BANK OF CANADA
BANQUE DU CANADA

Technical Report No. 103 / Rapport technique n° 103

2013 Methods-of-Payment Survey: Sample Calibration Analysis

by Kyle Vincent

www.bank-banque-canada.ca

April 2015

2013 Methods-of-Payment Survey: Sample Calibration Analysis

Kyle Vincent

Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
kvincent@bankofcanada.ca

The calibration analysis presented in this report pertains to the 2013 Methods-of-Payment survey questionnaire participants.

ISSN 1919-689X

Acknowledgements

The author would like to thank Marcos Sanches, Sasha Rozhnov, and Winton Klass for their review of the statistical methodology developed for the analysis presented in this technical report, and Marco Angrisani for his formal review of the report. The author would like to thank Chris Henry for obtaining the necessary marginal totals required for the calibration variables explored in this analysis, and Heng Chen and Rallye Shen for undertaking an analysis to address a dual-frame coverage issue. Finally, the author would like to thank Kim P. Huynh and Geoff Dunbar for providing helpful comments and suggestions during the time this work was undertaken.

Abstract

Sample calibration is a procedure that utilizes sample and national-level demographic distribution information to weight survey participants. The objective of calibration is to weight the sample so that it is demographically representative of the target population. This technical report details our calibration analysis for the 2013 Methods-of-Payment survey questionnaire sample. The analysis makes use of a variety of variables, with corresponding distributions from the 2011 National Household Survey and 2012 Canadian Internet Use Survey. Our primary objective is to seek a sensible set of variables for calibration and to propose a set of final weights that meet a validation criterion.

A raking ratio calibration procedure is used in the analysis. We base calibration on candidate variables and nesting of pairs of variables chosen within the context of the study. An imputation strategy is implemented to account for the relatively few missing observations. Three samples are obtained for the survey and we summarize an analysis that suggests that calibration should be based on the full/collapsed data set. We describe our research on several validation criteria and, after testing the calibration procedure, report our proposed set of final weights.

JEL classification: C81, C83
Bank classification: Central bank research

Résumé

Le calage est une méthode de redressement qui utilise de l’information sur la distribution d’un échantillon et de la population nationale pour déterminer la pondération des participants à une enquête. Le calage vise à pondérer un échantillon afin que sa composition démographique soit représentative de la population cible. Ce rapport technique expose l’analyse du calage effectué à l’égard de l’échantillon adopté pour l’enquête sur les modes de paiement de 2013. L’analyse se sert de diverses variables ainsi que des distributions correspondantes tirées de l’Enquête nationale auprès des ménages (2011) et de l’Enquête canadienne sur l’utilisation d’Internet (2012). Le principal objectif consiste à établir un ensemble pertinent de variables pour le calage et à proposer des pondérations définitives qui répondent à un critère de validation déterminé.

Aux fins de l’analyse, un calage, fondé sur les variables admissibles et sur l’imbrication de paires de variables choisies dans le contexte de l’étude, est réalisé selon la méthode itérative du quotient. Une stratégie d’imputation est appliquée pour tenir compte du nombre relativement peu élevé d’observations manquantes. Trois échantillons sont obtenus pour l’enquête; l’analyse indique que le calage devrait être basé sur l’ensemble complet de données (perçu aussi comme un ensemble de données regroupées). La recherche de plusieurs critères de validation est décrite et, après avoir testé la méthode de redressement retenue, des pondérations définitives sont proposées.

Classification JEL : C81, C83
Classification de la Banque : Recherches menées par les banques centrales

1 Introduction and Scope

The survey team,[1] situated within the Currency Department of the Bank of Canada, is responsible for administering the 2013 Methods-of-Payment (MOP) survey. The focus of the survey is to measure the Canadian population's usage of cash and adoption patterns of payment innovations. Henry et al. (2015) detail the results of the survey; all tables and figures presented in their paper are based on the final weights proposed in the current technical report. All computations performed in this analysis are achieved with the aid of the R "survey" package (Lumley, 2012, 2010). All analyses have been cross-checked in the Stata programming language; more details can be found in Chen and Shen (2015).

Fieldwork for the 2013 MOP survey was conducted by Ipsos Reid, a survey-based marketing research firm. The firm maintains two marketing access panels; *online* and *offline*. Recruitment for the 2013 MOP was conducted through three primary sources. The first sample comes directly from the online panel and the second sample directly from the offline panel. Recruitment for the third sample is based on a subsampling approach: individuals who had recently completed the Canadian Financial Monitor (CFM) survey, also conducted by Ipsos Reid and where recruitment is based on the offline panel, were invited to participate in the MOP survey. A total of 3,663 individuals filled out the survey questionnaire (SQ): 1,563 from the online panel, 728 from the offline panel, and 1,372 via subsampling the CFM. Online participants were required to complete the survey electronically, and offline participants by paper.

Target sample compositions for each of the three samples are based on demographic counts from the 2011 National Household Survey[2] (NHS). Three key demographics are identified, namely province (or region in the case of the Atlantic provinces), gender, and age, in order to obtain final samples representative, in terms of these features, of the national population.

[1] The survey team consists of Heng Chen, Chris Henry, Kim Huynh, Rallye Shen, Ye Tao, and Kyle Vincent.

[2] For more information, see `http://www12.statcan.gc.ca/NHS-ENM/index-eng.cfm`.

This study's primary goal is to determine a suitable set of calibration variables and propose a set of final weights that meet a validation criterion. We proceed as follows. We first seek a suitable set of demographic and technology-based variables to base calibration of the sample upon. We also seek to determine a choice of initial sampling weights, in order to achieve a final set of weights. Our final objective is to validate the weights using methods suggested in the literature.

This report is organized as follows. Section 2 provides an overview of the raking ratio calibration procedure. Section 3 outlines the candidate calibration variables chosen for the analysis. Section 4 discusses the method used for imputing the missing calibration variables in the data set. Section 5 provides a rationale for combining the three samples into one for the calibration analysis.

Section 6 proposes nestings of pairs of variables suitable for the calibration analysis of the 2013 MOP SQ. Section 7 details several combinations of calibration variables suitable for this analysis.

