VARIANCE ESTIMATION FOR THE REDESIGNED SURVEY OF HOUSEHOLD SPENDING

Joanne Charlebois[1] and Guylaine Dubreuil[2]

ABSTRACT

A major redesign of the Survey of Household Spending conducted annually by Statistics Canada has been undertaken over the last few years. A new continuous data collection model combining the use of a diary and a recall interview was developed. We give an overview of the variance estimation methods used, which are based on the Rao-Wu bootstrap for a stratified multi-stage sample design. The changes brought about by the redesign and their effects on the estimated coefficients of variation of survey variables, which may now have components from both the interview and diary sources, are highlighted.

KEY WORDS: Variance, bootstrap, multi-modal survey, redesign, expenditure diary.

RÉSUMÉ

L'Enquête sur les dépenses des ménages menée chaque année par Statistique Canada vient de faire l'objet d'un remaniement majeur. Un nouveau modèle de collecte continue utilisant un journal ainsi qu'une entrevue personnelle a été développé. Nous présentons les méthodes d'estimation de la variance utilisées pour l'enquête, méthodes qui sont basées sur le bootstrap de Rao-Wu pour plans d'enquête stratifiés à plusieurs degrés. Les changements engendrés par le remaniement sont mis en évidence ainsi que leurs effets sur les coefficients de variation des variables d'enquête, qui peuvent maintenant comporter des composantes de l'entrevue et du journal.

MOTS CLÉS : Variance, bootstrap, enquête multimodale, remaniement, journal de dépenses.

1. INTRODUCTION

The Survey of Household Spending has recently undergone a major redesign. A new continuous data collection model, combining the use of a recall interview and a diary, was developed to replace the one-year recall interview conducted in the first quarter of each year. The content of the periodic Food Expenditure Survey was also integrated. A pilot survey was conducted over a one-year period in 2007 and 2008 to evaluate this new collection model. This was followed in 2009 with a parallel run of the original and redesigned collection models at reduced sample sizes in order to investigate the impact of the change on the survey products. The redesigned survey was fully implemented in January 2010.

The major changes incorporated in the new collection methodology have an effect on all of the data processing and estimation steps, and have led to the development of new methods to deal with them. In particular, the variance estimation methodology was modified in order to handle the two collection modes, as well as to incorporate some other improvements not directly related to the redesign.

This paper will present a summary of the new variance estimation methodology. The redesign of the survey is briefly described in the next section followed by a description of the main steps of the weighting process in Section 3. The variance estimation process and some results from the parallel surveys of 2009 are presented in Sections 4 and 5 respectively. Section 6 concludes the paper by describing some of the challenges and future work we anticipate.

[1] Joanne Charlebois, Statistics Canada, 16 Q, R.-H.-Coats Building, Ottawa, Ontario, Canada, K1A 0T6, Joanne.charlebois@statcan.gc.ca
[2] Guylaine Dubreuil, Statistics Canada, 16 B, R.-H.-Coats Building, Ottawa, Ontario, Canada, K1A 0T6, Guylaine.dubreuil@statcan.gc.ca

2. REDESIGN OF THE SURVEY

The Expenditure Survey Program serves mainly as a data source for the System of National Accounts, the update of the basket used in the computation of the Consumer Price Index and a variety of social research applications, including the development of social policy simulation models. Prior to the redesign, the program was comprised of two surveys: The Survey of Household Spending (SHS) and the Food Expenditure Survey (FES). The SHS was conducted annually on a sample of about 21,000 households between January and March. Respondents were asked during a personal interview to report expenditures made over the previous calendar year for a comprehensive set of expenditure categories. The FES was conducted on a periodic basis and asked selected households to report their detailed expenditures on food in a diary for two consecutive weeks.

