A Nonresponse Adjustment Strategy using Modeling in the Canadian Labour Force Survey

Asma Alavi and Jean-François Beaumont, Statistics Canada
R.H. Coats Bldg., 16th floor, Tunney's Pasture, 120 Parkdale, Ottawa, Ontario, Canada, K1A 0T6
asma.alavi@statcan.ca

KEY WORDS: Nonresponse Adjustment Classes, Reweighting; Modeling, Logistic Regression, Contact Information

Abstract
Despite good sampling design and surveyors' best efforts to avoid nonresponse, it is an inevitable part of survey sampling. To offset the effects of nonresponse, reweighting of respondents is done to take into account the nonresponding part of the sample. The current reweighting method in the Canadian Labour Force Survey (LFS) creates nonresponse adjustment classes based on design variables and calculates a nonresponse adjustment factor within each class.

Our approach uses logistic regression modeling to form nonresponse adjustment classes (see for reference Little 1986, Eltinge and Yansaneh 1997, and Haziza *and al.* 2001). We model the sampled dwellings using contact information, such as the number of attempts to contact a sampling unit and time of the attempt, along with the design information. The results show significant improvement over the current method used in the LFS, in terms of several diagnostics.

We have also investigated the effect of separately modeling the probability of being contacted and the probability of response given a contact. With this approach, the final response probability is obtained as the product of the probabilities resulting from the two models. We have used different diagnostics to evaluate if this approach performs better than the more traditional one.

1. Introduction
Nonresponse and methods dealing with nonresponse have increasingly become a standard part of survey sampling. The ideal situation of all sampled units responding has little practical relevance. The prevalence of sample surveys have increased the response burden and hence occurrence of more nonresponse in the surveys. A nonresponse occurs in a survey when, for any reason, a selected unit does not respond. The usual methods of estimation, in the presence of nonresponse, give biased results, as it is generally the case that nonrespondents differ from the respondents in terms of the variables of interest.

The best way to deal with the issue of nonresponse is to make every effort, such as allowance of follow-ups and call-backs, at the design and development stages to avoid it. This comes with a high cost in terms of resources, either human or fiscal. The other possibility is the use of elaborate and extensive data collection and estimation methods, such as subsampling of respondents and randomized response, which make the effect of nonresponse negligible. These also have resource related issues. Hence, the most widely used scenario is to treat nonresponse, once it has been observed, in a way that results in estimators without too much bias. Under this scenario, which comes after the data collection phase, the aim is to get a complete data set and then use typical estimation methods.

Two main types of nonresponse are: unit nonresponse and item nonresponse. A unit nonresponse occurs when, for a variety of reasons, the sampled unit fails to respond. So apart from the design or contact information, no data at all are available for the unit. An item nonresponse is the term used when the data are missing for at least one, but not all the survey components for a particular unit. In this paper we are looking at unit nonresponse only.

2. Background
The Canadian Labour Force Survey is a monthly rotating panel survey that interviews approximately 54,000 dwellings each month. The target population consists of all non-institutionalised civilians of at least 15 years of age, who live in the ten provinces of Canada. The selected dwellings remain in the sample for six consecutive months.

The LFS is the only official source of labour force statistics such as national and provincial unemployment rates. The LFS employs a stratified multi-stage design to sample dwellings. The first stage of sampling consists of selecting smaller geographical areas, called clusters, from within each stratum. The second stage of sampling consists of selecting dwellings from within each selected cluster. The LFS information is gathered for all the eligible household members residing in the selected dwelling, using proxy-based interviewing.

The LFS uses computer assisted interviewing (CAI), employing both telephone (CATI) and personal (CAPI) interviewing methods. Contact information such as the number of attempts to contact a dwelling, and the time and the day of attempt are recorded for each dwelling. Once the data have been collected, they are checked for discrepancies and omissions. For the non-responding dwellings which have data available from a previous month, the records are imputed using either carry-forward imputation or hot-deck imputation. Responding dwellings are reweighted to compensate for the remaining nonresponding dwellings. This reweighting is based on the assumption that the responding and nonresponding dwellings have same characteristics within reweighting or nonresponse adjustment classes.

The current strategy for reweighting in the LFS creates nonresponse adjustment classes called nonresponse areas based on the design information. The design variables are the high income strata, the employment insurance economic regions, the type of design, and the rotation number of the dwelling. Each high income stratum constitutes a nonresponse area, and the cross classification of the other three design variables forms the remaining nonresponse areas. Currently, there are about 900 such nonresponse areas in the LFS. Within each nonresponse area, the observed weighted response rate is determined and the design weight of the responding dwellings is divided by the observed weighted response rate. This way the weights of the responding dwellings are inflated to take into account the nonresponse. The nonresponding dwellings treated through reweighting are then dropped from the sample.

Although the current method of reweighting seems in general satisfactory, there is always room for refinement and improvement. Two drawbacks of the current method are the creation of a large number of classes, as mentioned above, and no use of contact information. One disadvantage of a large number of classes is the increased possibility of a low within-class response rate that would result in a large weight associated with the responding dwellings in that nonresponse adjustment class. Large weights associated with dwellings could substantially increase the sampling variance of the estimators. Currently nonresponse adjustment classes with nonresponse adjustment factor of more than 2 (response rate less than 50%) are merged to reduce the size of the adjustment factor and keep the weights under control. Another problem with a large number of classes is the non robustness of the results, as a small change in the configuration of class can substantially alter the weights. The disadvantage of not using the contact information is the potential waste of information that is very much related to the process of responding or nonresponding for the survey.