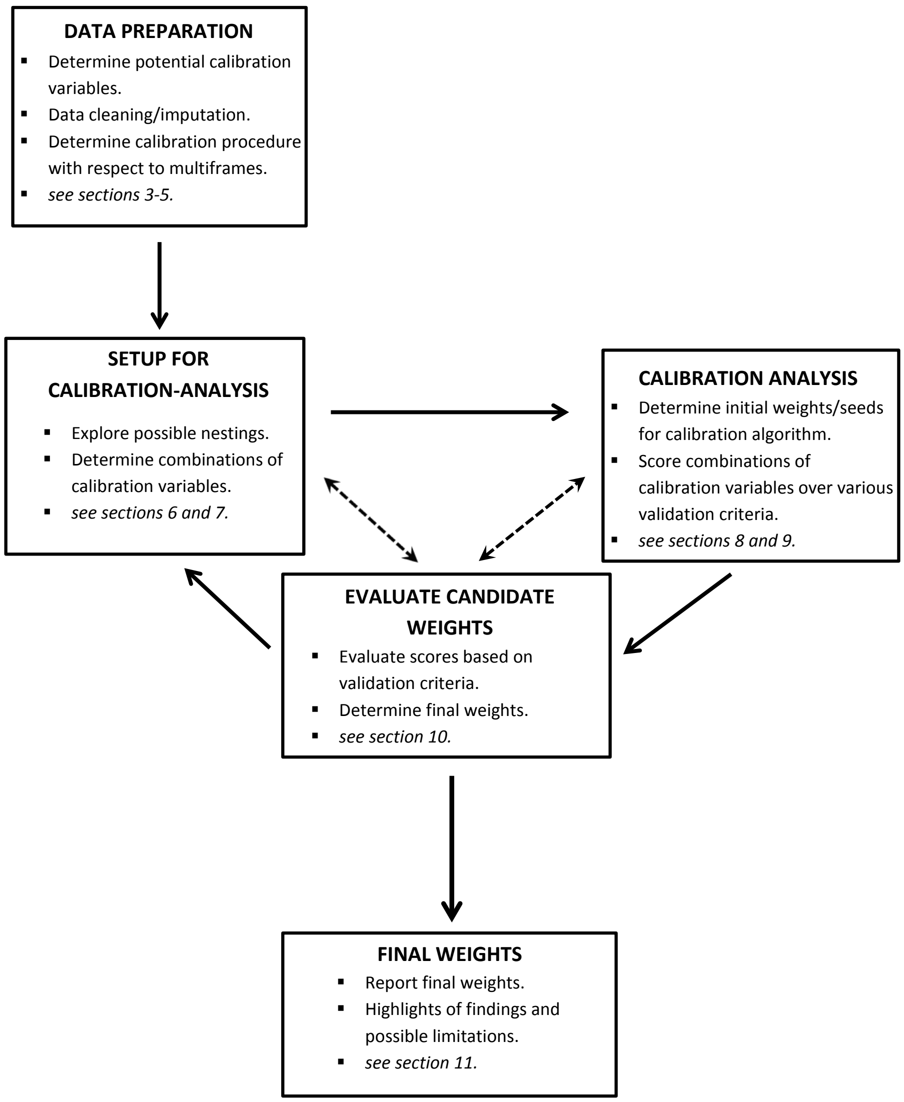
Section 8 provides a robustness analysis of the calibration procedure over the combinations of calibration variables to choices of initial seeds/weights for the calibration algorithm. Section 9 details the validation criteria and reports results when applying the calibration procedure to the various combinations of the (nested) calibration variables.

Section 10 provides an evaluation of the scores and rankings of the several combinations of calibration variables from the analysis. A proposal for the final weights is also made.

Section 11 reports the proposed final weights and discusses the limitations of the calibration analysis; it also provides recommendations on further validation.

Figure 1 provides a visual summary of the calibration analysis undertaken for this study.

Figure 1: Sample Calibration Flowchart



Notes: The flowchart illustrates the process of sample calibration for the 2013 Methods-of-Payment survey. Solid arrows indicate steps in the workflow. The dashed arrows indicate feedback between workflow steps; evaluation of candidate weights depends on the calibration algorithm as well as the subset of variables included. Corresponding sections in this report are identified in italics.

2 Overview of Raking Ratio Calibration Procedure

The raking ratio procedure (Deville et al., 1993), also known as iterative proportional fitting, the multiplicative method, or simply the *raking procedure*, is a commonly practiced method of sample calibration. A set of demographic variables, which we shall refer to as the calibration variables, are typically chosen based on the strength of their relationship with the survey response(s). At a marginal level, population-wide counts are required for each of the variables. The procedure makes use of a distance function that acts on a set of initial weights and what will be the final weights where this function is optimized in terms of the minimum distance between the initial and final weights (see Deville et al. (1993) for the mathematical details).

The procedure commences with the user specifying initial seeds/weights for the sampled individuals and then considering each demographic in a cyclic pattern, one-by-one, updating the sample weights so that they match the marginal totals of the demographic considered at each step. This step is repeated until the weights reach convergence, and thus the updated set of weights are taken to be the final weights (note that these weights will satisfy the aforementioned optimality criteria). The algorithm has an intuitive feel in that the weighted sample marginal totals for each of the raking variables should (nearly) coincide with those at the population level.

In this calibration analysis we use the raking procedure, since it is a popular calibration method and there is a tendency for its use in national statistical agencies (Särndal, 2007). Further, since the raking procedure was used to analyze the data from the 2010 Survey of Consumer Payment Choices (SCPC) (Angrisani et al., 2013) and the 2009 MOP survey (Sanches, 2010), using this procedure will permit future comparative analyses. Hence, the raking procedure serves as the calibration method for the 2013 MOP data.[3]

[3]The generalized regression procedure (Deville et al., 1993) is another commonly used calibration procedure. However, since this procedure is not as popular as the raking ratio procedure for empirical application, and is likely to return negative weights, this method is not explored in this analysis.

3 Candidate Calibration Variables

A candidate calibration variable is chosen based on two criteria: (1) the availability of its corresponding national-level marginal total counts via the 2012 Canadian Internet Use Survey[4] (CIUS), and (2) the strength of its association with key MOP SQ variables, in particular those related to cash usage and storage, as well as adoption and frequency of use of new methods of payment. We use the "polycor" package in R (Fox, 2010) to provide a general sense of which variables are correlated with such variables; polychoric correlation measures are based on Pearson product-moment correlations for pairs of numeric variables, polyserial correlations for pairs of numeric and ordinal variables, and polychoric correlations for pairs of ordinal variables (Drasgow, 2004).

Table 1 details the candidate calibration variables and corresponding categorical/ordinal responses; collapsing responses into cells is based on the criteria suggested by Battaglia et al. (2004), so that sample cells are arranged where both sample and corresponding population-based counts are greater than 5 per cent of the total count.

[4]For more information, see `http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4432`.

Table 1: Candidate calibration variables and corresponding responses

Gender	Male, Female
Age	18-24, 25-34, 35-44, 45-54, 55-64, 65 or more
Region (of residence)	Atlantic, Quebec, Ontario, Prairies, British Columbia
Income (household-level)	<25k, 25k-45k, 45k-65k, 65k-85k, 85k or more
Education	Some/completed public school, some high school Completed high school Some/completed technical school community college or university Completed bachelor's degree Completed graduate degree
Mobile[5]	Yes, No
Online[6]	Yes, No
Household size	1, 2, 3, 4, 5 or more
Marital status	Single, Married or common-law Separated or divorced, Widowed
Employment	Employed, Unemployed, Not in labour force
Home ownership	Own, Do not own

The Mobile variable is based on the response to Question 5 of the long version of the 2013 MOP SQ (Figure 2).

[5]Mobile is defined as ownership of a mobile phone.
[6]Online is defined as a payment made online in the past year.