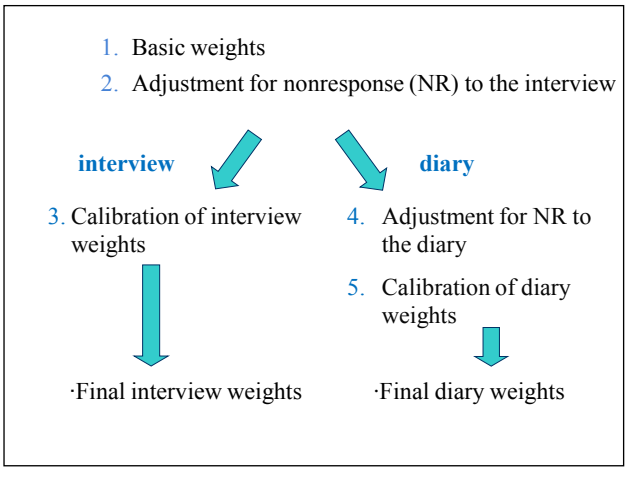
A new continuous data collection model, combining the use of a recall interview and a diary, was developed to address various issues with these two surveys; see Tremblay et al. (2010). Under this model, the annual sample is divided into twelve sub-samples assigned to monthly collection cycles. A questionnaire is administered to respondents to collect data on regular expenditures, such as rent and electricity, using a last payment approach, and on less frequent expenditures using recall periods of one, three or twelve months. Sampled households are also asked to report their expenditures, including detailed food expenditures, in a diary for a period of two weeks. Households are requested to include all of their expenditures with a few exceptions; however, diary data are mainly used to estimate the most frequent expenditures such as food and personal care, which are difficult to recall even for a short reference period.

3. WEIGHTING STEPS

The SHS-redesign sample is selected according to a stratified two-stage design from the Labour Force Survey (LFS) area frame. The annual sample allocation is done similarly to the original SHS. First the annual sample is allocated to the approximately 1,000 strata; the sample is then distributed among the 12 months of the year. Within strata, primary sampling units (PSUs) are selected, and, at the next stage, a sample of dwellings is selected from address lists using systematic sampling. On average, the number of PSUs selected per stratum is about 3, so each stratum may not be represented every quarter. For this reason strata were grouped to form "superstrata" with the goal of striving for quarterly sample representativity. Strata were grouped according to criteria based on geography as well as other characteristics such as high income strata; see Nadeau et al. (2007) for details. From 2009 to 2011, all interview respondents received or will receive a diary but, in the future, the diary data may be collected only for a subsample of the interview respondents.

According to this sampling plan, basic weights are calculated for each household as a function of the probability of selection. Interview nonresponse adjustments are applied to the basic weights as follows. Analysis is done on frame data, paradata and dwelling type in order to find a model, using variables correlated with expenditures, that explains the interview nonresponse. Nonresponse classes are then created using the score method (Alavi and Beaumont, 2004) to produce the interview nonresponse weight adjustments.

Figure 1 – Weighting steps



The nonresponse to the diary, combined with the possibility that in the future diary data may be collected only for a subsample of the interview respondents, have led to the production of two different sets of weights, one for interview respondents and one for diary respondents. For the diary weights, a second nonresponse adjustment was applied using an approach similar to that used for the interview, with the advantage that all the information provided by the interview respondents is available to build the diary nonresponse model. Both interview and diary weights are calibrated to certain household and population demographic estimates relevant to SHS; see Lessard (2005) for details. The weighting steps are summarized in Figure 1.

Before doing estimation, annualization factors are applied to address the variation in length among expenditures' reference periods and to produce annual expenditure aggregates. These

factors are calculated as the ratio of the length of a one year reference period over the length of the reference period used for data collection.

4. VARIANCE ESTIMATION

The Rao-Wu bootstrap for stratified multi-stage designs (Rao, Wu and Yue, 1992) is used to generate two sets of bootstrap weights, one set for the interview respondents and one set for the diary respondents. The process begins, for each stratum *h*, by selecting 1000 simple random subsamples (replicates) of n_h-*1* PSUs with replacement, one PSU fewer per stratum than in the original sample. Note that entire PSUs are selected; at each draw the surveyed households of a PSU are all in the replicate or none of them are in the replicate.

For a given replicate *b*, the initial bootstrap weight of household *k* in PSU *i* of stratum *h* is taken as

$$w_{k,b}^{init} = \left(\frac{n_h}{n_h - 1}\right) mult_{i,b} \cdot w_k \tag{4.1}$$

where $mult_{i,b}$ is the number of times PSU *i* was selected in replicate *b* ($mult_{i,b}$ ranges from 0 to n_h-*1*) and w_k is the original basic weight of the household. This initial bootstrap weight is calculated for all households in the sample, whether respondent or not, as we adjust for nonresponse in the next step.

For each replicate, an adjustment for nonresponse is made based on the respondent and nonrespondent households that appear in the replicate, using the nonresponse classes created during weighting. This nonresponse adjustment was not re-calculated in the original SHS; the adjustment obtained from the sample was simply applied to all replicates. This was done merely to simplify the calculations due to operational and informatics factors at the time of the original implementation of bootstrap variance estimation for SHS 2003. We therefore decided to take advantage of the redesign of the survey to improve the variance estimation with the addition of this step.

