**Research Paper**

2000 Survey of Household Spending Data Quality Indicators

by Jean-Claude Boies, Sylvie DeBlois, José Gaudet,
Bruno Lapierre, Christian Nadeau

Household Survey Methods Division
16 Jean Talon Building, Ottawa, Ottawa, K1A 0T6

Telephone: 613 951-7355

This paper represents the views of the authors and does not necessarily reflect the opinions of Statistics Canada.
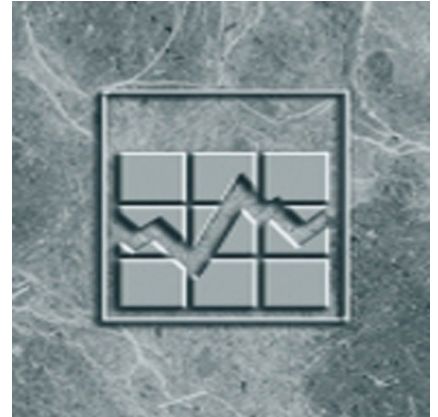
Statistics Canada   Statistique Canada

Canadä

Data in many forms

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered. Data are available on the Internet, compact disc, diskette, computer printouts, microfiche and microfilm, and magnetic tape. Maps and other geographic reference materials are available for some types of data. Direct online access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable database and retrieval system.

How to obtain more information

Inquiries about this product and related statistics or services should be directed to: Client Services, Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca) or to the Statistics Canada Regional Reference Centre in:

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

You can also visit our World Wide Web site: http://www.statcan.ca

Toll-free access is provided **for all users who reside outside the local dialing area** of any of the Regional Reference Centres.

National enquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Order-only line (Canada and United States)	**1 800 267-6677**

Ordering/Subscription information

All prices exclude sales tax

Catalogue no.62F0026MIE2002001, is available on internet for free. Users can obtain single issues at: http://www.statcan.ca/cgi-bin/downpub/research.cgi.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact your nearest Statistics Canada Regional Reference Centre.

2000 Survey of Household Spending
Data Quality Indicators

Note of appreciation

Abstract

This report describes the quality indicators produced for the 2000 Survey of Household Spending. It covers the usual quality indicators that help users interpret the data, such as coefficients of variation, nonresponse rates, slippage rates and imputation rates.

TABLE OF CONTENTS

TABLE OF CONTENTS (concluded)

HIGHLIGHTS

Sampling Errors

➢ The coefficient of variation (CV) of the estimate of average total expenditures per household varies between 1.2% and 1.6% depending on the province and is 0.8% at the national level.

➢ The coefficient of variation (CV) of the estimate of the average for the different categories of summary-level expenditures is less than or equal to 1.8% at the national level and generally below 5% at the provincial level. The results for the different categories of dwelling characteristics and household equipment are similar.

Nonresponse

➢ The nonresponse rate is 29.8%. Non-response is due to refusals (18.5%), households that could not be contacted (7.5%) and households for which the data were considered unusable (3.8%).

➢ The final nonresponse rate tends to increase with the level of urbanization. A nonresponse rate of 23.6% is observed in rural areas, and a rate of 37.6% in urban centres with one million residents or more. This tendency is also observed in collection nonresponse rates.

➢ According to an analysis of final response rates in strata consisting of high- and low-income geographic areas based on the sample design, the nonresponse rate in high-income strata (44.2%) is higher than those observed in regular strata (29.4%) and low-income strata (28.9%).

Coverage Errors

➢ For households, an undercoverage rate of 8.1% is observed at the national level. Undercoverage of households is also observed for all provinces, with rates varying from 5.2% to 13.2%.

➢ For persons, an undercoverage rate of 9.9% is observed at the national level. Undercoverage of persons is also observed for all provinces, with rates varying from 6.6% to 15.4%.

➢ The national slippage rates for children (0 to 6 years and 7 to 17 years) are quite different from those for other age groups. For children, overcoverage or a slight undercoverage is observed, while for adults there is always undercoverage. The undercoverage rate for children of all ages combined is 1.8%, whereas it is 12.2% for adults.

Response Errors

➢ Response errors include recall errors, telescopic error and errors due to proxy response. Because the SHS interview is lengthy, the response burden can lead to respondent fatigue and can affect data quality. Total interview time varies according to household characteristics; for some households the interview can take more than five hours.

Processing Errors Related to Imputation

i) Expenditure Variables

➢ For 8.9% of households, it was necessary to impute at least one expenditure variable (excluding clothing expenditures). For the majority of them, only one or two of the 237 expenditure variables were imputed.

➢ About 17% of individuals required imputation of clothing variables, but for the vast majority of these, the respondent supplied the total and only the components were imputed.

ii) Income Variables

➢ Less than 3% of individuals required imputation of at least one income variable. For just over 70% of these, the respondent provided total income but all components (income sources) were imputed.

INTRODUCTION

The Survey of Household Spending (SHS) is an annual survey that collects data on household income and expenditures using personal interviews. The 2000 SHS sample is made up of 24,180 dwellings[1] distributed throughout the ten provinces.[2] Collection takes place in January, February and March, and income and spending figures are obtained for the period from January 1 to December 31 of the previous year. Following a redesign that took place in 1997, this survey replaces the periodic Family Expenditure Survey and the Household Facilities and Equipment Survey (with adjustments to questionnaires and samples).

Like all surveys, the SHS is subject to errors, despite all the precautions taken at the different stages of the survey to control them. While there is no comprehensive measure of the quality of the data generated by a survey, some quality measures produced at the different stages of the survey can provide users with the information needed in order to interpret the data properly.

This report therefore seeks to describe the quality indicators produced for the 2000 Survey of Household Spending. It covers the usual quality indicators that are generally useful to users in interpreting data, such as coefficients of variation, nonresponse rates, slippage rates and imputation rates.

Quality indicators have been classified according to the main types of error encountered in a survey. Section 1 deals with sampling errors—that is, errors due to the fact that the inferences about the population drawn from the survey are based on information collected from a sample of the population, rather than the entire population. The sections that follow cover errors not due to sampling. Nonresponse and coverage errors are discussed first, in sections 2 and 3. Response errors and processing errors are dealt with in sections 4 and 5 respectively.

This report focuses on the quality of the data. For a detailed description of the survey methodology, see reference [1].

---

[1] Of these 24,180 dwellings, those that were ineligible had to be identified and excluded (see Section 2.1), leaving 20,989 households from which household income and expenditure data were collected.
[2] The territories were not covered by the 2000 SHS, but they will be in 2001. They will then be surveyed every second year.