Figure 2: Question 5 of the long version of the 2013 MOP SQ

5. Do you have a smartphone (e.g. iPhone, Blackberry, Android)?

Yes .... ☐

No .... ☐

The CIUS provides marginal counts based on binary responses to the question "Did you use the Internet on a smartphone/tablet in the past year?" Since responses to Question 5 give rise to a sample distribution reflective of the national distribution of this variable, we base our marginal counts for the Mobile variable on the corresponding weighted CIUS variable counts.[7]

The Online payment variable is based on the response to the last four subquestions of Question 1 of the long version of the 2013 MOP SQ (Figure 3).

Figure 3: Question 1 of the long version of the 2013 MOP SQ

1. In the **past year**, have you **used** any of the following **methods of payment** to make a purchase? *CHECK ONE BOX FOR EACH METHOD OF PAYMENT.*

	Used in past year	**Have not used in past year**
Prepaid card issued by Visa/MasterCard/AMEX........	☐	☐
Store-branded prepaid card	☐	☐
The 'Tap-and-Go' feature of a **credit** card	☐	☐
The 'Tap-and-Go' feature of a **debit** card	☐	☐
Mobile payment application	☐	☐
Online payment account (e.g. PayPal)	☐	☐
Interac online / Interac e-transfer	☐	☐
Online payment from a credit card account	☐	☐

The CIUS provides marginal counts based on binary responses to the question "Did you make an online purchase in the past year?" A response of "Yes" to this question is related to

[7]The 2012 CIUS is weighted to the 2011 NHS. See section 3.2 for more details.

a response of "Used in past year" to any one of the last four subquestions of Question 1 (note that in this analysis a mobile payment is regarded as an online payment). We therefore base our marginal counts for the Online payment variable on the corresponding weighted CIUS variable counts.

3.1 Rationale for candidacy of calibration variables

The choice of the candidate calibration variables is further strengthened with a theoretical rationale for each variable. The following list details this information.

Core variables commonly used in survey weighting:

1. Gender
2. Age
3. Region
4. Income
5. Education

Technology-based variables anticipated to both reflect heterogeneity in variables related to new payment technologies and correlate well with variables related to individual adoption tendency toward new methods of payment:

6. Mobile
7. Online

Variables associated with economies of scale and that are anticipated to correlate well with variables determined by wealth effects, for example consumption and payments:

8. Household size
9. Marital status

Status variables anticipated to correlate well with variables determined by income effects, for example frequency/quantities of monthly spending:

10. Employment
11. Home ownership

3.2 Marginal distribution of weighted 2012 CIUS sample and 2013 MOP SQ sample

The 2012 CIUS marginal counts are weighted to the 2011 NHS marginal counts so that little discrepancy exists when comparing the two features. We therefore base marginal counts on the timely availability of the weighted 2012 CIUS for all but the marital status variable, since details on that variable were not obtained at the time the analysis was undertaken. However, existing NHS-based details of the marital status variable were available to the survey team before the analysis was undertaken.

The appendix provides the corresponding approximated national and sample marginal counts of the calibration variables. Counts are coded to preserve a privacy clause with Ipsos Reid. For results based on an in-depth logistic regression-based analysis of how demographic variables influence participation rates of MOP invitees, see Shen (2014).

4 Imputation of Missing Observations

There are a relatively small number of missing observations for the calibration variables in the final data set. Table 2 identifies the proportion of missing observations over the candidate calibration variables.

Table 2: Proportion of missing observations for each calibration variable. The total number of survey participants is 3,663.

Gender	Age	Region	Income
0	0	0	0.00546

Mobile	Online payment	Education	Household size
0.01065	0.01775	0.01146	0.00437

Marital status	Employment	Home ownership	
0.00437	0.01392	0.00464	

The traditional approach to inference in such a case would be based on listwise deletion (Jones, 1996). Since we are performing a calibration analysis, such an approach would consist of removing all row entries from any participants not providing answers to all of the calibration variables. Discarding data corresponding with such survey participants can be considered wasteful and is seldom practiced nowadays (Kish, 1992). Instead, alternative methods, such as those based on imputation strategies, are quickly gaining justification and, hence, popularity (Little and Rubin, 2002).

Research on such imputation-based analyses has been explored by members of the survey team. For example, Henry and Vincent (2014) develop a Bayesian data augmentation routine to impute missing observations for a subset of the CFM survey financial data. Tao and Vincent (2013, 2014) explore the use of predictive mean matching and Poisson regression methods for imputing missing values in subsection 2.1 of the CFM. At this time, we opt to base imputation on the well-known "mice" R package (van Buuren and Groothuis-Oudshoorn, 2011; van Buuren, 2012). The canned function 'mice()' is used for imputation of missing data. In our analysis, this program makes use of a proportional odds model and logistic regression imputation. Alternative methods based on imputation research will continue to be explored in future analyses of the 2013 MOP survey data set.

5 Calibration on Full Data Set

Recall that invitations for the 2013 MOP survey are distributed amongst three sources: online, offline-paper, and offline-CFM. For such a case of recruitment based on several frames, approaches to sample weighting based on calibration procedures applied separately to combinations of the samples have been studied; for example, see Brick et al. (2006). Chen and Shen (2015) suggest that collapsing the three sources and then conducting the calibration analysis based on the non-overlapping dual frames is applicable for the current MOP SQ. Below is a summary of their analysis.

A comparative analysis is first undertaken to highlight key differences between the three sources in terms of missing data rates and the distribution of responses to the MOP questions. With respect to comparing missing data rates, it is found that few differences exist between the two offline-based samples. In contrast, it is found that statistically significant differences exist between the online and combined offline-based samples. With respect to comparing the distribution of responses to MOP questions, results from the Epps-Singleton test parallel those based on the comparison analysis for missing data rates; in general, there is no significant difference between the two offline-based samples but a difference exists between the online and combined offline-based samples. It is concluded that the samples obtained from the offline sources should be merged and treated as if they are obtained from the same frame. With respect to the sample obtained from the online source, invitations are initially drawn from a probability Internet panel. Hence, a classic dual-frame sample exists. Further, by construction, the two frames do not overlap. This is advantageous for this analysis, since complications with adjusting composite weights for the overlapping frames can be avoided (Callegaro et al., 2011).

For a dual-frame study such as that described above, Brick et al. (2006) suggest an analysis to compare estimates based on various tandem schemes of merging and weighting, and vice versa. Five such approaches are described as follows: determine weights based on (1) the sampling design induced weights, (2) merging the two frames and then calibrating, (3) calibrating the two frames separately and then merging, (4) merging the two frames and then calibrating with an "Internet access" calibration variable added to the procedure, and (5) calibrating only the offline paper-based sample. The analysis proceeds based on a comparison of the corresponding weighted estimates for the cash-on-hand variable, an MOP SQ variable of high interest to the survey team. It is found that point estimates from the five weighting schemes very nearly coincide. In contrast, it is found that option (2) gives rise to the lowest standard error of the point estimate, thus indicating that calibration after collapsing the dual frames should offer a significant increase in the precision of estimates (see D'Arrigo and Skinner (2010) for more details) of related MOP SQ variables and is therefore suggested for this calibration analysis.

