BANK OF CANADA
BANQUE DU CANADA

Technical Report No. 114 / Rapport technique nº 114

2017 Methods-of-Payment Survey: Sample Calibration and Variance Estimation

by Heng Chen, Marie-Hélène Felt and Christopher S. Henry

www.bank-banque-canada.ca

December 2018

2017 Methods-of-Payment Survey: Sample Calibration and Variance Estimation

Heng Chen, Marie-Hélène Felt and Christopher S. Henry

Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
hchen@bankofcanada.ca
mfelt@bankofcanada.ca
chenry@bankofcanada.ca

ISSN 1919-689X

Acknowledgements

We thank Kim P. Huynh, Jean-François Beaumont, Jack Gambino and participants in the 2018 Bank of Canada–Statistics Canada learning exchange for their comments and suggestions. We thank Gradon Nicholls for great research assistance. We acknowledge Shelley Edwards, Lisa Canini and the Ipsos Reid team for their superb collaboration on this project.

Abstract

This technical report describes sampling, weighting and variance estimation for the Bank of Canada's 2017 Methods-of-Payment Survey. Under quota sampling, a raking ratio method is implemented to generate weights with both post-stratification and nonparametric nonresponse weight adjustments. In the end, we estimate variances of weighted means and proportions using bootstrap replicate survey weights. Compared with probability sampling, we find that (i) strong assumptions are required to reduce bias when probabilities of selection are unknown, and (ii) multiple weight adjustments for bias reduction inflate variance. Therefore, it is important to focus more on bias than on variance in the context of nonprobability sampling.

Bank topic: Econometric and statistical methods
JEL codes: C81, C83

Résumé

Le présent rapport technique décrit les méthodes d'échantillonnage, de pondération et d'estimation de la variance qui ont été appliquées à l'enquête de la Banque du Canada sur les modes de paiement menée en 2017. Dans le cadre d'un échantillonnage non probabiliste reposant sur des quotas, nous mettons en œuvre la méthode itérative du quotient pour obtenir les poids de l'échantillon, en l'appliquant aux poids stratifiés a posteriori et corrigés non-paramétriquement pour la non-réponse. Pour finir, nous estimons la variance des moyennes et des proportions pondérées à l'aide d'une méthode de rééchantillonnage de type bootstrap. Comparativement à un échantillonnage probabiliste, nous constatons que (i) de solides hypothèses sont nécessaires pour atténuer le biais lorsque les probabilités de sélection sont inconnues et (ii) les multiples ajustements de poids réalisés pour atténuer le biais font augmenter la variance. Par conséquent, dans le cadre d'un échantillonnage non probabiliste, il vaut mieux se concentrer sur le biais que sur la variance.

Sujet : Méthodes économétriques et statistiques
Codes JEL : C81, C83

1 Introduction and scope

The Bank of Canada 2017 Methods-of-Payment (MOP) Survey is a follow-up of the 2009 and 2013 MOP surveys; see Arango and Welte (2012) and Henry et al. (2015). The main purpose of these surveys is to understand and monitor Canadian adults' demand for and attitudes toward cash and other payment instruments, as well as their adoption of payment innovations, such as the contactless feature of credit cards.

This report is a technical companion to the main survey results in Henry et al. (2018). It describes sampling, weighting and variance estimation for the 2017 MOP Survey. The methodologies employed are consistent with the 2009 and 2013 MOP surveys to ensure the comparability of the results: the sampling design is quota sampling, trimmed raking ratio weights are generated as in Vincent (2015), and variances are estimated using bootstrap replicate survey weights as in Chen and Shen (2015). See also Chen et al. (2016) for a summary of technical details of the 2013 MOP Survey.

Nevertheless, there are improvements upon previous iterations of the survey. For the sampling design, specific strategies are deployed to tackle declining response rates from 2013 to 2017, which include better incentivizing hard-to-reach strata and a front-loaded rollout plan.

For sample weighting, we explore the option of incorporating nonresponse adjustment into the raking ratio method. Furthermore, we present an extensive analysis of using different initial weights, where we evaluate the effectiveness of various raking alternatives for our nonprobability sample. Based on this analysis, we prefer a raking method with trimming using nonparametrically nonresponse-adjusted post-stratified initial weights.

In the end, we compute variance estimates for various raking approaches, and we find that the fewer the number of weight adjustments, the smaller the variance estimates. However, we caution that, compared with raking without either post-stratification or nonresponse adjustment, larger variance from raking with both post-stratification and nonresponse weight adjustment should not be interpreted as precision deterioration. Thus, in the case of a nonprobability sample, it is the bias that we should evaluate among various raking alternatives.

Figure 1 provides a visual summary of the sample collection, survey weight and variance estimation undertaken for this study. The remainder of this report is organized as follows. Section 2 describes the sampling strategy and the particular challenges we faced for sample collection. In Section 3 we provide the raked weights for our nonprobability sample. Section 4 discusses variance estimation of weighted means and proportions. Finally, Section 5 concludes.

2 Sample collection

2.1 Summary of sampling plan

Prior to conducting fieldwork, we developed a detailed plan to specify how the sampling and data collection would be conducted. Full details of the sampling plan can be found in Chen et al. (2017c); however, parts of that document contain the proprietary data owned by Ipsos. Hence, we give here a summary of the key components of the sampling plan and include only relevant statistics related to the sampling design.

2.1.1 Sample size

Sampling was conducted on three separate survey frames, which we refer to as "panels": the Online panel, the Offline panel and the CFM panel.[1] Data collection for the Online panel was conducted with an Internet-based survey instrument and email invitations, whereas for the other two panels data collection was conducted with paper-based survey instruments and mail-out invitations. This methodology is consistent with the sampling for the 2013 MOP Survey.

[1] The CFM panel is a sample from the Offline panel of individuals who recently responded to the Canadian Financial Monitor (CFM); it serves as a third frame. We treat these three panels as non-overlapping Canadian population frames, although they might have coverage errors due to the panel recruiting.

The desired sample sizes were calculated for each of the three panels. Sample size calculations were based on obtaining a sufficiently low coefficient of variation (10 per cent) for each panel when estimating the mean amount of cash holdings using the 2013 MOP data; see Table 1.

Note that the Online panel actually required a sample size of $n = 1,004$ to achieve the desired level of precision. However, in the end we set a target sample size of $n = 500$ for each panel, resulting in a total sample size of $n = 1,500$. This decision was based on cost considerations and the desire to have a balanced sample across the three panels, as well as the fact that we anticipated the possibility of boosting the sample using the Online frame during fieldwork; see the discussion below in Section 2.2.2.

2.1.2 Stratified quotas

For each panel the target sample size of $n = 500$ was further divided into stratified quotas, which were nested by region (Atlantic, Quebec, Ontario, Prairies and British Columbia), gender (male and female), and age (18–24, 25–34, 35–44, 45–54, 55–64, 65+) in that order; see Table 2. These quotas were designed to match the Canadian population totals from the 2016 Census.[2]

The survey company Ipsos shared with us the corresponding nested counts by region, gender and age for each of the three survey frames (Online, Offline and CFM). Using the 2013 MOP Survey response rates, we were therefore able to calculate the required number of invitations needed to meet the nested sampling targets, under the assumption that response rates to the 2017 MOP Survey would be similar to those in 2013.[3]

[2] We apply proportional sample size allocation. Alternatively, a straightforward square-root allocation could reduce the risk of having small cells and improve the precision of strata estimates. We thank Jack Gambino for pointing this out.

[3] In fact, our assumption was that survey response rates had declined over time; Ipsos advised us that that was a general trend across the industry. Therefore, to offset this decline we increased the level of monetary incentives for Online respondents to \$20 and offered the full range of non-pecuniary incentives (see Section 2.1.5) to all respondents. Online respondents in 2013 were randomly offered either \$5 or \$10, and non-pecuniary incentives were randomly offered to half the respondents in all three panels. The point of randomization in 2013 was to test which incentives were most effective. Offline and CFM respondents were offered \$20 in both 2013 and 2017.

2.1.3 Hard-to-reach strata

Based on the invitation list, we identified certain strata for which it would not be possible to meet the stratified quotas due to a lack of available respondents in the frame (called *deficient cells*), as well as strata for which even a small decline in response rates would result in not meeting the stratified quotas (called *risky cells*). In order to hit the stratified quotas for these deficient and risky cells, Ipsos used third-party vendor online frames to sample additional respondents. Both the Ipsos online panel and the third-party vendor online frames are aggregated together and treated as our Online panel; see "Multiple frames" in Section 3.2 for details.

2.1.4 Survey rollout plan

The distribution of survey invitations depended on the survey mode. For the Online panel, rolling invitations were sent out on a weekly basis, and based on real-time response rates, adjustments were made to calibrate the number of invitations needed to meet the stratified quotas. For the paper-based mode, invitations were mailed out in three waves. To mitigate risk, the invitations were front-loaded so that we would receive a bulk of the returns and be able to estimate strata-based response rates; see Table 3. This flexibility would allow us to send out an additional wave of invitations if the response rates were not in line with our initial predictions.

For both online and paper-based methodologies, the Ipsos proprietary survey frame was to be exhausted before utilizing vendor frames to compensate for deficient or risky cells. [4]

2.1.5 Incentives

Respondents were offered the following pecuniary and non-pecuniary incentives to complete the survey:

[4] In the end, 205 of the 1,889 online respondents were recruited from third-party vendor frames. All of them belong to the stratum of young (18–24 years old) males.

1. A letter from Bank of Canada Governor Stephen S. Poloz inviting respondents to complete the survey and explaining its importance for work at the Bank of Canada. The letter served to notify respondents in advance that the survey was to follow shortly (whether in the mail or by email) and to appeal to Canadians' sense of civic responsibility.

2. An accompanying letter contained in the survey package from the Managing Director of the Bank of Canada Currency Department, Richard Wall, thanking them in advance for filling out the survey and reminding them of its importance.

3. A reminder postcard, which followed receipt of the survey package.

4. A $20 reward for completing both the survey questionnaire (SQ) and the diary survey instrument (DSI).

5. For hard-to-reach cells, an additional $20 was offered.

Such incentive mechanisms follow Dillman's (2000) survey design principles and also incorporate findings from Shen and Vincent's (2014) analysis of the 2013 MOP Survey's randomized incentive scheme to determine which incentives were most effective.

2.2 Fieldwork

2.2.1 Monitoring process

The first wave of invitations was sent out on October 13, 2017. Once the survey was in the field, we received weekly updates from the survey company regarding the number of completions in each stratum, along with the response probabilities. These probabilities were used to predict the final number of returns in each stratum as the data collection proceeded.

2.2.2 Boost sample

On November 13, 2017, we met with Ipsos to assess the state of returns and discuss the possibility of conducting a boost sample. At this juncture we had received the bulk of returns from Wave 1 and Wave 2 along with early returns from Wave 3, and therefore felt confident projecting the final number of returns by stratum. There were two reasons to send out additional invitations beyond what was laid out in the original sampling plan: (i) certain strata had lower-than-expected response rates, and more invitations would be required to meet the stratified quotas; (ii) we had budget available to increase the number of Online respondents, which was a cost-effective way to meet the sampling targets for hard-to-reach strata.[5]

To avoid having zero returns for the Offline and CFM panels for certain strata, we also sent out a small number of ad hoc invitations. These invitations were sent directly to people in the social network of the survey team.

2.3 Description of the final sample

Data collection proceeded until the end of November 2017. By mid-December only a small number of returns were trickling in, and we officially ended the data collection. Table 4 shows the number of final returns broken down by panel. Note that the Online and CFM panels contain more than the planned 500 respondents each. This is mainly due to the boost sample described in Section 2.2.2. In addition, means and variances of the variable of interest *cash on hand* are reported across different frames. Compared with the 2013 results—reported in Table 1 from Chen et al. (2016)—the 2017 MOP variance estimates calculated using the bootstrap replicate survey weights method (VarBSRW) are larger, which is mainly due to extra nonresponse adjustment in the raking procedure; see Section 4 for details. Note that for online respondents, there is roughly a 1.5:1 ratio between the numbers of SQ and DSI

[5]Paper-based methods incur costs that are not incurred for web-based methods, the most significant of which is data capture costs, i.e., translating the data on paper returns into an electronic format.

completed. This is because online respondents might fill out the SQ but not the DSI, whereas paper-based returns were considered complete only if the returned package contained both the SQ and the DSI.[6]

Table 5 shows unit response rates for the 2017 and 2013 MOP surveys, overall and by demographic categories, for the Offline/CFM and Online panels, respectively. They are computed based on all individuals invited (the information contained in the invitation list consists of basic demographic variables as well as participation in past surveys). Although the total response rate of the merged Offline/CFM panels has declined from 40.5 per cent to 32.4 per cent, the overall rates for the combined three frames (Offline, CFM and Online) are very similar across the two survey iterations. This is driven mainly by increased response rates from the Online frame, which are likely due to better incentivization for online participants.

2.4 Data edit and imputation

To evaluate the data quality and implement survey weighting, it is crucial to investigate item nonresponse of calibration variables. Ideally, calibration requires respondents to provide complete calibration variables, so that all respondents' information is used. Item nonresponse for the calibration variables in the final MOP data set is very low; see Table 6. This is because, for most individuals, frame data with demographic information are available. Hence, basic demographic variables are imputed (backfilled) using frame information.

3 Survey weighting

Sample weights serve two purposes: weights can increase precision (Deville et al., 1993), and in cases in which the sample is not representative of the target population, they shift the sample distribution toward the target population distribution (Hellerstein and Imbens, 1999).

[6] In either case, respondents were compensated only if they returned both the SQ and the DSI.

Weights are usually created through three general methods: post-stratification, generalized regression estimation (GREG) and raking. However, the post-stratification method can result in unstable weights from adjusting a multi-way table, while the GREG requires extra modification to ensure positive weights (Huang and Fuller, 1978; Rao and Singh, 2009). In contrast, the raking method adjusts only marginals, or low-level interacted cells, and always guarantees positive weights. Thus the raking ratio method is chosen for the 2017 MOP Survey, similar to what was done for the 2009 and 2013 MOP surveys.

