REDESIGNING THE MONTHLY SURVEYS OF RETAIL AND WHOLESALE TRADE FOR THE YEAR 2000

Patrick J. Cantwell and Jock R. Black, Bureau of the Census*
Patrick J. Cantwell, Bureau of the Census, DSSD, Washington, DC 20233

Key words: Advance survey, early reporting, respondent burden, sample allocation, unduplication.

1. Introduction

The Census Bureau will soon select the samples for our monthly surveys of retail and wholesale trade. At this time we are re-examining the designs of these surveys and the procedures for selecting the sample--as well as some of the most serious problems we encounter.

(1) *Design of the Advance Monthly Retail Trade Survey.* The Advance survey suffers from several problems: a small sample, early reporting deadlines, heavy respondent burden, and occasional large revisions from the Advance estimate to that of the Monthly Retail Trade Survey (MRTS), published later. Further, our remedy to ease the burden creates concerns about the quality of the data used in the MRTS for some firms.

We propose two alternatives to the current design. One option--combining the Advance survey with the larger MRTS--would lower the sampling variance of the Advance estimates, would potentially decrease the revisions in the estimates we publish, and may bring with it operational efficiencies. Under a different option, we would select the samples so that the Advance survey and the MRTS have no small or medium-sized firms in common. The result would be reduced individual reporting burden, better sales reports from some sample units, and perhaps a higher response rate.

(2) *Sample design improvements.* We investigate several issues regarding the statistical procedures used by the Census Bureau to select the sample: how to obtain the proper sample size, which variable to use to allocate the sample, and how to automate the stratification procedures.

(3) *Coordinating the sample selection to reduce burden.* Currently we allow all frame units to be reselected when we determine new samples. This can place a heavy burden on units that fall in consecutive samples, and may lead to lower response. We propose a simple method for coordinating the sample selection from one instance to the next that would prevent smaller firms from being selected into consecutive samples, thus reducing their potential burden.

A much broader discussion of these topics is found in the full-length version of the paper available from the authors.

2. The Monthly Trade Surveys

The Bureau of the Census conducts several monthly surveys in retail and wholesale trade.

✦ The Monthly Retail Trade Survey (MRTS) measures sales in the kinds of business designated by Standard Industrial Classification (SIC) codes 52 through 59. These codes include retail automobiles, food and drink, furniture, and other goods.

✦ The Advance Monthly Retail Trade Survey (called the Advance, for short) is a subsample of the MRTS, conducted only a few days after the end of the month, to produce an early indication of retail sales for the month just completed.

✦ In the Monthly Wholesale Trade Survey (MWTS) the Census Bureau collects sales and inventory data from merchant wholesalers in SICs 50 and 51. SIC 50 covers durable goods; SIC 51 covers nondurable goods. (There is no Advance survey or advance report in wholesale.)

Because the design of the wholesale survey is very similar to that of its retail counterpart, the MRTS, we will focus on the retail surveys in this paper. Normally, on the ninth weekday of each month, the Bureau of the Census releases its first estimates of retail sales for the previous month as measured by the Advance. The most important estimated quantities are the level of sales for the month, and the change in sales from the prior month (expressed as a percent increase or decrease).

At the time the Advance estimates are made available, the Census Bureau also releases the preliminary estimates for the prior month, and the final estimates for the month before that. The latter two estimates are derived from the MRTS. Thus, within 75 days after the end of the reference month, we revise the Advance estimate with the preliminary estimate, and further revise the preliminary with the final estimate. For more information, see Cantwell and Caldwell (1998).

3. Should We Change the Design of the Advance Survey?

We are considering three alternative designs for the Advance survey.

(1) Retain the *current design.* That is, for the Advance

sample, draw a *subsample* of firms from the MRTS.

(2) Change to a *one-sample design*. That is, combine the Advance survey and the MRTS. Under this plan, we would mail or fax a questionnaire to the full sample and produce Advance estimates from those who respond by the Advance response deadline.

(3) Select a *nonoverlapping sample* for the Advance survey. After retaining some subset of the MRTS certainty cases (those selected with probability 1; typically, very large units) in the Advance sample, the Advance noncertainties would be selected from the remainder of the frame. There would be no noncertainty cases in common between the surveys.

Based on the following issues and other not discussed here, we feel the Census Bureau should seriously consider options (2) and (3) when redesigning the Advance survey.

