SSC Annual Meeting, June 1999
Proceedings of the Survey Methods Section

ESTIMATION AND BENCHMARKING FOR MONTHLY SURVEYS

M.Majkowski and R.Kaushal[1]

ABSTRACT

Monthly business surveys are usually benchmarked to total to the more detailed and accurate annual survey. The most current annual estimates are usually about two years out of date. If auxiliary data from the annual survey in the form of a regression type of estimator is used, discontinuities will occur in the monthly series whenever new auxiliary data from the annual survey is introduced. Benchmarking is unable to distinguish between the resulting biases and the true month-to-month change. Various options to handle such a situation are considered in this paper. The redesign of the Monthly Survey of Manufacturing and its benchmarking to the Annual Survey of Manufacturing is an example of this situation at Statistics Canada that will be referred to in this paper.

KEY WORDS: Benchmarking; Pro-rated approach; Denton approach; Regression estimator; Horvitz-Thompson estimator.

RÉSUMÉ

Les enquêtes-entreprises mensuelles sont habituellement étalonnées par rapport aux enquêtes annuelles plus détaillées et plus précises. Les estimations annuelles les plus actuelles sont habituellement vieilles de deux ans. L'utilisation de données auxiliaires de l'enquête annuelle sous la forme d'un estimateur du type régression produira des discontinuités dans la série mensuelle toutes les fois que de nouvelles données auxiliaires de l'enquête annuelle seront introduites. L'étalonnage ne peut distinguer entre les biais résultants et le changement réel d'un mois à l'autre. Dans cet article, on considère diverses options pour faire face à une telle situation. La nouvelle conception de l'Enquête mensuelle sur les industries manufacturières et de son étalonnage par rapport à l'Enquête annuelle sur les industries manufacturières est un exemple de cette situation à Statistique Canada qui sera abordé dans cet article.

MOTS CLÉS : Étalonnage; approche au pro-rata; approche de Denton; estimateur par la régression; estimateur Horvitz-Thompson.

1. INTRODUCTION

Two Statistics Canada surveys that collect data from manufacturing units in Canada will be used as examples in this paper to describe the issues that arise in estimation as a consequence of benchmarking. The first survey is the Monthly Survey of Manufacturing (MSM). It is a monthly sample survey of about 11,000 units and it collects basic information on a unit's shipments, inventories and orders. The second survey is the Annual Survey of Manufacturing (ASM). It is a census of the manufacturing units in Canada conducted once a year. The ASM collects detailed commodity information in addition to the basic information collected by the MSM on a monthly basis. Recently, the MSM has been undergoing a complete redesign of its survey methodology and systems to make it more efficient. Another goal of the redesigned MSM is for it to be consistent with the results produced by the ASM. The documentation of the redesigned MSM is found in Kaushal, Majkowski and Thomas (1999).

Section 2 describes some methods that can be used for estimation of a monthly survey. The use of good auxiliary data in the estimation method may yield more reliable estimates. This is described in Section 3. Section 4 describes the requirement of sub-annual estimates being forced to agree with an annual survey for the same reference period. This is our definition of "benchmarking" in this paper. Section 5 emphasizes the importance of considering both estimation and benchmarking at the same time. Section 6 concludes

[1] M.Majkowski, and R.Kaushal, Statistics Canada, 11th floor, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario, Canada, K1A 0T6.

the paper by summarizing the strategy that was decided upon for the redesigned MSM.

2. ESTIMATION OF A MONTHLY SURVEY

For monthly surveys, various estimators can be used depending on the auxiliary information available and the type of estimate required. The auxiliary information can often exist in either an annual or sub-annual (i.e. monthly) form. For a monthly survey, auxiliary information that is available in a monthly form is considered ideal because it is a source of continuous data that matches the frequency of the undertaken survey. However, monthly auxiliary data is often not available. Any data that exists may not contain a variable that is well correlated with the variable of interest from the survey. Annual data, however, are more frequently available and their data may be well correlated with the variable of interest.

If the auxiliary data, annual or sub-annual, are not well correlated with the variable to be estimated, only the simple expansion estimator (Horvitz-Thompson (HT) estimator) can be used. The estimates from such an estimator are unbiased.