6 Nesting Structures

Nesting demographics is a popular strategy for a sample calibration analysis, and was used for calibration of the 2012 SCPC data set by Angrisani et al. (2013) and the 2009 MOP survey data set by Sanches (2010). This strategy permits several advantages: analysts can more efficiently account for disagreements between the observed sample distribution and the national-level counts with nested pairs of variables (so that the procedure approaches a post-stratification one), and they can more efficiently employ the calibration algorithm while retaining the ability to obtain sample weights that approximate well the corresponding marginal count totals.

Calibration analyses based on sets of variables with pairings that interact would typically call for a nesting of the two in weaker cases and a removal of one in stronger cases. The primary reason for this is that such occurrences could present multicollinearity-based issues like those commonly encountered in regression analyses. In particular, raking makes use of a main effects type of approach and hence the calibration procedure can result in biased estimates if interactions exist between pairs of calibration variables (Deville et al., 1993). For these reasons, we opt to use calibration variables based on the following list of proposed nestings. A rationale for each pairing is provided.

1. Mobile within Age, and

2. Online within Age:

   Since attributes related to the ownership of a mobile device and online payments are expected to be concomitant with an individual's age, it is anticipated that finer category assignments of this variable will better capture the heterogeneity in responses related to new payment technologies.

3. Income within Education: Income and Education responses are known to be well-related, and it is anticipated that category assignments of this variable will better capture the heterogeneity in responses related to payment choices.

4. Marital status within Region: Marital status and Region are both expected to capture well the heterogeneity in responses related to wealth and hence are nested to more efficiently account for disagreements between the observed sample distribution and the national-level counts.

5. Employment within Region: Employment and Region are both expected to capture well the heterogeneity in responses related to payment adoptions, and hence are nested to more efficiently account for disagreements between the observed sample distribution and the national-level counts.

7 Choice of Variable Combinations for Calibration

Four combinations of calibration variables are considered for the analysis. The combinations are listed below with a rationale for their choice. Essentially, each combination is chosen to serve a unique role in the analysis. As a whole, the four combinations are expected to make for a suitable set of candidate combinations, since they are based on intuitively appealing combinations, from an economics perspective, of the variables at both the nested and unnested levels.

1. C.All: Gender, Age, Region, Income, Education, Mobile, Online, Household size, Marital status, Employment, Home ownership; this combination uses all candidate calibration variables at the unnested level, and its performance in the testing and validation section will serve as a benchmark for the other combinations.

2. C.Nine: Gender, Age, Region, Income, Education, Household size, Marital status, Employment, Home ownership; this combination retains all candidate calibration variables except for Mobile and Online, and will assist in determining the added benefit of including the variables related to technology.

3. C.Econ: Mobile nested within Age, Online nested within Age, Income nested within Education, Gender, Region, Home ownership; this combination aims to make efficient use of a variety of combination variables through nesting.

4. C.Econ.Plus: Mobile nested within Age, Online nested within Age, Income nested within Education, Marital status nested within Region, Employment nested within Region, Gender, Home ownership; this combination aims to make efficient use of more variables than C.Econ through nesting.

Exploring the use of trimmed weights is a method supported by many practitioners of sample calibration[8] (Särndal, 2007). Estimates based on trimmed weights are less likely to

[8]Trimming in the R programming language is achieved by first setting a threshold value. Any extreme final weights are reset to the threshold value, and the mass associated with the extreme weights is allocated

give rise to an inflated sampling distribution corresponding with estimators. In this analysis, we opt to explore trimming the resulting distributions of the weights at five times their mean, as suggested by Battaglia et al. (2004) and DeBell and Krosnick (2009). We also explore trimming the resulting distributions of the weights at six times their mean; this option explores the sensitivity of estimators to trimming at a small scale and is a direct competitor to trimming at five times the mean. These combinations will be referred to as C.All.5, C.Nine.5, C.Econ.5, C.Econ.Plus.5, and C.All.6, C.Nine.6, C.Econ.6, C.Econ.Plus.6.

8 Initial Weights

The raking procedure requires a specification for a set of initial weights for the calibration algorithm. Quite often, the choice of the inverse of the sample inclusion probabilities[9] of the survey participants is tacitly assumed in the literature. With a large-scale survey such as the MOP, it is likely that these can only be approximated, since the invitees choose whether to participate.

We explore the potential utility of using approximated inclusion probabilities by comparing the final weights generated from two different sets of initial weights. The first is based on the assumption of a simple random-sampling design so that the inclusion probabilities are homogeneous, and the second is based on the assumption of a stratified random-sampling design based on the region, gender and age variables so that the inclusion probabilities are heterogeneous over the induced strata. With respect to the former case, initial weights are calculated so that each individual has a probability n/N of being sampled, where n is the sample size and N is the population size; the corresponding initial weights are taken to be N/n. With respect to the latter case, initial weights are calculated so that each individual has a probability n_h/N_h of being sampled, where n_h is the sample size of stratum h and N_h is the population size of stratum h; the corresponding initial weights are taken to be N_h/n_h.

proportionally over the final weights less than the threshold value.

[9] An individual's inclusion probability is defined as the probability of that individual being selected for the sample.

The distribution of the initial weights based on the heterogeneous assignments can be found in Figure 4: each bar corresponds with the individuals from a specific stratum.

Figure 4: Scaled initial weights for survey participants corresponding with the stratified sampling design based on a combination of region, gender and age demographics. The weights are scaled about the mean weight.