Sample size, sampling variance. If the same sample design is retained, increasing the sample size in a statistically appropriate manner generally increases survey costs but lowers the sampling variance. The sample for the Advance survey now starts at about 4900 units. After removing refusals and other cases, we mail about 4100 questionnaires. In a typical month, about 2400 or 2500 responses are tabulated. Under a one-sample design (option (2)), the sample size for the Advance survey would be considerably larger than under options (1) and (3). The MRTS currently starts with about 13,000 sample cases; a one-sample design might start with 10,000 or 12,000 cases. If we could maintain a similar rate of response, we would tabulate about 5,000 cases--yielding smaller weights for most cases and a smaller sampling variance of the estimates.

Although the cost of producing the Advance estimates would rise with only one sample, the increase may not be proportional. As it becomes clear which cases can report before the deadline for data collection in the Advance survey, staff can concentrate on them--especially on firms with large (weighted) sales--and ignore those who cannot report in time. Further, some cost efficiencies may be realized in collecting and analyzing data under a one-sample design. For example, the Census Bureau currently has separate staffs of survey statisticians for retail trade--one to analyze data for the Advance survey, and another to analyze data for the MRTS.

Revisions to the estimates. For a specific data month, we first release an Advance estimate. We then revise this number with the Preliminary estimate one month later, and revise again with the Final estimate after another month. Naturally, we and our data users wish to avoid large revisions in the estimates. But to our knowledge, the following question has not been seriously raised and addressed within the Census Bureau: *What is more important--providing better estimates, or keeping the revisions smaller?* Obviously, if we can do both with one design, there is little need for argument.

Our current designs for the Advance and the MRTS apply different procedures for many important activities. Data collection is quite different for the two surveys, due mainly to the different time constraints. Editing methods differ greatly between the surveys. The Advance estimate includes only those units that report for the current month *and* the prior month; in the MRTS, a sample unit is included in the estimate without regard to its response status for the prior month. The Advance estimates do not capture the component of change due to births and deaths in the frame; in the MRTS, a sample of births is added--and deaths are removed from the sample--on a quarterly basis. In the Advance, nonresponse is addressed essentially by weighting up within the kind of business those units that respond in the current and prior months; in the MRTS, sales for nonrespondents are imputed. Even the methods of estimation differ between the two surveys.

The revisions from the Advance to the Preliminary (or Final) estimates can likely be reduced by standardizing some of these procedures, without adversely affecting the quality of the estimates. Under any of the three options, for example, a sample of births could be added to the Advance survey. But the one-sample option would provide for the greatest amount of standardization between the Advance and Preliminary (or Final) estimates. With all or most of these procedures done the same, and with the large sample size for the Advance estimates, it is likely that this option would produce the smallest revisions between Advance and Preliminary (or Final).

Early reporting and respondent burden. We address these two issues together, as their resolution may be intertwined. Because the Advance estimates for a given data month are usually released on the ninth weekday of the next month, all sample reports to be included must be received by the seventh weekday. It is not surprising then that many sample units cannot or will not report this early, leading to a response rate as low as 65% of the cases mailed (after some time in the sample). Further, many other reporters estimate their sales, possibly inducing bias or response variance in the estimates. (See Cantwell and Caldwell 1998.)

(One remedy for the problems of data quality caused by early reporting is to push back the release of the Advance estimates--and, with it, the deadline for data capture--by a week or so. Presumably, such a change would have the dual effect of increasing the response rate and, among those who respond, allowing time to provide a more accurate report. The obvious drawback here would be releasing these important estimates a week later.

When the Census Bureau asked a number of key data users their thoughts on postponement about seven years ago, the reactions were divided--some for delaying the release of the Advance data, some against. The issue has not been revisited since then. It appears that this strategy will not be considered further.)

For many cases that do not respond to the Advance survey, or for which we think the sales figure given is an estimate (perhaps based on a rounded sales number or comparison with prior reports from the firm), we contact them again a couple of weeks later, hoping to get their sales figure for use in the larger MRTS (for the same data month as the Advance just released). Each of these contacts consists of as many as three telephone calls or fax transmissions. This added reporting burden may well lead to lower response.