After the interview nonresponse adjustment, calibration is applied to the sample of interview respondents in each replicate, in the same way as during the weighting, to get the final interview bootstrap weights. For the diary, the starting point is the interview nonresponse-adjusted bootstrap weights; for each replicate a diary nonresponse adjustment is made and calibration in the same manner as during the weighting is applied.

Thus, there are 1,000 sets of bootstrap weights $w_{k,b,X}^{boot}$ for each interview respondent household *k*, and 1,000 sets of bootstrap weights $w_{k,b,Y}^{boot}$ for each diary respondent household *k*. For details on the theory behind the application of nonresponse adjustments and calibration to the bootstrap replicates, see Girard (2007).

Given an expenditure variable θ that is estimated using interview data only, in order to estimate the variance of $\hat{\theta}$, the set of interview bootstrap weights are first used to calculate the 1,000 bootstrap estimates $\hat{\theta}_{boot,b}$ of θ. If θ is a total, this would be $\hat{\theta}_{boot,b} = \sum_{k=1}^{n_{b,X}} w_{k,b,X}^{boot} \cdot \theta_k$ where θ_k is the value reported by household *k* and n_b is the number of households in replicate *b* (which varies since it is the PSUs that are resampled, and not all PSUs contain the same number of sampled households). If θ is a mean, the estimates would have the form

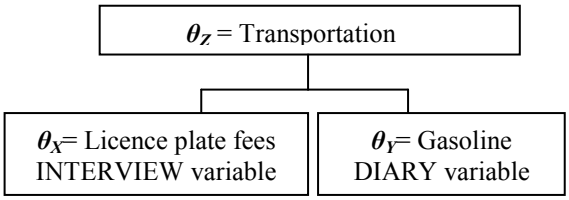
$$\hat{\theta}_{boot,b} = \frac{\sum_{k=1}^{n_{b,X}} w_{k,b,X}^{boot} \cdot \theta_k}{\sum_{k=1}^{n_{b,X}} w_{k,b,X}^{boot}}. \tag{4.2}$$

The variance estimate is then calculated as

$$\hat{\mathrm{Var}}_{boot}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_{boot,b} - \bar{\hat{\theta}}_{boot}\right)^2, \tag{4.3}$$

where B=1000 and $\hat{\bar{\theta}}$ is the mean of the bootstrap estimates. The procedure is analogous for a variable estimated from diary data only, but using the diary bootstrap weights in the above formulae.

Figure 3 – Example of the combination of diary and interview expenditure variables



Some categories of expenditures now have components from each of the two sources. For example, as shown in Figure 3, "Transportation" is a summary variable including, among others, both licence plate fees, a variable collected in the interview, and gasoline, a variable collected in the diary. Of course, we do not have the exact same set of respondents for these two components.

If θ_Z is a total, $\theta_Z = \theta_X + \theta_Y$ is obviously estimated as $\hat{\theta}_Z = \hat{\theta}_X + \hat{\theta}_Y$. However, in the case of a mean, the denominators of the two components may be different when estimating means over domains. One possibility is to use

$$\hat{\bar{\theta}}'_Z = \frac{\left(\sum_{k=1}^{n_X} w_{k,X} \cdot \theta_{k,X} + \sum_{k=1}^{n_Y} w_{k,Y} \cdot \theta_{k,Y}\right)}{\sum_{k=1}^{n_X} w_{k,X}}, \tag{4.4}$$

but following consultation with users of the survey products, $\hat{\bar{\theta}}_Z = \hat{\bar{\theta}}_X + \hat{\bar{\theta}}_Y$ was chosen as it was desirable to have consistency when adding up individual estimates from the two sources. In order to estimate the variances of these combined estimators, a covariance term was added to the bootstrap variance estimation process as follows:

$$\hat{\mathrm{Var}}_{boot}(\hat{\theta}_Z) = \hat{\mathrm{Var}}_{boot}(\hat{\theta}_X) + \hat{\mathrm{Var}}_{boot}(\hat{\theta}_Y) + 2\hat{\mathrm{Cov}}_{boot}(\hat{\theta}_X, \hat{\theta}_Y) \tag{4.5}$$

where $\hat{\mathrm{Cov}}_{boot}(\hat{\theta}_X, \hat{\theta}_Y) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_{X,boot,b} - \hat{\bar{\theta}}_{X,boot}\right)\left(\hat{\theta}_{Y,boot,b} - \hat{\bar{\theta}}_{Y,boot}\right)$, and similarly for estimators of the means.