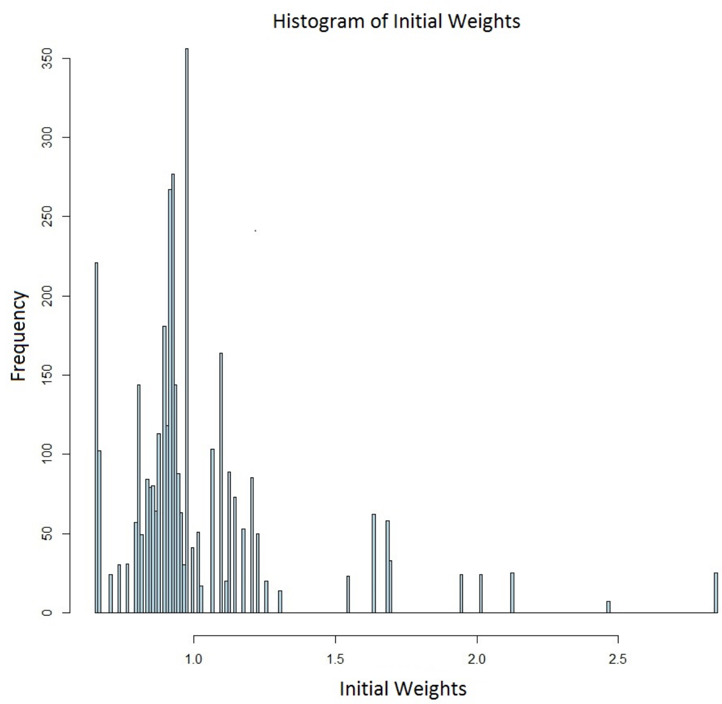
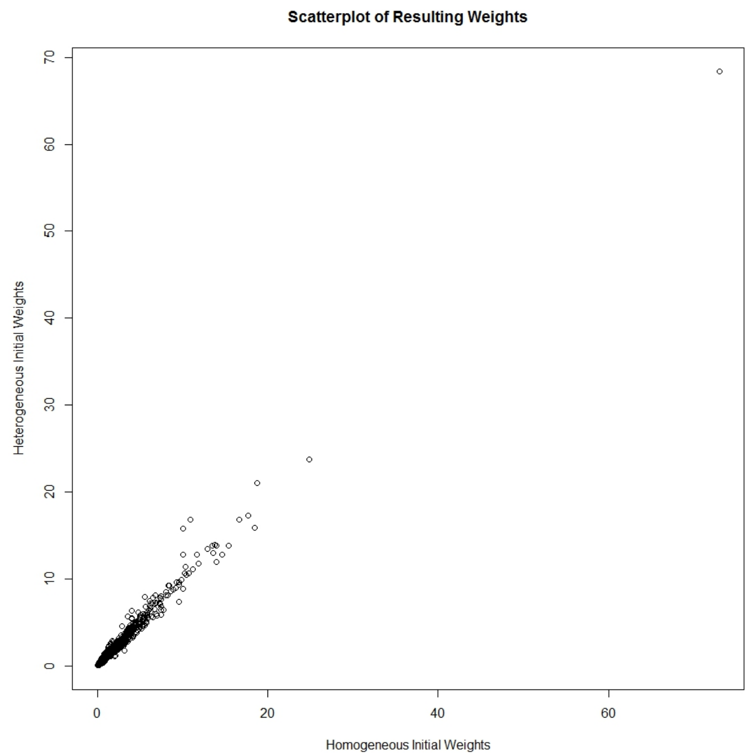


Figure 5 provides a scatterplot of the resulting weights standardized by their mean and based on the two assignments of initial weights and the raking calibration procedure where calibration is based on the C.All combination. The pattern of the resulting weights based on the other combinations is found to be similar. The correlation measure of the two generated sets of weights based on the C.All, C.Nine, C.Econ and C.Econ.Plus are evaluated at 0.991, 0.980, 0.990 and 0.990, respectively. Further, two-sample Kolmogorov-Smirnov tests report high p-values, indicating that the null hypothesis, which states that the resulting weights arise from the same distribution, should not be rejected. It is therefore decided, for simplicity, to base calibration on homogeneous initial sampling weights.

Figure 5: Scatterplot of standardized final weights of individuals based on homogeneous by heterogeneous initial sample seeds where the C.All combination of weights is used



9 Validation Analysis

To choose an appropriate combination of calibration variables, we use the criteria detailed below within each module. Since the survey team has a strong interest in the evolving usage of cash and its direct competitors, the criteria are predominantly based on measures associated with cash, debit and credit card variables:

1. Estimation ability; a set of criteria based on a mean-squared error (MSE)-type score,
2. Design/Misspecification effect; a set of criteria based on the design and misspecification

effects outlined by Lumley (2010) and (Skinner et al., 1989), where consideration is given to the mean and median of the scores, and

3. Distribution of resulting weights; a set of criteria based on dispersion measures, in particular the standard deviation and ratio of the maximum and minimum values of the resulting distribution of weights.

The combinations outlined in section 7 are assigned a score based on their performance under each criteria, upon which rankings based on each set of scores are used to determine a proposed set of final weights. The following three subsections detail the validation analyses regarding the use of estimation ability, design effect and distribution of resulting weights-based criteria. The following section summarizes the results and presents a proposed set of calibration variables.

9.1 Estimation ability

We use weighted point estimates provided by Ipsos Reid and based on the Canadian Financial Monitor from the final quarter of 2013, in particular those that provide a measure of the average amount of cash-on-hand per individual and the withdrawal frequency of cash from an automated banking machine (ABM). These values serve as proxies toward the true values, and we implement a mean-squared error type of criterion in that scores are based on the sum of the squared difference between the point estimates and the CFM-based estimate with the squared value of the standard error of the estimate. In a similar fashion, we implement a mean-squared error type of criterion based on estimates for other MOP SQ variables, namely:

1. The typical number of times an individual uses cash in a month,
2. The number of credit cards an individual has access to,
3. The typical number of times an individual uses the tap-and-go feature of a credit card,
4. The number of debit cards an individual has access to, and

5. The typical number of times an individual uses the tap-and-go feature of a debit card.

The point estimate provided by using the weights based on the C.All combination is taken to be a proxy for the true value. A rationale for using this value is provided as follows. We conjecture that, for the aforementioned responses, the more calibration variables that are used in the procedure the less biased the estimate (note that there would typically be a trade-off with the standard error).

We use this criterion since it will allow one to directly compare the efficiency, in terms of precision, of the estimators that each combination of calibration variables gives rise to.

Table 3 provides the MSE-type scores standardized by the lowest score for each of the seven variables. On average, it can be seen that C.Econ and C.Econ.Plus-based combinations score better than the C.All and C.Nine combinations, with C.Econ.Plus performing best. Further, scores based on the trimmed distribution with the C.Econ and C.Econ.Plus combinations appear to serve as a strong competitor to their untrimmed counterparts. Finally, it is interesting that using the Mobile and Online variables in the C.All combination gives rise to a better performance, on average, than the C.Nine combination.

Table 3: Standardized MSE-type scores

Combination	Cash-on-hand	ABM withdrawals	Cash usage	Number of credit cards	Tap-and-go credit use
C.All	1.000	1.257	10.440	23.717	10.669
C.All.5	1.543	1.868	4.433	22.690	5.964
C.All.6	1.444	1.735	5.016	24.044	5.933
C.Nine	2.321	2.670	8.308	7.319	6.709
C.Nine.5	2.674	3.340	3.942	5.405	4.031
C.Nine.6	2.608	3.189	4.016	4.134	4.166
C.Econ	1.257	2.022	1.283	1.479	1.230
C.Econ.5	1.507	2.254	1.000	1.952	1.000
C.Econ.6	1.476	2.156	1.062	1.823	1.057
C.Econ.Plus	1.385	1.102	1.609	1.000	1.124
C.Econ.Plus.5	1.769	1.002	1.038	1.446	1.036
C.Econ.Plus.6	1.714	1.000	1.173	1.326	1.092

Combination	Number of debit cards	Tap-and-go debit use
C.All	1.190	3.378
C.All.5	1.174	2.155
C.All.6	1.222	2.133
C.Nine	1.000	7.407
C.Nine.5	1.860	28.256
C.Nine.6	1.801	25.251
C.Econ	5.714	1.182
C.Econ.5	5.371	1.094
C.Econ.6	5.737	1.094
C.Econ.Plus	4.904	1.173
C.Econ.Plus.5	4.716	1.014
C.Econ.Plus.6	4.986	1.000

9.2 Design/Misspecification effect

The *design effect* (Kish, 1965; Park and Lee, 2001) is defined as the ratio of the variance of an estimate, typically obtained under a complex sampling design, to the variance of that estimate based on a simple random sample. For example, suppose $\hat{y}_{HT}$ is the Horvitz-Thompson estimator for the population mean based on an unequal probability sampling design. The design effect is then

$$\text{var}(\hat{y}_{HT})/\text{var}_0(\hat{y}_{HT}), \tag{1}$$

where $\text{var}_0(\cdot)$ refers to the variance of a statistic when simple random sampling is employed.