Obviously, we at the Census Bureau attempt to minimize these contacts to ease the burden on the respondents and our analysts. When we feel that the Advance figure is accurate--not an error and not merely a rough estimate--we usually transfer the sales figure directly from the Advance database to the MRTS database for use in the Preliminary and Final estimates. This procedure eliminates the second contact (or series of contacts) for the same data month. Unfortunately, for many of these cases the sales figure would have been changed had we made a later contact for the MRTS. Based on recent callbacks made to study this issue, we have seen that many early sales figures would change with a recontact--sometimes substantially. Thus, using data collected in the Advance for the MRTS--while reducing burden on the respondent--can lead to diminished quality in the Preliminary and Final estimates.

Option (3) addresses these issues satisfactorily--at least for the noncertainty cases. By selecting a nonoverlapping sample for the Advance survey, the units need only be contacted for the Advance survey. Further, there is no transfer of early (perhaps questionable) data to the MRTS, and thus no concern about the effect on the Preliminary and Final estimates.

As an additional benefit, under option (3) we can retire Advance sample cases at any time without affecting the MRTS. Because we feel that asking respondents to provide data earlier adds to their burden, we generally try to introduce a new Advance sample every two to three years. Under the current design (option (1)), their burden is only slightly reduced if these retired cases must continue to report in the MRTS. Under the one-sample design (option (2)), a respondent cannot be removed from only the Advance survey.

4. Sample Design and Stratification Issues

In this section we present several issues encountered when planning the sample selection.

Sample size and allocation concerns. The Census Bureau estimates sales and inventories for particular kinds of business such as automobile dealers, department stores, and restaurants. To help ensure that these detailed estimates meet sampling variability constraints, the frame is first stratified by kind of business. Then we further substratify within each kind of business before selecting a sample. At the risk of oversimplifying some activities and omitting important details, we describe five steps in this process.

1. *Determine substratum boundaries.* Based on the estimated volume of sales (more detail below) for the individual frame units and the variability among the units, we apply Dalenius and Hodges' cumulative $\sqrt{f}$ rule.

2. *Assign frame units to substrata.* The boundaries in step 1 are used.

3. *Determine the sample size for the kind of business.* According to the variability constraint placed on the estimate of the monthly level for this kind of business, the required sample size is computed. For some kinds of business, additional constraints are applied for estimates of aggregates or month-to-month change.

4. *Allocate the sample to the substrata.* Given the sample size from step 3, a portion is allocated to each substratum using Neyman allocation.

5. *Select a sample within each substratum.* In each substratum, we select a simple random sample without replacement based on the results of step 4.

The ideal variable for the measure of size in the first four steps--if we knew it--would be the monthly sales the unit *would* report, denoted by Y. However, as future sales are not available, we substitute a proxy. Among variables that are known when we determine the sampling parameters, let X_1 represent the one most closely associated with Y. At the Census Bureau, we consider this to be the *sales as reported by the unit in the prior economic census.* (Again, it is more complex than this. For some units in the frame, census data are not available.) We use X_1 to determine the stratum boundaries (step 1) and to assign the units to strata (step 2). Obviously, the resulting stratum boundaries are not as efficient as if we had known and used the Y values.

When determining the optimal sample size in step 3, there are other implications of not knowing Y. For a sample unit i with corresponding sales (X_{1i}, Y_i), the unit is placed in the appropriate sampling stratum based on X_{1i}. Yet the true variance of the estimate is a function of the variability of the Y_i *within strata determined by the* X_{1i}. For simplicity, we denote this variance of the estimator of the population total for Y by Var(Y; str X_1), where "; str X_1" indicates that the stratification used in the design and subsequent sample selection is based on X_1. Var(Y; str X_1) will usually be somewhat larger than Var(X_1; str X_1) because Y_i can fall in the same stratum as X_{1i} *or a different stratum*. In other words, the variance of the estimator for Y is generally increased because--not knowing Y when designing the sample--we had to stratify on X_1.

To anticipate this increased variance in the estimator and still satisfy variability constraints for the kind of business, we want to inflate the sample size when completing step 3. Under current procedures, the Census Bureau has used a second variable (call it X_2) in steps 3 and 4. The idea is to select X_2 such that Var(X_2; str X_1) is similar to Var(Y; str X_1). At the Bureau, the characteristic *payroll as reported in the previous census*, multiplied by a constant factor to put X_1 and X_2 on a comparable scale, has been used as X_2 to obtain the larger sample size.