The raking ratio procedure is also known as iterative proportional fitting (IPF), or simply raking. For example, consider the estimation of a population total T_Y of a survey variable Y taking values y_i for units i in a population U:

$$T_Y = \sum_U y_i, \tag{1}$$

with three sets of post-strata to be used for raking. Let x_i denote the vector of indicator variables for these categories:

$$x_i = \left(\delta_{1..i}, \cdots, \delta_{A..i}, \delta_{.1.i}, \cdots, \delta_{.B.i}, \delta_{..1i}, \cdots, \delta_{..Ci}\right)^\top, \tag{2}$$

where $\delta_{a..i} = 1$ if unit i is in category a of the first auxiliary variable and 0 otherwise, $\delta_{.b.i} = 1$ if unit i is in category b of the second auxiliary variable and 0 otherwise, and so on. The population total T_X of this vector thus contains the population counts in each of the (marginal) categories for each of the three auxiliary variables. It is assumed that T_X is given and that x_i is known for unit i in sample s. The raking adjustment involves iterative modifications of initial weights ω_i to adjusted weights w_i so that:

$$\sum_s w_i x_i = T_X. \tag{3}$$

The resulting raking estimator of T_Y is

$$\widehat{T}_Y = \sum_s w_i y_i. \tag{4}$$

The adjustment depends only upon the cell in the contingency table formed by the auxiliary variables, that is $w_i = \omega_i h(x_i)$, where the multiplicative adjustment factor $h(x_i)$ is fixed for all units with common values of the auxiliary variables. Let $\widehat{N}_\omega [h(x)]$ and $\widehat{N}_w [h(x)]$ denote the weighted estimates of the population counts in the cell of the table defined by x using the weights ω_i and w_i, respectively. Then we may write $\widehat{N}_w [h(x)] = h(x)\widehat{N}_\omega [h(x)]$, where IPF makes use of standard post-stratification. Ireland and Kullback (1968) demonstrate that this method converges to a solution that minimizes

$$\sum \widehat{N}_w \log\left(\widehat{N}_w / \widehat{N}_\omega\right), \tag{5}$$

subject to the calibration equations, when the sum is over all cells defined by x. This objective function may alternatively be expressed as

$$\sum_s w_i \log(w_i/\omega_i), \tag{6}$$

that is, under convergence of the iterative algorithm, w_i minimizes the above function, subject to the calibration equations.

3.1 Raking for nonprobability samples

To better understand the differences between probability and nonprobability sampling, note that probability sampling has a sampling frame linked to the target population, with every sampled unit having a known probability of being selected, and design-based theory is used for statistical inference; in contrast, nonprobability sampling does not have a target population sampling frame, and selection probabilities are unknown, so inference relies heavily on model-based assumptions.

According to Baker et al. (2013b), there are three types of nonprobability sampling:

(i) *Convenience sampling*: Participants are recruited because they are easy to reach. It includes mall intercepts, volunteer samples, river samples, observational studies, snowball samples.

(ii) *Quota matching*: Members of the sample are selected to match a set of important population characteristics.

(iii) *Network sampling*: Members of some population are asked to identify other members of the population with whom they are somehow connected. It includes response-driven sampling.

As discussed in Section 2, our 2017 MOP sampling belongs to the quota-matching type.

In the context of nonprobability sampling, there are three approaches to address this sample-selection issue. One requires modelling the outcome regression under the superpopulation set-up, while the other two are based on quasi-randomization following Rosenbaum and Rubin (1983).[7]

(i) *Regression-based approach* (Beaumont et al., 2018): This approach usually assumes that a common model can be used to predict the values of y for both probability and nonprobability samples. It imputes the unobserved y_i in the probability sample from the nonprobability one, and computes $\widehat{T}_Y = \sum_{ps} w_i \widehat{y}_i$, where ps denotes the probability sample, based on probability sample weights w_i and imputed $\widehat{y}_i$ values. However, diagnosing whether a model holds for both samples may be difficult or impossible.

(ii) *Pseudo-weights approach* (Elliott and Valliant, 2017): Since a nonprobability sample is not selected randomly from an explicit sampling frame, selection probabilities cannot be obtained directly. This approach combines the nonprobability sample with a

[7]It is also possible to have a double robust estimator by combing these two approaches, such that only one of the two need be correctly specified to obtain an unbiased estimator (Bang and Robins, 2005).

probability sample to estimate the propensity of being a nonprobability survey respondent. Thus it imputes the unobserved w_i for the nonprobability sample, and computes $\widehat{T}_Y = \sum_{nps} \widehat{w}_i y_i$, where $\widehat{w}_i$ is the imputed probability sample weight and *nps* refers to the nonprobability sample. Similar to the regression-based approach, the pseudo-weights approach requires the existence of a probability sample. Note that the probability sample must not be subject to coverage or other types of bias. However, many probability samples are now subject to high nonresponse rates and are tantamount to nonprobability samples themselves. In addition, both regression-based and pseudo-weights approaches need a comprehensive variable list common to both samples to validate the imputation assumption.

(iii) *Raking approach* (Elliott and Valliant, 2017): This method, similar to the raking procedure for probability samples, is based on implied weights from the empirical-likelihood method. Let $f_{nps}(y, x)$ and $f_U(y, x)$ be the sampled population density (associated with the nonprobability sample) and target population density, respectively. Borrowing the terminology of the pseudo-true value (White, 1982), let T_g be the population total given a sample drawn randomly from the population with probability density function $f_g(y, x)$. Following Hellerstein and Imbens (1999) and Nevo (2002), we have

$$\widehat{T}_Y \equiv \sum_{nps} w_i y_i \xrightarrow{p} T_g, \tag{7}$$

where

$$f_g(y_i, x_i) \equiv f_s(y_i, x_i) \cdot w_i,$$

and w_i is the implied weights from raking the nonprobability sample *nps*. They further show that T_g is generally different from T_U. Unlike the regression-based and pseudo-weights approaches, the raking approach does not need a probability sample.[8]

[8] Lee and Valliant (2009) discuss a combination of pseudo-weights and raking and show that this hybrid method reduces bias in volunteer panel web surveys.

Since we do not have a probability sample available, we implement the raking approach. Yeager et al. (2011), Tourangeau et al. (2013) and Wang et al. (2015) show that raking can remove or reduce selection bias in nonprobability Internet surveys.

In addition to unknown selection probabilities due to nonprobability sampling, nonresponse behaviour is also unknown and needs to be modelled. The unknown sampling design and unknown response behaviour create double selection, as discussed in Chen et al. (2017a), where they apply a pseudo-weights approach to account for the unknown sampling design.

In Table 5, we can observe that response rates vary substantially across demographic categories. Also, response rates are much smaller in the Online panel than in the merged Offline/CFM panels. This motivates the need for nonresponse adjustments. Note that, though very common in the context of probability samples, nonresponse adjustment is seldom implemented with nonprobability samples because of the absence of the information on nonrespondents (such as our invitation list).

Four different raking procedures are considered, which differ with respect to their initial weights:

(i) Rk,1: raking procedure with $\omega_i = 1$ (Elliott and Valliant, 2017);

(ii) Rk,PS: raking procedure with ω_i equal to post-stratified (PS) weights. Following Chen and Shen (2015), we use post-strata that are identical to the strata used for quota stratification, defined by gender, age and region. PS weights are taken to be N_h/n_h, for each unit in stratum h, where n_h and N_h are respectively the sample and population sizes of stratum h. Population sizes are taken from the 2016 Canadian Census.

(iii) Rk,NR$^{\mathrm{P}}$: raking procedure with ω_i equal to nonresponse-adjusted PS weights, where nonresponse adjustment is based on individual predicted probabilities $\hat{p}_i$ from a parametric nonresponse model. That is,

$$\omega_i^{\text{NR-P}} \equiv \omega_i^{\text{PS}} f_i^{\text{NR-P}}, \tag{8}$$

where $f_i^{\text{NR-P}}$ is the parametric nonresponse adjustment factor defined as the inverse of the predicted response propensity. Table 7 presents the estimation results for our response propensity model estimated for the Offline/CFM and Online frames.[9] Allowing for different response behaviour between offline and online improves the model fit and so captures the response heterogeneity.

(iv) Rk,NR$^{\text{NP}}$: raking procedure with ω_i equal to nonresponse-adjusted PS weights, where nonresponse adjustment is based on nonparametric methods. We first group the individual predicted response propensities $\hat{p}_i$ into homogeneous classes, and then calculate the nonparametric nonresponse adjustment factor $f_i^{\text{NR-NP}}$ for respondent i in a given class as the inverse of the mean response rate within that class; see Little (1986) and Haziza and Lesage (2016). For a given number of classes, grouping is performed using the k-means classification algorithm. To determine the number of classes to use, we follow Haziza and Beaumont (2007) and rely on the coefficient of determination criterion. Figure 2 shows the squared coefficient of correlation R^2 (resulting from an analysis of variance between $\hat{p}_i$ and the class identifier variable) for a number of classes from 4 to 20. The Rk,NR$^{\text{NP}}$ weights are obtained with 8 classes, which is the smallest number of classes that brings a value of R^2 above 97.5 per cent.

3.2 Set-up for raking

The raking parameters selected by Vincent (2015) and Chen and Shen (2015) for the 2013 MOP Survey serve as a benchmark for the 2017 MOP Survey raking analysis.

Multiple frames

For the 2013 MOP Survey, Chen and Shen (2015) show that, given the sampling design based on multiple frames, it was preferable to perform calibration on the full combined data

[9]In column 3 of Table 7, we also compute the online SQ unit nonresponse using the sequential logistic model, and their marginal effects are similar. Note that the $20 reward was only for respondents who completed both the SQ and the DSI, but some SQ participants did not fill out the DSI.

set instead of on the three subsamples separately. People have advocated using blended panels in which the same survey is administered to members of several different frames so that the results are aggregated to protect against unusual results. In fact, the single most important characteristic for an unbiased sample is that it reflects the heterogeneity of the target population. Thus, increasing heterogeneity by blending samples from many different sources should improve sample quality. Therefore, we merge the Offline, CFM, and Online subsamples and then rake. As in 2013, we also implement the Epps-Singleton test for comparing distributions across subsamples. In Table 8, differences in the distributions of the Online and Offline/CFM subsamples are statistically significant for most of the raked variables as well as *day of week* (the first day of the three-day DSI), whereas there are no significance differences between the Offline and CFM subsamples. Therefore, the benefit of merging Online and Offline/CFM samples is to improve the coverage and reduce the potential bias, while that of merging the Offline and CFM samples is to increase the sample size and reduce the variance.

We also compare online respondents recruited from third-party vendor frames to the respondents from Ipsos's online panel. Detailed results are not reported, but we summarize here our main findings. The response rates, demographic profiles and (unweighted) *cash on hand* and *CTC usage* of the respondents are considered, and we observe differences across both sources in all these aspects. In that respect, merging the vendor and Ipsos online respondents should also ensure a better coverage of the target population, and in particular young males. To assess the influence of third-party vendor respondents on our final results, we also rake the sample without using the vendor online respondents. We find that weighted estimates with and without respondents from the vendor frames are very similar, but including those respondents helps avoid empty or sparse strata so that post-stratification adjustment does not generate extreme weights for these strata. Therefore, using respondents recruited from third-party vendor frames also helps reduce the variance of estimates and improve the precision of small-area estimation.

Raking variables

Calibration variables are chosen based on two criteria: (i) the availability of their corresponding national-level total counts, and (ii) the strength of their association with key MOP variables, such as *cash on hand*.

For the 2013 MOP Survey, national counts were obtained from the 2012 Canadian Internet Use Survey (CIUS). Empirical and theoretical rationales for the choice of raking variables, as well as nesting, are provided in Vincent (2015). Therefore, the 2013 MOP Survey was raked on the following combination of variables: marital status nested within region; age category nested within mobile phone ownership; age category nested within online purchase; income category nested within education, gender, home ownership; and employment status nested within region.

To calibrate the 2017 sample weights, we take advantage of the availability of the Canadian 2016 Census data, as well as the 2016 Survey of Household Spending (SHS).[10] The former provides marginal and nested distributions of core demographic variables (gender, age, region, income, education, household size, marital status, employment and home ownership), while the latter provides information on Internet access from home.[11] Thus, the 2017 MOP raking variables are gender, age, home ownership, household size, marital status nested within region, income category nested within education, employment status nested within region, and Internet access from home.

Trimming

To avoid extreme weights, we trim weights at five times their mean, which was also used in Vincent (2015) for the 2013 MOP Survey. Figure 3 depicts the change in the distribution of raked weights induced by trimming at five times the mean. Although trimming at five times the mean is used as a benchmark, we also compare with results obtained using untrimmed

[10]The most recent CIUS was conducted in 2012; therefore, it cannot be used to calibrate the 2017 MOP Survey sample.

[11]The variable *Internet access from home* replaces *online payment* and *mobile ownership*, the technology-based variables used in the 2013 MOP Survey weighting to account for the fact that more tech-savvy individuals might be oversampled in the online survey.

weights.[12]

Convergence criteria

We construct raked weights using the *ipfraking* command in Stata. As our benchmark convergence criteria, we set the tolerance parameter, the relative difference of the weights in two successive iterations, to 0.01.[13] To test for robustness, we also rake with the tolerance at 0.001. However, in general we observe little impact of a stricter convergence criterion on raked weights. For example, Figure 4 plots $\mathrm{Rk,NR}^{\mathrm{NP}}$ weights from tolerance at 0.01 and 0.001. Both sets of weights almost align on the 45-degree line, which implies an almost perfect linear correlation between them. Therefore, in the remainder of the analysis we use everywhere the tolerance value 0.01.

3.3 Results

Descriptive statistics of raked weights

In Table 9, we compare unweighted and weighted demographic variables in the 2017 MOP Survey sample to the 2016 Canadian Census and the 2016 SHS.[14] The unweighted sample is biased in terms of age, income, education, marital status, employment status and Internet access. Furthermore, there are obvious discrepancies in several key variables across the three subsamples (online, CFM and offline). Although the simple PS weights compensate for the sample imbalances in terms of age and income, they do not perform well with respect to other variables. However, raked weights $\mathrm{Rk,NR}^{\mathrm{NP}}$ match the weighted sample distributions to the population ones almost perfectly for all raking variables.

To better understand how the various weighting adjustments affect the final weights,

[12] In the *ipfraking* command in Stata, we set the trim frequency to "often," which means that trimming is performed after each marginal adjustment; see Kolenikov (2014) for details. Through iterative raking and trimming operations, final raked weights are able to match weighted distributions very close to external sources.

[13] Convergence will be declared if the largest relative difference of the weights in two successive iterations (a full cycle over all raking variables) does not exceed this value; see Kolenikov (2014).

[14] For other nested variables (not reported), $\mathrm{Rk,NR}^{\mathrm{NP}}$ also matches the population targets.

pairwise comparisons are performed in Figures 5 to 8. Figure 5 compares PS and Rk,1 weights. They differ greatly, because PS accounts only for the region, gender and age strata, which are a subset of the raking variables. The difference between PS and Rk,1 weights is due to the extra information brought by the additional calibration variables. Figure 6 illustrates the correlation between two sets of raked weights, Rk,1 and Rk,PS. Note that most of the weights are overlapping. This is not surprising because both sets of weights rely on the same external information. In fact, the information embedded during the post-stratification process is used again in the raking procedure, so there is no essential gain from using PS initial weights over $\omega_i = 1$.

The important impact of nonresponse adjustment on raked weights can be seen in Figure 7, a scatter plot of Rk,PS vs. Rk,NR^{P}. PS and nonresponse-adjusted PS weights generate substantially different sets of final weights. Therefore, it is expected that incorporating nonresponse adjustment into the calibration procedure should affect weighted means and proportions of variables of interest. Although Rk,PS can, to some extent, implicitly account for nonresponse, the choice of the raking function is generally important when used for treating nonresponse. By choosing a given raking function, one is effectively making a strong statement about the underlying nonresponse mechanism. Therefore, using explicit rather than implicit nonresponse adjustment has the advantage of making the nonresponse modelling assumptions more transparent. However, because the parametric method is vulnerable to mispecification of the response model, a nonparametric method is often preferred. Comparing Rk,NR^{P} and Rk,NR^{NP}, the scatter plot in Figure 8 shows that mild differences exist between the two sets of weights.

