Investigation of Variance Estimators for the Survey of Business Owners (SBO)

Marilyn Balogh[1], **Sandy Peterson**

U.S. Census Bureau
4600 Silver Hill Rd, Washington, D.C. 20233

Abstract

The Survey of Business Owners and Self-Employed Persons (SBO) provides, on a quinquennial basis, selected economic and demographic characteristics for businesses and business owners by gender, ethnicity, race, and veteran status. The SBO sample is a stratified systematic sample that employs the random group variance estimator to estimate sampling error. This paper reports the results of a simulation study conducted to compare the random group variance estimator to the delete-a-group jackknife variance estimator and the stratified jackknife variance estimator. The methods were compared by examining the relative bias and the coefficient of variation of each variance estimate for five states (Florida, Georgia, Kansas, New York, and North Dakota) for different characteristic data (public ownership, race, ethnicity, sex, and veteran status). Another factor taken into consideration was computer processing time, which is not trivial for the stratified jackknife applied to a large survey. The results of this study suggest that the delete-a-group jackknife variance estimator may provide an improvement over the random group estimator, with little differences in processing time.

Keywords: variance estimation, stratified jackknife variance estimator, delete-a-group jackknife variance estimator, random group variance estimator

1 Introduction

The Survey of Business Owners and Self-Employed Persons (SBO) provides, on a quinquennial basis, selected economic and demographic characteristics for businesses and business owners. Currently, SBO employs the random group variance estimator to estimate sampling error. The primary purpose of this study was to determine if there is a more efficient variance estimator compared to the current method. This study compares the random group variance estimator to the delete-a-group jackknife variance estimator and the stratified jackknife variance estimator to determine which is best for estimating SBO sampling error.

We compared variance estimators by simulating data using the 2007 SBO universe, and examining the relative bias and the coefficient of variation (CV) of each variance estimate for five states (Florida, Georgia, Kansas, New York, and North Dakota) for different characteristic data (public ownership, race, ethnicity, sex, and veteran status).

2 Background on SBO

The SBO provides economic and demographic data for businesses and business owners as part of the Economic Census every five years, for years ending in "2" and "7," by Title 13 of the United States Code. The SBO samples and publishes data on a firm basis where "a firm is a business consisting of one or more domestic establishments that the reporting firm specified under its ownership or control" (Survey of Business Owners[2]). The SBO publishes the number of employer and nonemployer firms, sales and receipts, annual payroll, and employment by gender, ethnicity, race, and veteran status of the majority business owners.

The SBO uses administrative data to place each firm into one of nine sampling frames: American Indian, Asian, Black or African American, Hispanic, Non-Hispanic white men, Native Hawaiian and Other Pacific Islander, Other,

---

[1] This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress). Any views expressed on statistical, methodological, technical, or operational issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.

[2] http://www.census.gov/econ/sbo/

Publicly owned, and Women. The SBO universe is stratified by frame, state, industry code (NAICS), and whether the firm has paid employees. Firms that operate in multiple states or whose receipts, payroll, or employment exceed a stratum specific cutoff are selected with certainty and represent only themselves. We subject all other firms to systematic sampling within each stratum (Survey of Business Owners Methods[3]).

The SBO performs hot-deck donor imputation for unit and item non-response on selected demographic characteristics for business owners (gender, ethnicity, race, veteran status, or publicly held). We then tabulate each firm as publicly held or primarily owned by members of each of these selected demographic groups. The publicly held category includes publicly owned firms and any firms that are not classifiable by gender, ethnicity, race, and veteran status. The SBO calculates the estimates using the Horvitz-Thompson estimator, with the inverse probability of selection as the sampling weight.

Currently, the SBO estimates the sampling error using the random group (RG) variance estimator (simple method) with 10 non-certainty random groups and a finite population correction (fpc) factor adjustment. This study does not change the number of random groups or the use of the fpc for the alternate variance estimators.

3 Variance Estimation Methodology

We considered three variance estimators: the RG variance estimator, the delete-a-group jackknife (DAG) variance estimator, and the stratified jackknife (SJK) variance estimator. Additionally, we considered two reweighting procedures for the RG and DAG variance estimators, specifically the simple and the stratum-specific methods.