However, if auxiliary information is available from a single source for all units then a regression estimator can be used. Regression estimators gain efficiency over the HT estimator by taking advantage of the correlation between the variable being estimated and the auxiliary variable. However, these estimators are biased, although the bias is negligible in large samples. It should be noted that there will be situations when the auxiliary information from the same source is not available for all units. In this case, other sources can be used to supply the auxiliary data. When data from multiple sources are used as auxiliary information, their use can cause various problems. For instance, the data common to the various sources may not be consistent (different reference periods, different concepts, etc.). In cases of multiple sources, it is possible to use a combination of estimators for subsets of the sample. This combination of estimators results in a compound estimator. A compound estimator studied during the redesign of the MSM included two estimators (the ratio and the Hàjek estimator) to estimate the population total Y. The ratio estimator (Hidiroglou, Choudhry and Lavallée, 1991) was used when the correlation of the variable of interest with the annual auxiliary data (ASM) was good. The Hàjek estimator was used in cases where the correlation was not as good. The compound estimator in this case for the population total Y took the following form:

$$\hat{Y} = \frac{X}{\sum_{s_1} a_k x_k} \sum_{s_1} a_k y_k + \frac{N}{\hat{N}} \sum_{s_2} w_k y_k \qquad (2.1)$$

where s_1 refers to that part of the sample where the ratio estimator is applied and s_2 refers to the part where the Hàjek estimator is used. The survey variable for each unit k is denoted y_k. In the ratio estimator, X refers to the total of the auxiliary data for all units in the population, x_k refers to the auxiliary data value for the sampled unit k and a_k is the design weight for the sampled unit k. In the Hàjek estimator, N is the size of the population, $\hat{N}$ is the size of the population that is estimated from the sample and w_k is the design weight of the sampled unit k. This method takes advantage of good auxiliary information where possible by using the ratio estimator and uses the Hàjek estimator where the population counts of the units are accurate but the auxiliary information itself may not be adequate. However, one drawback of this method is that there might be too few units in one or both of the partitions when the sample size at the cell level is small. This may lead to some problems in calculating the variance.

3. USING AUXILIARY DATA FOR MONTHLY SURVEYS

As stated previously, auxiliary data may be available from different sources. Of the many monthly series in existence, the MSM will be used as an example to give an idea of what auxiliary data is available, how it can be used and its impacts on the estimates. For the MSM, there are four possible sources that one could consider. The four possible sources are:

1) Counts of manufacturing units from the Business Register (BR).
2) Manufacturing unit's annual revenue from the BR called Gross Business Income (GBI).
3) Manufacturing unit's annual shipment value from the ASM called Goods of Own Manufacture (GOM).
4) Manufacturing unit's monthly revenue from Revenue Canada's GST files (not currently available).

Sources 1) and 4) are examples of auxiliary data that are updated each month and where the data is continuous. The use of continuous auxiliary data implies that the monthly survey estimates will have no breaks. Source 1) may be used in the form of a Hàjek estimator. Source 4), though not yet available, may be used in the form of a regression estimator. If the monthly GST data are well correlated with the variable of interest from the survey, the estimates will have a smaller variance.

Sources 2) and 3) are examples of annual auxiliary data that are updated once a year. However, the annual data lags the monthly data by 18 to 24 months. There are two issues to keep in mind with annual data. One is due to the fact that the annual survey is far behind the monthly survey, thereby causing the variables from the monthly and annual to be less correlated. Secondly, breaks will be introduced in the monthly series when the annual data is updated. The series will be disrupted when moving from December to January because of the different annual auxiliary data being used for the two months. For instance, the estimate for December 1996 could be derived using 1996 annual data as the auxiliary information while 1997 annual data could be used as the auxiliary information for January 1997. It should be noted that a combination of sources 2) and 3) was used when the ratio estimator was calculated. Source 3 was the auxiliary information when it was available for a particular unit and source 2) was used when it was not.

Figure 1 displays a monthly time series of MSM data (monthly shipments by establishments in the electrical and electronic products industry) from January 1994 to January 1999. It should be noted that all four figures in this paper pertain to the same industry. The dotted line in Figure 1 represents the estimates obtained by using the simple expansion estimator (Horvitz-Thompson (HT)), while the solid line represents the estimates obtained by using a combined ratio estimator with ASM data (annual shipments) being used as the auxiliary variable.