Finally, Figure 9 presents box plots of PS, Rk,1, Rk,PS, Rk,NR^{P} and Rk,NR^{NP}, with and without trimming at five times the mean. These two figures show that PS weights are less dispersed but have larger means than other weights. The first graph shows how nonresponse adjustment increases the dispersion of the raked weights. The second graph shows that trimming removes differences in terms of dispersion across various raked weights.

In addition, nonresponse-adjusted raked weights have smaller means than the ones without nonresponse adjustments.

Mean estimates

Table 10 and Table 11 present mean and proportion estimates for two variables of interest, *cash on hand* and *contactless credit card (CTC) usage*, obtained with various weighting schemes. The first variable is the amount of cash the respondent has in his or her wallet, purse or pockets when completing the survey. This continuous variable is important for the Bank of Canada to understand Canadian demand for cash; see Chen et al. (2016). The second is a binary variable indicating whether the respondent has used the contactless feature of a credit card in the past year, so that its weighted mean corresponds to the proportion of the population that have used it. *CTC usage* represents an important payment innovation, and is studied in Chen et al. (2017b) for its effect on cash usage. All weighted results in Table 10 and Table 11 are obtained with weights trimmed at five times their mean. Also, to improve the clarity of the results, all columns from the second to the last are standardized by the values in the first column.

Although trimming can help avoid a few observations having unduly large influence in the weighted results, it may also introduce bias. In Table 12 we present descriptive statistics about the respondents who are assigned extreme (untrimmed) weights. To a large majority, these respondents were sampled from the Online frame. Compared with respondents that receive non-extreme weights, they are more likely to live in households with high income and of large size, and to own their home. Yet, they are also less likely to have high education levels, to have access to Internet from home or to have answered correctly the second financial literacy question in the 2017 MOP Survey. This analysis provides some evidence about the type of respondents that are under-represented in the 2017 MOP sample. In terms of the two variables of interest, *cash on hand* and *CTC usage*, it is interesting to note that individuals who would be assigned extreme untrimmed weights hold on average about \$160, compared with average cash holdings of \$93 in the overall sample. Further, respondents with very large weights present much smaller CTC adoption than the rest of the sample. Hence, this analysis indicates the direction of the bias that trimming introduces in the weighted estimates presented in Table 10 and Table 11.

Weighted results obtained with untrimmed weights are presented in Table 13 and Table 14. To formally investigate the impact of trimming, we apply outlier robust tests to compare the difference between trimmed and untrimmed weighted results with standard errors (Kaji, 2018). These tests indicate that there are no statistically significant differences between trimmed and untrimmed weighted results. Therefore, although trimming does introduce some bias, its magnitude seems statistically negligible.

Also note that only a few observations seem to be affected by trimming.[15] In Figure 3, we see that the distribution of untrimmed Rk,NR^{NP} weights is very skewed; in fact, only 107 respondents are assigned untrimmed weights larger than the trimming threshold of five times the mean. Figure 10 depicts the linear correlation between trimmed and untrimmed Rk,NR^{NP} weights. Only 53 respondents have final trimmed weights equal to the trimming threshold; they are scattered on a horizontal line at the y-axis value of 5. For most respondents however, the untrimmed and trimmed weights are very similar, being on or very close to the 45-degree line. Therefore, it seems that the bulk of the observations are not affected by our trimming rule.

Validation analysis

To evaluate the performance of various raking alternatives, we compare weighted MOP estimates to external sources. First we compare with national statistics on credit card ownership from the Survey of Financial Security (SFS). The SFS proportions as well as the MOP estimates under various raking schemes are presented in Table 15. This validation

[15] Because of the iterative nature of the raking and trimming process, we cannot say exactly how many observations are trimmed.

exercise focuses on assessing the bias of various weighted estimates.[16]

From Table 15, estimates based on trimmed weights often outperform untrimmed ones. Also, among the trimmed weights, it is Rk,NR^{NP} that provides sample estimates closest to the SFS estimates, on average. However, in some domains, such as age between 18 and 24 years old, household income between $65,000 and $85,000 and household size of 1, all raked weights perform relatively poorly, regardless of whether they are trimmed or not.

We can also compare our weighted mean of cash on hand of about $105 (obtained with Rk,NR^{NP}) with the average Canadian bank notes and coins in circulation, around $2,600$ based on TSI (2018).[17] Ours is a clear underestimate, but also note that the amount from TSI (2018) consists of cash held in both wallet and safety box. As for the *CTC usage* variable, our weighted proportion of 61 per cent (obtained with Rk,NR^{NP}) is close to what can be computed from TSI (2018).[18]

4 Variance estimation using bootstrap

Variance estimates are crucial for building confidence intervals to assess dispersion, and for implementing statistical inferences to test various hypotheses. In general, survey variance estimates depend on the specific weighting procedure, not just on the numerical values of the weights; variance estimates that disregard the weighting procedure are often biased. Hence an unbiased estimation method must incorporate two sources of randomness from the weighting procedure: (i) the sampling design, which, in our case, is the design weights induced by complicated sampling; and (ii) the raking procedure, which involves adjusting the sample counts to match the population counts through raked weights. If we ignore either

[16] One limitation of this analysis is that the SFS is a household survey, while the MOP is an individual survey; hence, when we compare proportions of households with at least one credit card with individuals with at least one credit card, the difference might be due to the unit of measurement.

[17] The total notes in circulation divided by the Canadian population, or 90 billion divided by 36 million persons, implies about $2,500 per person. However, such computation ignores the cash held by non-individuals, e.g., retailers and bank branches.

[18] A predicted 16.2 million CTC users divided by the Canadian adult population, or about 16.2 million divided by 28 million, implies a proportion of users of about 58 per cent.

source of randomness, the variance estimates will be incorrect.

To account for the randomness from the sampling design, it is important to understand the design-based inference. While the units in the population as well as their characteristics are assumed to be fixed, the randomness in the design-based statistics comes from randomization performed at the sample-selection stage. The design-based distributions are obtained by enumerating all samples possible under a given design scheme and associating the numeric values of the statistics of interest with the probabilities of the samples they are based on.

As for the randomness from the raking procedure, adjusting design weights makes final weights depend on the particular raking method, and such adjustments affect the variances of weighted estimates. In contrast with the non-random design weights, raked weights are usually random (Lu and Gelman, 2003).

4.1 Methodology

To capture the variability from both sampling design and calibration, we follow Chen and Shen (2015) and use a resampling method. We choose resampling over linearization because resampling has the advantage of accounting for the calibration procedure more easily. Though quota sampling was used, the resampling method approximates the 2017 MOP Survey sampling design with a stratified simple random sampling design, where the population is divided into non-overlapping strata.[19]

Among existing resampling methods, we prefer the bootstrap over the balanced repeated replication (BRR) and the jackknife (Rust and Rao, 1996). We do not use BRR because it is more suitable for a stratified clustered sampling design, where $n_h = 2$ in all strata. The main reason we do not employ the jackknife is that the traditional delete-1 jackknife variance estimator will be inconsistent for non-smooth functions (e.g., sample quantiles).

[19]Treating quota sampling as stratified simple random sampling is a convenient assumption for bootstrapped variance estimation. However, this assumption ignores the fact that the replicated nonprobability sample is selected using quotas. We leave the validity and effect of the assumption to future research.

Instead of recreating the sample in each replication, we implement the more practical method of generating bootstrap replicate survey weights (BRSW). The construction of the replicate weights $w_i^{(t)}$ involves first taking the initial weights ω_i and, from these, constructing a set of initial replication weights $\omega_i^{(t)}$, $t = 1, ..., T$ according to the replication method and the sampling scheme. Next, the raking adjustment method is applied to each of these T sets of weights separately. This generates the required weights $w_i^{(t)}$, $t = 1, ..., T$.

To construct the t^{th} replicate bootstrap sample under stratified sampling, we follow the steps below:

- Step 1: Take a simple random sample with replacement of n_h units from the original data in stratum h, repeating independently across strata.

- Step 2: Modify the initial weights as in the rescaling booststrap from Rao and Wu (1988) by applying the following formula (before the raking procedure):

$$\omega_{hj}^{(t)} = \left\{1 - m_h^{1/2} \left(n_h - 1\right)^{-1/2} + m_h^{1/2} \left(n_h - 1\right)^{-1/2} \frac{n_h}{m_h} m_{hj}^{(t)}\right\} \omega_{hj}, \tag{9}$$

  where $m_{hj}^{(t)}$ is the bootstrap frequency of unit hj, that is, the number of times $_{hj}$ was used in forming the t^{th} bootstrap replicate.[20]

- Step 3: Implement the raking produce to obtain the replicate weight $w_{hj}^{(t)}$. Note that the weight adjustment takes place after the internal scaling of Step 2.

- Step 4: Estimate the parameter of interest, $\widehat{T}_Y^{(t)}$. Repeat T times, and estimate variance using

$$\widehat{V}_B\left(\widehat{T}_Y\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{T}_Y^{(t)} - \widehat{T}_Y\right)^2, \tag{10}$$

  where we choose the number of bootstrap replicates $T = 1,000$.

[20]When m_{hj} is chosen to be $n_h - 1$, the above expression can be simplified following Rao et al. (1992).

4.2 Results

Table 10 and Table 11 show variance computations for mean estimates based on different weights. The variances in the *Linearization* columns are calculated using Stata's linearization method, which does not take calibration into account. The variances in the *Resampling* columns are calculated by the bootstrap resampling method. We follow the American Association for Public Opinion Research (AAPOR) guidance on reporting precision for nonprobability samples (Baker et al., 2013a). Recall that the bootstrapping resampling is carried out under the stratified random sampling without clustering. Also, survey weights are computed via the trimmed raking method and calculated separately for each replicate.

Comparison of linearization and resampling variance estimates

When comparing the linearization and resampling variance of weighted estimates, we observe that the latter is almost always smaller than the former. This feature was also observed in Chen and Shen (2015) for the 2013 MOP Survey. It is because the raking ratio estimator makes use of the correlation between the raking variables and the outcome variable of interest, and thus produces efficiency gains over the estimator, which does not exploit this correlation (Graham, 2011).

When comparing the resampling variances based on the raking procedure with the variances based on the uniform weights, we observe that the ratios are almost always larger than one. As discussed in Chen and Shen (2015), this happens when there is sizeable nonresponse and noncoverage, in which case the estimates based on the raked weights reduce bias at the expense of an increase in variance. For example, the fourth column in Table 9 shows potential coverage issues with respect to different demographics; and the third column in Table 5 illustrates the degree of unit nonresponse. It should be noted that the ratios above one are much more numerous in the 2017 results than in the 2013 results. This may be symptomatic of an increasing difficulty in recruiting survey respondents, as well as a higher reliance on online respondents in 2017 relative to 2013.

Comparison of variance estimates with and without nonresponse adjustment

Applying nonresponse adjustment to PS weights systematically and significantly increases the associated variance estimates. This is consistent with Figure 7, which demonstrates an increased variation in raked weights with nonresponse adjustment, as more points are above the 45-degree line than below. Note also that, since the marginal effect of British Columbia (B.C.) in the merged Offline/CFM frames is statistically insignificant in Table 7, its variance estimate is larger than other regions in Table 10. This illustrates why we should choose variables for the nonresponse model to be closely correlated to the response behaviour.

From Table 10 and Table 11, we find that the fewer weight adjustments, the smaller the variance estimates. However, we caution that, compared with raking without either post-stratification or nonresponse weight adjustment, larger variance from raking with both post-stratification and nonresponse weight adjustments should not be interpreted as precision deterioration. Thus, in the presence of the nonprobability sample, it is the bias that should serve as criterion to discriminate among various raking alternatives.

5 Summary and discussion

The 2017 MOP Survey uses nonprobability sampling with data collection based on stratified quotas. We outline various raking ratio methods to obtain survey weights and compute weighted means and proportions and their corresponding variances. The proposed final raked weights are obtained with both post-stratified and nonresponse-adjusted initial weights, where nonparametric nonresponse adjustment is based on eight homogeneous response classes. They are trimmed at five times their mean. Our bootstrap variance estimation is able to account for both the sampling design and the raking procedure. Figure 11 summarizes the above technical details, and complements Figure 1 with specific choices made during the sampling, weighting and variance estimation processes.

Compared with probability sampling, it is important to focus more on bias than on variance in the context of nonprobability sampling. This is because reducing bias from unknown selection probabilities is challenging, and various weight adjustments for bias reduction would inflate variance. It would be ideal to further validate our weighting approach by comparing our weighted estimates with ones based on a probability sample, especially with variables related to cash and payment behaviours. In that respect, further investigation of the financial data from the SFS is a potential avenue for future research.

References

Arango, C. and A. Welte (2012): "The Bank of Canada's 2009 Methods-of-Payment Survey: Methodology and Key Results," Staff Discussion Paper No. 2012-6, Bank of Canada.

Baker, R., J. M. Brick, N. A. Bates, M. Battaglia, M. P. Couper, J. A. Dever, K. J. Gile, and R. Tourangeau (2013a): "Non-Probability Sampling. Report of the AAPOR Task Forceon Non-Probability Sampling," Report, American Association for Public Opinion Research.

——— (2013b): "Summary Report of the AAPOR Task Force on Non-Probability Sampling," *Journal of Survey Statistics and Methodology*, 1, 90–143.

Bang, H. and J. M. Robins (2005): "Doubly robust estimation in missing data and causal inference models," *Biometrics*, 61, 962–972.

Beaumont, J.-F., C. Bocci, G. Chatrchi, and D. Haziza (2018): "Two Experiments using Nonprobability Survey Data at Statistics Canada," *Working paper.*

Chen, H., G. Dunbar, and Q. R. Shen (2017a): "The Mode is the Message: Using Predata as Exclusion Restrictions to Evaluate Survey Design," Staff Working Paper No. 2017-43, Bank of Canada.

Chen, H., M.-H. Felt, and K. P. Huynh (2017b): "Retail Payment Innovations and Cash Usage: Accounting for Attrition by using Refreshment Samples," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180, 503–530.

Chen, H., C. Henry, K. P. Huynh, Q. R. Shen, and K. Vincent (2016): "Measuring Consumer Cash Holdings: Lessons from the 2013 Bank of Canada Methods-of-Payment Survey," *Survey Practice*, 9.

Chen, H., C. Henry, and V. D. Jiongo (2017c): “The 2017 MOP Sampling Strategy,” Internal report, Bank of Canada.

Chen, H. and Q. R. Shen (2015): “Variance Estimation for Survey-Weighted Data using Bootstrap Resampling Methods: 2013 Methods-of-Payment Survey Questionnaire,” Technical Report No. 104, Bank of Canada.

