Creating an Automated Edit and Imputation System for the Survey on Quebec Accommodation Establishment Occupancy

Catherine Fontaine[1], Luc Côté[1]
[1]Institut de la statistique du Québec, 200, chemin Ste-Foy Québec (Canada) G1R 5T4

Abstract
The Survey on Quebec Accommodation Establishment Occupancy is used to collect monthly data about accommodation establishments. Following the survey redesign that took place in 2011, an automated edit and imputation system (SIVEMEH) was created to handle item nonresponse for total rental income. One of the requirements of this improved system was to consider, when available, income from a previous month, current income (by category) and establishment type as auxiliary variables. These requirements have led to a system using composite imputation. The main imputation methods are nearest neighbour imputation and historical imputation. In some cases, the number of rooms occupied and the number of rooms available are also imputed to preserve the relationship between these variables and rental income. Large establishments received special attention during the creation of the system. This paper discusses the steps leading to the development of this improved imputation system and the first results obtained since its implementation in 2012. Future work is also discussed.

Key Words: Item nonresponse, imputation, accommodation establishment

1. Introduction

The Institut de la statistique du Québec (ISQ) is the government agency responsible for producing, analysing and disseminating official, objective and quality statistical information for Québec. This information enhances knowledge, enlightens debate and supports decision-making by the various players in Québec society.[1]

The ISQ also conducts statistical surveys on behalf of other government agencies to help them carry out their own mandates. This is the case with the *Survey on Québec Accommodation Establishment Occupancy*, which provides information on the vitality of the hotel sector. The results of the survey are produced and released monthly by the ISQ and are used by Tourisme Québec for publication, planning and analysis purposes.

The item nonresponse imputation strategy for this survey had not been reviewed in the past few years. As part of a redesign of various aspects of the survey between 2010 and 2012, this review was performed and is the subject of this article. Section 2 presents the survey; section 3 describes the development of the imputation strategy; section 4 provides

[1] See the ISQ website at http://www.stat.gouv.qc.ca/organisa/mission_an.htm.

the main steps of the imputation system; section 5 details the results and future developments; and section 6 concludes the article.

2. Survey description

The monthly *Survey on Québec Accommodation Establishment Occupancy* is aimed at accommodation establishments with 4 or more rooms in Québec. An accommodation establishment can be a hotel, an inn, a motel or a tourist accommodation. The sample frame is provided by Tourisme Québec, who manages the list of permits issued to active establishments in the industry. The population size is approximately 2,100 establishments per year. A stratified sample is selected at random once a year, in March, and is surveyed from May to April. This sample is made up of about 1,300 establishments per year. The goal is to have 75% of units sampled in year A-1 selected in year A. Moreover, all establishments with 150 or more rental units are surveyed.

The survey questionnaire collects information on four variables. Estimated totals are produced for three of them at the provincial and regional level: the number of rooms available during the month, the number of rooms occupied during the month and detailed income rental. Other fields of analysis at the subregional level are used to produce estimates. Ratios are also estimated based on these three variables. Important measures include the occupancy rate (ratio of rooms occupied to rooms available), the average price per room occupied (ratio of income to number of rooms occupied) and the average price per room available (ratio of income to number of rooms available).

At the beginning of each month, Tourisme Québec sends the ISQ the list of establishments that were in operation in the previous month. For a given month t, we survey the establishments selected in March that were part of the active population in the previous month. Data collection is done by telephone, but a number of establishments prefer to send in their questionnaires by fax or by mail. To be considered as a respondent, an establishment must report at least two variables: the number of rooms available and the number of rooms occupied. Therefore, item nonresponse is limited to responding establishments that do not report income. Between 2008 and 2011, the item nonresponse rate varied between 2% and 8% (at the provincial level). However, it can be much higher in certain tourism regions (20% in May and June 2011 for the Baie-James region) (Côté, 2011).