---

1.   SAMPLING ERRORS

Sampling errors result from the fact that inferences drawn from the survey about the population are based on information collected from a sample, rather than from the entire population. In addition to the sample design and the estimation method used in the Survey of Household Spending, the sample size and the variability of each characteristic are factors that affect sampling error. Characteristics that are rare or are distributed very unevenly in the population will have greater sampling error than characteristics that are observed more frequently or are more homogeneous in the population.

1.1   Measures of Sampling Error

The standard error is a commonly used measure of sampling error. The standard error is the degree of variation of the estimate considering that a particular sample was selected instead of another among all the possible samples of the same size under the same sample design. Since the SHS uses a complex sample design and estimation method, the standard error is estimated using a resampling method known as the jackknife technique. For more details on this method, see reference [2].

The coefficient of variation (CV) is also a frequently used measure of the reliability of the estimate. It merely expresses the standard error as a percentage of the estimate. Thus, if an estimate Y is obtained for a certain characteristic and SE is the estimated standard error, then the CV will be (SE/Y) x 100.

Finally, either the standard error or the coefficient of variation may be used to derive another measure of the accuracy of estimates, namely the confidence interval. This measure indicates the level of confidence with which it can be stated that for a characteristic observed, the real value in the population lies within specified limits. An interval with a confidence level of 95% corresponds to the estimate obtained from the sample $\pm$ 2 standard errors (Y $\pm$ 2 SE).[3] This means that if the sampling were repeated a large number of times, each sample would provide a different interval and 95% of the intervals would contain the true value of the characteristic. Similarly, if the sampling were repeated, the interval Y $\pm$ SE would contain the true value in 68% of cases.

1.2   Coefficients of Variation

Estimates of coefficients of variation are calculated for estimates of many characteristics collected in the SHS. CVs for the estimates of average household expenditure (for detailed expenditure categories), as well for dwelling characteristics and household equipment, are available at the national and provincial levels in the *User Guide – Survey of Household Spending* (see reference [3]).

---

[3] The confidence interval is calculated directly from the CV in similar fashion, namely Y $\pm$ 2 (CV x Y)/100.

It should be noted that the estimated CVs do not take account of the fact that some of the data were imputed; thus the CVs shown may underestimate the true values. For most variables, the imputation rate is low (see Section 5) and the CVs provided are good estimates of the true CVs. However, it is important to consider both the CV and the imputation rate if one is interested in the reliability of detailed expenditures that have a high imputation rate.

Table 1.1 gives an overview of the CVs of the estimates of household averages at the provincial level as well as at the national level for the estimation of selected summary-level expenditure categories and income.

Table 1.1
Coefficients of Variation (%) by Province and at the National Level for the Estimation of Average Household Expenditures for Selected Summary-level Expenditure Categories and Average Income

Summary-level expenditure category	Can.	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta	B.C.
Total expenditure	0.8	1.4	1.6	1.5	1.2	1.6	1.6	1.3	1.6	1.2	1.6
Consumer expenditures	0.6	1.3	1.6	1.5	1.2	1.2	1.1	1.2	1.5	1.2	1.4
Food	0.6	1.1	1.6	1.1	1.2	1.1	1.2	1.2	1.9	1.2	1.1
Shelter	0.8	1.6	2.2	1.8	1.7	1.7	1.5	1.7	1.9	1.5	2.2
Household operation	0.9	1.7	2.1	2.0	2.1	2.1	1.8	1.7	1.9	2.0	1.9
Furnishings	1.8	3.7	5.3	4.8	4.5	3.6	3.4	3.6	3.8	3.6	3.5
Clothing	1.1	2.8	3.6	2.9	2.4	2.1	2.2	2.6	3.6	2.4	2.1
Transportation	1.5	3.2	4.7	3.6	3.1	3.1	2.9	3.6	3.7	3.0	3.2
Health care	1.3	3.1	3.4	3.1	2.9	2.4	3.2	2.5	4.6	2.4	2.6
Personal care	1.0	2.0	3.2	2.2	2.1	2.0	2.0	2.6	2.6	2.1	2.1
Recreation	1.4	3.1	3.9	5.6	4.0	2.9	2.7	2.7	3.7	3.7	3.5
Reading & printed materials	1.5	3.8	3.9	3.7	3.6	3.3	2.9	3.5	3.7	3.1	3.0
Education	3.3	6.4	10.8	11.0	10.4	6.0	6.2	7.2	7.5	7.5	6.9
Tobacco, alcoholic beverages	1.7	3.7	5.0	4.8	4.2	3.4	3.6	4.1	4.7	3.8	3.8
Games of chance (net)	4.9	8.2	14.0	7.8	7.9	11.1	9.2	6.6	11.5	11.4	5.5
Miscellaneous expenditures	2.6	7.0	11.0	6.9	11.4	5.2	5.2	6.5	5.4	5.8	6.9
Personal income tax	1.8	2.9	4.1	3.1	2.3	2.9	3.6	3.2	2.9	2.3	3.7
Personal insurance	6.1	2.5	2.9	2.3	2.3	1.6	13.7	3.0	2.6	6.1	2.8
Gifts and contributions	5.3	5.1	12.7	12.2	10.8	17.4	9.3	5.8	7.5	9.6	10.9
Income	0.9	1.0	1.6	1.2	1.0	1.1	1.9	1.1	0.9	0.8	1.0

The coefficient of variation of the estimate of average total expenditures per household varies between 1.2% and 1.6% depending on the province and is 0.8% at the national level.

For summary-level expenditure categories, the CVs at the national level are less than or equal to 1.8%, except for the following categories: *education*, *games of chance*, *miscellaneous expenditures, personal insurance payments and pension contribution,* and *gifts of money and contributions*. These expenditure categories represent respectively 1.5%, 0.5%, 1.5%, 5.6% and 2.3 % of total expenditure (data not shown). Also, with the exception of these categories, the CVs are generally less than or equal to 5% at the provincial level.

Table 1.2 gives an overview of the CVs of estimates for selected categories of dwelling characteristics and household equipment at the provincial level and the national level.