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

Dillman, D. A. (2000): *Mail and Internet Surveys: The Tailored Design Method.*, John Wiley & Sons.

Elliott, M. R. and R. Valliant (2017): “Inference for Nonprobability Samples,” *Statistical Science*, 32, 249–264.

Graham, B. S. (2011): “Efficiency Bounds for Missing Data Models with Semiparametric Restrictions,” *Econometrica*, 79, 437–452.

Haziza, D. and J.-F. Beaumont (2007): “On the Construction of Imputation Classes in Surveys,” *International Statistical Review/Revue Internationale de Statistique*, 75, 25–43.

Haziza, D. and É. Lesage (2016): “A Discussion of Weighting Procedures for Unit Nonresponse,” *Journal of Official Statistics*, 32, 129–145.

Hellerstein, J. K. and G. W. Imbens (1999): “Imposing Moment Restrictions from Auxiliary Data by Weighting,” *Review of Economics and Statistics*, 81, 1–14.

Henry, C., K. Huynh, and R. Shen (2015): “2013 Methods-of-Payment Survey Results,” Staff Discussion Paper No. 2015-4, Bank of Canada.

Henry, C., K. P. Huynh, and A. Welte (2018): “2017 Methods-of-Payment Survey Report,” Mimeo.

Huang, E. T.-J. H. and W. Fuller (1978): "Nonnegative Regression Estimation for Sample Survey Data," *Proceedings of the Section on Social Statistics*, 300–305.

Ireland, C. T. and S. Kullback (1968): "Contingency Tables with Given Marginals," *Biometrika*, 55, 179–188.

Kaji, T. (2018): "Switching to the New Norm: From Heuristics to Formal Tests using Integrable Empirical Processes," *Working paper.*

Kolenikov, S. (2014): "Calibrating Survey Data using Iterative Proportional Fitting (Raking)," *The Stata Journal*, 14, 22–59.

Lee, S. and R. Valliant (2009): "Estimation for Volunteer Panel Web Surveys using Propensity Score Adjustment and Calibration Adjustment," *Sociological Methods & Research*, 37, 319–343.

Little, R. J. (1986): "Survey Nonresponse Adjustments for Estimates of Means," *International Statistical Review/Revue Internationale de Statistique*, 54, 139–157.

Lu, H. and A. Gelman (2003): "A Method for Estimating Design-Based Sampling Variances for Surveys with Weighting, Poststratification, and Raking," *Journal of Official Statistics*, 19, 133–151.

Nevo, A. (2002): "Sample Selection and Information-Theoretic Alternatives to GMM," *Journal of Econometrics*, 107, 149–157.

Rao, J. N. and A. Singh (2009): "Range Restricted Weight Calibration for Survey Data using Ridge Regression." *Pakistan Journal of Statistics*, 25.

Rao, J. N. and C. Wu (1988): "Resampling Inference with Complex Survey Data," *Journal of the american statistical association*, 83, 231–241.

Rao, J. N., C. Wu, and K. Yue (1992): "Some Recent Work on Resampling Methods for Complex Surveys," *Survey methodology*, 18, 209–217.

Rosenbaum, P. R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rust, K. F. and J. N. Rao (1996): "Variance Estimation for Complex Surveys using Replication Techniques," *Statistical methods in medical research*, 5, 283–310.

Shen, Q. R. and K. Vincent (2014): "Survey Recruitment Using Financial and Non-Financial Incentives: Results from the 2013 Methods-of-Payment Survey," Mimeo.

Tourangeau, R., F. G. Conrad, and M. P. Couper (2013): *The Science of Web Surveys*, Oxford University Press.

TSI (2018): "Canadian Payments Forecast 2018," *Technology Strategies International Report*.

Vincent, K. (2015): "2013 Methods-of-Payment Survey: Sample Calibration Analysis," Technical Report No. 103, Bank of Canada.

Wang, W., D. Rothschild, S. Goel, and A. Gelman (2015): "Forecasting Elections with Non-Representative Polls," *International Journal of Forecasting*, 31, 980–991.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica: Journal of the Econometric Society*, 50, 1–25.

Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang (2011): "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples," *Public opinion quarterly*, 75, 709–747.

Table 1: Sample size calculations based on *cash on hand* variable

	Total	Online	Offline	CFM
2013 MOP data:				
Cash on hand - mean	89.78	87.73	96.9	87.98
Cash on hand - VarBSRW	13.39	53.62	61.16	27.49
Number of observations	3,413	1,441	680	1,292
Coefficient of variation (CV)	4.08	8.35	8.07	5.96
2017 planning:				
Sample size required for 10% CV		1,004	443	459
Sample size chosen for 2017 MOP Survey	1,500	500	500	500

Notes: Data are from the 2013 MOP Survey. VarBSRW refers to the variance estimates calculated using the bootstrap replicate survey weights method; these estimates were computed in Chen and Shen (2015).

Table 2: Stratified quotas

Region	Gender	Age	Target	Region	Gender	Age	Target
Atlantic	Male	18–24	3	Prairies	Male	18–24	5
Atlantic	Male	25–34	2	Prairies	Male	25–34	9
Atlantic	Male	35–44	2	Prairies	Male	35–44	8
Atlantic	Male	45–54	3	Prairies	Male	45–54	8
Atlantic	Male	55–64	3	Prairies	Male	55–64	7
Atlantic	Male	65+	4	Prairies	Male	65+	7
Atlantic	Female	18–24	3	Prairies	Female	18–24	5
Atlantic	Female	25–34	2	Prairies	Female	25–34	9
Atlantic	Female	35–44	3	Prairies	Female	35–44	8
Atlantic	Female	45–54	3	Prairies	Female	45–54	8
Atlantic	Female	55–64	3	Prairies	Female	55–64	7
Atlantic	Female	65+	4	Prairies	Female	65+	8
Quebec	Male	18–24	6	B.C.	Male	18–24	4
Quebec	Male	25–34	9	B.C.	Male	25–34	5
Quebec	Male	35–44	9	B.C.	Male	35–44	5
Quebec	Male	45–54	10	B.C.	Male	45–54	6
Quebec	Male	55–64	11	B.C.	Male	55–64	6
Quebec	Male	65+	12	B.C.	Male	65+	7
Quebec	Female	18–24	6	B.C.	Female	18–24	3
Quebec	Female	25–34	9	B.C.	Female	25–34	6
Quebec	Female	35–44	9	B.C.	Female	35–44	5
Quebec	Female	45–54	10	B.C.	Female	45–54	6
Quebec	Female	55–64	11	B.C.	Female	55–64	6
Quebec	Female	65+	15	B.C.	Female	65+	8
Ontario	Male	18–24	11		Total		500
Ontario	Male	25–34	15				
Ontario	Male	35–44	15				
Ontario	Male	45–54	17				
Ontario	Male	55–64	16				
Ontario	Male	65+	18				
Ontario	Female	18–24	11				
Ontario	Female	25–34	16				
Ontario	Female	35–44	16				
Ontario	Female	45–54	18				
Ontario	Female	55–64	17				
Ontario	Female	65+	22				

Notes: This table shows the stratified quotas for each of the Online, Offline and CFM panels. Stratified quotas match proportions found in the Canadian population as measured in the 2016 Canadian Census. All three panels have the same targets.

Table 3: Survey rollout plan: Offline and CFM panels

	Wave 1	Wave 2	Wave 3
Proposed mail-out date	13-Oct-17	19-Oct-17	23-Oct-17
Proportion of total invitations sent	0.61	0.25	0.14

Notes: This table shows the scheduled proportion of invitations per mail-out wave. Sampling for the Online panel was conducted on a rolling basis with returns updated in real time, negating the need to specify a strict mail-out schedule.

Table 4: Final sample: Number of respondents and *cash on hand* estimate

			Paper-based		
	Overall	Online	Offline	CFM	Ad hoc
SQ (observations)	3,123	1,889	510	709	15
Cash on hand - mean	104.84	110.46	89.55	84.76	
Cash on hand - VarLin	56.96	93.13	110.73	83.34	
Cash on hand - VarBSRW	21.45	38.88	77.08	31.83	
DSI (observations)	2,187	953	510	709	15

Notes: This table shows the final sample sizes from each of the three panels plus the ad hoc invitations. For the Online panel we collected more SQ than DSI responses because not all respondents who completed the questionnaire also completed the diary. Ad hoc participants are merged with the Offine respondents for computing the *cash on hand* variable estimates. VarLin is computed using linearization, which does not take into account the weighting procedure; VarBSRW refers to the variance estimation using bootstrap replicate survey weights.

Table 5: Unit response rates

	2017 MOP Survey			2013 MOP Survey		
	Offline/CFM	Online	Overall	Offline/CFM	Online	Overall
Total	32.4	4.9	7.4	40.5	3.5	7.3
Gender						
Male	33.8	5.1	7.3	39.8	3.4	7.0
Female	31.3	4.7	7.4	41.2	3.7	7.6
Age						
18–34	25.2	2.9	3.7	32.4	2.5	5.2
35–64	31.3	7.2	11.7	43.0	4.0	8.5
65+	47.2	13.8	19.6	52.9	4.7	8.7
Region						
Atlantic	35.5	4.4	6.5	36.0	3.3	6.1
Quebec	35.4	5.2	7.6	45.1	2.7	5.7
Ontario	32.3	5.3	8	39.7	4.3	8.9
Prairies	28.5	4	6.5	36.5	3.7	7.3
B.C.	33.1	5	7.4	42.3	3.9	8.2
Household income						
<$25K	29.4	2.9	4.8	38.3	2.4	4.7
$25–45K	31.7	4	7.1	42.0	3.6	7.3
$45–70K	32.9	6.3	9.6	40.5	4.2	8.2
$70–100K	35.3	6.4	8.6	42.2	4.2	9.1
$100K+	34.3	6.7	8.6	39.0	4.1	9.7
Education						
High school	29.1	3.1	4.5	38.9	2.5	5.0
College	35.5	4.9	6.9	36.9	3.7	7.8
University	33.8	9.2	13.6	48.3	4.7	9.3
Household size						
1	35.2	6.7	12	41.4	3.5	7.3
2–4	31.3	4.8	6.6	40.2	3.6	7.3
5+	21	2.8	3.6	37.2	2.8	6.8
Employment status						
Employed	49.4	5.2	8	39.9	3.6	7.4
Unemployed	28	3.1	5.7	37.6	2.6	4.6
Not in labour force	38.1	6.6	9.7	44.8	3.7	7.2

Notes: The first row presents total response rates for the 2017 and 2013 MOP surveys. The following rows show response rates by demographic categories for the 2017 and 2013 MOP surveys. The columns are computed from the merged Offline and CFM invitation lists, the Online invitation list and the whole invitation list, respectively.

Table 6: Item nonresponse of raking variables

	Original sample		Back-filled sample	
	observations	%	observations	%
Gender	0	0.00	0	0.00
Age	123	3.93	0	0.00
Region	0	0.00	0	0.00
Income	8	0.26	3	0.10
Education	30	0.96	5	0.16
Household size	4	0.13	1	0.03
Marital status	7	0.22	3	0.10
Employment status	18	0.58	7	0.22
Own their home	9	0.29	4	0.13
Internet access	4	0.13	4	0.13

Notes: Proportion of missing observations for each calibration variable. The total number of survey participants is 3,123. *Original sample* refers to the raw sample of respondents. *Backfilled sample* refers to the edited sample where missing items have been imputed with frame data, when available.

Table 7: SQ unit nonresponse analysis with logit and sequential logit models

	Offline/CFM	Online	Online (seqlogit)
Female	0.002	0.005**	0.002**
Age	-0.002	0.005***	0.002**
Age squared	0.000	-0.000***	-0.000***
Quebec	0.014	0.016***	0.007***
Ontario	-0.018	0.019***	0.008**
Prairies	-0.060**	0.012***	0.005***
B.C.	-0.037	0.013***	0.006**
Hh income: $25–45K	-0.012	-0.003	-0.001
$45–70K	0.008	0.011***	0.004**
$70–100K	0.004	0.014***	0.006***
$100K+	-0.004	0.007	0.003
Some high school	-0.040	0.009*	0.004**
Graduated high school	0.055***	0.023***	0.011 ***
Some college	-0.008	0.036***	0.020***
Graduated college	0.064***	0.049***	0.031***
Some university	0.033	0.117***	0.171***
Hh size: 2–4	-0.013	-0.018***	-0.005 **
5+	-0.090**	-0.029***	-0.008**
Married	0.004	0.002	0.001
Common law	-0.106***	0.005	0.002
Div./sep./wid.	0.003	-0.011**	-0.003**
Unemployed	-0.104***	0.009	0.004
Not in labour force	-0.050***	0.010***	0.004***
Number of past surveys taken	0.005***	0.000***	0.000***
Observations	3766	38498	38498
Pseudo R-squared	0.100	0.114	0.120

Notes: Logistic regressions estimate SQ unit response propensities for the Offline/CFM and Online frames. Not all Online SQ participants also decided to fill out a DSI. Therefore, a sequential logit regression is estimated for the Online frame. In contrast, because all Offline/CFM SQ participants are also DSI participants, their decisions of responding to SQ and then DSI boil down to a one-step decision. Marginal effects are shown. The base categories are: Gender: Male; Region: Atlantic; Household income: less than $25,000; Education: Primary school or less; Household size: 1; Marital status: Single; Employment status: Full-time. Marginal effects for the sequential logit model are computed with the Stata command *seqlogit*. *Observations* means the total number of invitations sent out.

Table 8: Epps-Singleton test of homogeneous distributions among subsamples

	Online vs. Offline/CFM		Offline vs. CFM	
	Test statistic	P-value	Test statistic	P-value
Gender	-0.61	0.54	0.76	0.44
Age	249.10	0.00	6.23	0.18
Region	7.3	0.12	5.80	0.22
Income	12.67	0.01	5.44	0.25
Education	6.72	0.15	2.08	0.72
Household size	164.26	0.00	1.51	0.83
Employment status	368.25	0.00	0.786	0.94
Cash on hand	3.36	0.49	0.164	0.80
CTC usage	2.76	0.29	0.77	0.44
Day of week	39.07	0.00	5.81	0.21

Notes: The Epps-Singleton test is carried out to compare distributions across different subsamples. *Day of week* refers to the start day of the three-day DSI, e.g., Monday.