3. Development of the imputation strategy

3.1 Background

The survey manager was the professional in charge of operating the imputation system. Certain signs led him to believe that the statistical model used for the imputation strategy was outdated, as a growing number of imputed incomes were deemed inconsistent. Indeed, the donor imputation strategy did not take establishment type into account. Since hotels and tourist accommodation may have a very different average price per room occupied, this had an impact on the estimated ratio involving the income and number of rooms occupied. In addition, no historical data on establishments were used in the imputation strategy. This led to inconsistencies between imputed incomes and incomes reported at time t-12 (comparisons with estimates for the same month in the previous year). These two situations required manual corrections that had to be made by the project manager. Moreover, the item nonresponse imputation strategy was based on an

Access application which integrated several computer languages. The application format left little room for modifications, even simple ones. Finally, the project manager had noticed that the same donor was often used many times. The redesign gave us the opportunity to address all of the problems mentioned, which affected quality and increased production time.

The Institut has not developed a generalized edit and imputation system such as BANFF at Statistics Canada (read the article by Kozak, 2005). It was therefore necessary to create a new imputation system for this survey while respecting budget and time constraints.

3.2 Required elements

Certain elements were required by the project manager and had to be taken into account during the development of the new imputation system. The system had to allow for the automated imputation of detailed rental income using the SAS software, without any intervention from the project manager. Additionally, historical information and establishment type had to be taken into account in the system. This imputation system had to be easily and quickly integrated into the production line without any direct intervention from the methodologist (who could however provide ad hoc support). Furthermore, given that the ISQ adopted its General Quality Management Policy in 2006, good imputation practices had to be respected as much as possible. It was important to measure variance due to imputation during the production of estimates using SEVANI (System for the Estimation of Variance due to Nonresponse and Imputation), a software provided to the ISQ by Statistics Canada (read the article by Beaumont and Bissonnette, 2011). Finally, the system had to include documentation that outlined the reasons behind the choices made during system development, as well as guides to help the project manager during imputation and quality verification.

3.3 Situation assessment

A report on the situation was first produced before developing the imputation strategy. Since financial and human resources were limited, the characteristics of item nonrespondents had to be determined to focus efforts wisely. Modeling efforts were targeted according to the frequency of the cases needing imputation, while making sure that in the future, all establishments, <u>no matter their characteristics</u>, would be processed by the system.

The reference year 2010 was the most recent year for which data were available to test the new system. Due to time constraints, two months of the year 2010 were selected for modeling: the months of February and August. February is a month when the number of active establishments is low (“low” season), while August is a month when that number is high (“high” season). The following independent variables available for the monthly or annual sample frame as well as for the collection files were analysed:

Table 1: Available Auxiliary Variables by Type

CATEGORICAL	*CONTINUOUS*
o Classification	o Income at time t-v, for v=1,....,12
o Establishment type	o Number of rooms occupied at times t to t-v, for v-1,....,12
o Establishment size	o Number of rooms available at times t to t-v, for v-1,....,12
o Tourism region	o Establishment size

Note 1: Although establishment size is expressed as the number of rental units, in the definition of the sample strata, it is expressed as categories (5).
Note 2: An establishment's classification is an indicator of the level of comfort and services provided by that establishment.

One of the objectives pursued was to include historical information into the imputation system. The variable expressing income for the same month in the previous year (i.e. time t-12) was selected, as is usually the case in other recurrent surveys. Since comparisons between a given month and the same month in the previous year are important, and the concept of seasonal activity is an integral part of this type of industry, this choice was self-evident. However, a given establishment might not have been part of the sample in the same month of the previous year or was nonrespondent. This limits the availability of information at time t-12. About 25% of establishments to be imputed for item nonresponse for the months analysed had reported income at time t-12, compared with about 70% of establishments without item nonresponse.