Table 1.2
Coefficients of Variation (%) by Province and at the National Level for Selected Categories of Dwelling Characteristics and Household Equipment

CATEGORY	Can.	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta	B.C.
Owner	1.0	1.8	2.9	2.1	1.6	2.2	2.1	1.8	2.0	1.8	1.8
Renter	1.8	5.3	6.8	4.9	4.6	2.8	4.0	4.4	4.9	4.6	3.3
Automatic washing machine	0.7	1.2	2.5	1.6	1.2	1.0	1.7	1.5	1.3	1.1	1.4
Clothes dryer	0.7	1.2	2.5	1.9	1.2	1.1	1.8	1.6	1.2	1.2	1.4
Built-in dishwasher	1.4	4.3	6.2	4.4	3.7	2.5	3.3	3.3	4.0	2.5	2.4
Freezer	1.0	1.3	2.6	2.2	1.8	2.5	2.2	1.6	1.6	1.7	2.1
Microwave oven	0.4	1.0	1.3	0.9	0.8	0.9	0.8	0.9	0.9	0.6	0.8
Cellular phone	1.4	4.5	5.7	4.4	4.2	4.0	2.6	3.3	3.5	2.5	2.7
CD player	0.6	1.7	2.1	1.8	1.7	1.4	1.3	1.8	1.8	1.2	1.3
Cable TV	1.0	1.6	4.3	2.2	2.5	2.1	2.1	1.8	2.4	2.0	1.3
Personal computer	1.0	3.0	4.3	2.7	3.3	2.4	1.9	2.4	2.6	2.0	1.7
Modem	1.2	3.7	5.0	3.2	3.7	2.9	2.2	2.9	3.2	2.5	2.1
Use of internet (home)	1.3	4.0	5.5	3.4	4.2	3.3	2.4	3.6	3.9	2.7	2.3
Owned vehicles (one)	1.4	3.3	4.5	3.4	3.6	3.0	2.7	3.2	3.1	3.2	2.7
Owned vehicles (2 or more)	1.5	4.6	3.3	3.5	3.1	3.8	3.2	3.4	2.9	2.5	2.7

The coefficients of variation for dwelling characteristics and household equipment are generally below 5% for the Atlantic provinces, and generally below 4% for the other provinces. The *renter* category is an exception, with several provincial CVs exceeding these thresholds.

The coefficients of variation for dwelling characteristics and household equipment at the national level are less than or equal to 1.8%.

1.3 Model for Deriving an Approximation of the CV

Estimates for different domains of interest (e.g., by income quintile) for summary-level expenditure categories are available in *Spending Patterns in Canada* (reference [4]). Estimates for different domains of interest for detailed expenditure categories are also available upon request from the Income Statistics Division. (For more details on tables available upon request from the Income Statistics Division, see reference [3] or [4]). For operational reasons, it is not possible to produce CVs for all the different levels of aggregation that might interest users.

1.3.1 Approximation of the CV for Estimates of Domains

However, it is possible to calculate an approximation of the CV by using a relationship between the number of households in the sample that reported expenditures for a given category and the CV at an aggregated level. This relationship, based on the tendency of the CV to increase in proportion to a decrease in the square root of the number of households reporting an expenditure, is illustrated below.

Formula for Approximating the CV for a Domain (Subgroup of the Population)

If CV(Y) is the CV for the estimate of the average per household of a given characteristic for the entire population, then an approximation of the CV of the estimate of that characteristic can be calculated for a domain (which may be considered as a subgroup of the population, such as a household type, an income quintile, an urbanization level, etc.) according to the following equation:

$$CV(Y_d) = CV(Y) \times \sqrt{\frac{nP}{n_d P_d}}$$

where

n: *number of households in the sample*
P: *estimate of the proportion of households reporting a value > 0 for this characteristic in the population*
n_d: *number of households in the sample in domain d*
P_d: *estimate of the proportion of households reporting a value > 0 for this characteristic in domain d*

Generally, approximations for the different domains are calculated using the CV, size *n* and proportion *P* at the national level. If an approximation of a CV is desired for a domain that is entirely contained within a single province (e.g., a metropolitan area), then it is preferable to use the values at the provincial level, since provincial CVs are published for the 2000 SHS (reference [3]). It should be kept in mind that the value obtained using this approach is only an approximation of the CV.

1.3.2 Approximation of the CV Using the Microdata File

Users of microdata files can obtain an approximation of the CV of the estimates using another approach that will generally provide better results than the one described in the previous section for CVs of detailed expenditure categories. This approach is described in detail in the documentation provided with the 2000 microdata file. It can be used only with the microdata file, since it is necessary to have data and weights for each household to calculate this approximation.

The document on data quality for the 1997 SHS contains the results of an evaluation of the performance of these two CV approximation methods.

1.4 Suppression of Unreliable Data in Estimation Tables

Since the coefficient of variation is an indicator of the reliability of the data, we would like to use it to determine whether or not the estimates should be published. Estimates in which the CV is estimated at more than 33% are not considered sufficiently reliable to be published. However, CV estimates are not calculated for many of the published estimates. The suppression rule for expenditure estimates is therefore based on the number of households reporting a value greater than 0.[4]

It can be shown that CVs are usually below 33% if the number of households reporting an expenditure is greater than 30. Since this is an approximate rule, some estimates may be published even if the CV is greater than 33%, and some estimates will not be published even if the CV is less than 33%. The document on data quality for the 1997 SHS gives the results of the evaluation of risks of error in the use of the suppression rule.

2. NONRESPONSE

Errors due to nonresponse result from the fact that some potential respondents do not provide the necessary information, or the information provided proves to be unusable. When the respondent has failed to respond to only some questions, this is referred to as partial nonresponse. In such a case, the missing data are imputed. Errors associated with imputation are described in Section 5, which deals with processing errors. In the present section, nonresponse includes collection nonresponse—which is mainly due to the inability to contact the household or to members of the household refusing to participate in the survey either partially or completely—as well as households for which data were collected but are unusable.

The main impact of nonresponse on data quality is that it can introduce a bias in the estimates if the characteristics of respondents and nonrespondents differ and this difference has an impact on the characteristics studied. Nonresponse rates are easy to calculate, but they have only an indicative value with regard to data quality, since they do not allow estimation of the bias associated with the estimates. The scope of nonresponse may be considered as an indicator of the risks of bias in the estimates.

2.1 Nonresponse Rates and Vacancy Rates

In the SHS, since the units selected are dwellings, interviewers must first identify ineligible dwellings. These are dwellings occupied by persons who are not part of the target population, dwellings that no longer exist (e.g., demolished dwelling, relocated mobile home or dwelling converted to business use) and vacant dwellings (unoccupied, seasonal or under construction).

---

[4] In practice, we use an estimate of the proportion of households reporting an expenditure, which is multiplied by the sample size.