Table 9: Unweighted and weighted sample proportions vs. population proportions

	Online	CFM	Non-CFM	Sample	PS	Rk,NR^{NP}	Population
Female	47.8	50.1	47.2	48.2	51.3	51.3	51.3
Age: 18–24	22.8	6.3	7.4	16.5	11.1	11.1	11.1
25–34	18.3	16.2	10.7	16.6	16.7	16.7	16.7
35–44	11.8	17.5	16.6	13.9	16.4	16.4	16.4
45–54	12.8	19.3	21.1	15.7	18.2	18.2	18.2
55–64	14.7	18.9	20.6	16.6	17.7	17.7	17.7
65+	19.7	21.7	23.6	20.8	20.0	20.0	20.0
Atlantic	9.3	8.3	9.7	9.1	6.8	6.8	6.8
Quebec	22.9	21.7	18.7	21.9	23.3	23.3	23.3
Ontario	36.9	37.7	36.2	37.0	38.6	38.6	38.6
Prairies	16.6	17.9	22.9	17.9	17.7	17.6	17.7
B.C.	14.4	14.4	12.6	14.1	13.6	13.6	13.6
Hh income: $<$ $25K	14.6	20.2	15.3	16.0	15.6	9.4	9.4
$25–45K	18.2	19.2	23.3	19.3	19.1	15.6	15.6
$45–65K	19.5	18.4	17.0	18.8	18.9	17.4	17.4
$65–85K	16.8	14.1	15.1	15.9	15.9	15.9	15.9
$85K+	30.9	28.0	29.4	30.0	30.5	41.7	41.7
High school	21.4	20.0	18.0	20.5	19.8	42.3	42.4
College	26.4	33.9	31.0	28.9	29.9	30.4	30.4
University	52.2	46.2	51.1	50.6	50.3	27.3	27.2
Household size: 1	24.5	40.7	40.0	30.8	31.2	14.4	14.4
2	39.2	35.3	34.9	37.6	37.8	34.2	34.2
3	17.4	11.3	11.2	15.0	14.6	18.7	18.7
4	12.2	8.5	9.3	10.9	11.0	18.6	18.5
5+	6.7	4.2	4.6	5.8	5.4	14.1	14.1
Married/common law	54.2	43.7	45.6	50.4	52.3	61.0	60.9
Employed	55.3	54.7	56.0	55.3	57.0	61.7	61.7
Unemployed	3.8	4.7	2.7	3.8	3.7	5.0	5.0
Not in labour force	40.9	40.6	41.3	40.9	39.3	33.3	33.3
Own their home	54.5	60.7	62.5	57.2	59.8	73.3	73.0
No Internet from home	1.0	3.7	2.5	1.9	1.9	9.1	9.1

Notes: Numbers are percentages. Columns 1 to 3 show unweighted proportions obtained from the Online, CFM and Offline subsamples, respectively. Ad hoc participants are merged to the Offline subsample. Column 4 shows unweighted estimates for the overall sample. Columns 5 and 6 show weighted results, where PS is for post-stratification weights, while Rk,NR^{NP} is for trimmed raked weights based on nonparametrically nonresponse-adjusted PS initial weights. Population distributions presented in the last column come from the 2016 Canadian Census (based on residents of Canadian provinces aged 18 years or older), except for *Internet access from home*. The latter is based on the 2016 Survey of Household Spending, where household counts have been scaled up to individual counts according to the household type. In the case of binary variables, only one category is shown.

Table 10: Mean and variance estimates for *cash on hand* with trimmed raked weights

			Linearization					Resampling		
	Uniform	PS	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}
Overall	92.82	0.98	1.08	1.08	1.13	1.13	1.08	1.08	1.13	1.13
	(12.93)	*(0.88)*	*(3.60)*	*(3.57)*	*(4.48)*	*(4.40)*	*(1.41)*	*(1.37)*	*(1.63)*	*(1.66)*
Female	106.33	0.97	1.10	1.09	1.14	1.14	1.10	1.09	1.14	1.14
	(32.07)	*(0.83)*	*(4.60)*	*(4.59)*	*(5.34)*	*(5.39)*	*(1.64)*	*(1.63)*	*(1.83)*	*(1.88)*
Male	78.28	1.01	1.09	1.09	1.15	1.14	1.09	1.09	1.15	1.14
	(18.62)	*(1.04)*	*(2.37)*	*(2.30)*	*(3.54)*	*(3.27)*	*(1.17)*	*(1.09)*	*(1.64)*	*(1.55)*
18–34	87.20	0.96	1.22	1.21	1.37	1.35	1.22	1.21	1.37	1.35
	(62.39)	*(0.89)*	*(6.42)*	*(6.33)*	*(7.89)*	*(7.49)*	*(2.23)*	*(2.16)*	*(2.69)*	*(2.70)*
34–54	86.19	0.99	1.05	1.04	1.10	1.11	1.05	1.04	1.10	1.11
	(31.20)	*(0.88)*	*(2.34)*	*(2.27)*	*(3.04)*	*(3.38)*	*(1.08)*	*(1.07)*	*(1.46)*	*(1.49)*
55+	103.03	0.99	1.02	1.03	1.01	1.01	1.02	1.03	1.01	1.01
	(24.28)	*(1.12)*	*(1.91)*	*(1.98)*	*(2.39)*	*(2.31)*	*(0.99)*	*(1.02)*	*(0.95)*	*(0.94)*
<$45K	69.17	1.00	1.01	1.01	0.99	0.98	1.01	1.01	0.99	0.98
	(12.93)	*(0.98)*	*(1.68)*	*(1.69)*	*(2.60)*	*(2.26)*	*(1.38)*	*(1.37)*	*(2.16)*	*(1.98)*
$45–85K	97.46	0.97	0.96	0.96	1.03	1.02	0.96	0.96	1.03	1.02
	(38.38)	*(1.02)*	*(1.48)*	*(1.49)*	*(3.10)*	*(2.77)*	*(1.06)*	*(1.08)*	*(1.61)*	*(1.54)*
$85K+	115.37	0.97	1.08	1.07	1.13	1.14	1.08	1.07	1.13	1.14
	(73.83)	*(0.74)*	*(2.97)*	*(2.94)*	*(3.24)*	*(3.30)*	*(1.07)*	*(1.03)*	*(1.14)*	*(1.17)*
Atlantic	83.65	0.99	1.00	1.00	0.89	0.95	1.00	1.00	0.89	0.95
	(62.63)	*(1.16)*	*(2.49)*	*(2.35)*	*(2.98)*	*(5.63)*	*(2.48)*	*(2.60)*	*(2.62)*	*(2.67)*
Quebec	87.18	0.98	1.07	1.07	1.12	1.12	1.07	1.07	1.12	1.12
	(24.09)	*(0.94)*	*(2.13)*	*(2.16)*	*(3.66)*	*(3.57)*	*(1.66)*	*(1.63)*	*(1.86)*	*(1.92)*
Ontario	94.80	0.99	0.98	0.98	1.07	1.04	0.98	0.98	1.07	1.04
	(37.81)	*(1.03)*	*(1.70)*	*(1.69)*	*(2.81)*	*(2.28)*	*(0.95)*	*(0.95)*	*(1.40)*	*(1.28)*
Prairies	94.30	0.94	0.98	0.98	0.95	0.96	0.98	0.98	0.95	0.96
	(115.48)	*(0.50)*	*(1.26)*	*(1.27)*	*(1.36)*	*(1.38)*	*(0.53)*	*(0.54)*	*(0.55)*	*(0.56)*
B.C.	100.37	0.98	1.53	1.51	1.62	1.64	1.53	1.51	1.62	1.64
	(119.90)	*(1.02)*	*(12.52)*	*(12.38)*	*(13.68)*	*(14.28)*	*(4.22)*	*(4.02)*	*(4.70)*	*(4.92)*
CFM	79.93	0.99	1.01	1.03	1.05	1.06	1.01	1.03	1.05	1.06
	(17.33)	*(0.99)*	*(2.28)*	*(2.34)*	*(4.63)*	*(4.81)*	*(1.18)*	*(1.19)*	*(1.79)*	*(1.84)*
Offline	96.95	1.00	0.89	0.89	0.93	0.92	0.89	0.89	0.93	0.92
	(84.53)	*(1.12)*	*(0.75)*	*(0.78)*	*(1.34)*	*(1.31)*	*(0.64)*	*(0.65)*	*(0.90)*	*(0.91)*
Online	96.46	0.97	1.16	1.15	1.15	1.15	1.16	1.15	1.15	1.15
	(26.37)	*(0.84)*	*(4.47)*	*(4.48)*	*(3.61)*	*(3.53)*	*(1.66)*	*(1.63)*	*(1.45)*	*(1.47)*

Notes: All columns after the first have been divided by the value in the first column. Each row shows weighted point estimates, with corresponding variance estimates below in parentheses. Linearization estimates are produced with the linearization procedure in Stata, which simply assumes a stratified random sample and does not take into account the weighting procedure. Resampling estimates are computed as described in Section 4.1 using bootstrap replicated survey weights. All the raked weights are trimmed at five times their mean and obtained with a tolerance level of 0.01 as convergence criteria. Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Table 11: Mean and variance estimates for *CTC usage* with trimmed raked weights

			Linearization					Resampling		
	Uniform	PS	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}
Overall	0.64	1.00	0.95	0.95	0.96	0.95	0.95	0.95	0.96	0.95
	(7.24E-05)	*(1.06)*	*(2.83)*	*(2.86)*	*(3.20)*	*(3.25)*	*(1.37)*	*(1.37)*	*(1.41)*	*(1.45)*
Female	0.64	1.01	0.95	0.94	0.95	0.94	0.95	0.94	0.95	0.94
	(1.40E-04)	*(1.04)*	*(3.07)*	*(3.08)*	*(3.50)*	*(3.55)*	*(1.50)*	*(1.49)*	*(1.47)*	*(1.49)*
Male	0.64	1.00	0.96	0.96	0.97	0.96	0.96	0.96	0.97	0.96
	(1.50E-04)	*(1.07)*	*(2.61)*	*(2.65)*	*(2.93)*	*(2.97)*	*(1.31)*	*(1.31)*	*(1.38)*	*(1.41)*
18–34	0.69	1.03	0.89	0.89	0.89	0.88	0.89	0.89	0.89	0.88
	(2.02E-04)	*(1.01)*	*(3.59)*	*(3.64)*	*(4.09)*	*(4.17)*	*(1.77)*	*(1.76)*	*(1.84)*	*(1.91)*
34–54	0.63	1.00	1.00	0.99	0.98	0.99	1.00	0.99	0.98	0.99
	(2.50E-04)	*(1.02)*	*(2.43)*	*(2.43)*	*(2.71)*	*(2.75)*	*(1.27)*	*(1.27)*	*(1.29)*	*(1.30)*
55+	0.61	1.00	0.98	0.98	1.00	0.99	0.98	0.98	1.00	0.99
	(2.04E-04)	*(1.05)*	*(2.64)*	*(2.69)*	*(3.02)*	*(3.04)*	*(1.31)*	*(1.31)*	*(1.36)*	*(1.38)*
<$45K	0.53	0.99	0.89	0.89	0.91	0.91	0.89	0.89	0.91	0.91
	(2.27E-04)	*(1.06)*	*(3.21)*	*(3.24)*	*(3.74)*	*(3.75)*	*(1.65)*	*(1.67)*	*(1.70)*	*(1.74)*
$45–85K	0.66	1.00	0.94	0.94	0.94	0.93	0.94	0.94	0.94	0.93
	(2.04E-04)	*(1.06)*	*(2.75)*	*(2.79)*	*(3.27)*	*(3.32)*	*(1.56)*	*(1.56)*	*(1.57)*	*(1.59)*
$85K+	0.76	1.00	0.92	0.92	0.91	0.91	0.92	0.92	0.91	0.91
	(1.98E-04)	*(1.05)*	*(2.70)*	*(2.72)*	*(2.93)*	*(3.00)*	*(1.30)*	*(1.29)*	*(1.38)*	*(1.39)*
Atlantic	0.57	1.00	0.94	0.92	0.98	0.95	0.94	0.92	0.98	0.95
	(8.60E-04)	*(1.19)*	*(3.01)*	*(3.08)*	*(3.72)*	*(3.87)*	*(1.94)*	*(2.00)*	*(2.01)*	*(2.01)*
Quebec	0.62	1.00	0.94	0.94	0.93	0.92	0.94	0.94	0.93	0.92
	(3.41E-04)	*(1.05)*	*(2.60)*	*(2.61)*	*(2.99)*	*(3.02)*	*(1.31)*	*(1.31)*	*(1.37)*	*(1.40)*
Ontario	0.68	1.01	0.92	0.93	0.94	0.94	0.92	0.93	0.94	0.94
	(1.86E-04)	*(1.02)*	*(2.91)*	*(2.91)*	*(3.23)*	*(3.26)*	*(1.40)*	*(1.41)*	*(1.42)*	*(1.43)*
Prairies	0.60	0.99	1.02	1.01	1.01	1.01	1.02	1.01	1.01	1.01
	(4.22E-04)	*(1.06)*	*(2.73)*	*(2.77)*	*(3.08)*	*(3.10)*	*(1.35)*	*(1.34)*	*(1.40)*	*(1.40)*
B.C.	0.67	1.00	0.97	0.97	0.98	0.96	0.97	0.97	0.98	0.96
	(4.98E-04)	*(1.11)*	*(3.01)*	*(3.13)*	*(3.38)*	*(3.48)*	*(1.40)*	*(1.40)*	*(1.48)*	*(1.50)*
CFM	0.61	1.00	0.94	0.94	0.88	0.87	0.94	0.94	0.88	0.87
	(3.35E-04)	*(1.04)*	*(2.76)*	*(2.76)*	*(5.16)*	*(5.21)*	*(1.43)*	*(1.44)*	*(2.20)*	*(2.22)*
Offline	0.64	1.00	0.95	0.94	0.90	0.89	0.95	0.94	0.90	0.89
	(4.43E-04)	*(1.04)*	*(2.93)*	*(2.89)*	*(4.81)*	*(4.86)*	*(1.50)*	*(1.48)*	*(2.02)*	*(2.04)*
Online	0.65	1.00	0.96	0.96	0.97	0.96	0.96	0.96	0.97	0.96
	(1.18E-04)	*(1.07)*	*(2.75)*	*(2.81)*	*(2.54)*	*(2.58)*	*(1.36)*	*(1.37)*	*(1.18)*	*(1.21)*

Notes: All columns after the first have been divided by the value in the first column. Each row shows weighted point estimates, with corresponding variance estimates below in parentheses. Linearization estimates are produced with the linearization procedure in Stata, which simply assumes a stratified random sample and does not take into account the weighting procedure. Resampling estimates are computed as described in Section 4.1 using bootstrap replicated survey weights. All the raked weights are trimmed at five times their mean and obtained with a tolerance level of 0.01 as convergence criteria. Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Table 12: Analysis of extreme weights

	Sample	<99pct	$\geq$ 99pct
Female	48.2	48.1	53.1
Age: 18–34	33.0	33.0	31.3
35–64	29.6	29.5	34.4
65+	37.4	37.5	34.4
Hh income: <$45K	35.3	35.4	25.0
$45–85K	34.7	34.8	21.9
$85K+	30.0	29.8	53.1
Atlantic	9.1	9.1	9.4
Quebec	21.9	21.9	25.0
Ontario	37.0	36.9	37.5
Prairies	17.9	18.0	15.6
B.C.	14.1	14.1	12.5
High school	20.5	20.0	68.8
College	28.9	28.9	25.0
University	50.6	51.1	6.3
Own their home	57.2	56.9	90.6
Hh size: 1	30.8	31.0	6.3
2	37.6	37.7	21.9
3+	31.6	31.2	71.9
Single	40.0	40.2	21.9
Married/common law	50.4	50.3	62.5
Div./Sep./Wid.	9.6	9.6	15.6
Full-time	42.8	42.6	59.4
Part-time	12.5	12.5	15.6
Not employed	44.7	44.9	25.0
Born in Canada	84.5	84.5	87.5
No Internet	0.9	0.7	12.5
Rural	13.2	13.3	9.4
Fin. Lit. Q1 Correct	85.3	85.3	84.4
Fin. Lit. Q2 Correct	69.8	70.1	46.9
Fin. Lit. Q3 Correct	59.9	59.9	62.5
CFM	22.7	22.9	3.1
Offline	16.8	16.9	6.3
Online	60.5	60.2	90.6
Mean cash on hand	92.8	92.1	161.4
Median cash on hand	40	40	55
Mean CTC usage	0.642	0.644	0.438
Median CTC usage	1	1	0
N	3,123	3,091	32

Notes: Numbers are proportions in the first part of the table. Column 1 shows unweighted estimates for the overall sample. Columns 2 and 3 describe respondents with weights respectively below and above the 99^{th} percentile of the untrimmed Rk,NR^{NP} weights distribution. *Fin. Lit. Q1* (*Q2* and *Q3*, respectively) *Correct* refers to answering the first (second and third, respectively) financial literacy question correctly. In the case of binary variables, only one category is shown.