Other analyses highlighted the importance of the type of establishment, because of both its significant association with item nonresponse and its link to income at time t. This confirmed that imputation had to be done separately for the different establishment types (which was not done before). An establishment's classification is not significantly associated with item nonresponse, but it is linked to income. As for the continuous variables, income at time t-12 was predominant compared with the other two variables available at time t (number of rooms available and occupied). The importance of the number of rooms occupied was greater than that of the number of rooms available, whether or not income at time t-12 was an independent variable in the model. All three of these variables are strongly associated with income.

Establishment size (i.e. the number of rental units) is strongly correlated with the number of rooms available. Consequently, the results obtained for this variable generally apply to establishment size. Income analysis has shown that, in certain regions, average incomes are either always high, always low or varying, depending on the month examined (low or high season). This finding makes it more difficult to group tourism regions by income, as these groups would need to be stable over time.

At this point, it was decided that additional information should be requested from establishments that refused to provide detailed rental income, seeing as the questionnaire was already being redesigned. These establishments would have the possibility to provide income by category. This new variable could then be taken into account in the development of the imputation strategy, knowing that it would be strongly associated with income.

3.4 Constraints and methodological choices

Certain constraints justified the selection of specific imputation methods and of a general strategy. First, we had to favour methods based on "real" and validated incomes, i.e. donor imputation methods, for most cases needing imputation. An imputation method based on income modeling could have been implemented if there had been a sufficient number of months to be analysed and if it had been possible to determine the requirements for using an income modeling method without monthly validation by the methodologist. Second, we had to take into account the fact that most cases needing imputation concerned small or medium establishments. Therefore, fewer efforts were made with regard to modeling for large establishments. Finally, we had to keep in mind that most establishments needing imputation had no income at time t-12. Consequently, any extra time available for modeling had to be devoted to establishments without available income[2] at time t-12.

3.4.1 Imputation methods for small and medium establishments

The imputation strategy for small and medium establishments is based on the availability of the two auxiliary variables with the strongest link to income at time t: income at time t-12 and income category at time t. Various imputation methods were available, and their benefits and drawbacks are well described in Kalton and Kasprzyk (1986). Three imputation methods were selected for SIVEMEH. The nearest-neighbour (NN) imputation method is described in subsection 3.4.1.1; the unit-trend imputation method in subsection 3.4.1.2; and the mean imputation method in subsection 3.4.1.3.

3.4.1.1 Nearest-neighbour imputation

This is the first method attempted to impute missing incomes, except when **both** income category and income at time t-12 are available. In that particular case, it is used only as an alternative if the first method fails validation (see section 4. for a description of the validation rules). NN imputation is often used in business surveys, as it takes into consideration the continuous nature of variables that are strongly associated with income, namely historical income, the number of rooms occupied and the number of rooms available, in the distance measure. The variables taken into account in the distance measure all have the same weight when calculating distance.

NN imputation is performed within imputation classes, which are formed by cross-tabulating the categorical auxiliary variables associated with income. The reader is invited to refer to Haziza and Beaumont's article (2007) to learn about the benefits of using imputation classes. Auxiliary variables associated with responding or not to questions about income must be prioritized to minimize nonresponse bias and variance due to nonresponse. Establishment type is the main cross-tabulation variable. The project manager wanted to take it into account in order to decrease the number of inconsistent imputations. This variable is linked to income and item nonresponse for income. Therefore, establishment type was considered as the most important categorical variable.

Two types of imputation classes were formed, depending on whether income category was available for the establishment needing imputation. When income category is available, it is used as the cross-tabulation variable to form a first type of imputation classes (establishment type by income category). If income category is not provided, the

[2] For the rest of the article, "available" income at time t-12 refers to income reported by a sampled, respondent, eligible and non-imputed establishment for the same month of the previous year.

establishment's classification is used as a cross-tabulation variable to form a second type of imputation classes (establishment type by classification).

