Measuring Consumer Cash Holdings: Lessons from the 2013 Bank of Canada Methods-of-Payment Survey
Heng Chen, Chris Henry, Kim P. Huynh, Q. Rallye Shen, Kyle Vincent

Keywords: and variance estimation, raking, calibration, sample weighting

Introduction

The Bank of Canada has an interest in understanding the levels of cash holding as it is the sole issuer of Canadian bank notes. Measuring the amount of cash holdings is difficult, however, as cash is an anonymous payment method. Therefore, in 2009, the Bank of Canada undertook a Methods-of-Payment (MOP) survey. The 2009 MOP survey introduced a three-day payment diary, which served as a memory aid to record all payments. This payment diary methodology has been successfully used in some other Organisation for Economic Co-operation and Development (OECD) countries; more details are available in Bagnall et al. (forthcoming). Further, Arango, Huynh, and Sabetti (2015) find that demographic factors are strongly correlated with cash usage.

In 2013, the Bank of Canada conducted another MOP survey to measure cash usage in Canada. The 2009 and 2013 MOP surveys involve both paper and online collection methods. The paper or offline sample was recruited via regular post mail while the online sample was selected from a market access panel accessible by e-mail. The 2013 MOP contains a subsample of the offline sample taken from another comprehensive annual household survey, the Canadian Financial Monitor (CFM), with approximately 12,000 Canadian households per annum. The CFM survey instrument collects information that complements and overlaps with the MOP, providing a complete picture of household finances. About 3,600 surveys were collected for the 2013 MOP from across the country and then weighted to ensure that the sample is proportionally representative of the Canadian population with respect to certain demographic variables.

The next section discusses survey planning and sample weighting for the 2013 MOP. The Variance estimation section discusses bootstrap resampling methods for variance estimation. Finally, the article ends with the Conclusion.
Planning

The 2013 MOP survey is an update to the 2009 MOP. For further information on the 2009 MOP survey, refer to Arango and Welte (2012); for the 2013 MOP, see Henry, Huynh, and Shen (2015). One important lesson from the 2009 MOP is that certain subpopulations were underrepresented in the final sample. This led to empty/low cell counts, which, in turn, gave rise to extreme weights. While sampling targets were achieved for marginal demographic counts, missing cell counts at a nested level caused difficulties for the weighting process.

Several measures were implemented to ensure that the sampling procedure in the 2013 MOP would avoid this problem. First, established minimum quota sample sizes were nested by region, age, and gender; for example, we specified the minimum number of males, aged 18–24, and from the Prairies region of Canada, that were required for the sample to reflect the composition of the Canadian population and meet size calculations. These predefined targets, built into the statement of services for the survey company, facilitated the ongoing monitoring of returns during data collection. Second, various levels and types of incentives were randomly offered to potential respondents, which allowed us to determine the most effective combination.

Collaboration with the survey company was important to ensure that these tools were effectively employed to hit the nested sampling targets. The survey company provided almost daily updates to establish up-to-date projections for the final returns. During data collection, certain cells were identified as potentially being under-represented in the final sample. Through timely collaboration, an additional sampling wave was added for the offline recruitment. As a result, we were able to hit (and exceed) all nested targets and ensure that no empty cells would impede the calibration.

The other main innovation of the 2013 MOP was to leverage the existing CFM survey via the survey instrument and method of data collection. Some topics in the 2009 survey, such as cash usage, were maintained in the 2013 MOP but made directly comparable to questions in the CFM. This allowed us to shorten the 2013 MOP questionnaire for such potential recruits; together with the subsampling approach of CFM respondents this proved very successful, and we were able to obtain a response rate of over 50 percent. Furthermore, the CFM survey provides an external benchmark with which to compare measures of consumer cash holdings/usage. As a result of these efforts, 90 percent of respondents were satisfied with the survey experience, and there was about 1–3 percent of item nonresponse.

Sample Weighting

Unbalanced samples typically result from a lack of a full frame or control oversampling. Sample calibration can be used to re-weight the sample to make it conform to known auxiliary totals, which, in turn, reduces nonresponse/
coverage issues. Consequently, the resulting estimators will be more efficient (Kish 1992; Särndal 2007). We make use of national-level counts based on the 2011 National Household Survey and 2012 Canadian Internet Use Survey. Below we detail the steps used to determine the calibration weights (see Vincent 2015 for details) and discuss issues of nonresponse and the utility of online-based samples.

**Calibration Analysis**

The raking ratio calibration method (Deville, Särndal, and Sautory 1993) is chosen for (1) its popularity amongst statistical agencies (Särndal 2007), and (2) its prior use in the 2009 MOP (Sanches 2010).

Stage A: We consider a set of candidate calibration variables based on their conjectured relationship with important survey questions. Missing entries are imputed with the R “mic e” package (van Buuren 2012). The sample is drawn from three distinct frames. Hence, we must determine a suitable method of combining the subsamples. Due to similarities in missing rates and response distributions of the offline and CFM subsamples, we directly combine these. As invites are initially sent to these subsamples and de-duplicated for the online invites, we have a nonoverlapping dual frame study. We perform an analysis suggested by Brick et al. (2006) and Young et al. (2012) and find that “merge and rake together” is preferred to the “rake separately and then merge together” method, via comparing the means and variances of these two options.

Stage B: We use the polychoric correlation measure (Drasgow 2004; Fox 2010), to measure the association between candidate variables and key survey responses. Combinations of variables that associate well with responses are retained, and we nest pairs of well-correlated variables.

Stage C: Initial design weights based on a simple and stratified random sampling design are proposed. The correlation of the resulting sets of weights was high. Hence, we conclude that there is minimal benefit to exploring the use of heterogeneous initial weights.