The vertical lines in Figure 1 indicate the months of January in the time series. This is where the false breaks can be introduced. The figure shows that the Horvitz-Thompson estimator certainly gives a much more consistent December to January movement over the years as compared to the movement given by the ratio estimator. The estimates from the Horvitz-Thompson estimator are generally at a higher level each month compared to the estimates from the combined ratio estimator. In the MSM, the sample over time has become somewhat biased as small units have been replaced with larger units. The replacing units assume the weights of the smaller discarded units. This has resulted in the Horvitz-Thompson estimate yielding a larger estimate than reality. The combined ratio estimate is less affected by this biased sample due to the use of the auxiliary information from the ASM.

4. BENCHMARKING

Benchmarking is often required for sub-annual surveys. Benchmarking in this paper refers to the process of forcing sub-annual estimates to add up to the estimates from an annual survey for the same reference period. In many business surveys, the use of benchmarking implies that the annual survey data and the sub-annual survey data collected from the same unit can be compared. If a difference between the two surveys exists, the monthly data is revised to reduce the response error in the sub-annual survey data. The reason for this is that the annual data is judged to be of better quality. The end result of benchmarking is that users of the two surveys will have consistent totals from both surveys of the annual estimate. Regression estimators do not guarantee that the sum of 12 monthly estimates will add exactly to the corresponding totals from the annual survey. However, benchmarking does guarantee that this consistency happens.

There are several techniques used for benchmarking a monthly survey to its annual census. The simplest approach is to multiply the monthly figures with the ratio of the annual to the twelve months total and revise the figures. This approach is known as the pro-rated approach. The pro-rated approach is simple and each yearly benchmark is independent of others. The drawback of this method is that jumps can occur in the series when a new benchmark factor is used. In this regard, the problem is similar to using a regression estimator, where the jumps occur when new size measures (auxiliary data) are introduced.

Figure 2 displays a benchmarked and non-benchmarked monthly time series of MSM data from January 1993 to January 1999. The benchmarking has been done using the pro-rated approach. Also on the graph are horizontal lines representing 1/12 of the ASM value for the respective year. Horizontal lines for the months in 1998 and 1999 do not exist because those ASM values are not yet available. For the months in the years of 1993 to 1995, the MSM values have been benchmarked to their respective years. For the months following 1995, the MSM values have been benchmarked to the 1995 ASM. During the benchmarking of the 1995 data, a benchmark factor (based on the ratio of the 1995 annual to the sum of the 12 months) was calculated and this factor was applied each month following 1995. For the months following 1995, this benchmark factor will be replaced by a factor calculated from a subsequent benchmarking to a more recent ASM (i.e. 1996 ASM, 1997 ASM, etc.).