Analyses were conducted to find a way to group tourism regions by income, and to integrate these groups into the second type of imputation classes (by classification). However, results were not conclusive and this option was rejected. There were other interesting methods for creating imputation classes, such as the score method, but it is based on the creation of a model that we did not want to automate without having performed the necessary tests beforehand and examined how to form groups. It seemed simpler to create imputation classes by cross-tabulating the auxiliary variables. However, the score method could have led to such homogeneous classes in terms of income that it would have been enough to impute the mean of incomes or to use hot deck imputation (Rancourt, 2004) within classes.

The *minmax* distance measure (Sande, 1979 and Rancourt, 1999), calculated after adequately standardizing variables, is defined as:

distance(i,j) = MAX(| $(x_{1i} - x_{1j})$| , |$(x_{2i} - x_{2j})$| , .., |$(x_{ki} - x_{kj})$|.)

where j represents a potential donor and i the recipient for variables x_1 , x_2 , ...

For each potential donor, absolute deviation between the value of a variable for the recipient and the donor is calculated for variables 1 to k. The distance selected (distance(i,j)) for a potential donor is the maximum absolute deviation for one of the variables in the measure. Then, we select the donor with the smallest distance from the recipient and its income is attributed to the establishment whose income is missing.

When income at t-12 is available, it is included in the distance measure, in addition to the number of rooms occupied at time t and the number of rooms available at time t. Donor imputation "overwrites" (for estimation purposes only) the number of rooms occupied and the number of rooms available to maintain intra-record consistency. The inclusion of historical income in the distance measure ensures longitudinal consistency, i.e. continuity with what was reported at time t-12.

With the new imputation system, a key goal was to decrease the frequency of cases where the same donor is used repeatedly, as was the case in the previous system. One of the methods chosen was to use a penalized distance (distance$_p$(i,j)), meaning that the distance is increased based on a penalty and on the number of times the potential donor was previously used in the imputation system (see article by Colledge et al., 1978).

distance$_p$(i,j) =distance(i,j) + (distance(i,j) *p*n)

where p has a value of 0.3 (derived from empirical tests) and n represents the number of times establishment j was chosen as a donor for other recipients already imputed, at the time when establishment i is to be imputed.

Two conditions must be met regarding the use of NN imputation within an imputation class. The ratio of recipients to donors must be less than or equal to 0.5 per imputation class, and there must be at least 10 donors per class. When both these conditions are met, NN imputation can be used. Otherwise, combining of classes is attempted, and these two conditions must be met in the combined classes for NN imputation to be performed.

3.4.1.2 Unit-trend imputation

This imputation method is used first when both income at time t-12 and income category are available for the establishment to be imputed.

$$y^{*}_{t,i} = y_{t-12,i} * \left(\frac{x_{t,i}}{x_{t-12,i}} \right)$$

where i represents an establishment with a missing income value (recipient).

$y_{t,i}$* represents imputed income at time t (current month);
$y_{t-12,i}$ represents income at time t-12 (same month in the previous year) ;
$x_{t,i}$ represents the number of rooms occupied at time t;
$x_{t-12,i}$ represents the number of rooms occupied at time t-12.

This method is equivalent to ratio imputation with one unit per imputation class. It is used when all variables on the right of the equation are higher than 0 and if they have not been imputed. It is also very easy to program. It the trend (ratio) between month t-12 and month t is very similar for the number of rooms occupied and for income, a good imputation should be obtained.

Since different methods can be used to integrate historical information into an imputation strategy, questions arose regarding the best way to maintain longitudinal consistency between incomes at time t-12 and at time t. Empirical tests were conducted to compare the NN method using income at time t-12 in the distance measure with the unit-trend method. These tests demonstrated that:

a) Unit-trend imputation works better when income category at time t is also available. It has been shown that using income category as a validation boundary minimizes the difference between imputed income and real income.
b) This method better maintains the difference between imputed and real income than the NN method. However, NN imputation will be attempted secondly if validation fails with the unit-trend method (does not respect the income category) or if requirements for its use are not met (for example, the number of rooms occupied at month t-12 is equal to 0).