Scores were given for different combinations of calibration variables based on:

1. Estimation ability of well-approximated population totals (from larger surveys),
2. Design effect, and
3. Distribution of resulting weights.

A final set of calibration variables/weights is determined based on ordinal rankings of the scores.
**Weighting in the Presence of Nonresponse**

The issue of nonresponse is a common concern among survey practitioners. If nonresponse is rare, then imputation should be considered. However, determining a suitable imputation strategy can be resource intensive for comprehensive surveys.

If responses are missing completely at random (MCAR) (Rubin 1976) and occurrences are rare, then one approach is to rescale the calibration weights. One may further assume MCAR within each stratum. This assumption, missing at random (MAR), permits estimation of response probabilities.

A respondent $i$ will receive a base weight $1/P_i$ (via the sampling design) and response probability $r_{hi}$ (possibly depending on demographic profiles). The initial weight is then $w_i=1/(P_i r_{hi})$.

However, as noted by Särndal (2007), an inherited bias is likely. However, rigorous methods have been developed to approximate nonresponse bias as a function of national-level covariate information (Särndal and Lundstrom 2008).

**Online Samples**

Online surveys can be expected to oversample the more tech savvy individuals. Typical demographic variables for calibration may not account for such bias. We address this by including “online payment” and “mobile ownership” variables so that the MAR assumption is credible. For alternative methods of correcting for nonrandom selection in online surveys, Schonlau, Van Soest, and Kapteyn (2007) propose using ‘Webographic’ questions to adjust online estimates using the propensity score.

**Variance Estimation**

In order to capture the variability from both sampling design and calibration, we propose a resampling method, specifically the bootstrap replicate survey weights (BRSW) method. We prefer resampling over both the linearization estimator which requires including all strata variables, calibrated weights and design weights in the final survey (Kolenikov 2010), and jackknife methods which are inconsistent for nonsmooth functions (e.g., delete-1 jackknife for the median estimate).

The construction of the BRSW involves first re-creating the sample in each replicate and then adjusting the associated calibrated weights. For example, if a unit from a replicate is not sampled, a zero weight is assigned to it, and then the weights of the other units in the same stratum are expanded to compensate. Next, weight calibration is applied to these adjusted weights, and generating the bootstrap replicate weights. Chen and Shen (2015) provide exact details of the technical implementation.
Table 1  Mean and variance estimates for average cash holdings in 2013 MOP.

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Total</td>
</tr>
<tr>
<td>Mean</td>
<td>87.68</td>
<td>89.78</td>
</tr>
<tr>
<td>VarLin</td>
<td>15.90</td>
<td>23.58</td>
</tr>
<tr>
<td>VarBRSW</td>
<td>N/A</td>
<td>13.39</td>
</tr>
<tr>
<td>R-squared</td>
<td>N/A</td>
<td>0.033</td>
</tr>
<tr>
<td>Observations</td>
<td>3,413</td>
<td>3,413</td>
</tr>
</tbody>
</table>

Average cash holdings are measured in Canadian dollars. Statistics are based on respondents reporting positive figures (i.e., excluding zero responses). VarLin and VarBRSW are the linearized and the bootstrap replicate survey weights variance estimates, respectively. R-squared is the goodness of fit from regressing cash holdings against the raking variables.

Table 1 shows means and variances computed from the 2013 MOP data for the overall, online, offline non-CFM, and CFM samples. The overall weighted mean cash holdings estimate is $89.78. The mean cash holding estimate for online is $87.73, CFM is $87.98, and offline mean being highest at $96.90.

For comparison, we compute the variance based on linearization (VarLin) and BRSW (VarBRSW). The VarBRSW is smaller than VarLin because the resampling method takes into account the weight calibration procedure that is applied after the sample is collected. This explanation is related to a well-known fact in the literature: estimators based on the inverse of the nonparametric estimates of the propensity score, rather than based on the true propensity score, achieve the semiparametric efficiency bound (Hirano, Imbens, and Ridder 2003).

For the online and CFM subsamples, the VarBRSW are about 40 percent lower than VarLin, whereas for the offline non-CFM subsample, VarBRSW is only 4 percent lower. Although the correlation between cash on hand and the raking variables (as measured by R-squared) is stronger for offline non-CFM respondents, the larger sample sizes might drive the improvement in variance for the online and CFM subsamples.

Conclusion

Our experience from the 2013 MOP survey highlights the fact that a well-thought-out survey design and preparation along with rigorous survey weighting methodology will go a long way to ensuring that the survey is representative. Namely, we suggest that survey teams:

1. Use mixed-mode survey methods so that each mode can be used to validate and verify with external data.

2. Use the methods espoused by Dillman (2009) to induce higher response rates.
Acknowledgements

We thank our colleagues at the Bank of Canada, Marco Angrisani, Kevin Foster, Geoff Gerdes, Catherine Haggerty, Arthur Kennickell, Xumei Liu, and Marcos Sanches for their useful comments and encouragement in undertaking this survey. Maren Hansen provided excellent editorial assistance. We acknowledge the collaboration and support from Shelley Edwards, Jessica Wu, and Ipsos Reid for their dedication to this study. Finally, we thank Statistics Canada for providing access to cross-tabulations of the 2011 National Household Survey and 2012 Canadian Internet Usage Survey. The views of this paper are those of the authors and do not represent the views of the Bank of Canada.

3. Work closely with the survey firm to ensure that the objectives are laid out in advance.

4. Conduct post-stratification using a variety of methods, but ensure that the methods are robust and sensible. Use a variety of external data to conduct post-stratification.

5. Estimate variances using resampling methods, which result in lower variance estimates compared to the linearization method and provides a way to anonymize the sample data.

Survey Practice
REFERENCES