Calculating this covariance separately rather than directly using the 1000 values of $\hat{\theta}_Z$ to estimate the combined variance was done because for the particular statistics under consideration (totals, means calculated in the above manner) it was no more or less complicated to do so, and it explicitly shows the covariance, which was of interest for analysis purposes. In the future, if combined variance estimates are required for other types of statistics such as ratios and percentiles, the bootstrap will be applied directly to these combined statistics.

The estimated coefficient of variation (CV) for each variable is then calculated as $CV(\hat{\theta}_Z) = \sqrt{\hat{\mathrm{Var}}_{boot}(\hat{\theta}_Z)} \Big/ \hat{\theta}_Z$.

5. RESULTS FROM THE 2009 SURVEYS

5.1 Interview, Diary and Combined Mean Estimates for SHS-Redesign 2009

Examples of mean estimates from SHS-redesign 2009 as well as their coefficients of variation are shown in Table 1. In addition to the global mean and CV, the portion of the mean and the individual CV for both the interview and diary components of the expenditure category are presented. As we can see, in general the interview CVs are lower, and the CV of a combined category tends to be in between the CVs of the two components most of the time. The many factors

Table 1 – Mean Estimates for Selected Expenditures – SHS-Redesign 2009

Expenditure	Interview Component		Diary Component		Total	
	Mean ($)	CV (%)	Mean ($)	CV (%)	Mean ($)	CV (%)
Clothing	2,342	3.2	110	8.2	2,452	3.2
Health Care	2,041	2.2	333	5.6	2,374	2.2
Household Furnishings, Art and Antiques	667	5.1	166	10.2	833	4.3
Food	70	8.3	7,239	1.3	7,309	1.3
Transportation	8,634	2.7	2,710	4.5	11,344	2.4

having an influence on the CVs of SHS-redesign 2009 will be discussed in a future document.

Since we are now recalculating the nonresponse adjustment during the bootstrap when previously we did not, variance estimates of some large expenditure categories of SHS-redesign 2009 were compared under the previous and the improved methods. Although a detailed discussion is beyond the scope of this paper, on average, there was increase in variance for the interview variables using the improved variance estimation method. When comparing variance estimates of means of nine large interview expenditure categories at the national and provincial levels, the majority of variances increased by less than 30%, and about half of those increased by less than 15%. The main exception was in Manitoba where most categories showed increases of over 30% and up to 77%. A handful of cases showed equal or slightly smaller variances using the new method, especially in Quebec. For the diary on the other hand, variances generally did not increase due to the addition of the nonresponse adjustment calculation. When comparing the variance estimates of means of twelve large diary expenditure categories at the national and provincial levels, the new method led to an increase of variance no more than 11% for all cases, and for the vast majority of cases the increase was no more than 2%. There were also instances where the variance estimate under the new method was in fact slightly lower, especially for the provinces of Prince Edward Island and Saskatchewan.

5.2 Comparison with SHS 2009

As mentioned, for the reference year 2009 a parallel run of the original SHS and the SHS-redesign was conducted in order to investigate the impact of the new collection model and estimation methodologies on the survey products. Collection for SHS 2009 was done from January to March of 2010 with a sample size of 16,000 households, whereas collection for SHS-redesign 2009 was done monthly during 2009 with a sample size of 9,000 households.

Table 2 shows a comparison of mean household expenditure estimates, CVs and ratios of variances, at the national level, for a selection of expenditure variables. The variances rather than the CVs are used for comparison purposes since the estimates from the two sources are not equal. The SHS-redesign variance estimates have been adjusted for the differences in overall sample size and provincial allocation between the two surveys, so that the adjusted national variance estimate is a weighted sum of the provincial variance estimates multiplied by the ratio of their respective provincial sample sizes in SHS-redesign and SHS. Note also that in this table, the SHS-redesign variances have been calculated using the original nonresponse adjustment factor from the weighting process, as is done for SHS, rather than recalculating the adjustment factor during the bootstrap process. Both of these modifications were made so that the SHS 2009 and SHS-redesign 2009 variances are more comparable.