All sampled firms have an initial sampling weight, w_i, which is the inverse probability of selection. To account for the non-negligible sampling fractions, Wolter (1985, Ch. 2) suggests using an fpc-adjusted weight $w_i' = \sqrt{1 - f_i} w_i$ in place of the sampling weight w_i, where f_i is equal to the probability of selection. This is currently used for SBO and we used this adjustment in all the methods; however, for the SJK method it is incorporated in the variance calculation instead of the weights.

All firms selected with certainty have a sampling weight of one, so their adjusted weight is always zero and they are not included in the RG and DAG variance calculations.

3.1 Random Group Method

The RG method divides the non-certainty firms into *R* random groups, creates *R* replicate estimates by reweighting the random group's sample to represent the full sample, then determines the simple variance of the *R* replicate estimates.

The difference between the two reweighting procedures (simple and stratum-specific) of the RG variance estimator is in the creation of the replicate-*r* weights. Replicate-*r* weights are the fpc adjusted sampling weights multiplied by a reweighting factor and are used to calculate the RG replicate *r* estimates $\hat{\theta}_{r(RG)}$ (Thompson, 2003, pp. 4227-4228).

For the RG simple method (RG_S), the replicate-r weight is:

$$w_{ri} = \begin{cases} R * w_i' & i \in \text{replicate group } r \\ 0 & \text{otherwise} \end{cases}$$

where w_i' is the fpc-adjusted sampling weight of unit *i* and *R* is the number of non-certainty random groups.

For the RG stratum-specific method (RG_ST), the replicate-r weight is:

$$w_{rhi} = \begin{cases} \frac{n_h}{n_{hr}} * w_{hi}' & i \in \text{replicate group } r \\ 0 & \text{otherwise} \end{cases}$$

[3] http://www.census.gov/econ/sbo/about.html

where w'_{hi} is the fpc-adjusted sampling weight of unit *i* in stratum h, n_h is the total number of non-certainty sampled units in stratum *h*, and n_{hr} is the total number of non-certainty sampled units in stratum *h* and replicate group *r*.

The RG replicate *r* estimates are $\hat{\theta}_{r(RG)} = \sum_i w_{ri}\theta_i$ for RG_S and $\hat{\theta}_{r(RG)} = \sum_{h\in H}\sum_{i\in h} w_{rhi}\theta_i$ for RG_ST and the full sample estimate is $\hat{\theta}' = \sum_i w'_i\theta_i$. Since we are using an fpc-adjusted sampling weight, w'_i, $\hat{\theta}'$ and $\hat{\theta}$ (the tabulated full sample estimate) are not equivalent. The RG variance for any estimate $\hat{\theta}$ is then:

$$\widehat{var}_{RG}(\hat{\theta}) = \frac{1}{R(R-1)}\sum_{1}^{R}\left(\hat{\theta}_{r(RG)} - \hat{\theta}'\right)^2$$

3.2 Delete-a-group Jackknife Method

The DAG method divides non-certainty firms into *R* random groups, creates *R* replicate estimates by deleting one group at a time from the sample, then determines the simple variance of the *R* replicate estimates.

Similar to the RG method, the key to the two reweighting procedures (simple and stratum-specific) of the DAG variance estimator is in the creation of the replicate-*r* weights, which are used in the calculation of the DAG replicate *r* estimate $\hat{\theta}_{r(DAG)}$.

For the DAG simple method (DAG_S), the replicate-*r* weight is:

$$w_{ri} = \begin{cases} 0 & i \in \text{replicate group } r \\ \frac{R}{R-1} * w'_i & \text{otherwise} \end{cases}$$

where w'_i is the fpc-adjusted sampling weight of unit *i* and *R* is the number of random groups (Thompson 2003).

For the DAG stratum-specific method (DAG_ST), it is assumed that all stratum sample sizes are large ($n_h \geq 5$), and violation of this can cause the DAG variance estimate to be biased upward (Kott 1998). For SBO, this assumption does not hold because many strata have fewer than five non-certainty firms. For this reason, we used the extended DAG method (Kott 2001) in the creation of the replicate-*r* weights.

Let:

w'_{hi}	= the fpc-adjusted sampling weight of unit *i* in stratum h
n_h	= the number of non-certainty sampled units in stratum h
R	= the number of random groups
S_{hr}	= the set of n_{hr} non-certainty sampled units in stratum h and random group r

When $1 < n_h < R$, then the replicate-*r* weight is

$$w_{rhi} = \begin{cases} w'_{hi} & S_{hr} \text{ is empty} \\ w'_{hi} * (1 - [n_h - 1]Z) & i \in S_{hr} \\ w'_{hi} * (1 + Z) & \text{otherwise} \end{cases}$$

where $Z^2 = R/[(R-1)n_h(n_h-1)]$. For this study, we collapse strata to ensure that there are at least two non-certainty units in each stratum.