Table 13: Mean and variance estimates for *cash on hand* with untrimmed raked weights

		Linearization					Resampling			
	Uniform	PS	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}
Overall	92.82	0.98	1.09	1.09	1.26	1.25	1.09	1.09	1.26	1.25
	(12.93)	*(0.88)*	*(2.67)*	*(2.99)*	*(11.33)*	*(11.34)*	*(2.31)*	*(2.59)*	*(9.29)*	*(8.73)*
Female	106.33	0.97	1.05	1.05	1.19	1.18	1.05	1.05	1.19	1.18
	(32.07)	*(0.83)*	*(2.44)*	*(2.73)*	*(11.94)*	*(10.54)*	*(2.43)*	*(2.78)*	*(10.82)*	*(9.98)*
Male	78.28	1.01	1.15	1.16	1.36	1.37	1.15	1.16	1.36	1.37
	(18.62)	*(1.04)*	*(3.31)*	*(3.72)*	*(11.53)*	*(13.81)*	*(2.59)*	*(2.84)*	*(8.73)*	*(8.83)*
18–34	87.20	0.96	1.18	1.20	1.57	1.62	1.18	1.20	1.57	1.62
	(62.39)	*(0.89)*	*(3.77)*	*(4.69)*	*(18.41)*	*(19.20)*	*(3.44)*	*(4.23)*	*(16.12)*	*(15.64)*
34–54	86.19	0.99	1.05	1.04	1.18	1.12	1.05	1.04	1.18	1.12
	(31.20)	*(0.88)*	*(2.03)*	*(1.97)*	*(8.37)*	*(6.55)*	*(1.90)*	*(1.84)*	*(7.46)*	*(6.17)*
55+	103.03	0.99	1.05	1.06	1.12	1.12	1.05	1.06	1.12	1.12
	(24.28)	*(1.12)*	*(2.53)*	*(2.48)*	*(7.36)*	*(7.50)*	*(2.18)*	*(2.15)*	*(5.59)*	*(5.72)*
<$45K	69.17	1.00	1.13	1.13	1.12	1.05	1.13	1.13	1.12	1.05
	(12.93)	*(0.98)*	*(4.04)*	*(4.40)*	*(5.60)*	*(3.94)*	*(3.04)*	*(3.30)*	*(4.79)*	*(3.59)*
$45–85K	97.46	0.97	0.98	0.98	1.09	1.06	0.98	0.98	1.09	1.06
	(38.38)	*(1.02)*	*(1.78)*	*(1.66)*	*(5.68)*	*(5.66)*	*(1.79)*	*(1.70)*	*(4.77)*	*(4.59)*
$85K+	115.37	0.97	1.02	1.03	1.29	1.32	1.02	1.03	1.29	1.32
	(73.83)	*(0.74)*	*(1.85)*	*(2.17)*	*(8.95)*	*(8.89)*	*(1.69)*	*(2.00)*	*(7.63)*	*(7.34)*
Atlantic	83.65	0.99	0.98	0.98	0.76	0.78	0.98	0.98	0.76	0.78
	(62.63)	*(1.16)*	*(2.05)*	*(2.29)*	*(1.38)*	*(1.96)*	*(1.80)*	*(1.94)*	*(1.51)*	*(2.21)*
Quebec	87.18	0.98	1.12	1.11	1.15	1.12	1.12	1.11	1.15	1.12
	(24.09)	*(0.94)*	*(3.05)*	*(2.98)*	*(4.52)*	*(4.07)*	*(2.81)*	*(2.80)*	*(4.34)*	*(3.90)*
Ontario	94.80	0.99	1.02	1.03	1.23	1.22	1.02	1.03	1.23	1.22
	(37.81)	*(1.03)*	*(2.31)*	*(2.65)*	*(9.05)*	*(10.93)*	*(1.99)*	*(2.21)*	*(7.41)*	*(7.51)*
Prairies	94.30	0.94	1.03	1.03	1.20	1.24	1.03	1.03	1.20	1.24
	(115.48)	*(0.50)*	*(1.53)*	*(1.40)*	*(5.51)*	*(5.87)*	*(1.42)*	*(1.31)*	*(4.54)*	*(4.87)*
B.C.	100.37	0.98	1.30	1.31	1.74	1.73	1.30	1.31	1.74	1.73
	(119.90)	*(1.02)*	*(5.07)*	*(6.21)*	*(30.16)*	*(25.32)*	*(5.36)*	*(6.59)*	*(28.63)*	*(25.03)*
CFM	79.93	0.99	0.97	0.98	1.03	1.06	0.97	0.98	1.03	1.06
	(17.33)	*(0.99)*	*(1.76)*	*(1.89)*	*(3.70)*	*(4.38)*	*(1.74)*	*(1.86)*	*(3.92)*	*(4.58)*
Offline	96.95	1.00	0.99	1.00	1.11	1.12	0.99	1.00	1.11	1.12
	(84.53)	*(1.12)*	*(1.35)*	*(1.44)*	*(2.24)*	*(2.37)*	*(1.32)*	*(1.38)*	*(2.29)*	*(2.36)*
Online	96.46	0.97	1.15	1.15	1.26	1.25	1.15	1.15	1.26	1.25
	(26.37)	*(0.84)*	*(3.17)*	*(3.61)*	*(8.08)*	*(8.01)*	*(2.84)*	*(3.22)*	*(6.73)*	*(6.29)*

Notes: All columns after the first have been divided by the value in the first column. Each row shows weighted point estimates, with corresponding variance estimates below in parentheses. Linearization estimates are produced with the linearization procedure in Stata, which simply assumes a stratified random sample and does not take into account the weighting procedure. Resampling estimates are computed as described in Section 4.1 using bootstrap replicated survey weights. All the raked weights are trimmed at five times their mean and obtained with a tolerance level of 0.01 as convergence criteria. Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Table 14: Mean and variance estimates for *CTC usage* with untrimmed raked weights

			Linearization					Resampling		
	Uniform	PS	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}
Overall	0.64	1.00	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	(7.24E-05)	*(1.06)*	*(3.36)*	*(3.44)*	*(4.86)*	*(5.13)*	*(2.85)*	*(2.98)*	*(3.81)*	*(4.06)*
Female	0.64	1.01	0.94	0.94	0.91	0.91	0.94	0.94	0.91	0.91
	(1.40E-04)	*(1.04)*	*(3.38)*	*(3.33)*	*(5.28)*	*(5.35)*	*(3.08)*	*(3.08)*	*(4.26)*	*(4.47)*
Male	0.64	1.00	0.96	0.95	0.98	0.98	0.96	0.95	0.98	0.98
	(1.50E-04)	*(1.07)*	*(3.31)*	*(3.50)*	*(4.44)*	*(4.90)*	*(2.84)*	*(2.95)*	*(3.78)*	*(3.85)*
18–34	0.69	1.03	0.91	0.90	0.88	0.87	0.91	0.90	0.88	0.87
	(2.02E-04)	*(1.01)*	*(3.50)*	*(3.76)*	*(7.75)*	*(8.29)*	*(3.54)*	*(3.71)*	*(6.84)*	*(7.07)*
34–54	0.63	1.00	0.99	0.99	1.00	1.01	0.99	0.99	1.00	1.01
	(2.50E-04)	*(1.02)*	*(2.89)*	*(2.85)*	*(3.49)*	*(3.84)*	*(2.55)*	*(2.55)*	*(3.14)*	*(3.43)*
55+	0.61	1.00	0.96	0.95	0.96	0.96	0.96	0.95	0.96	0.96
	(2.04E-04)	*(1.05)*	*(3.49)*	*(3.60)*	*(4.23)*	*(4.12)*	*(2.79)*	*(2.88)*	*(3.44)*	*(3.36)*
<$45K	0.53	0.99	0.90	0.90	0.97	0.95	0.90	0.90	0.97	0.95
	(2.27E-04)	*(1.06)*	*(4.11)*	*(4.27)*	*(4.47)*	*(4.63)*	*(3.59)*	*(3.72)*	*(4.59)*	*(4.48)*
$45–85K	0.66	1.00	0.94	0.93	0.93	0.94	0.94	0.93	0.93	0.94
	(2.04E-04)	*(1.06)*	*(2.56)*	*(2.61)*	*(4.06)*	*(4.14)*	*(2.44)*	*(2.50)*	*(3.09)*	*(3.18)*
$85K+	0.76	1.00	0.90	0.90	0.87	0.87	0.90	0.90	0.87	0.87
	(1.98E-04)	*(1.05)*	*(3.47)*	*(3.58)*	*(6.09)*	*(6.43)*	*(2.79)*	*(2.82)*	*(4.39)*	*(4.83)*
Atlantic	0.57	1.00	1.00	0.98	0.97	0.94	1.00	0.98	0.97	0.94
	(8.60E-04)	*(1.19)*	*(2.52)*	*(3.04)*	*(4.88)*	*(5.14)*	*(2.24)*	*(2.77)*	*(4.37)*	*(4.27)*
Quebec	0.62	1.00	0.94	0.94	0.96	0.97	0.94	0.94	0.96	0.97
	(3.41E-04)	*(1.05)*	*(3.12)*	*(3.05)*	*(4.24)*	*(3.99)*	*(2.99)*	*(3.00)*	*(3.79)*	*(3.77)*
Ontario	0.68	1.01	0.91	0.91	0.92	0.93	0.91	0.91	0.92	0.93
	(1.86E-04)	*(1.02)*	*(3.78)*	*(3.91)*	*(5.58)*	*(6.30)*	*(3.31)*	*(3.38)*	*(4.27)*	*(4.61)*
Prairies	0.60	0.99	1.02	1.01	0.98	0.98	1.02	1.01	0.98	0.98
	(4.22E-04)	*(1.06)*	*(3.17)*	*(3.04)*	*(4.54)*	*(4.25)*	*(2.90)*	*(2.86)*	*(3.96)*	*(3.73)*
B.C.	0.67	1.00	0.95	0.95	0.92	0.89	0.95	0.95	0.92	0.89
	(4.98E-04)	*(1.11)*	*(3.18)*	*(3.55)*	*(4.51)*	*(5.14)*	*(2.91)*	*(3.11)*	*(4.25)*	*(4.73)*
CFM	0.61	1.00	0.93	0.94	0.97	0.97	0.93	0.94	0.97	0.97
	(3.35E-04)	*(1.04)*	*(3.73)*	*(3.78)*	*(4.80)*	*(5.08)*	*(3.27)*	*(3.36)*	*(5.14)*	*(5.33)*
Offline	0.64	1.00	0.95	0.94	0.86	0.85	0.95	0.94	0.86	0.85
	(4.43E-04)	*(1.04)*	*(4.17)*	*(4.06)*	*(7.55)*	*(7.03)*	*(3.32)*	*(3.21)*	*(6.04)*	*(5.86)*
Online	0.65	1.00	0.96	0.95	0.94	0.94	0.96	0.95	0.94	0.94
	(1.18E-04)	*(1.07)*	*(2.77)*	*(2.92)*	*(3.92)*	*(4.16)*	*(2.41)*	*(2.54)*	*(3.06)*	*(3.36)*

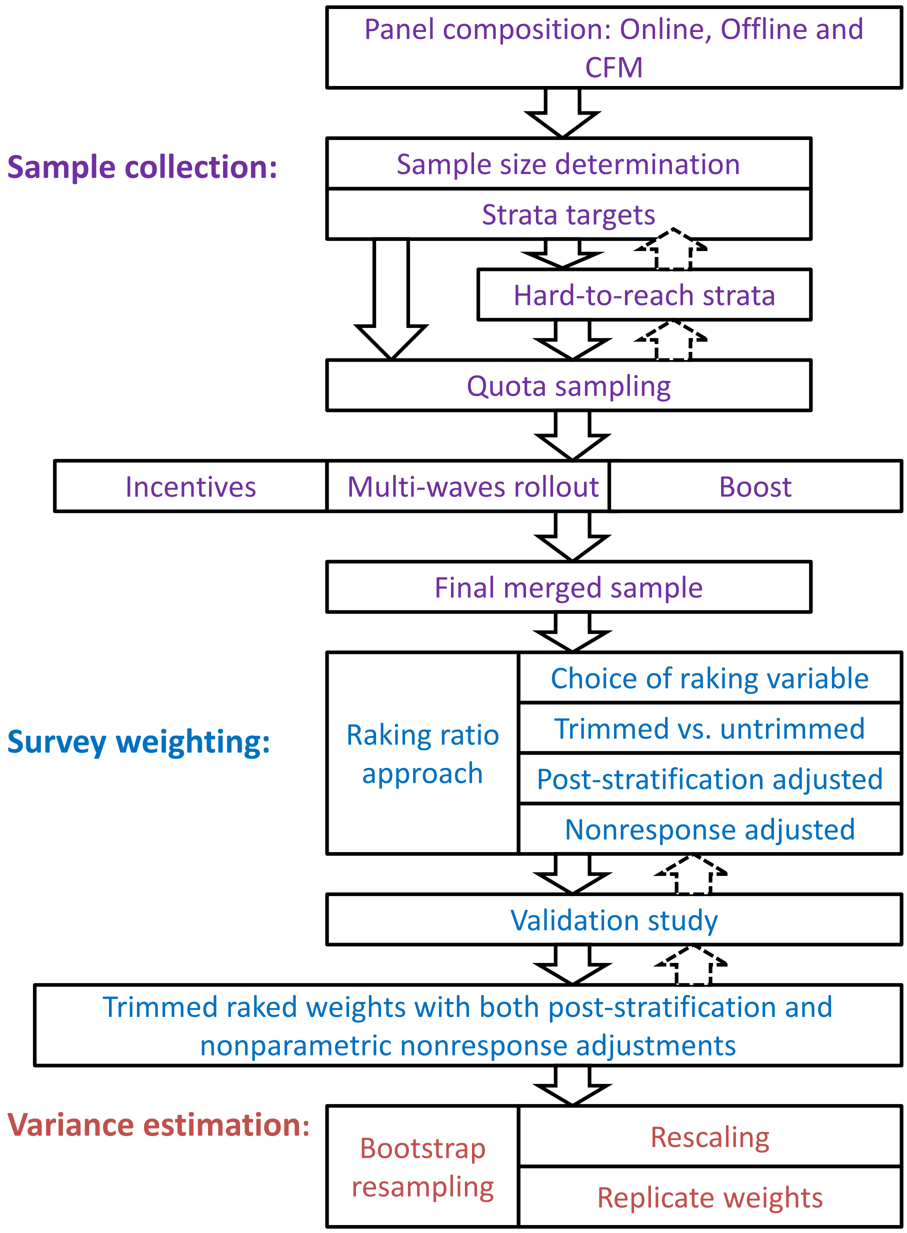
Notes: All columns after the first have been divided by the value in the first column. Each row shows weighted point estimates, with corresponding variance estimates below in parentheses. Linearization estimates are produced with the linearization procedure in Stata, which simply assumes a stratified random sample and does not take into account the weighting procedure. Resampling estimates are computed as described in Section 4.1 using bootstrap replicated survey weights. All the raked weights are trimmed at five times their mean and obtained with a tolerance level of 0.01 as convergence criteria. Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Table 15: Credit card ownership in SFS and the 2017 MOP Survey weighted sample