Among eligible dwellings, we next evaluate the proportion of households that did not respond to the survey, which is called the collection nonresponse rate. These include households that refused to participate in the survey and households where the respondents could not be contacted, either because they were away or because of special circumstances (language problem, illness, death).

Again, among eligible dwellings, the rate of unusable data is determined. Unusable data refers to the number of households for which the questionnaire was at least partially completed but was rejected during the processing of the data. There are two main causes of rejection. First, when many of the questions on income or expenditures have been left unanswered, the questionnaire is classified as incomplete and is not used. The other source of rejection consists of questionnaires in which the difference between receipts (income and other sources of money received by the household) and disbursements (expenditures and net change in assets and liabilities) is greater than 20%. These questionnaires are also excluded from the estimation and are considered as nonresponse.

Table 2.1 shows the final nonresponse rate, the collection nonresponse rate (broken down by refusals and no-contacts), and the unusable data rate (broken down into incomplete and out-of-balance data). The vacancy rate is also included. These rates are provided at the national level as well as at the provincial level.

Table 2.1
Nonresponse Rates (%) and Vacancy Rates (%) by Province and at the National Level

Province	Vacancy rate	Collection nonresponse rate			Unusable data rate			Final nonresponse rate (at estimation stage)
		TOTAL	No contact	Refusal	TOTAL	Incom-plete	Out-of - balance	
Canada	11.0	26.0	7.5	18.5	3.8	2.2	1.6	29.8
Newfoundland	19.1	21.0	8.8	12.2	2.0	0.8	1.2	22.9
P.E.I.	20.0	14.9	4.3	10.6	0.5	0.1	0.4	15.4
N.S.	14.1	27.6	9.3	18.4	6.6	3.4	3.2	34.3
N.B.	13.6	23.9	5.7	18.2	3.7	2.1	1.7	27.7
Quebec	7.0	25.9	6.2	19.7	1.1	0.7	0.4	26.9
Ontario	7.0	34.3	10.1	24.2	5.4	2.1	3.3	39.7
Manitoba	12.3	18.8	4.4	14.4	3.2	2.0	1.1	22.0
Saskatchewan	10.5	23.7	7.6	16.1	3.5	1.9	1.6	27.2
Alberta	7.0	29.3	6.8	22.5	2.5	1.3	1.2	31.8
B.C.	8.9	26.9	8.4	18.5	6.3	5.5	0.8	33.2

The final nonresponse rate at the national level is 29.8%. It is due to refusals (18.5%), to households that could not be contacted (7.5%), and finally to households for which the data were unusable (3.8%). In all provinces, the main cause of nonresponse is refusals, followed by no contact and then unusable data.

The final nonresponse rate varies from one province to another. The rate is lower than 20% in Prince Edward Island, while rates in excess of 30% are observed in Nova Scotia, Ontario, Alberta and British Columbia. Ontario's nonresponse rate is particularly high at almost 40%. The highest rates of no contact (10.1%) and refusal (24.2%) are also observed in Ontario.

The vacancy rate is shown in Table 2.1, but it should be kept in mind that vacant dwellings do not contribute to the bias of the sample if they are correctly identified. By analysing vacancy rates, we can detect dwelling identification problems associated with the collection process. The national vacancy rate for the 2000 SHS is 11.0%, which is slightly lower than for the Labour Force Survey (LFS) for the same period.

2.2 Nonresponse According to Urbanization Level

Nonresponse varies according to urbanization level. The various rates at the national level are shown by urbanization level in Table 2.2.[5]

Table 2.2
Nonresponse Rates (%) and Vacancy Rates (%) by Urbanization Level

Urbanization category	Vacancy rate	Collection nonresponse rate			Unusable data rate			Total nonresponse rate (at estimation stage)
		TOTAL	No contact	Refusal	TOTAL	Incomplete	Out-of-balance	
URBAN								
1,000,000 or more	5.1	32.8	9.9	22.9	4.8	3.0	1.9	37.6
500,000 to 999,999	4.7	28.2	6.9	21.3	2.6	1.6	1.0	30.8
250,000 to 499,999	6.3	29.4	9.7	19.7	6.5	3.1	3.4	35.9
100,000 to 249,999	7.4	29.8	8.7	21.1	2.7	1.5	1.1	32.5
30,000 to 99,999	4.8	25.1	7.0	18.1	2.8	2.1	0.8	27.9
Less than 30,000	7.9	20.2	6.2	14.0	3.8	2.3	1.5	24.0
RURAL	23.4	19.7	5.5	14.2	3.8	2.1	1.7	23.6
TOTAL	11.0	26.0	7.5	18.5	3.8	2.2	1.6	29.8

The final nonresponse rate generally increases with the urbanization level. According to the results in Table 2.2, only the 500,000 to 999,999 group does not follow this pattern.

The collection nonresponse rate also tends to increase with the urbanization level. There is a sizable 12.6% difference between the urbanization categories "less than 30,000" and "1,000,000 or more." Refusals account for more than 50% of the total nonresponse at every level of urbanization.

An examination of the vacancy rate by urbanization level reveals that the vacancy rate in rural areas (23.4%) is almost three times higher than in urban

---

[5] Tables on nonresponse rates according to urbanization level by province are available on request from the Household Survey Methods Division.

areas with small populations (7.9%). The latter areas also exhibit a higher rate than high-population urban areas. This phenomenon is also observed in the LFS and may be explained by a greater number of seasonal dwellings in rural areas. Since the SHS sample is more concentrated in high-population urban areas than the LFS is, the national vacancy rate for the SHS can be expected to be slightly lower.

2.3   Nonresponse According to Income Strata

Since income information is not available for nonrespondents, it is not possible to compare nonresponse rates according to income. However, the LFS sample design, used for the SHS, was developed in such a way that in seven large cities there are strata consisting of geographic areas where the average household income exceeds $100,000 as well as strata consisting of apartments inhabited by households with an average income of less than $20,000. Even though such strata are few in number and they account for only a small number of dwellings in the SHS sample (approximately 560 and 160 for high-income and low-income strata respectively, or less than 3% of the sample), a comparison of nonresponse rates in these two groups in relation to the other strata is revealing. Table 2.3 shows these results.