When $n_h \geq R$, then the replicate-*r* weight is

$$w_{rhi} = \begin{cases} 0 & i \in S_{hr} \\ w'_{hi} * \left(\frac{n_h}{n_h - n_{hr}}\right) & \text{otherwise} \end{cases}$$

The DAG replicate *r* estimates are $\hat{\theta}_{r(DAG)} = \sum_i w_{ri}\theta_i$ for DAG_S and $\hat{\theta}_{r(DAG)} = \sum_{h \in H} \sum_{i \in h} w_{rhi}\theta_i$ for DAG_ST and the full sample estimate is $\hat{\theta}' = \sum_i w'_i \theta_i$. The DAG variance for any estimate $\hat{\theta}$ is then:

$$\widehat{var}_{DAG}(\hat{\theta}) = \frac{R-1}{R} \sum_{r=1}^{R} \left(\hat{\theta}_{r(DAG)} - \hat{\theta}'\right)^2$$

3.3 Stratified Jackknife Method

The SJK method constructs one replicate estimate per sample unit by dropping one unit at a time from the stratum and multiplying the remaining units in the stratum by $n_h/(n_h - 1)$.

The replicate-*k* weight for the SJK variance estimate is:

$$w_{khi} = \begin{cases} w_i & \text{if unit } i \text{ is not in stratum } h \\ \frac{n_h}{n_h - 1} w_i & \text{if unit } i \text{ is in stratum h but not unit } k \\ 0 & \text{otherwise} \end{cases}$$

where w_i is the sampling weight of unit *i* and n_h is the number of non-certainty sampled units in stratum *h* (Lohr 1999). For this study, we collapse strata to ensure that there are at least two non-certainty units in each stratum.

The SJK replicate estimates are $\hat{\theta}_{k(SJK)} = \sum_{h \in H} \sum_{i \in h} w_{khi}\theta_i$ and the full sample estimate is $\hat{\theta} = \sum_i w_i \theta_i$. The SJK variance for any estimate $\hat{\theta}$ is:

$$\widehat{var}_{SJK}(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{k=1}^{n_h} (1 - f_{hi}) \left(\hat{\theta}_{k(SJK)} - \hat{\theta}\right)^2$$

where f_{hi} is the probability of selection for firm *i* in stratum *h* and *H* is the number of strata (Steel 2009).

4 Simulation Study

The first step in running simulations was to create the population. We started with the 2007 SBO universe in five selected states (Florida, Georgia, Kansas, New York, and North Dakota) and assigned race, gender, ethnicity, veteran status, and publicly held for the non-selected 2007 firms based on the 2007 SBO estimates. These five states were chosen for their variation of data, from small number of firms with little demographic diversity (North Dakota) to large number of firms with great demographic diversity (New York) and the rest somewhere in between (Kansas, Florida, and Georgia).

We selected 5,000 different stratified systematic samples from the simulated 2007 population using the SBO stratification and sampling design. For each of the 5,000 samples, we assigned sampled units to 11 random groups to be used in the calculation of the RG and DAG variance estimators. The SBO sampling design assigns all cases selected with certainty (these cases are in all of the 5,000 samples) to random group 0 and all non-certainty cases to random groups 1 to 10.

For each of the 5,000 samples, we calculated the five variance estimators for the firm count and total receipts: RG simple (RG_S), RG stratum-specific (RG_ST), DAG simple (DAG_S), DAG stratum-specific extended method (DAG_ST), and stratified jackknife (SJK). We then used all 5,000 variance estimates to calculate the relative bias and the CV for each variance estimator. "The relative bias is a measure of the bias of the variance estimate as a proportion of the true variance and the CV measures the variance of the variance estimate" (Thompson 2003).

To calculate the relative bias and the CV of the variance estimate, we first measured the true variance of the statistic that we were estimating. The true variance is:

$$var(\hat{\theta}) = \frac{1}{5000}\sum_{m=1}^{5000}\left(\hat{\theta}_m - \bar{\theta}\right)^2$$

where $\hat{\theta}_m$ is the estimate from sample m (e.g., the weighted total of firms or receipts), and $\bar{\theta}$ is the average estimate across the 5,000 samples (NOT the true population value).