		Trimmed				Untrimmed			
	SFS	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}	Rk,1	Rk,PS	Rk,NR^{P}	Rk,NR^{NP}
Overall	89.1	88.6	88.7	88.7	88.7	88.2	88.2	88.4	88.8
Age									
18–24	85.3	68.6	69.2	69.3	69.0	71.3	71.9	73.7	75.8
25–34	89.9	90.1	89.6	89.7	89.7	91.2	90.1	88.0	87.6
35–44	90.9	90.2	90.4	90.1	90.4	90.3	90.4	89.9	90.9
45–54	90.2	92.7	92.8	92.8	92.5	92.4	92.5	93.2	93.2
55–64	90.2	92.6	92.8	92.6	92.6	91.7	91.9	89.7	89.7
65+	86.2	90.0	90.2	90.1	90.2	86.6	86.7	90.4	90.4
Region									
Atlantic	82.1	81.8	82.4	82.8	82.1	84.0	83.4	78.9	77.8
Quebec	87.3	88.3	88.3	88.5	88.6	88.8	89.2	89.5	90.3
Ontario	90.1	89.2	89.3	88.9	88.8	87.8	87.9	88.5	89.4
Prairies	91.1	89.4	89.6	89.8	90.3	90.2	90.2	92.0	92.1
B.C.	90.9	90.0	89.8	89.8	89.6	87.9	87.4	86.7	85.6
Household income									
<$25K	64.5	62.5	62.1	61.6	61.5	64.0	63.8	64.0	62.8
$25–45K	84.3	85.3	85.8	86.2	85.7	82.3	82.5	84.6	84.3
$45–65K	93.6	91.0	90.9	90.5	91.1	90.3	90.4	91.0	91.7
$65–85K	95.3	87.4	87.6	87.8	87.9	87.2	87.1	85.7	87.5
$85K+	98.7	95.3	95.4	95.3	95.2	95.4	95.4	95.3	95.5
Household size									
1	78.4	86.5	86.6	86.1	85.6	87.6	87.8	85.8	85.0
2	92.4	93.0	93.1	93.3	93.2	91.2	91.5	93.6	93.3
3	93.2	86.5	86.5	87.1	87.1	86.1	85.9	84.8	85.4
4	95.8	88.0	88.0	86.1	87.1	88.8	89.1	87.9	89.1
5+	93.6	83.7	83.9	85.3	84.8	83.3	82.4	83.9	85.5

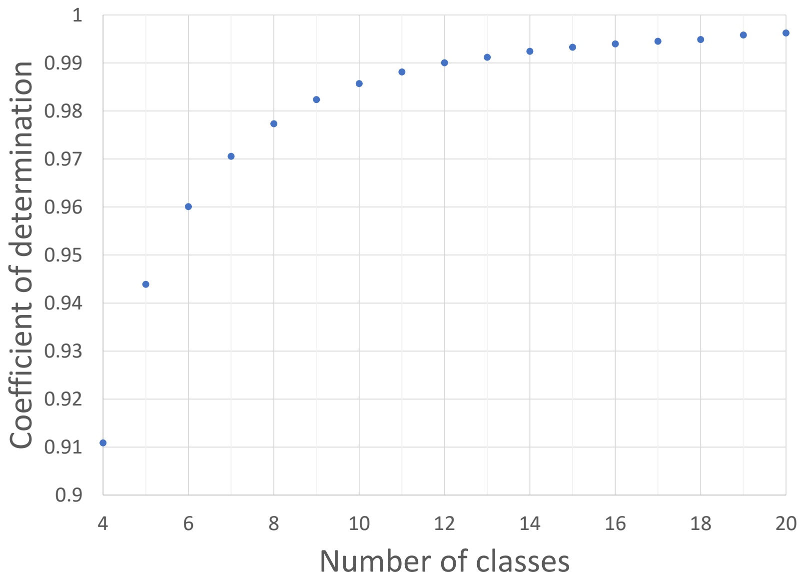
Notes: This table presents proportion estimates of credit card ownership. Proportions presented in the first column are based on household counts from the 2016 Survey of Financial Security (SFS). Columns 2 to 5 show estimates from the 2017 MOP Survey obtained with trimmed (at five times their mean) raked weights, while the following columns present estimates based on untrimmed raked weights. Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Figure 1: The 2017 MOP Survey methodology flow chart



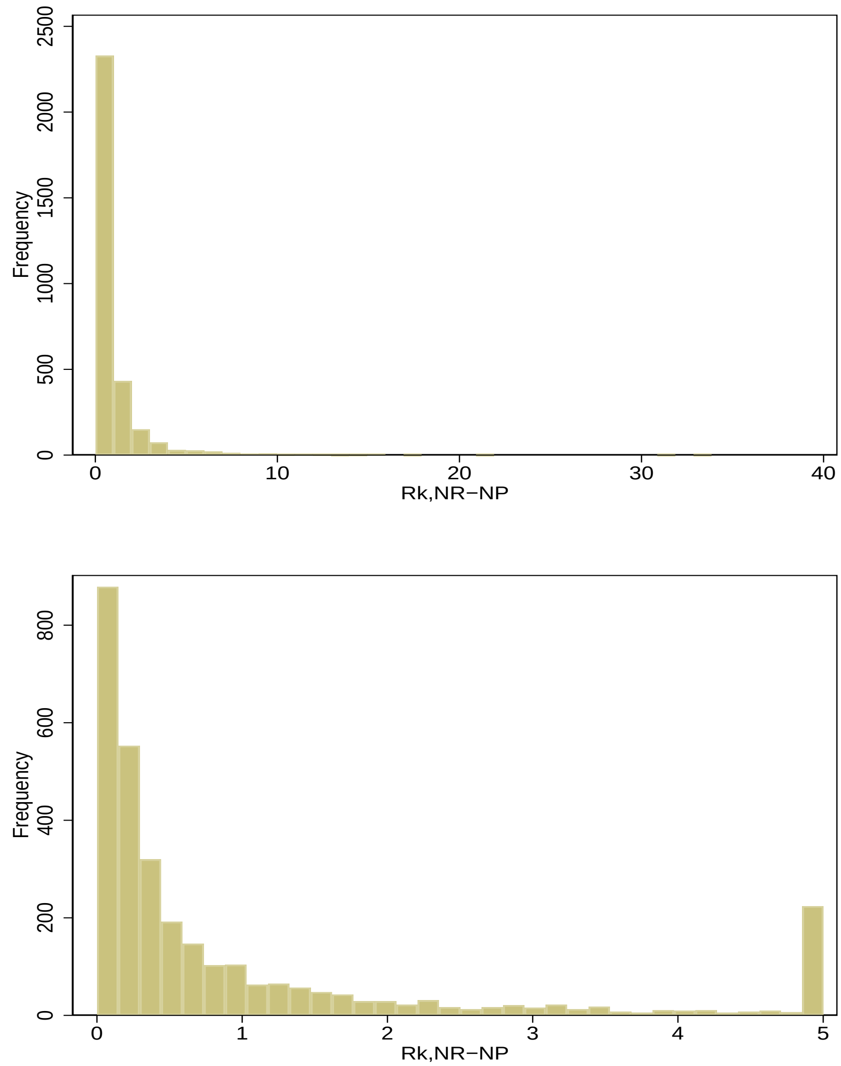
Notes: The flowchart illustrates the process of sample collection, survey weighting and variance estimation for the 2017 MOP Survey. The solid arrows indicate steps in the workflow, while the dashed arrows indicate feedback between workflow steps.

Figure 2: Empirical determination of the number of homogeneous classes for nonparametric nonresponse adjustment



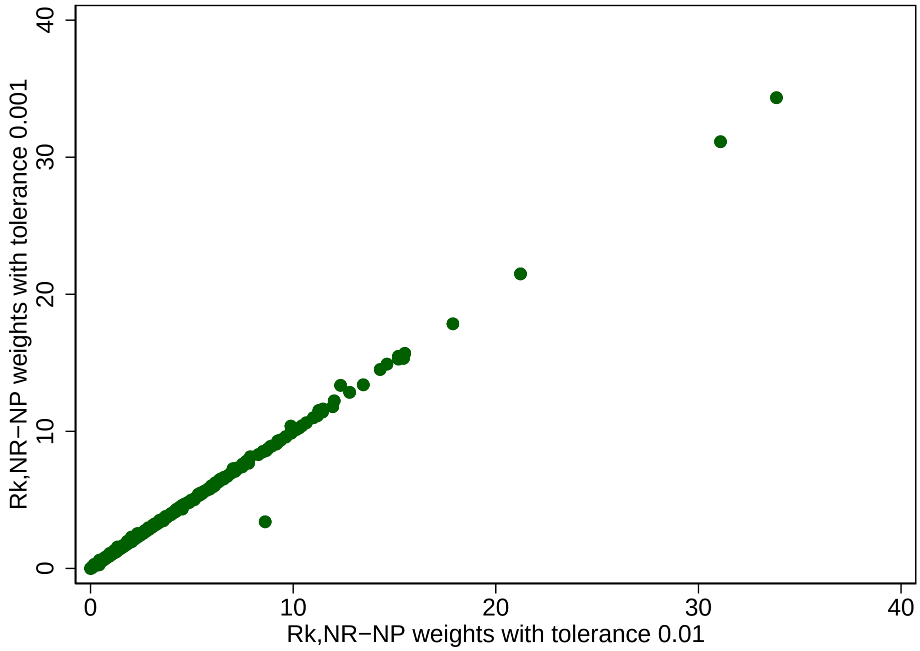
Notes: This figure shows the coefficient of determination resulting from an analysis of variance between the individual predicted response probability $\hat{p}_i$ and the class identifier variable, for a number of classes from 4 to 20. As the number of classes increases, the classes become increasingly homogeneous with respect to $\hat{p}_i$. We look for a small number of classes that leads to a sufficiently large value of R^2; see Haziza and Beaumont (2007) for more details. To derive Rk,NR$^{\mathrm{NP}}$, we use 8 classes, which gives an R^2 value just above 97.5 per cent.

Figure 3: Distribution of untrimmed vs. trimmed Rk,NR^{NP} weights



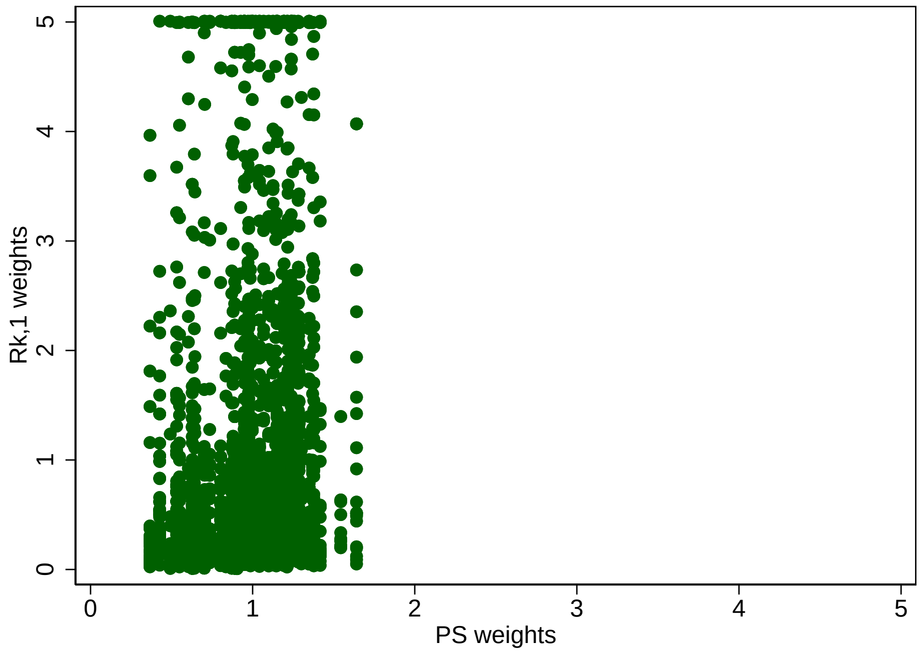
Notes: The top histogram shows raked weights Rk,NR^{NP} obtained without trimming and standardized by their mean. The bottom histogram shows raked weights Rk,NR^{NP} trimmed at five times their mean and standardized by their mean.