Figure 1. Comparison of HT and Ratio

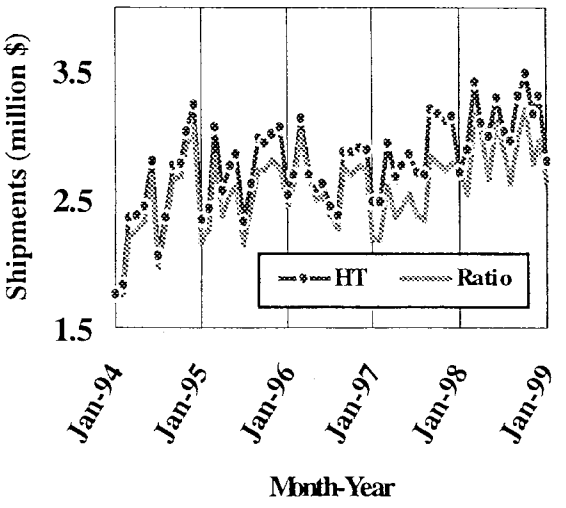

Figure 2. MSM (pro-rated and non-benchmarked) series

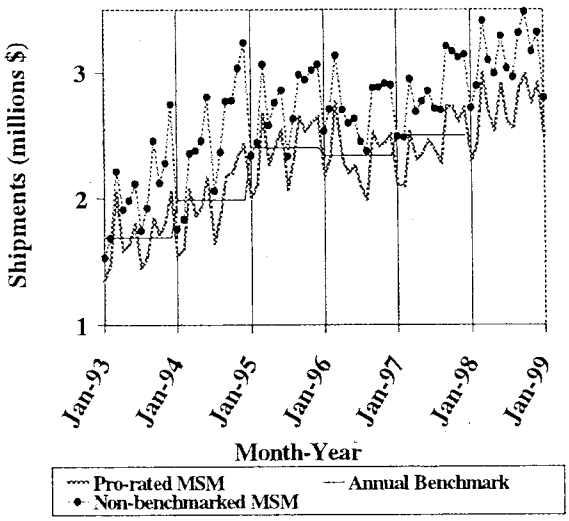

Figure 3. Benchmarked and Non-benchmarked (HT and Ratio) MSM Series

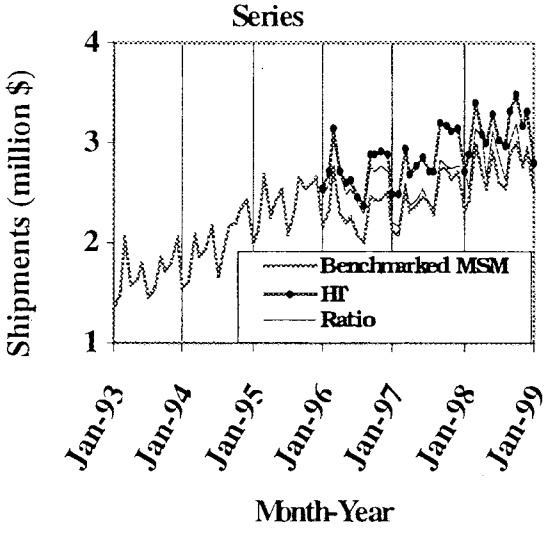

Figure 4. Benchmarking MSM with Denton Approach

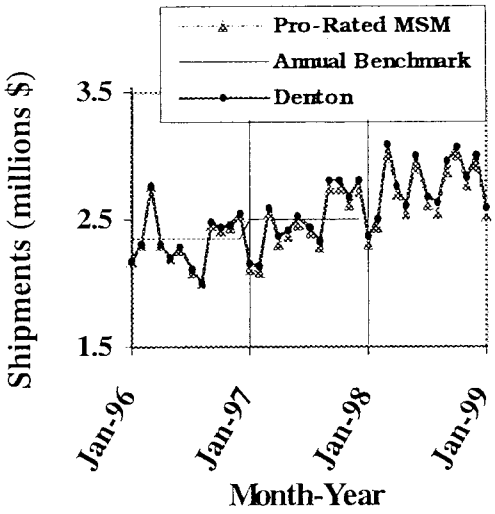

Figure 2 shows that the non-benchmarked series can differ substantially from the benchmarked series. In many of the years that are graphed, the non-benchmarked series is always at a higher level than the line representing 1/12 of the annual benchmark value. This shows that a large difference does exist between the MSM and the ASM in these years. Through benchmarking, this difference is eliminated within the particular year. Figure 2 also shows that the month-to-month movement that exists in the non-benchmarked series is largely preserved after benchmarking. The only month-to-month movement that can be disrupted is the December to January movement. This movement is largely affected by the change in the annual benchmark value that is being used for the benchmarking.

In the case of a monthly series that is benchmarked to an annual series, months can be estimated where no current annual series is available. For instance, the monthly estimate for June 1999 is required and the most recent annual data available is from 1995. The months of 1995 are benchmarked to this 1995 annual data, but what about the months that follow? They are also benchmarked to 1995 in the sense that the discrepancy between the 1995 annual estimate and the 12 months of 1995 is assumed to exist in the following years. A benchmark factor, based on the ratio of the 1995 annual to the sum of the 12 months, is applied to the months that follow. Applying the benchmark factor to the months following the benchmarked year results in the avoidance of a series break. If the benchmark factors between years differ, there is a jump in the series at the end of the benchmarked portion. Often with benchmarking, the previous benchmark factor is used for the following years. Revisions going back to the benchmark year are published every time a benchmark is produced.

