, J., Wasserman,W. and Kutner, M.H. (1989). Applied Linear Regression Models. Homewood, IL: Richard D. Irwin, Inc.

Nordholt, E.S. (1998). Imputation: Methods, Simulation Experiments and Practical Examples. *International Statistical Review*, **66**, pp. 157-180.

Thompson, K.J., Sausman, K., Walkup, M., Dahl, S., King, C., and Adeshiyan, S. (2001). *Developing Ratio Edits and Imputation Parameters for the Services Sector Censuses (SSSD) Plain Vanilla Ratio Edit Module Test*. Washington, D.C.: U.S. Census Bureau (Technical report #ESM-0101, available from the Economic Statistical Methods and Programming Division).

Thompson, K.J. and Sigman, R.S. (1996). *Evaluation of Statistical Methods for Developing Ratio Edit Module Parameters.* Washington, D.C.: U.S. Census Bureau (Technical report #ESM-9601, available from the Economic Statistical Methods and Programming Division).

Thompson, K.J. and Williams, Q. (2003). *Developing Imputation Models for the Services Sectors Portion of the Economic Census.* Federal Committee on Statistical Methodology, Statistical Policy Working Paper 37, available from www.fcsm.gov.