This method of using X_2 in step 3 to inflate the sample size appears to have worked quite well. But we raise several questions. Is there one variable X_2 (census payroll?) such that Var(X_2; str X_1) approximates Var(Y; str X_1) well *in each of the many kinds of business* in retail and wholesale? Over the life of the sample (generally five years), as the association between Y and X_2 (as well as X_1) diminishes and cases drop out of sample, what happens to the relationship between Var(X_2; str X_1) and Var(Y; str X_1)? How does this affect the resulting variance of the estimator? Is there a special relationship between census payroll and reported sales that makes the former the proper choice as X_2? We were not able to find any research at the Census Bureau that addresses these issues.

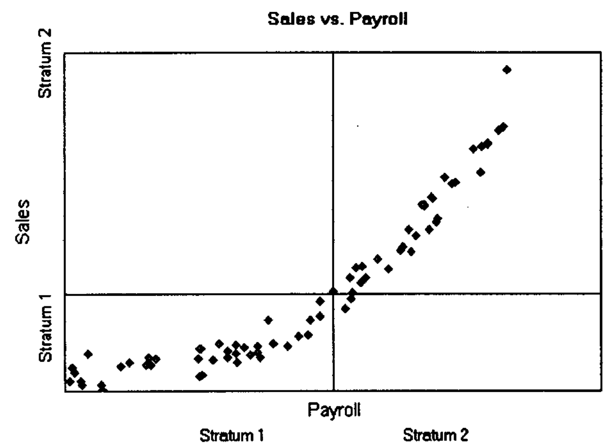
We wonder if a more direct approach to the problem might work better. Instead of relying on the variable X_2, why not treat the increased variance (due to the unknown Y) similar to a design effect and estimate the ratio

$$\frac{\mathrm{Var}\,(Y;\mathrm{str}\,X_1)}{\mathrm{Var}\,(X_1;\mathrm{str}\,X_1)}$$

The true value of Var(Y; str X_1) is unknown because we never have the entire population of Y values. Still, it can be estimated from current or past samples. Using the ratio above to inflate the sample size, different factors can be applied in the various kinds of business to reflect differing relationships between X_1 and Y.

Allocating the sample. Just as important, we question the Bureau's use of the variable X_2 in step 4--allocating the sample. Given the sample size determined in step 3, we want to allocate the sample to the substrata according to the number of units in the substrata and the variability of Y. As Y is unknown, it is then appropriate to allocate the sample using the variable most closely associated with Y. But by our definition, this is X_1.

Sometimes there is little difference in the efficiencies of two samples, one allocated according to X_1, the other allocated according to X_2. Yet the graph below illustrates a potential problem where X_1 (sales) is used to determine the substratum boundaries, and X_2 (payroll) to allocate the sample to the substrata. In this instance, Neyman allocation based on X_2 (payroll) would assign more sample cases to stratum 1 because the payroll of its units is more variable and stratum 1 has more units. Basing the allocation on X_1 (sales), however, the greater variability in stratum 2 tells us to allocate more of the sample there.



Automating the design decisions. One problem that arises in severely skewed populations such as ours is setting up the initial cells necessary for deriving stratum boundaries. Here we face a problem of requiring either many cells with no observations or fewer cells of unequal lengths. We decided to use unequal-length cells to form frequency counts. This led to a problem caused by the extreme skewness of the distributions.

The Dalenius and Hodges cumulative $\sqrt{f}$ procedure is modified when using cells of unequal length. Here, f represents the number of sampling units in each cell of the frequency distribution. As Cochran (1977, p.130) states, "when the interval changes from one of length d to one of length ud, the value of root-f for the second interval is multiplied by root u when forming" the cumulative total. This adjustment essentially adjusts the cumulative totals assuming that there is the same uniform distribution over the entire range of the cells. But for extremely skewed distributions this assumption is not warranted.

As an example, for a particularly skewed distribution, the adjustment factor, u, was set to approximately 400,000,000 for the final cell which had an actual frequency of one. When implementing the cum $\sqrt{f}$ rule, this is equivalent to adding 20,000 sampling units in 20,000 distinct cells of unit length. However, it seems to be unreasonable to expect 20,000 sampling units when we know that there is only one.

In addition, when one frequency cell contributes 20,000 to the cum $\sqrt{f}$ scale, setting up equal-length intervals on that scale requires each interval to be very large. For instance, if adjacent points on the cum $\sqrt{f}$ scale are x and $x + 20{,}000$, attempting to create strata with endpoints corresponding to x, $x + 10{,}000$, and $x + 20{,}000$ will result in one empty stratum. In some cases, this phenomenon resulted in populations of more than 10,000 sampling units accommodating only three strata. To surmount these problems, we modified the procedure so that the adjustment factor, u, would not exceed 1% of the number of sampling units in the frame.