Figure 3 displays the effect on the time series of not carrying forward the benchmark factor after the 1995 year. The dotted line represents the HT estimation while the lightly dashed line represents the ratio estimates. The line representing the HT estimates differs the most from the benchmarked MSM line.

Denton (1971) proposed a method to provide year-to-year continuity while minimizing the impact on the month-to-month change. This method was expanded by Cholette and Dagum (1994) to account for the possible bias in the series. Quadratic minimization methods are utilized in these two proposals. Such methods yield smooth series from one year to the next. However, the cost of preserving the smoothness is that monthly estimates of previously benchmarked years can be changed resulting in revisions to the series. The Denton type approach is based on the assumption that the month-to-month change estimates are representative of true changes before benchmarking occurs. This assumption does not hold in the case of regression estimators when auxiliary data is updated annually. In the month that the update occurs, the month-to-month change estimate reflects, in part, the change in auxiliary information. Hence, if such an estimation strategy is used with a Denton type approach, the changes due to an auxiliary data switch become a part of the series. Therefore, even before benchmarking is attempted there is a problem with using regression estimators.

5. ESTIMATION TOGETHER WITH BENCHMARKING

When determining the methodology for a monthly survey requiring benchmarked estimates, estimation and benchmarking strategies have to be viewed together because they have an impact on the time series. Depending on the constraints and the needs of the survey, various approaches can be used.

The Horvitz-Thompson (HT) estimator provides an unbiased estimate of level and month-to-month change. The regression estimator with a large enough sample will provide a lower variance than the HT estimator. However, the regression estimator can provide an inaccurate month-to-month change when a switch to another year of auxiliary data has occurred. HT can be used with either a pro-rated or the Denton benchmarking. HT with the pro-rated approach results in time series that can have discontinuities (false breaks) from year-to-year. Both HT and regression estimators can be improved at the sample design stage (i.e. stratification) with good auxiliary information. Using HT with the Denton approach is the best combination to avoid discontinuities since any annual breaks are smoothed into the rest of the year. This combination moves the use of updated auxiliary data from the estimation stage to the sample design stage. Updates to auxiliary data can be incorporated into the restratification or resampling procedure. The maximization of the sample overlap (using the Kish-Scott or the Keyfitz approach) will minimize the effect on the estimates from applying these procedures.
Figure 4 compares the HT-Denton approach with the past method that was estimated with the HT estimator and benchmarked using the pro-rated approach. The figure has the lines from the Denton approach and the pro-rated approach mostly coinciding with each other. This is due to the fact that the annual benchmarks for 1996 and 1997 are

fairly similar and a small time period (Jan. 1996 to Jan. 1999) of MSM data had the Denton approach applied to it.

6. CONCLUSION

For the MSM, unbiased current estimates are of utmost importance. The HT estimator with stratification by an annual auxiliary variable (from the ASM) in conjunction with the Denton approach has been chosen as the method. The most recent benchmark factor from the Denton approach will be used to bridge the jump between benchmarked years and current estimates. This factor is constant and hence does not affect current estimates of change. Once the new benchmarking is done (2 years later), the current estimates of change will remain unbiased. The benchmarked series is revised but the revisions to the monthly movements will be minimized.

ACKNOWLEDGEMENTS

The authors are grateful to Mike Hidiroglou for helpful comments on the topic and on the paper. They are also thankful to Marietta Morry and Mark Armstrong for reviewing the paper prior to submission.

REFERENCES

Cholette, P. A., and Dagum, E. B. (1994), "Benchmarking Time Series with Autocorrelated Survey Errors," *International Statistical Review*, 62, 365-377.

Denton, F. T. (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," *Journal of the American Statistical Association*, 66, 99-102.

Hidiroglou, M. A., Choudhry, G.H., and Lavallée, P. (1991), "A Sampling and Estimation Methodology for Sub-annual Surveys," *Survey Methodology*, 17, 195-210.

Kaushal, R., Majkowski, M., and Thomas, S. (1999), "Redesign of the Monthly Survey of Manufacturing," *Proceedings of the Section on Survey Research Methods, American Statistical Association Conference.*