Table 2.3
Comparison of Nonresponse and Vacancy Rates (%) in High-income and Low-income Strata in Relation to Other Strata

Stratum type based on income	Vacancy rate	Collection nonresponse rate			Unusable data rate			Final nonresponse rate (at estimation stage)
		TOTAL	No contact	Refusal	TOTAL	Incom-plete	Out-of-balance	
High-income	3.9	38.5	8.3	30.3	5.6	2.4	3.2	44.2
Regular	11.2	25.7	7.5	18.2	3.8	2.2	1.6	29.4
Low-income	4.8	24.8	12.1	12.8	4.0	3.4	0.7	28.9
TOTAL	11.0	26.0	7.5	18.5	3.8	2.2	1.6	29.8

The final nonresponse rate in high-income strata is much higher than in regular strata, at 44.2%. The refusal rate for high-income strata is just over 30%, a rate much higher than that for the other types of strata.

Households in low-income strata have a final nonresponse rate and a collection nonresponse rate similar to those for the regular strata. However, the causes of nonresponse are different for the two types of strata. For the regular strata, collection nonresponse is mainly due to refusals, while for the low-income strata it is equally attributable to refusals and no contact.

As in the 1997 to 1999 surveys, the vacancy rate is higher for regular strata than for each of the other two types of strata. In high-income and low-income strata, the vacancy rates are similar.

2.4 Adjustment for Nonresponse

To compensate for nonresponse, the weights in the SHS are inflated by the inverse of the weighted response rate within certain groups defined on the basis of the different urbanization levels in each province. The weighted rates differ from the rates presented in this section, since they take each household's sampling weight into account. An algebraic description of the adjustment for nonresponse is provided in Appendix A.

The adjustment of weights for nonresponse takes account of the differences in nonresponse by urbanization level as described in Section 2.2. It will reduce the bias to the extent that the characteristics of respondents and nonrespondents are similar for a given urbanization level.

3. COVERAGE ERRORS

In the design of the survey, the target population was defined. It is useful to go over the definition of the target population for the SHS, since a good understanding of the target population is necessary in order to properly interpret the survey data.

Target population

> The target population consists of individuals living in private households. It therefore excludes residents of institutions such as prisons, chronic care hospitals or senior citizens' homes, as well as members of religious orders and other groups living communally, members of the Armed Forces living in military compounds, and individuals residing permanently in hotels or rooming houses. Also excluded are foreign countries' official representatives residing in Canada and their families as well as individuals residing on Indian reserves or public lands. With these exclusions, the survey covers nearly 98% of the population in the ten provinces. Territories are excluded from the target population for the 2000 SHS, as the survey covers this region only every other year.

Coverage errors result from inadequate representation of the target population based on the units of the sampling frame. Some units of the target population may be omitted from the sampling frame, in which case there is undercoverage. Other units that are not in the target population may be included by error, or some units may be included more than once; these units are responsible for overcoverage.

3.1 Undercoverage and Overcoverage: Slippage Rates

In the SHS, the sample is selected using a list of dwellings in each selected cluster. Factors contributing to undercoverage are the omission of dwellings in

the creation of the list, new dwellings that are added between the creation of the list and the interviewer's visit (mainly in developing areas) as well as the misclassification of vacant dwellings. The inclusion of dwellings that are not within the boundaries of the cluster is a source of overcoverage. Similarly, errors can take place during data collection, when the persons who are members of selected households are identified. These errors contribute to both undercoverage and overcoverage.

A good representation of the target population is essential to the production of realistic expenditure estimates. The number of persons per household is also an important characteristic in estimating households' average expenditures. Therefore, it is necessary for the sample to adequately represent not only the individuals in the target population but also the distribution of households according to their size.

In 1999, a new weighting method was introduced, utilising new control groups. With this new method, the representation of the target population can be corrected more effectively by using better-defined age-sex groups than those used previously, for which the coverage varies from one group to the other.

There is generally net undercoverage of the number of persons and the number of households in the SHS. This is corrected by adjusting the weights using auxiliary data, which are based on post-censal demographic estimates. The slippage rate (see Appendix A) is a measure of the percentage difference between these auxiliary data and the survey estimates calculated using weights not adjusted with these data.[6] Tables 3.1 and 3.2 show slippage rates by age-sex group at the national level and at the provincial level respectively, while Table 3.3 presents these rates for the household size categories used for the weight adjustment. A positive rate indicates overcoverage of the number of persons or households in the survey.

Table 3.1
National Slippage Rates by Age-Sex Group

National Slippage Rates by Age-Sex Group				
	Age	Sex		Total
		Male	Female	
Canada	0-6 years	2.7	-2.3	0.3
	7-17 years	-5.0	-0.9	-3.0
	18-24 years	-11.6	-15.5	-13.5
	25-34 years	-16.7	-12.7	-14.7
	35-54 years	-15.5	-11.3	-13.4
	55-59 years	-7.8	-10.0	-8.9
	60-64 years	-10.2	-7.6	-8.8
	65-69 years	-13.7	-11.9	-12.8
	70 years and +	-7.8	-4.7	-6.0
	Total	**-10.9**	**-8.8**	**-9.9**

---

[6] These are subweights, that is, survey weights adjusted for nonresponse (see Appendix A)

For the 2000 SHS, the rate of undercoverage of the population was 9.9%. If Table 3.1 is analysed with respect to age groups, it may be seen that national slippage rates for children (aged 0 to 6 and 7 to 17) are quite different from those for other age groups. Overcoverage or a slight undercoverage is observed with respect to children, while for adults there is always undercoverage. The undercoverage rate for children of all ages combined is 1.8%, while it is 12.2% for adults (data not shown). The highest rates at the national level are observed among males aged 25 to 34 and 35 to 54 and among females aged 18 to 24.

The SHS uses the LFS sampling frame. Over the same period, the LFS slippage rate at the national level was 9.7% (Reference [5]), which is slightly lower than the 12.2% SHS rate for persons aged 15 years and over (data not shown).