Focusing on the properties of the cum $\sqrt{f}$ algorithm allowed for an automated method for setting the number of strata. The rule creates equal-length intervals on the cum $\sqrt{f}$ scale, with the length of each interval being the cum $\sqrt{f}$ divided by the number of strata desired. It is apparent that the number of strata is a function of the cum $\sqrt{f}$ and the length of each interval. By determining the minimum length of these intervals, we can determine the maximum number of strata a particular frequency distribution will allow. For our surveys, we prefer to use 13 strata wherever possible. Because the software we used would error-terminate if we requested too many strata, it was important to know *a priori* whether 13 strata could be fit based on the calculated frequency distribution. We determined the maximum number of strata allowed using the following logic.

In order to create approximately equal-length intervals with the property that at least one value on the cum $\sqrt{f}$ scale is in each interval, the minimum length of any interval must be less than half of the maximum value of uf for any cell. Thus we calculated the required minimum length of the intervals and divided the cum $\sqrt{f}$ by this minimum. This yielded the maximum number of strata possible. If the maximum was less than 13, we set the number of strata requested to the maximum. Otherwise, we used 13 strata.

5. Coordinating the Sample Selection to Reduce Burden

We considered several ways to lessen the reporting burden in our samples. One strategy is to prevent the smaller firms from being selected to report in *the same survey* two consecutive times. We started by considering continuous sample rotation. Under such a procedure, one introduces new sample units and retires others on a regular basis, perhaps every month or once a year (Srinath and Carpenter 1995).

We decided not to pursue this option, as several other major changes to our monthly trade surveys must be made by 2001--specifically, the shift in classification systems from SIC coding to the North American Industry Classification System (NAICS), the possible redesign of the Advance survey, and the move to a standardized processing system for the Census Bureau's economic surveys. Implementing a continuous sample rotation would require a dramatic change in the procedures we use for introducing units into the sample, and a serious study of the implications on sampling, data processing, the treatment of births and deaths, and other issues.

For now, we propose instead to implement a simple procedure we call "unduplication." The method is simple to apply, and appropriate statistical adjustments can be made to the estimation weights. Briefly, we try to ensure that a unit selected for the 1997 sample is not reselected for the 2001 sample. All noncertainty units--or all those below a specific level of sales--that were selected for the 1997 sample would be matched to the 2001 frame and removed from the frame; they would have *no* chance of being selected for 2001. All units remaining in the sampling substratum would have an increased conditional chance of selection. Later, we would make sure a unit selected for sample in 2001 is not reselected for 2006. There would be *no* attempt to prevent units from being selected into the sample for 1997 and later for 2006; this would happen to a small number of units, but they would be guaranteed five years out of sample between selections. We exempt certainty units from unduplication because of their importance to the survey estimates.

To see how the procedure might work, we start by dividing a kind of business stratum into several sampling substrata, determined by measure-of-size boundaries. Restricting our attention to a specific substratum, suppose there are initially N_A frame units. After matching the 1997 sample to the frame (including this substratum) and removing common units, there are N_B ($\leq N_A$) units remaining in the substratum. Let X_A and $\bar{X}_A$ represent the total and mean, respectively, of the *measures of size used for sampling* in the substratum, where we include all N_A units--matched and unmatched. X may be retail sales as determined in the recent economic census. Let X_B and $\bar{X}_B$ be the total and mean for the frame with matched units removed. Finally, define Y_A, $\bar{Y}_A$, Y_B, and $\bar{Y}_B$ analogously, where the Y terms correspond to the units' sales values *at the time the survey is conducted.* We select n sample units from the N_B units in the unduplicated frame using simple random sampling without replacement. Note that for X (unlike for Y), all values in the frame are known when we select the sample and compute the estimates.