Next, we calculated five variance estimates (var_{METH}) per domain from the 5000 samples. We compared these variance estimates by the relative bias and the CV (also referred to as stability) (Thompson 2003).

The relative bias is:

$$RB_{METH} = \frac{\overline{\widehat{var}_{METH}(\hat{\theta})}}{var(\hat{\theta})} - 1$$

The coefficient of variation is:

$$CV_{METH} = \frac{\sqrt{\frac{1}{5000}\sum_{i=1}^{5000}[\widehat{var}_{METH}(\hat{\theta}_i) - var(\hat{\theta})]^2}}{var(\hat{\theta})}$$

5 Results

The best variance estimator is one where both the relative bias and the CV are near zero (Thompson 2003).
To determine the best variance estimator, we first used the sign test on the pairwise differences of each method's CV across all domains[4] by estimate type (firm count and total receipts). We used a one-sided sign test to test the null hypothesis that there is no difference in median. The alternative is that the median of the X_i variables tends to be bigger than the median of the Y_i variables. We paired the data for two methods within each domain and calculated the difference ($D_i = X_i - Y_i$) of the CV between the two methods. Next, we calculated the test statistic T, which is the number of times the difference was greater than zero (with $\alpha = 0.05$) (Conover 1998).

Table 1 presents the results of the sign test for the CVs across all domains by estimate type (firm count and total receipts).

[4] The 255 domains are all possible cross tabulations of 17 demographic characteristics, 5 states, and 3 employer statuses (firms with employees, firms without employees, and all firms). The 17 demographic characteristics are: female, male, equally male/female, Hispanic, non-Hispanic, equally Hispanic/non-Hispanic, White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, some other race, veteran, nonveteran, equally veteran/nonveteran, publicly held, and all firms.

Table 1: Sign Test Comparing the Coefficient of Variations between the Different Variance Estimators

Estimate	Variance Estimators	T	p-value	Superior Method
Firm Count	RG_S - RG_ST	148	0.0061	RG_ST
	RG_S - DAG_S	126	0.3508	None
	RG_S - DAG_ST	198	<0.0001	DAG_ST
	RG_S - SJK	253	<0.0001	SJK
	RG_ST - DAG_S	107	0.0061	RG_ST
	RG_ST - DAG_ST	111	0.0225	RG_ST
	RG_ST - SJK	249	<0.0001	SJK
	DAG_S - DAG_ST	198	<0.0001	DAG_ST
	DAG_S - SJK	253	<0.0001	SJK
	DAG_ST - SJK	250	<0.0001	SJK
Total Receipts	RG_S - RG_ST	171	<0.0001	RG_ST
	RG_S - DAG_S	105	0.0518	None
	RG_S - DAG_ST	150	0.0019	DAG_ST
	RG_S - SJK	249	<0.0001	SJK
	RG_ST - DAG_S	84	<0.0001	RG_ST
	RG_ST - DAG_ST	93	<0.0001	RG_ST
	RG_ST - SJK	221	<0.0001	SJK
	DAG_S - DAG_ST	150	0.0019	DAG_ST
	DAG_S - SJK	249	<0.0001	SJK
	DAG_ST - SJK	250	<0.0001	SJK

The median of the CVs of the SJK method are smaller than the median of all other methods. The RG_ST median CVs are smaller than all other methods except SJK, and the DAG_ST median CVs are smaller than both simple methods. The median CVs of the stratum-specific method are smaller than the simple methods.

One limitation of the sign test is that it just looks at the sign instead of the magnitude of the difference. To find a superior method, we also looked at the actual values of the relative biases and the CVs for each of the variance estimation methods across all domains.

Table 2 shows the relative bias of the five variance estimation methods for New York across the 17 demographic characteristics for all firms (firms with and without paid employees) by firm count. The other states and employer status categories show similar results.