As the variance of an estimator will likely not be known in advance of an empirical study, so too will the design effect not be known. However, we can use an estimate of the *misspecification effect* to approximate the design effect. The misspecification effect is defined as the ratio of the variance of the complex estimator to the expectation of the estimated variance of the sample mean, $\bar{y}$, where data collection is based on the complicated sampling design (Skinner et al., 1989). With respect to the Horvitz-Thompson estimator, the misspecification effect is

$$\text{var}(\hat{y}_{HT})/E[\hat{\text{var}}(\bar{y})]. \tag{2}$$

An estimate of the misspecification effect can be made by replacing the numerator with an estimate of the variance of the complicated estimator and replacing the denominator with the estimated variance of the sample mean under a simple random-sampling design. The estimator is thus

$$\hat{\text{var}}(\hat{y}_{HT})/\hat{\text{var}}_0(\bar{y}). \tag{3}$$

Use of these criteria ultimately allows one to gauge and compare the quality of strategies (i.e., sampling design and calibration-based estimator) with that based on the simple random-sampling design and sample mean estimator. The criteria allow one to compare the variability of estimates based on calibration across the combinations.

We use estimated misspecification effect scores based on estimates for the following MOP SQ variables:

1. What was the total amount you charged to your main credit card last month?

   In a typical month, how often do you use each of the following methods of payment:

2. Cash,

3. Debit, and

4. Credit.

5. How much cash do you have in your purse, wallet, or pockets right now?

6. What is the total value of your household-based cash holdings?

Table 4 provides the mean and median statistics of the estimated misspecification scores over all six variables at their unstandardized and standardized values by the corresponding lowest score. Though there is no generally accepted rule of what misspecification effect to aim for when performing a calibration analysis, a value between 1.5 and 2 appears reasonable; Lumley (2010) reports values between 1.4 and 2 based on analyses for the California Health Interview Survey and mentions that "Design effects for large studies are usually greater than 1.0" [p. 6]. In this analysis, the C.Econ and C.Econ.Plus combinations consistently gave rise to values less than 2. Further, on average, the C.Econ and C.Econ.Plus combinations outperform the other combinations, indicating that they are more likely to give rise to efficient estimates. Notice the additional benefit in trimming the combinations, particularly at five times the mean.

Table 4: Unstandardized and standardized estimated misspecification effect mean and median scores

Combination	Mean	Median	Mean	Median
C.All	2.528	2.408	1.517	1.391
C.All.5	2.160	1.952	1.296	1.128
C.All.6	2.268	1.993	1.361	1.151
C.Nine	2.747	1.963	1.649	1.134
C.Nine.5	1.851	1.752	1.111	1.012
C.Nine.6	1.988	1.817	1.193	1.050
C.Econ	1.831	1.809	1.099	1.045
C.Econ.5	1.666	1.731	1.000	1.000
C.Econ.6	1.716	1.751	1.030	1.012
C.Econ.Plus	2.005	2.035	1.203	1.176
C.Econ.Plus.5	1.817	1.855	1.090	1.071
C.Econ.Plus.6	1.869	1.914	1.122	1.105

9.3 Distribution of resulting weights

We consider two dispersion measures of the resulting distributions of the final weights based on each combination: the standard deviation and ratio of the maximum and minimum values. We use these criteria since distributions of weights with smaller scores are desirable: they are more likely to give rise to stable estimates and standard errors.

A visual illustration of the distribution of the final weights can aid in comparing the combinations of calibration variables. Figure 6 provides histograms of the untrimmed weights standardized by their means based on each of the four calibration combinations. Especially evident is the long tail associated with each, which is more prominent when more calibration variables are used.

Figure 6: Histograms of final weights standardized by their means and corresponding with the four combinations of calibration variables. The vertical line shows the 95th percentile of the distribution. The corresponding weight associated with each individual can be interpreted loosely as the amount of weight they contribute to an estimator relative to other survey participants.