Figure 4: Rk,NR$^{\mathrm{NP}}$ weights with convergence tolerances at 0.01 and 0.001



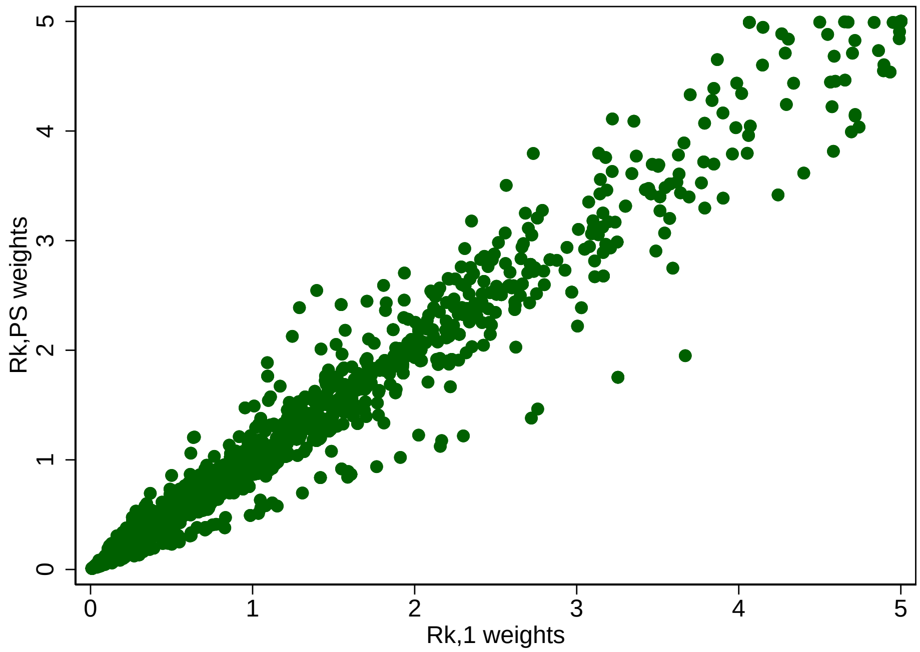
Notes: Rk,NR$^{\mathrm{NP}}$ are raked weights based on nonparametrically nonresponse-adjusted PS weights. Both sets of weights are untrimmed and have been standardized by the mean of the x-axis variable.

Figure 5: Correlation between PS and Rk,1 weights



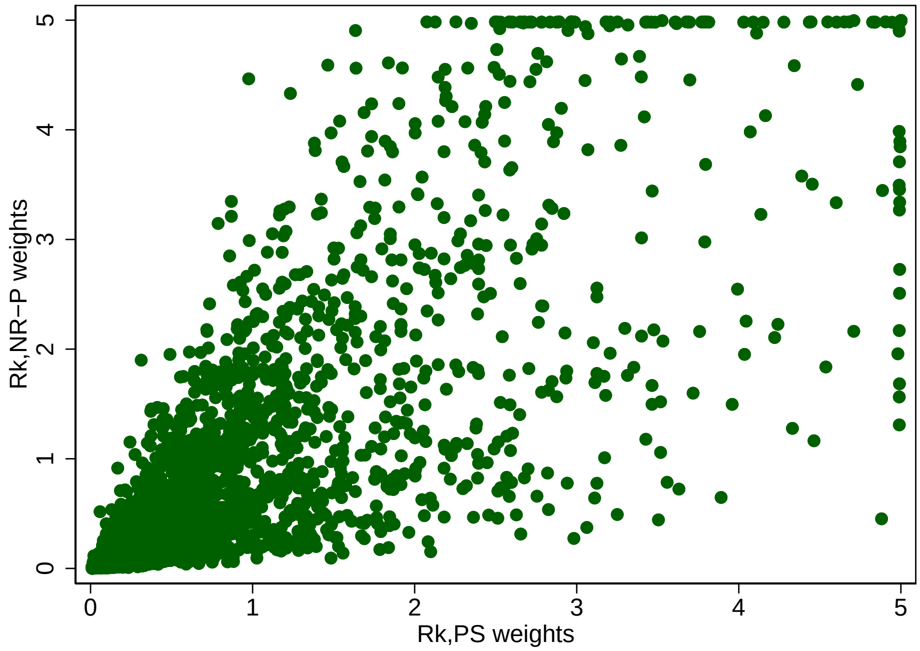
Notes: PS are the post-stratified weights without raking. Rk,1 are raked weights with initial weights equal to one and trimmed at five times their mean. Both sets of weights are standardized by the mean of the x-axis variable. The linear correlation coefficient is 0.28.

Figure 6: Correlation between Rk,1 and Rk,PS weights



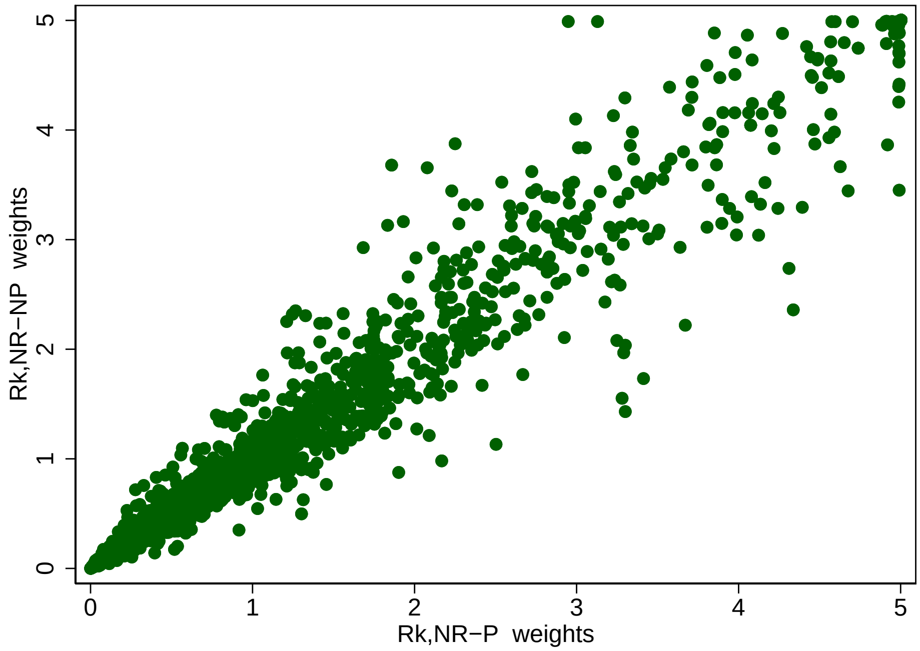
Notes: Rk,1 are raked weights with base weights equal to one. Rk,PS are raked weights based on PS initial weights. Both sets of weights are trimmed at five times their mean and are standardized by the mean of the x-axis variable. The linear correlation coefficient is 0.92.

Figure 7: Correlation between Rk,PS and Rk,NR^{P} weights



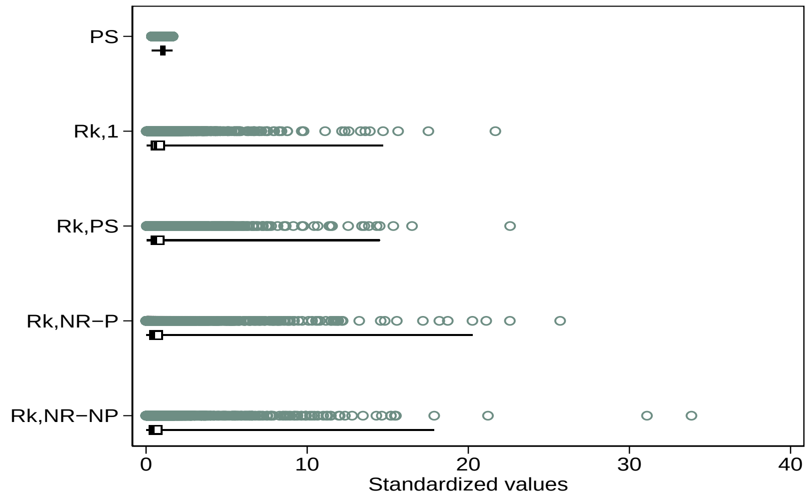
Notes: Rk,PS are raked weights based on PS weights. Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights. Both sets of weights are trimmed at five times their mean and are standardized by the mean of the x-axis variable. The linear correlation coefficient is 0.64.

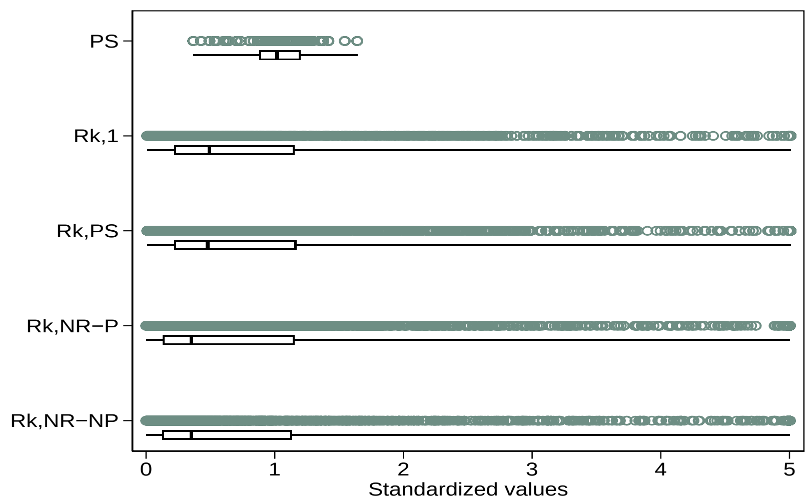
Figure 8: Correlation between Rk,NR^{P} and Rk,NR^{NP} weights



Notes: Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights. Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights. Both sets of weights are trimmed at five times their mean and are standardized by the mean of the x-axis variable. The linear correlation coefficient is 0.91.

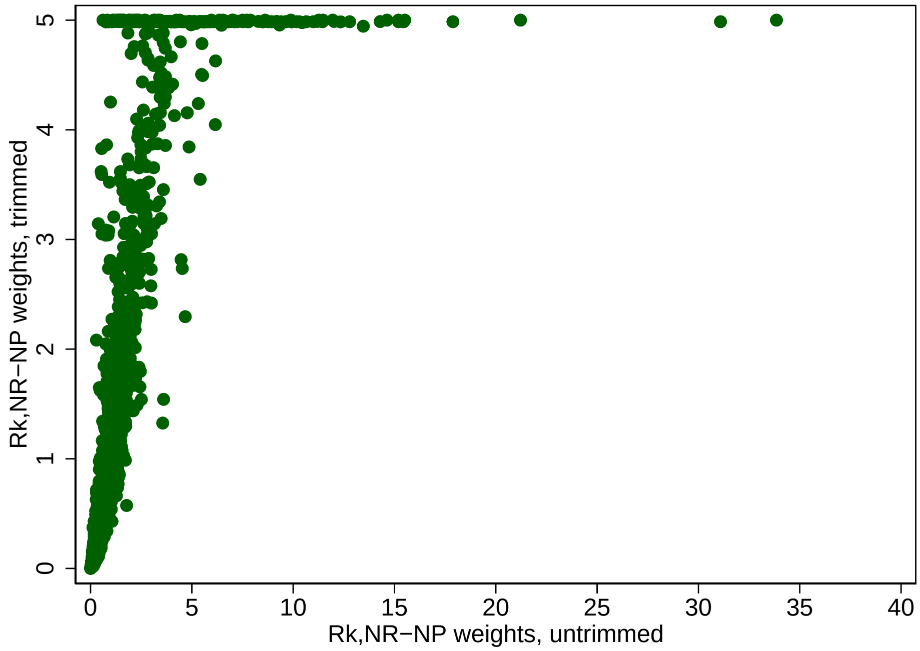
Figure 9: Box plots of untrimmed vs. trimmed weights





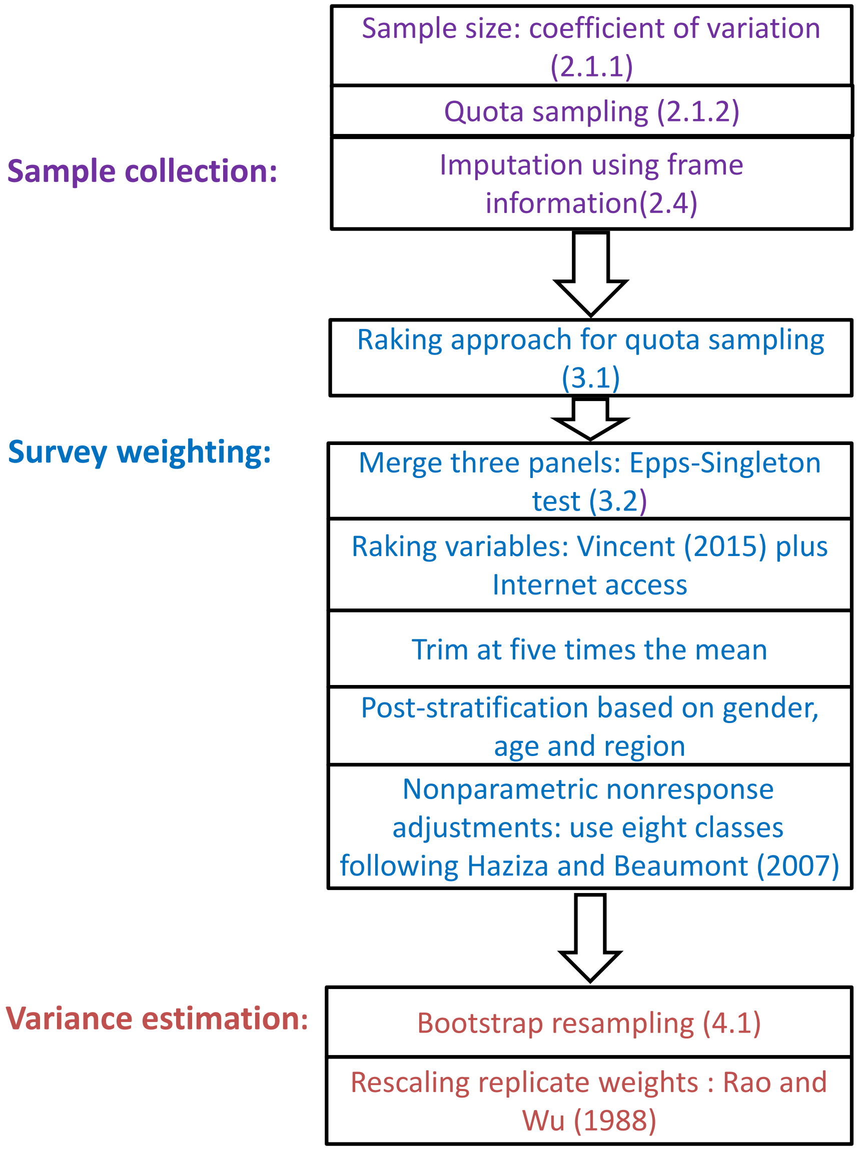
Notes: For each set of weights, a dot plot of the values is shown above a box-and-whisker plot. The whiskers end at the 0.1 and 99.9 percentiles. The top set of box plots shows untrimmed weights, while the bottom one shows trimmed weights. In each case, the scale of the x-axis has been standardized by the mean of the Rk,NR^{NP} weights. PS are the post-stratified weights without raking; Rk,1 are raked weights with base weights equal to one; Rk,PS are raked weights based on PS initial weights; Rk,NR^{P} are raked weights based on parametrically nonresponse-adjusted PS initial weights; and Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS initial weights.

Figure 10: Correlation between trimmed and untrimmed Rk,NR^{NP} weights



Notes: Rk,NR^{NP} are raked weights based on nonparametrically nonresponse-adjusted PS weights. Both sets of weights have been standardized by the mean of the x-axis variable.

Figure 11: The 2017 MOP Survey technical details



Notes: This flowchart summarizes details of the implementation of sample collection, survey weighting and variance estimation for the 2017 MOP Survey. The solid arrows indicate steps in the workflow.