Table 3.2
Slippage Rates for Provinces by Age-Sex Group

Sex	Age	Newfoundland	Prince Edward Island	Nova Scotia	New Brunswick	Quebec	Ontario	Manitoba	Saskatchewan	Alberta	British Columbia
						Slippage Rates by Age-Sex Group					
Male	0-6	-8.0	-4.9	1.1	6.4	6.9	8.0	-8.5	14.3	-1.4	-16.5
	7-17	-16.2	-11.5	9.6	3.1	-7.0	-3.2	-12.7	8.9	-0.8	-15.4
	18-24	-22.2	-9.5	-28.8	-25.9	-3.7	-8.3	-2.7	-32.3	-15.0	-21.3
	25-34	-27.5	-32.0	-18.2	-16.0	-9.8	-21.3	-14.7	-9.2	-11.3	-19.5
	35-54	-12.7	-12.7	-15.4	-13.3	-14.3	-16.3	-10.9	-17.7	-16.1	-17.0
	55-59	3.8	5.3	-10.6	-25.5	2.2	-9.1	2.2	-21.5	-11.1	-19.9
	60-64	-26.7	-14.5	-8.9	-21.2	-6.2	-10.2	-15.8	-16.6	-18.3	-5.2
	65-69	-17.3	6.0	2.4	-16.9	-14.6	-10.3	-16.9	-3.3	-22.2	-21.4
	70 +	-3.9	4.5	-12.0	-0.1	-0.8	-14.7	6.0	-5.2	-6.2	-7.2
	Total	**-15.2**	**-11.2**	**-10.4**	**-11.1**	**-7.7**	**-11.2**	**-9.1**	**-9.2**	**-10.9**	**-16.6**
Female	0-6	-0.4	-15.0	12.3	17.8	-2.9	-10.1	0.8	-0.3	16.2	-0.8
	7-17	6.2	-11.2	-6.6	-0.3	6.4	-1.3	-14.7	-2.9	-4.1	-4.8
	18-24	-7.4	-19.4	-20.5	-25.8	-6.1	-17.7	-9.0	-11.6	-10.3	-31.4
	25-34	-16.8	-21.1	-21.6	-5.0	-4.7	-17.3	-14.1	-4.5	-0.3	-21.6
	35-54	-7.0	-11.4	-4.3	-9.4	-7.6	-12.0	-9.5	-9.0	-14.2	-17.4
	55-59	-21.1	2.9	-13.0	-23.2	-18.1	-1.9	-12.0	-1.1	-12.6	-12.0
	60-64	-16.9	9.9	-19.4	1.9	-2.7	-5.4	-14.0	-18.8	-9.2	-16.5
	65-69	1.9	16.0	-2.1	0.7	-15.1	-12.9	-1.1	-9.9	-15.0	-11.2
	70 +	-7.3	-15.6	-6.3	-5.9	-7.5	-7.6	2.5	1.6	10.2	-2.5
	Total	**-7.0**	**-11.4**	**-8.5**	**-6.5**	**-5.5**	**-10.4**	**-8.6**	**-5.8**	**-5.4**	**-14.3**
Total		**-11.1**	**-11.3**	**-9.4**	**-8.8**	**-6.6**	**-10.8**	**-8.9**	**-7.5**	**-8.1**	**-15.4**

We observe net undercoverage for all provinces, with the rates varying from 6.6% to 15.4%. However, a low overall rate of undercoverage does not necessarily mean better coverage. For example, Saskatchewan's overall slippage rate (-7.5%) conceals the worst case of undercoverage for a provincial

age-sex group (32.3% for males aged 18-24) and the fourth worst case of overcoverage (14.3% for males aged 0-6). It is also worth noting that the pattern of variation in slippage rates by age-sex group differs substantially from one province to the next.

Table 3.3
Slippage Rates for Provinces by Household Size

Province	Slippage Rate			
	Households	One-person households	Two-person households	Households of three persons or more
Canada	**-8.1**	**-7.4**	**-4.7**	**-11.2**
Newfoundland	-8.3	4.5	-4.6	-14.8
Prince Edward Island	-9.0	-19.6	10.2	-16.7
Nova Scotia	-8.5	-8.4	-3.8	-12.4
New Brunswick	-6.6	-4.8	-0.5	-12.2
Quebec	-5.4	-8.8	3.6	-10.3
Ontario	-9.2	-6.5	-9.4	-10.6
Manitoba	-5.2	4.5	-5.2	-11.9
Saskatchewan	-5.7	2.9	-11.0	-7.3
Alberta	-6.2	-6.2	-3.4	-8.4
British Columbia	-13.2	-14.0	-8.4	-16.7

Nationally, we observe that the number of households is underestimated by 8.1%. This slippage rate is comparable to the -9.9% slippage rate for individuals.

Among the provinces, there is a substantial variation in the slippage rate for one-person households. This rate ranges from -19.6% for Prince Edward Island to 4.5% for Newfoundland and Manitoba. The same phenomenon is observable with respect to two-person households, where the slippage rate ranges from -11.0% for Saskatchewan to 10.2% for Prince Edward Island. The slippage rate for households of three or more persons is less variable, with values ranging from -7.3% for Saskatchewan to -16.7% for Prince Edward Island. For all provinces except Prince Edward Island and Saskatchewan, the greatest undercoverage is observed for households of three or more persons.

3.2 Adjustment for Coverage Error Relating to the Population and Number of Households

To correct the coverage problem illustrated in tables 3.1 and 3.2 and reduce the resulting bias, the survey data are adjusted for each province during weighting by using demographic estimates for the age-sex groups defined in these tables. For more information on the adjustment methodology, see reference [1]. This adjustment greatly reduces the bias caused by coverage errors, but it does not completely eliminate bias if the characteristics of the individuals omitted from the survey differ from those of the included individuals for a given age group in a province.

It should also be noted that the effectiveness of the coverage adjustment based on the demographic estimates greatly depends on the quality of those estimates and their accuracy in representing the target population of the survey. The demographic estimates are not error-free. They are post-censal estimates based on the population counts from the 1996 Census adjusted for net undercoverage, and they take into account recent statistics on migration, births, deaths, etc. These demographic estimates are adjusted to take account of certain exclusions specific to household surveys, such as persons living in institutions. Conceptually, they differ slightly from the SHS target population in that they include persons living in non-institutional collective dwellings, such as members of groups living communally and individuals permanently residing in hotels or rooming houses. However, this difference is considered negligible, since such individuals represent less than 0.4% of the Canadian population.

To remedy the problem of the representativeness of the sample in terms of the number of households by size as illustrated in Table 3.3, the survey data are adjusted using supplementary data. By adjusting the SHS weights to reflect post-censal estimates of the number of households by size, the goal is to compensate for the bias resulting from inadequate representation of households. However, this bias will not necessarily be eliminated if the characteristics of households not interviewed (i.e., omitted or non-respondent households) differ from those of respondent households for a given household size. Also, the effectiveness of the adjustment will depend on the quality of the supplementary data.

In addition to demographic estimates for age-sex groups by province, three other groups of supplementary data are used to adjust the survey data during weighting and thereby improve their representativeness. The first set of data is used to control for the number of children and adults in certain major cities. The second is designed to control for the number of single-parent households and couples with children by province. Finally, counts for major categories of income from wages and salary are used when adjusting weights to ensure a certain degree of consistency between the income distributions from the SHS and those from outside sources.