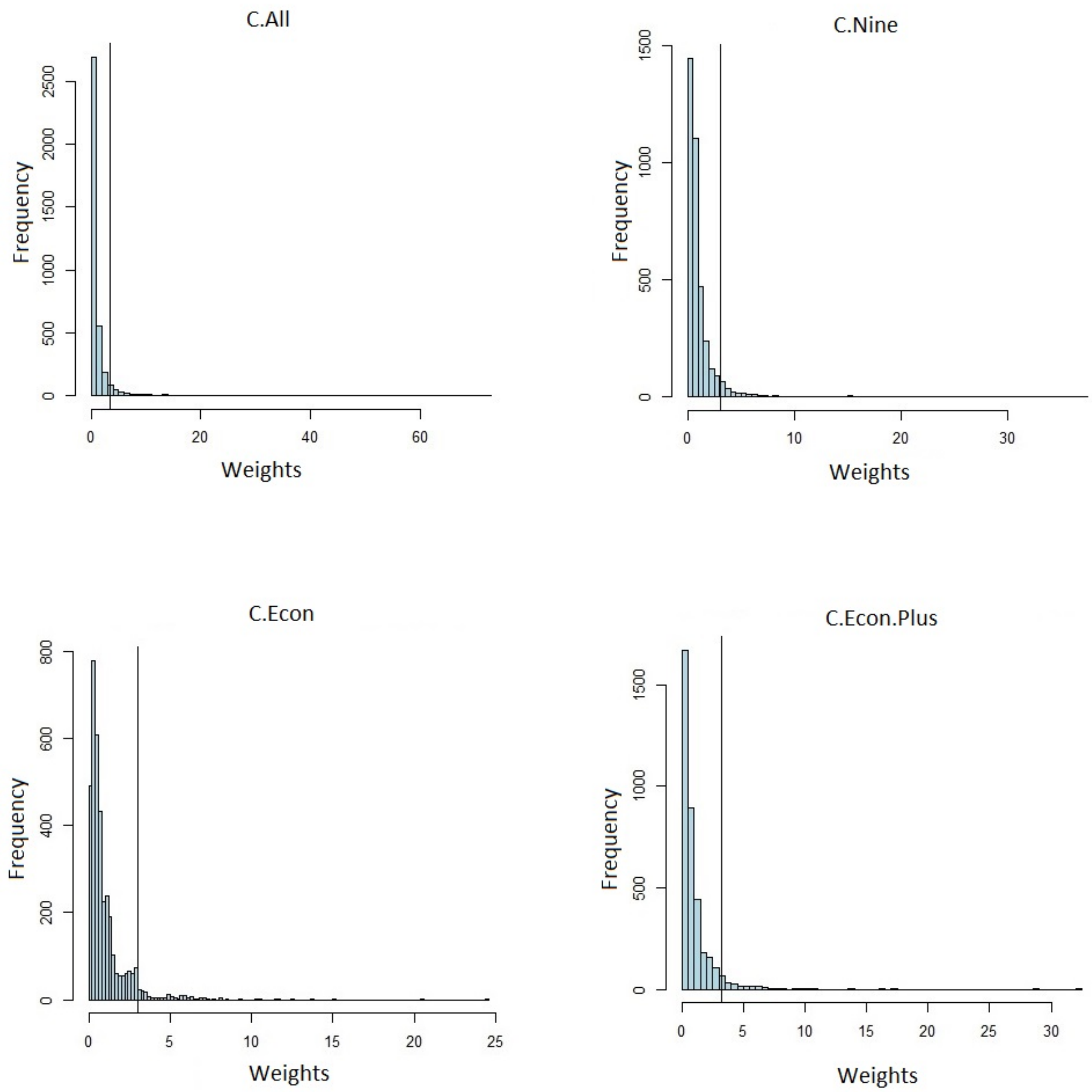


Table 5 provides the scores standardized by the corresponding lowest value. Clearly, trimming has given rise to distributions with dispersion measures much more reasonable than their untrimmed counterparts. Though neither combination performs uniformly better than any other, scores based on trimming at five times the mean are all approximately on the same order of magnitude, thus indicating that stability in resulting estimators can likely be achieved with any combination trimmed at this value.

Table 5: Standardized dispersion measures

Calibration variables	Standard deviation	Max/Min
All	2.030	93.944
All.5	1.088	1.000
All.6	1.182	1.432
Nine	1.394	25.885
Nine.5	1.000	1.509
Nine.6	1.058	2.134
Econ	1.375	18.447
Econ.5	1.035	1.514
Econ.6	1.108	2.272
Econ.Plus	1.493	45.621
Econ.Plus.5	1.065	1.665
Econ.Plus.6	1.144	2.612

10 Evaluation

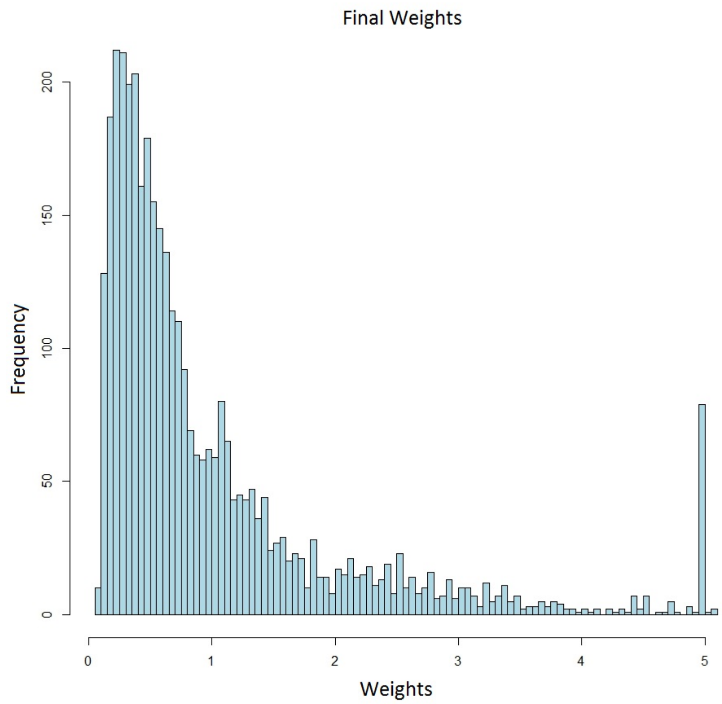
Table 6 provides the distribution of counts for each combination based on rankings when considering the criteria outlined in the previous section. It is evident that the C.Econ and C.Econ.Plus combinations consistently score higher, especially when based on trimmed values. Further, these combinations, when based on trimmed values, avoid low-ranking scores, thus indicating their strong and consistent performance.

Table 6: Count of ranking scores over all combinations. The vertical line partitions the scores in half.

Combination	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
C.All	1	0	1	1	0	0	0	0	1	0	2	5
C.All.5	1	1	0	0	1	1	1	2	2	2	0	0
C.All.6	0	1	0	2	1	0	1	1	1	3	0	1
C.Nine	1	0	0	0	0	0	0	0	2	5	2	1
C.Nine.5	1	0	2	0	1	1	2	1	0	0	0	3
C.Nine.6	0	0	1	0	2	1	2	2	0	0	3	0
C.Econ	0	1	0	3	1	2	1	0	2	0	1	0
C.Econ.5	4	1	0	2	0	2	0	0	1	1	0	0
C.Econ.6	0	2	3	0	2	1	1	1	0	0	0	1
C.Econ.Plus	1	0	2	0	2	1	0	2	0	0	3	0
C.Econ.Plus.5	0	4	2	1	1	1	1	0	1	0	0	0
C.Econ.Plus.6	2	1	0	2	0	1	2	2	1	0	0	0

We propose the use of C.Econ.Plus.5, since (1) most of its mass is concentrated in the top portion of the rankings, (2) it makes efficient use of the most calibration variables, and (3) it is trimmed at five times the mean, a conventional choice among practitioners (Battaglia et al., 2004; DeBell and Krosnick, 2009). A histogram of the standardized weights of C.Econ.Plus.5 can be found in Figure 7.

Figure 7: Histograms of proposed final weights, C.Econ.Plus.5, standardized by the mean weight



11 Discussion and Recommendation

We have outlined the statistical methods used to obtain the proposed final weights. Via statistical and visual comparisons, we also validate the proposed final weights of those based on the raking ratio procedure. The combination of calibration variables, after trimming at five times its mean weight, are: Mobile nested within Age, Online nested within Age, Income nested within Education, Marital Status nested within Region, Employment nested within Region, Gender, and Home Ownership. All computations performed in this analysis are achieved with the aid of the R "survey" package (Lumley, 2012, 2010). All analyses have been cross-checked in the Stata programming language; more details can be found in Chen and Shen (2015).

The topic of trimming was discussed extensively in this analysis, and future surveys may benefit from the corresponding observation. The majority of individuals assigned larger weights are found to be younger, with a lower education status and higher income. Therefore, future surveys based on similar sampling strategies and access panels may benefit from allocating additional effort to recruiting such individuals and/or making more stringent measurements regarding them, perhaps with suitable follow-up procedures.