3.4.1.3 Mean imputation

This imputation method is used as a last-resort option: the weighted mean of incomes is imputed per class. It is used when no other class combination respects the conditions of use of NN imputation, or when imputation does not meet the validation rules.

3.4.2 Imputation methods for large establishments

For establishments with more than 150 units, i.e. large establishments, the number of cases to be imputed monthly is expected to be small. Donor imputation is not appropriate because there are fewer establishments of this size, meaning that the pool of potential donors is insufficient. In addition, since all large establishments are surveyed and have response rate higher than smaller establishments, historical income is more readily available than for small or medium establishments. Since the tourism activity of large establishments is more stable than that of other establishments, the method selected was historical imputation by substitution.

$$y^*_{t,i} = y_{t-v,i}$$

$y_{t,i}$* represents imputed income at time t (current month);
$y_{t-v,i}$ represents income at time t-v (v=1, ...,12).

One month is selected among the 12 months preceding the survey, according to a predetermined order, to assign the income reported for that month to the current month (the first one being month t-12). This method underestimates the changes between income at month t-v and month t, but the impact on the comparisons between these two months is expected to be minimal because of the low imputation rate among large establishments. After imputation, the average price per room occupied is examined. If the average price exceeds a certain threshold, the two other variables in the questionnaire are imputed to maintain intra-record consistency.

4. Description of the imputation system

The item nonresponse imputation system was named SIVEMEH (imputation and validation system for the monthly Survey on Québec Accommodation Establishment Occupancy). This system follows four major sequential steps:

1. Preparation of the input file
2. Imputation for small and medium establishments
3. Imputation for large establishments
4. Quality verification

The first step is done by the project manager who prepares the file containing the variables necessary to the proper operation of the imputation system. The second step is at the heart of SIVEMEH: using the input file, an output file containing imputed incomes for small and medium establishments is produced. The main SAS program distributes cases to be imputed among the different blocks in the system, which are chosen according to the availability of the auxiliary variables (see columns 2 and 3 in the figure below). These blocks define the imputation method, imputation classes and variables used for the distance measure (if applicable). Imputation of cases to be imputed according to these blocks is done sequentially, from 1 to 4 (block 5 is processed separately). Only block 1 has an exception in the application of the sequence. If validation of imputation at block 1 fails, NN imputation is attempted.[3] If the number of rooms occupied reported by the establishment is zero, deductive imputation of income ($0) is done at the beginning.

[3] With imputation classes defined in the same way as in block 3 and a distance measure defined in the same way as in block 2.

Block	**Income at t-12?**	**Income by category?**	**Type of imputation**	**Definition of imputation classes**	**Distance measure**
SMALL AND MEDIUM ESTABLISHMENTS					
1	YES	YES	UNIT-TREND AT T-12	*NONE*	*NONE*
2		NO	NN WITHIN AN IMPUTATION CLASS	TYPE OF ESTABLISHMENT * CLASSIFICATION	1. INCOME AT **T-12** 2. ROOMS OCCUPIED AT **T** 3. ROOMS AVAILABLE AT **T**
3	NO	YES		TYPE OF ESTABLISHMENT * INCOME BY CATEGORY	1. ROOMS OCCUPIED AT **T** 2. ROOMS AVAILABLE AT **T**
4		NO		TYPE OF ESTABLISHMENT * CLASSIFICATION	
LARGE ESTABLISHMENTS					
5	SELECTION OF AN INCOME BETWEEN T-12 AND T-1		HISTORICAL (SUBSTITUTION)	*NONE*	*NONE*

Figure 1: Description of the block-based imputation system

When NN imputation is used, imputed variables must be validated. The same validation rules as those developed for the data collection software (CATI) are used, namely a comparison between imputed income at time t and income at time t-12; a comparison between the imputed number of rooms occupied at time t and the number of rooms occupied at time t-12; as well as a comparison between income and the number of rooms occupied at time t, if income at time t-12 is not available (or was imputed). For block 1, prior validation is carried out. The purpose of this validation is to verify that the imputed income at time t falls into the income category reported by the establishment.