For this substratum, Y_A can be estimated in many ways; here we consider only two methods. In what follows, Σ_s indicates the sum over all sample observations arising from the substratum. First, let

$$\hat{Y}_A^{(1)} = \sum_s w^{(1)} y_j ,$$

where $w^{(1)} = \dfrac{N_A}{n} = \dfrac{N_A}{N_B} \times \dfrac{N_B}{n}$. (1)

As an alternative, we consider a ratio-type estimator:

$$\hat{Y}_A^{(2)} = \sum_s w^{(2)} y_j,$$

where $w^{(2)} = \dfrac{N_A}{n} \times \dfrac{\bar{X}_A}{\bar{X}_B} = \dfrac{X_A}{X_B} \times \dfrac{N_B}{n}$. (2)

Each estimator, in weighting up the sample, accounts for the two stages of "selection": first determining the units eligible for sampling, and second randomly selecting units from those eligible. However, while the first estimator accounts for the "first stage of selection" using a ratio of the *number of units retained* in the frame (N_A / N_B), the second uses instead a ratio of the total *volume of sales retained* as given by the measure of size (X_A / X_B).

Because the sample for 1997 was selected by simple random sampling within substrata, it is reasonable to assume that the units matched and removed from the frame for the 2001 sample represent a randomly determined subset of the frame. (The stochastic mechanism for determining the matches is actually much more complex. The substrata used for selection in 2001 are not necessary the same as those used for 1997; many frame units will have changed in size or in their kind of business by 2001; the 2001 substrata are a mix of units subjected to sampling in 1997, and births that were subjected to birth sampling between 1997 and 2001. Still, we feel that treating the determination of matches as a simple first-stage of sampling is a reasonable model.)

With this assumption, and for any fixed number of matched units (N_A - N_B), the substratum estimator of total $\hat{Y}_A^{(1)}$ is *unconditionally* unbiased, being the usual Horvitz-Thompson estimator for a two-stage random sampling process. (By unconditionally, we mean over all possible sets of $N_A - N_B$ matches with equal probability.) $\hat{Y}_A^{(2)}$ has a small but negligible unconditional bias.

What is more important is the *conditional* expectation of the estimators. After all, when it is time to match the frame units to the 1997 sample and unduplicate, only the one realized match will matter--not all the possible random combinations of matches. Given a specific set of matched units represented in the values of N_B , X_B , and Y_B, the conditional expectation of the estimator $\hat{Y}_A^{(1)}$ is easily shown to be $N_A \times \bar{Y}_B$. Thus, its conditional bias is

$$N_A (\bar{Y}_B - \bar{Y}_A). \qquad (3)$$

The conditional expectation of $\hat{Y}_A^{(2)}$ is $\dfrac{X_A}{X_B} \times N_B \times \bar{Y}_B$. Its conditional bias is then

$$N_A \left(\frac{\bar{X}_A}{\bar{X}_B} \times \bar{Y}_B - \bar{Y}_A \right). \qquad (4)$$

From (3), one sees that the conditional bias of $\hat{Y}_A^{(1)}$ can be large in absolute value if the mean sales value of a matched unit is a bit different from that across the entire substratum. However, as seen in (2), the estimator $\hat{Y}_A^{(2)}$ adjusts for this possibility by using the ratio of the mean measures of size; the results are reflected in (4).

It might be noted that the estimators in (1) and (2) are only two in a class of such estimators; others can be considered. Further, adjusting the weight as in (2) is not always desirable. If the relationship between the measure of size and the reported data is not strong enough, the bias reduction can be small while the variance of the estimator increases. Finally, if there are many characteristics to be estimated in the survey, it may be difficult to determine one variable to adjust the weights that works well for all characteristics. (We assume that, to maintain simplicity and consistency across estimates, one would prefer to assign each sample unit just one weight for all characteristics.) However, in the monthly trade surveys, *because we only estimate sales and inventories,* using census sales to adjust the weights should work well.

Acknowledgment

The authors thank Nelvy Viguera-Mejia for providing computer support for much of this work, and Dennis Schwanz and Michael Ikeda for reviewing the paper and providing helpful comments.

References

CANTWELL, P. and C.V. CALDWELL (1998). "Examining the Revisions in Monthly Retail and Wholesale Trade Surveys Under a Rotating Panel Design," *Journal of Official Statistics, 14, No. 1,* Statistics Sweden, 47-59.

COCHRAN, W.G. (1977). *Sampling Techniques,* 3rd ed., J. Wiley & Sons, New York, NY.

SRINATH, K.P. and R. CARPENTER (1995). "Sampling Methods for Repeated Business Surveys," in *Business Survey Methods,* ed. by Brenda J. Cox, et al., J. Wiley & Sons, New York, NY, 171-183.

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.