Some limitations exist in that census-level financial data, such as those based on the Survey of Financial Security[10] (SFS) and the Survey of Household Spending[11] (SHS), are not available to assist in the analysis. Hence, we cannot provide as meaningful a comparison of calibration procedures with the NHS- and CIUS-based data available; we can only semi-compare the estimation ability of combinations for a small number of key demographics. Additional validation of the calibration procedure and proposed final weights should proceed using data based on surveys such as the SFS and SHS.

[10] For more information, see `http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2620`.

[11] For more information, see `http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3508`.

References

Angrisani, M., Foster, K., and Hitczenko, M. (2013). The 2010 survey of consumer payment choice: Technical appendix. Technical report, Federal Reserve Bank of Boston.

Battaglia, M. P., Izrael, D., Hoaglin, D. C., and Frankel, M. R. (2004). Tips and tricks for raking survey data (aka sample balancing). Technical report, Abt Associates.

Brick, J. M., Dipko, S., Presser, S., Tucker, C., and Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opinion Quarterly* **70,** 780–793.

Callegaro, M., Ayhan, O., Gabler, S., Haeder, S., and Villar, A. (2011). *Combining landline and mobile phone samples: a dual frame approach*, volume 2011/13.

Chen, H. and Shen, R. (2015). Variance estimation for survey-weighted data using bootstrap resampling methods: 2013 methods-of-payment survey questionnaire. Technical report, Bank of Canada, forthcoming.

D'Arrigo, J. and Skinner, C. J. (2010). Linearization variance estimation for generalized raking estimators in the presence of nonresponse. *Survey Methodology* **36,** 181–192.

DeBell, M. and Krosnick, J. (2009). Computing weights for American national election study survey data. ANES Technical Report Series nes01242, Stanford University.

Deville, J. C., Särndal, C. E., and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association* **88,** 1013–1020.

Drasgow, F. (2004). *Polychoric and Polyserial Correlations*. John Wiley and Sons, Inc.

Fox, J. (2010). *polycor: Polychoric and Polyserial Correlations*. R package version 0.7-8.

Henry, C., Huynh, K., and Shen, Q. R. (2015). 2013 methods-of-payment survey results. Bank of Canada Discussion Paper No. 2015-4.

Henry, C. and Vincent, K. (2014). A Bayesian data augmentation procedure for missing data analysis: An application to the Canadian Financial Monitor Survey. Draft, Bank of Canada.

Jones, M. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association* **91,** 222–230.

Kish, L. (1965). *Survey Sampling.* John Wiley & Sons, Ltd., New York.

Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistics* **8,** 183–200.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Wiley Series in Probability and Statistics. Wiley, New York, 2nd edition.

Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R.* John Wiley & Sons, Ltd.

Lumley, T. (2012). *survey: analysis of complex survey samples.* R package version 3.28.2.

Park, I. and Lee, H. (2001). The design effect: Do we know all about it? In *Proceedings of the Annual Meeting of the American Statistical Association.*

Sanches, M. (2010). 2009 method of payment survey weighting manual. Bank of Canada.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* **33,** 99–119.

Shen, R. (2014). Logistic regression on invite list data.

Skinner, C. J., Holt, D., and Smith, T. M., editors (1989). *Analysis of Complex Surveys.* Wiley, New York.

Tao, Y. and Vincent, K. (2013). Multiple imputation of section 2.1 of CFM - cash purchases. Draft, Bank of Canada.

Tao, Y. and Vincent, K. (2014). Multiple imputation of section 2.1 of CFM - cash holdings. Draft, Bank of Canada.

van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman and Hall/CRC Press.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software* **45,** 1–67.

Appendix: Approximated CIUS and MOP Sample Distributions

National counts are based on residents of Canada aged 18 years or older, excluding: residents of the Yukon, Northwest Territories and Nunavut, inmates of institutions, persons living on Indian reserves, and full-time members of the Canadian Armed Forces. The national counts are based on the 2012 Canadian Internet Use Survey (CIUS) that have been weighted to the 2011 National Household Survey (NHS). According to the 2012 CIUS, the total size of this population is 28,057,000.

The online sample corresponds with those individuals recruited from Ipsos Reid's online panel. The offline sample corresponds with those individuals recruited either directly from Ipsos Reid's offline panel or via subsampling the Canadian Financial Monitor.

Due to the sensitivity of the data obtained for the 2013 Methods-of-Payment survey, a privacy clause exists between the Bank of Canada and Ipsos Reid. To preserve the confidentiality of the data, the counts presented in the tables of the appendix are coded and based on the following legend. For a description of the variables see section 3.

Legend:

+: indicates that the corresponding cell count of the sample is within 2.5 per cent of the CIUS cell count,

++: indicates that the corresponding cell count of the sample is between 2.5 per cent and 7.5 per cent, inclusive, of the CIUS cell count, and

+++: indicates that the corresponding cell count of the sample is outside 7.5 per cent of the CIUS cell count.

Table 7: Gender

	CIUS	Online	Offline
Male	0.49	+	+
Female	0.51	+	+

Table 8: Age

	CIUS	Online	Offline
18-24	0.15	++	++
25-34	0.17	+	+
35-44	0.16	+	+
45-54	0.19	+	++
55-64	0.16	+	++
65 or more	0.18	+	+

Table 9: Region

	CIUS	Online	Offline
Atlantic	0.07	+	+
Quebec	0.23	++	++
Ontario	0.39	++	+
Prairies	0.17	+	+
British Columbia	0.13	+	+

Table 10: Income

	CIUS	Online	Offline
<25k	0.13	++	++
25k-45k	0.18	+++	++
45k-65k	0.18	++	+
65k-85k	0.15	+	+
85k or more	0.36	+++	+++

Table 11: Education

	CIUS	Online	Offline
Some/completed public school, some high school	0.18	+++	+++
Completed high school	0.19	+	+
Some/completed technical school, community college, or university	0.39	++	+++
Completed bachelor's degree	0.17	+++	+++
Completed graduate degree	0.07	+	++

Table 12: Mobile

	CIUS	Online	Offline
Yes	0.48	++	+
No	0.52	++	+

Table 13: Online

	CIUS	Online	Offline
Yes	0.46	+++	+++
No	0.54	+++	+++

Table 14: Household size

	CIUS	Online	Offline
1	0.14	+++	+++
2	0.35	++	+
3	0.20	++	++
4	0.19	+++	++
5 or more	0.12	++	++

Table 15: Marital status*

	NHS	Online	Offline
Single	0.33	+	+
Married or Common-Law	0.50	+	+
Separated or Divorced	0.12	+	+
Widowed	0.05	+	+

*Note that the Marital Status national counts are based on the 2011 NHS.

Table 16: Employment

	CIUS	Online	Offline
Employed	0.64	+++	++
Unemployed	0.05	+	+
Not in labour force	0.31	+++	++

Table 17: Home ownership

	CIUS	Online	Offline
Own	0.73	+++	+++
Do not own	0.27	+++	+++