4.   RESPONSE ERRORS

Response errors represent a lack of accuracy in responses to questions. They may be attributable to different factors, including deficiencies in the questionnaire, misinterpretation of questions by interviewers or respondents, and errors in respondents' statements.

In the SHS, there can be various reasons for errors in respondents' statements. First, there are recall errors, which occur when a respondent forgets expenditures made during the survey period (which corresponds to the calendar year) or provides an erroneous value because of the time interval that has elapsed between the time of purchase and the date of the interview. Recall errors are probably the survey's largest source of response error, since the reference period is long (12 months) and a great range of information is requested.

One of the main measures taken to minimize recall error in the SHS is to calculate, for each household, the difference between receipts (income and other amounts received by the household) and disbursements (expenditures plus net change in assets and liabilities). When the difference exceeds 10% of either receipts or disbursements, whichever figure is higher, respondents are contacted again in order to obtain additional information and try to identify errors or omissions. The respondent is also encouraged to consult various documents (bills, bank statements, etc.) in order to provide more accurate data. To determine expenditures for small items purchased on a regular basis, interviewers generally suggest to respondents that they estimate the frequency of the purchases and the price generally paid in order to derive expenditures over a twelve-month period.

A second source of error in respondents' reporting is telescopic error, which consists of including in the reference period events that occurred before or after it. In the SHS, the use of the calendar year is considered to provide a good marker for the start of the reference period. Furthermore, since the reference period is a long one, telescopic error has less impact.

Responses by proxy can also contribute to response error. The household member who made an expenditure is generally the one who is able to report it most accurately. This is definitely the case with, say, personal purchases. Expenditures reported by an intermediary are more likely to be tainted by response error, and this type of error tends to have a greater effect on certain types of expenditures.

Among other sources of response error, the extent of the respondent's co-operation should not be overlooked. For personal reasons, the respondent may decide not to mention particular expenditures or twist the facts.

In the SHS, consideration is also given to the response burden, in light of the length of the interview and the great variety of items to be reported, as well as the pace of the interview. These factors can lead to respondent fatigue and affect the quality of the responses obtained. The interview time varies greatly from one household to another, depending on household size, income and various other characteristics. For some households, the interview can take more than 5 hours.

While response errors are a major source of error in a retrospective interview, they are the aspect of data quality that is hardest to measure. Generally, to attempt to measure them, it is necessary to conduct quite costly special studies. Efforts are made to reduce response error by using survey techniques designed for this purpose.

5.   PROCESSING ERRORS

Processing errors can occur in all types of data handling. The main stages of data processing are response coding, data entry, editing, imputation of partial nonresponse and weighting. In the SHS, different procedures are applied at each stage in order to minimize processing errors, and the survey estimates are

compared with other data sources prior to release. Errors that may result from adjustments made at the weighting stage have been described in sections 2 and 3. This section primarily focuses on the other types of processing errors.

Very few questions require response coding. This task is performed by the interviewer and is subsequently checked by a senior interviewer. Data entry is checked by means of an automated edit system that groups the questionnaires into batches and selects some questionnaires from each batch to be entered a second time. Any errors found are then corrected. If the number of errors in a batch exceeds a certain threshold, the entire batch is again submitted for entry.

The first stage of automated edit is carried out after each questionnaire has been checked manually by both the interviewer and the senior interviewer. The edit ensures that the responses recorded on the questionnaire satisfy certain essential consistency rules. Unusual situations that might justify corrections are also identified. This edit stage is carried out in Statistics Canada regional offices, so that respondents can be recontacted if more information is required to resolve inconsistencies in their answers. Any problems identified during this edit are solved by members of specially trained edit resolution teams. Subsequently, other edits are conducted at head office, and invalid responses are corrected.

The processing of SHS data also involves imputation for partial nonresponse. Partial nonresponse occurs when the respondent refuses to answer or does not know the answer to specified questions. The imputation approach differs depending on whether the data are categorical or continuous. Categorical data take on only specific values (as in yes/no questions or questions on the type of dwelling inhabited), while continuous data can take any numerical value (e.g., income and expenditure data).

Categorical data, which are obtained mainly in the sections on dwelling characteristics and household equipment, are imputed using a "hot deck" imputation technique that randomly chooses a donor from a group of respondent households with similar characteristics.

Income and expenditure data are imputed using the nearest neighbour technique. The imputation is done on one group of related variables at a time. A group generally corresponds to a section of the questionnaire. For every group, the missing values of a recipient (a household that has some missing data for at least one of these variables) are imputed using data from the most similar of all the donors (households that have no missing values for these variables). For each recipient, the nearest donor is the one that has the lowest value for a particular distance function. This function is based on matching variables chosen because they are correlated with the variables to be imputed. For example, the total income of the household is one of the matching variables for almost all sections pertaining to expenditures. Another consideration in selecting the donor is that after the missing values are imputed, the recipient household must satisfy certain consistency rules. For most sections, the imputation is done at the household level, but in some cases, e.g., for income and for clothing expenditures, the imputation is done at the individual level since the original data for these variables is collected at that level.

The bias in the estimates that is caused by imputation of partial nonresponse is difficult to evaluate. It depends on the difference between respondents and nonrespondents as well as the effectiveness of the imputation method in producing unbiased estimates. Nevertheless, the imputation rates give an indication of the magnitude of partial nonresponse. They are presented in the following section.[7]

5.1 Proportion of Households or Persons Requiring Imputation at the National and Provincial Levels

A primary indicator of the magnitude of the partial nonresponse is the proportion of households requiring imputation and the number of variables imputed per household. The questionnaire can be divided into two major groups of variables: those collected at the household level and those collected at the individual level such as income and clothing expenditure. For this second type of variable, it is important to note that it is acceptable for the respondent to provide only the total income or total expenditures if he/she is unable to provide the breakdowns by source of income or type of expenditure. Thus, the amount of imputation for the components of income and clothing expenditure is larger but does not affect total income, total clothing expenditures or total expenditures.

The percentage of households requiring imputation for household expenditures (excluding clothing expenditures) is presented in the following sub-section. Then the next sub-section presents the percentage of persons requiring imputation for at least one clothing expenditure variable and those requiring imputation for at least one income variable. The results are provided at both the national and provincial levels. This gives an indication of which provinces are the most affected by imputation.