The third step is imputation for large establishments (block 5). At the end of this step, imputed incomes are integrated into the file derived from the imputation of establishments in blocks 1 to 4, and quality verification can be performed.

The fourth and last step is quality verification. Several indicators of imputation quality are calculated, which are all linked to good imputation practices. Since a large volume of results are produced and production time is short, some of these verifications are done on a monthly basis, and others on an annual basis (or as needed).

i) Imputation rate at the provincial level and by cross-tabulation variable
ii) Frequency of use of imputation methods
iii) Number of times a donor is used
iv) Number of donors used
v) Ratio of imputed total income to total income

To support quality verification, each imputed income is assigned an 8-digit code indicating the path it followed in the system. For example, a case initially assigned to one of the blocks could be imputed without any problem according to the intended method. However, it could be imputed then fail validation, or be assigned to the NN method but undergo mean imputation due to an insufficient number of donors in the imputation class. The assigned code makes it possible to verify if the system is functioning properly by providing:

- the number of cases that failed validation (when performed[4]);
- the number of cases that were part of an imputation class that needed to be combined; and
- the distribution of cases needing imputation according to the block assigned and the final imputation method.

These codes allow for a certain validation of the imputation model. It is the case for block 1, where a high rate of incomes failing validation by income category would signal that unit-trend imputation needs to be reassessed. Similarly, if an imputation class often needs to be combined for NN imputation to be performed, this imputation classes may need to be modified.

After quality verification, statistical production resumes based on a data file containing incomes, imputed number of rooms occupied and available, as well as an imputation identifier associated with each of these three variables.

5. Results and future developments

SIVEMEH was implemented in May 2012. The results from the first months are necessary to adjust the system if need be and to monitor its performance. Data for the months of May 2012 to February 2013 were analysed in detail, and the most important findings are presented below along with the proposed developments.

i) Lower imputation rate

The (weighted) imputation rate varied between 0.2% and 2.4% over the first 10 months following system implementation. This is a vast improvement over the monthly imputations rates measured between 2008 and 2011. Although imputation rates are still high for some regions depending on the month, the outcome is still positive. There are now 10 to 20 cases to be imputed per month, whereas in 2010 (reference year for the development of the system), there were approximately 50 to 60 cases per month. Can this drop be attributed to the implementation of SIVEMEH? In reality, it is mostly due to the general system redesign. The redesign improved validation during data collection, leading to a decline in the number of cases deemed inconsistent post hoc, i.e. to be imputed. In addition, interviewers were reminded of the importance of this variable during collection, and new instructions were given to them. When a respondent does not provide detailed income, the interviewer lets him or her know that this information will be asked again the following month (which was not done before). This can increase response by allowing respondents to prepare.

[4] If a case to be imputed is assigned to NN imputation but application requirements are not met for the imputation class it belongs to, mean imputation is used. This imputation method is not followed by validation of imputed income.

ii) Control of the number of times a donor is used

While the same donor could be used up to seven times in the previous system, in SIVEMEH, the same donor has not been used more than twice for the same month.

iii) Distribution of cases needing imputation by block

Block 4 is assigned the largest proportion of cases to be imputed, month after month (from 7 out of 12 cases in December to 12 out of 13 cases in November). Block 3 follows in terms of frequency, although its use is less extensive (at most 5 out of 12 cases in December). These are the two blocks for which no income is available at time t-12. This confirms that any improvements to be made must focus on cases where income at time t-12 is unavailable. For example:

- Reviewing the definition of imputation classes using another method and/or by including tourism regions.
- Assessing if another ratio imputation method would be a better option than NN imputation.
- Verifying if a different month in the establishment's history could be used to maximize the number of cases imputed based on historical information.