Table 2: New York total firm count relative biases by demographic characteristic and variance estimator

Demographic Characteristic	Relative Bias by Variance Estimator				
	RG_S	RG_ST	DAG_S	DAG_ST	SJK
All firms	0.174	31.417	0.174	-0.997	-0.988
Female	-0.032	0.048	-0.032	-0.040	-0.012
Male	-0.069	0.222	-0.069	-0.086	-0.076
Equally male/female	-0.068	-0.048	-0.068	-0.068	-0.055
Hispanic	-0.050	-0.049	-0.050	-0.058	-0.047
Non-Hispanic	-0.054	1.664	-0.054	-0.136	-0.146
Equally Hispanic/non-Hispanic	-0.132	-0.133	-0.132	-0.132	-0.104
White	-0.245	0.414	-0.245	-0.281	-0.276
Black or African American	-0.095	-0.074	-0.095	-0.107	-0.098
American Indian and Alaska Native	-0.078	-0.064	-0.078	-0.079	-0.049
Asian	-0.552	-0.544	-0.552	-0.555	-0.560
Native Hawaiian and Other Pacific Islander	-0.011	-0.018	-0.011	-0.011	-0.020
Some other race	-0.002	-0.002	-0.002	-0.002	-0.021
Veteran	0.030	0.039	0.030	0.028	0.030
Nonveteran	-0.106	0.864	-0.106	-0.149	-0.141
Equally veteran/nonveteran	-0.040	-0.039	-0.040	-0.040	-0.052
Publicly held	-0.079	0.070	-0.079	-0.085	-0.096

In certain instances (all firms, male, non-Hispanic, White, and nonveteran), the relative bias of the RG_ST method is very high compared to the other methods. For example, for "all firms" the relative bias for the RG_ST method is 31.417, that is, the average of the RG_ST variance estimate is much bigger than the true variance. There is little difference between the other four methods (RG_S, DAG_S, DAG_ST, and SJK). All five methods tend to underestimate the true variance. The relative biases for the total receipts estimates found no differences between any of the estimators.

Tables 3 and 4 show the CVs of the five variance estimation methods for New York across the 17 demographic characteristics for all firms (firms with and without paid employees) by estimate type (firm counts and total receipts). The other states and employer status categories show similar results.

Table 3: New York total firm count CVs by demographic characteristic and variance estimator

Demographic Characteristic	Coefficient of Variation by Variance Estimator				
	RG_S	**RG_ST**	**DAG_S**	**DAG_ST**	**SJK**
All firms	0.643	31.419	0.643	0.997	0.988
Female	0.449	0.445	0.449	0.444	0.014
Male	0.446	0.487	0.446	0.443	0.077
Equally male/female	0.441	0.437	0.441	0.440	0.056
Hispanic	0.454	0.451	0.454	0.452	0.048
Non-Hispanic	0.449	1.714	0.449	0.429	0.146
Equally Hispanic/non-Hispanic	0.428	0.427	0.428	0.428	0.112
White	0.429	0.535	0.429	0.439	0.276
Black or African American	0.442	0.430	0.442	0.438	0.098
American Indian and Alaska Native	0.451	0.445	0.451	0.452	0.065
Asian	0.591	0.582	0.591	0.593	0.560
Native Hawaiian and Other Pacific Islander	0.490	0.483	0.490	0.491	0.124
Some other race	0.472	0.470	0.472	0.473	0.050
Veteran	0.490	0.489	0.490	0.489	0.033
Nonveteran	0.438	0.953	0.438	0.429	0.142
Equally veteran/nonveteran	0.458	0.457	0.458	0.458	0.058
Publicly held	0.442	0.435	0.442	0.441	0.097

Across almost all characteristics, the SJK method has the smallest CVs for the firm count, almost by half in some instances. As with the firm relative biases in Table 2, the RG_ST method tends to have much higher CVs for firm estimates in certain groups (all firms, male, non-Hispanic, White, and nonveteran). There is little difference among the other three methods (RG_S, DAG_S, DAG_ST) for the firm counts.

Table 4: New York total receipts CVs by demographic characteristic and variance estimator

Demographic Characteristic	Coefficient of Variation by Variance Estimator				
	RG_S	**RG_ST**	**DAG_S**	**DAG_ST**	**SJK**
All firms	0.698	0.695	0.698	0.667	0.352
Female	0.830	0.831	0.830	0.829	0.459
Male	0.506	0.458	0.506	0.498	0.259
Equally male/female	0.559	0.560	0.559	0.560	0.351
Hispanic	0.953	0.712	0.953	0.957	0.771
Non-Hispanic	0.599	0.557	0.599	0.583	0.297
Equally Hispanic/non-Hispanic	1.887	1.831	1.887	1.886	1.899
White	0.608	0.554	0.608	0.592	0.313
Black or African American	0.931	0.872	0.931	0.929	0.812
American Indian and Alaska Native	1.171	1.192	1.171	1.173	1.034
Asian	0.549	0.529	0.549	0.547	0.321
Native Hawaiian and Other Pacific Islander	1.622	1.592	1.622	1.619	1.550
Some other race	0.946	0.944	0.946	0.947	0.817
Veteran	0.539	0.532	0.539	0.540	0.316
Nonveteran	0.525	0.478	0.525	0.515	0.249
Equally veteran/nonveteran	0.670	0.674	0.670	0.673	0.363
Publicly held	0.670	0.524	0.670	0.647	0.357