5.1.1 Imputation of Household Expenditures by Province

The percentage of usable households that required imputation for at least one expenditure variable (excluding clothing expenditures) is shown in Table 5.2. The usable households are all households living in eligible dwellings, excluding households that could not be contacted, refused to participate in the survey or provided incomplete data or data that do not balance (see Section 2.1) The table provides a breakdown according to the number of imputed variables (out of 237) for a household.

---

[7] For operational reasons, these data quality indicators are not available for categorical data such as those on dwelling characteristics and household equipment.

Table 5.1
Households Requiring Imputation of Expenditures by Province

Province	Households (%) requiring imputation of **EXPENDITURE VARIABLES** (excluding clothing expenditures)			
	Number of variables imputed (out of 237)			
	1	2	3 or more	TOTAL
Canada	**6.9**	**1.2**	**0.7**	**8.9**
Newfoundland	12.4	0.8	0.5	13.6
Prince Edward Island	5.4	0.4	0.0	5.8
Nova Scotia	8.9	1.9	0.6	11.4
New Brunswick	7.9	1.4	0.9	10.3
Quebec	2.4	0.5	0.4	3.3
Ontario	7.1	1.9	1.2	10.2
Manitoba	4.7	0.7	0.9	6.3
Saskatchewan	6.1	0.8	0.3	7.2
Alberta	4.5	0.2	0.4	5.1
British Columbia	10.3	2.6	1.6	14.5

Table 5.1 shows that at the national level, 8.9% of households required some expenditure imputation (excluding the clothing section), but for nearly 80% of them, only one variable was imputed. There are very few households at the national level that had more than one variable imputed (1.9%). At the provincial level, Quebec (3.3%), Alberta (5.1%) and Prince Edward Island (5.8%) have the lowest imputation rates. The highest rates are observed in Newfoundland (13.6%) and British Columbia (14.5%). The small percentage of persons for whom variables had to be imputed, combined with the low number of variables that had to be imputed when imputation was necessary, suggests that the impact of the imputed values on the estimates should not be too great.

5.1.2   Imputation of Clothing Expenditures and Income by Province

Since some respondents provide only totals for clothing expenditure and income variables, these variables are imputed (at the individual level), in two stages. Individuals requiring imputation of only a few components are imputed first, followed by those for whom only totals are available and all components must be imputed (see reference [1] for a more detailed description of this process).

Table 5.2 shows, by province, the percentage of usable individuals (persons who are members of usable households) for whom an income variable had to be imputed. Included in the table is the percentage of persons for whom exactly one variable was imputed, those for whom two or more variables (but not all) were imputed and those for whom only total income was available, with the result that all components had to be imputed. The total percentage of persons requiring some form of income imputation is also shown. The last column of Table 5.2 indicates the corresponding total percentage of persons requiring some form of imputation for clothing expenditure variables.

Table 5.2
Persons Requiring Income Imputation and Persons Requiring Imputation of Clothing Expenditures, by Province

Province	Percentage of persons requiring imputation of INCOME VARIABLES				Percentage of persons requiring imputation of at least one of the 11 CLOTHING EXPENDITURE VARIABLES
	1 variable imputed	2 or more variables imputed (not all)	All variables imputed (total income known)	**TOTAL** (any form of income imputation)	
Canada	0.5	0.2	2.0	2.7	16.6
Newfoundland	0.3	0.1	1.2	1.6	13.4
Prince Edward Island	0.0	0.2	2.9	3.2	20.1
Nova Scotia	0.3	0.2	2.0	2.5	11.1
New Brunswick	0.0	0.2	4.4	4.6	19.3
Quebec	0.5	0.0	1.0	1.6	20.1
Ontario	1.1	0.3	1.8	3.2	11.9
Manitoba	1.5	0.2	1.9	3.6	14.7
Saskatchewan	0.2	0.2	0.7	1.1	20.9
Alberta	0.1	0.2	2.4	2.7	11.5
British Columbia	0.7	0.2	1.9	2.8	24.9

These results show that less than 3% of the persons in usable households required imputation of at least one income variable. For a little more than 70% of them, total income was available but all the components had to be imputed. For a very large proportion of the remaining persons requiring imputation, only one income component (one variable) was imputed. At the provincial level, the percentage of persons requiring imputation of at least one income variable are also small, ranging from a low of 1.1% for Saskatchewan to a high of 4.6% for New Brunswick.

The last column of the table shows that about 17% of persons required imputation for at least one of the clothing expenditure variables. The rates at the provincial level range from about 11% (Nova Scotia, Ontario and Alberta) to nearly 25% for British Columbia. Almost all these persons provided their total expenditure on clothing but the components had to be imputed. The higher level of imputation required on clothing expenditure components suggests that the estimates for these components might be greatly affected by imputation, whereas the effect on the estimates of total clothing expenditures will be negligible.

BIBLIOGRAPHY

[1] Tremblay, J. and Arsenault, S. (2001). Methodology of the Survey of Household Spending, Household Survey Methods Division, Statistics Canada.

[2] Wolter, K.M. (1985). Introduction to Variance Estimation, Springer-Verlag New-York Inc.

[3] User guide – Survey of Household Spending, 2000, Catalogue Number 62F0026MIE2001004.

[4] Spending Patterns in Canada (2000), Catalogue Number 62-202-XPE.

[5] Labour Force Survey, Monthly Operations Report, Survey 200012, Household Survey Methods Division.

APPENDIX A

Algebraic notation

1. Nonresponse Adjustment

The subweight (i.e. the design weight adjusted for nonresponse) for a household k, denoted as w_k^{NR}, is

$$w_k^{NR} = \pi_k^{-1} * \frac{1}{rate_g} \qquad with \qquad rate_g = \frac{\sum_{k \in s_{g,r}} \pi_k^{-1}}{\sum_{k \in s_{g,r}} \pi_k^{-1} + \sum_{k \in s_{g,nr}} \pi_k^{-1}}$$

where

$s_{g,r}$ is the set of respondents in nonresponse group g,

$s_{g,nr}$ is the set of nonrespondents (refusals, no contacts, unusable data) in nonresponse group g, and

π_k^{-1} is the design weight assigned to household k.

2. Calculation of the slippage rate

The slippage rate for a control group c, denoted as $rate_c$, is

$$rate_c = 100 * \frac{\left(\sum_{k \in s_{c,r}} w_k^{NR} \right) - t_c}{t_c}$$

where

$s_{c,r}$ is the set of respondents in control group c,

w_k^{NR} is the subweight of household k,

t_c is the total of the auxiliary data for the control group c.