iv) Mean imputation method

Donor imputation was the most frequent method used each month. However, mean imputation must be used for as few cases as possible each month since incomes imputed with this method are not validated. Mean imputation was used up to 5 out of 20 cases in May; 3 out of 13 cases in November and 3 out of 12 cases in December. The conditions for validation of NN or unit-trend imputation could be examined to determine if they are too strict, which would cause a large number of cases to be imputed using mean imputation. Mean imputation is also used when combined imputation classes do not meet the conditions of intra-class NN imputation. Class combining could be reviewed to enable NN imputation by establishment type only (in blocks 3 and 4). This would increase the number of cases imputed using this method.

v) Number of cases needing imputation by establishment size

Among the large establishments surveyed, only one needed to have its income imputed for one month out of the 10 analysed. This is a positive finding since it corresponds to what was observed in the development phase, and justifies the amount of effort invested in this step. However, special attention should be given to this establishment category in the future to assess if the imputation strategy should be modified in case of higher item nonresponse.

vi) Estimating variance due to imputation

When initial objectives were established, the idea of estimating variance due to imputation using SEVANI was discussed. Due to lack of time, this objective has not yet been achieved. However, the choices made during the development of the imputation strategy (methods, number of donors per class, etc.) aimed to make alignment possible between SIVEMEH and SEVANI. With version 2.3.4 of SEVANI, available at the ISQ, it is not possible to estimate variance for a ratio, but the survey uses three ratios as key measures. Developments on this front will be monitored in order to assess whether this option will eventually be integrated into the system.

6. Conclusion

This first system assessment aimed to ensure its proper overall operation. It highlighted possible improvements and elements to be monitored. This new system enables greater and more varied use of historical data, uses composite imputation, introduced a mechanism to control the number of times a donor is used, and ensured that certain choices were made in preparation for the future use of SEVANI. In short, most of the initial objectives were met, and imputation quality was increased. This result was also reported by the project manager, who saw the number of imputations deemed inconsistent drop significantly.

Acknowledgements

The author wished to thank all of her colleagues at the ISQ who contributed at some point to this article and the preparation of the poster. More specifically, she would like to thank Luc Côté, coauthor and methodologist, for his expert advice throughout this project; Francine Chercuitte, project manager for the redesign, for her support during the development of SIVEMEH, and Éric Gagnon, coordinator, for his relevant feedback.

References

BEAUMONT, J.-F., and J. Bissonnette. (2011). Variance estimation under composite imputation: The methodology behind SEVANI. *Survey Methodology*, 37: 171-179.

COLLEDGE, M.J., et al. (1978). Large Scale Imputation of Survey Data. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 431-436.

COTE, L. (2011). *Enquête sur la fréquentation des établissements d'hébergement du Québec: rapport méthodologique 2010-2011*, Institut de la statistique du Québec.

HAZIZA, D., and J.-F. Beaumont. (2007). On the Construction of Imputation Classes in Surveys. *International Statistical Review*, 75: 25-43.

INSTITUT de la statistique du Québec. (2006). *General Quality Management Policy of the Institut de la statistique du Québec*.

INSTITUT de la statistique du Québec. (2009). *Recueil de bonnes pratiques dans les enquêtes*, internal document of the Institut de la statistique du Québec.

KALTON, G., and D. Kasprzyk. (1986). The Treatment of Missing Survey Data. *Survey Methodology*, 12: 1-16.

KOZAK, R. (2005). The BANFF System for Automated Editing and Imputation. *Proceedings of the Survey Methods Section*, Statistical Society of Canada.

RANCOURT, E. (1999). Estimation with Nearest Neighbour Imputation at Statistics Canada. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

RANCOURT, E. (2004). *Le traitement de la non-réponse dans les enquêtes et recensements*. Document handed out during training given at the Institut de la statistique du Québec on February 18 and 19, 2004.

SANDE, I.G. (1979). A Personal View of Hot Deck Imputation Procedures. *Survey Methodology*, 5: 238-258.