Across almost all characteristics, the SJK method has the smallest CVs for the total receipts, almost by half in some instances. There is little difference among the other four methods (RG_S, RG_ST, DAG_S, DAG_ST).

Although the SJK method seems to have the lowest CV, the amount of time required to calculate the variance estimates is extremely high compared to the other methods. We compared the real time for running the current SBO RG_S method to the SJK method, using one sample from our simulated universe. Then we compared times using the full 2007 SBO sample, which includes about 2.3 million firms for all states across a small portion of what we tabulate.

For our small study sample, the SJK code took 12.6 times longer in real time than the RG_S method. When we compared the times using the full 2007 SBO sample, we found that the SJK method took 73 times longer to run in real time. Currently, the processing time to run all the estimates for SBO publications takes about 12 hours, if we used the SJK method it would take over a month. For this reason, we rejected SJK as a possible SBO variance estimator. The processing times of the other methods (RG_ST, DAG_S, and DAG_ST) are similar to the RG_S method.

6 Conclusion

The SJK variance estimator was the superior method compared to the other variance estimators due to consistently producing a low coefficient of variation and showing no difference in relative bias to the other methods. However, the processing would take too long to create all SBO estimates. We recommend future research into more efficient processing for the SJK variance estimator, to be able to implement the best variance estimator possible.

The RG_ST showed huge fluctuations in both the relative bias and the coefficient of variation for the firm count for specific groups, so we dropped it from consideration for future SBO estimates.

Since there appears to be no difference between the RG_S, DAG_S, and DAG_ST methods, we referred back to Kott (2001). Kott recommends developing stratum-specific replicate factors for DAG and since the DAG_ST handles strata with few records ($n_h \leq 5$), which is evident in many of SBO strata, we believe the best variance estimator at this time is the DAG_ST.

7 Acknowledgements

We would like to thank the team of mathematical statisticians that worked with us on the research that went into this project, specifically Jeffrey Dalzell, Robin Gibson, Terry Pennington, Beth Schlein, and Meijin Ye. We would also like to especially thank Maxwell Mitchell for all the work and effort he put into the research and coding needed to complete this project.

8 References

Balogh, M. "2007 SBO Estimation." Internal memo, U.S. Census Bureau.

Conover, W.J. (1998). Practical Nonparametric Statistics (3rd ed.). New York: John Wiley & Sons.

Kott, P. (2001). "The Delete-a-Group Jackknife." Journal of Official Statistics, Vol. 17, No.4, 2001, 521-526.

Kott, P. (1998) "Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys." *NASS Research Report*, Washington, DC, March 1998.

Lohr, Sharon L (1999). Sampling: Design and Analysis. Boston: Brooks/Cole.

Steel, Philip, McNerney, Victoria, and Slanta, John (2009). "An Investigation of Stratified Jackknife Estimators Using Simulated Establishment Data Under and Unequal Probability Sample Design," *Proceedings of the Research Conference of the Joint Statistical Meetings,* Washington, DC, August 1-6, 2009, 5222-5231.

Thompson, K. (2003). "Investigation of Alternative Replicate Variance Estimators for the Annual Capital

Expenditures Survey," *Proceedings of the Research Conference of the Joint Statistical Meetings*, San Francisco, CA, August 3-7, 2003, 4226-4233.

Thompson, K. (2005). "An Emperical Investigation into the Effects of Replicate Reweighting on Variance Estimates for the Annual Capital Expenditures Survey, *Proceedings of the Research Conference of the Federal Committee on Statistical Methodology,* Washington, DC, November 2005.

Wolter, K. M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.

"Survey of Business Owners (SBO)." *Survey of Business Owners.* US Census Bureau, n.d. Web. 12 Aug. 2013.