Table 2 – Comparison of SHS 2009 and SHS-Redesign 2009 Mean Estimates for Selected Expenditures

Expenditure	Main source	SHS 2009		SHS-Redesign 2009		
		Mean ($)	CV (%)	Mean ($)	CV (%)	Variance ratio SHS-R / SHS
Expenses related to the use of vehicles	both	4,814	1.1	4,621	2.3	4.00
Food	diary	7,262	0.7	7,308	1.1	2.18
Health Care	interview	2,004	1.8	2,374	1.8	1.44
Eye care (sub-component of Health Care)	interview	234	3.2	205	3.1	0.76
Household Furnishings, Art, Antiques	interview	932	3.3	833	3.5	0.90
Major Appliances	interview	394	3.7	343	4.3	1.03

As the table shows, the SHS-redesign variances are sometimes greater than the SHS 2009 variances; however they are also sometimes smaller, depending on the expenditure in question. For expenditure variables that are mainly or completely estimated from the diary, due to the short diary reference period, we expected to see increased variance for expenditure types that are sporadic, but to see a lesser effect on frequent expenditures. For example, "Expenses related to the use of vehicles" had more than half of its constituent components collected in the diary. Some of its components from the diary are frequent, such as gasoline, while others are much less frequent but may be quite large values, such as repairs and maintenance. The variance estimate is four times as high as the corresponding SHS 2009 variance[3]. "Food" is a

[3] It is to be noted that the vehicle repairs and maintenance question is in fact to be moved from the diary to the interview and given a 12 month reference period, beginning with SHS-redesign 2012.

variable collected almost entirely in the diary and whose variance increased twofold while still maintaining a relatively low CV of 1.1, since food is an extremely regular expenditure. On the other hand some variables such as "Eye care", "Household furnishings, art and antiques" and "Major appliances" had their variances improve or stay about the same under the redesign. These variables were mainly collected in the interview. In fact, most of the 12 month recall interview variables and last payment interview variables lead to very similar estimates of means as well as variances when comparing with SHS 2009.

6. FUTURE WORK

As detailed in this paper, the variance estimation method for the SHS has undergone several adaptations due to the significant changes in collection methodology. Other developments will likely be necessary to respond to emerging issues or specific needs of survey data users. A few examples follow.

As mentioned, so far every household in the sample is asked to fill out a diary but at some point subsampling for the diary may become necessary for budget reasons. This would entail greater complexity in the calculation of the bootstrap weights for the diary, as we would need to consider this second phase of sampling.

In addition, the combination of two years of survey data may be done in the future in order to satisfy the needs of some users of the data as far as precision is concerned. In theory, the samples are independent from year-to-year, however in practice there are some dependencies. Due to operational reasons, since we select the sample from the LFS survey frame and must therefore follow the LFS's PSU-selection schedule fairly closely, we are constrained significantly in the choice of PSUs within strata. These restrictions result in about 30% overlap in the PSUs that are selected from one year to the next. We anticipate a potential effect of positive covariance that would need to be investigated and accounted for when combining multiple years of data.

Finally, there may be some requirements by data users to produce sub-annual estimates for certain selected expenditures having reference periods less than 12 months (so that sub-annual estimation makes sense). We now have quarterly samples where many efforts have been made to make them as representative of the population as possible; however, the degree to which this representativity holds needs further study. Quarterly estimates also raise variance related challenges since obviously the sample size would be only one quarter of the yearly sample size. More study is needed to determine whether quarterly or semi-annual estimates can be sufficiently precise to maintain the publishability of the results.

ACKNOWLEDGEMENTS

The authors would like to thank Claude Girard, Denis Malo, and Johanne Tremblay for their very helpful comments.

REFERENCES

Alavi, A. and Beaumont, J.-F. (2004). Nonresponse Adjustment Plans for the Labour Force Survey. *Paper presented to the Statistics Canada Advisory Committee on Statistical Methods, Meeting no. 38*.

Girard, C. (2007). How to Avoid Getting All Tied Up Bootstrapping a Survey: A Walk-Through Featuring the National Longitudinal Survey of Children and Youth. *Statistics Canada Working Paper*.

Lessard, S. (2005). Revision of the Calibration Strategy of the Survey of Household Spending. *Statistics Canada Internal Document.*

Nadeau, C., Tremblay, J. and Lessard, S. (2007). Sampling and Estimation Issues in the Redesign of the Canadian Survey of Household Spending. *Joint Statistical Meetings 2007*.

Rao, J.N.K., Wu, C.F.J., and Yue, K. (1992). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology*, **18**, 209-217.

Tremblay, J., Lynch, J. and Dubreuil, G. (2010). Pilot Survey Results from the Canadian Survey of Household Spending Redesign. *Joint Statistical Meetings 2010*.