3. Modeling Approach

The current method has been in use for a couple of past redesigns of the LFS. A redesign of the LFS is in order after each decennial census of population. Since the last census occurred in 2001 a redesign of LFS is underway. This redesign provides a good opportunity to redefine reweighting methods used in the LFS, since we now have more information available regarding the collection stage, which can be used as auxiliary information.

We suggest the use of modeling to determine the nonresponse adjustment or weighting adjustment classes. Since the response variable is binary (a dwelling responds or not), we use logistic regression. Instead of just using the design information as mentioned above, we add two more variables: the number of attempts to contact a dwelling and the start time of the last attempt. Thus

$$log\left(\frac{\hat{p}(\mathbf{x})}{1-\hat{p}(\mathbf{x})}\right)=\hat{\boldsymbol{\beta}}'\mathbf{x},$$

where $\hat{p}(\mathbf{x})$ is the estimated conditional mean of response variable (probability or propensity of response) given $\mathbf{x}$, the vector of auxiliary variables. This is what we call the one-step modeling approach.

We started with a model with six main effects namely: province (10 categories); start time of the last attempt (5 categories); number of attempts (5 categories); the type of design (9 categories); rotation group (6 categories); whether a stratum is high income or not (2 categories); and all the first-order interactions (15). We used SAS to do a stepwise regression to choose the model. The process was repeated with the LFS data from several months. We chose the final model with main effects and interactions that were present and most significant in various months. The final model contained the five main effects (see appendix) excluding the high income stratum variable and four interactions (interaction between number of attempts and rotation, interaction between number of attempts and start time of the last attempt, interaction between number of attempts and province, and interaction between province and start time of the last attempt).

After the final model selection, we obtained the estimated response probabilities for each dwelling resulting from the final logistic model. We applied PROC FASTCLUS in SAS to form nonresponse adjustment classes homogeneous with respect to the estimated response probabilities. The process was conditioned to have at least 20 respondent dwellings in each class. Then, within each class, a weighted response rate was calculated to obtain the nonresponse adjustment factor. The nonresponding dwellings were dropped from the sample and the weights of the responding dwellings were inflated by their corresponding nonresponse adjustment factors. We made nonresponse adjustment classes to obtain robustness against a model failure. At the same time we wanted to retain the high predictive power of the original model.

We also investigated the effect of separately modeling the probability of being contacted and the probability of response given a contact. With this two-step approach, the final response probability is obtained as the product of the probabilities resulting from two separate models. The first step was the logistic regression of contact on various auxiliary variables similar to those defined for the one-step modeling of response, and then a separate modeling of response given contact was established in a second step. A detailed comparison of the two-step modeling with the one-step modeling was undertaken. Although the two-step model was better in terms of Cox-Snell maximum re-scaled (generalized coefficient of determination) R^2 (see Cox and Snell (1989)), and predictive power, etc. the differences were not large especially when we consider the complexity of the two-step modeling. Henceforth, we will only consider the one-step model.

4. Results

In this section we will compare the results obtained using the current methodology of nonresponse adjustment in the LFS to those based on modeling as described in Section 3.

A number of diagnostic measures have been obtained to compare the current (implicitly assuming it is a modeling method too with nonresponse areas as classes) and modeling methods of nonresponse adjustment in the LFS. These include generalized coefficient of determination (R^2), the Hosmer-Lemeshow goodness-of-fit test, and the distribution of the nonresponse adjustment factor. The Hosmer-Lemeshow goodness-of-fit statistic is obtained by calculating the Pearson chi-square statistic from the 2×10 table of observed and expected frequencies, where 10 is the number of groups used in SAS. Table 1 presents the R^2 values for the current and modeling methods for selected months along with the largest nonresponse adjustment factor.

Table 1: Diagnostics for Current and Modeling Methods

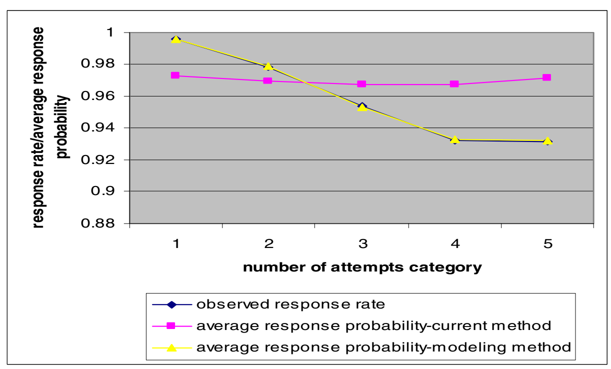
	Cox-Snell R^2		Maximum Nonresponse Adjustment Factor	
	current	modeling	current	modeling
Mar., 2001	0.1242	0.2085	3.6667	2.8323
Jun., 2001	0.1209	0.2145	1.8458	2.0850
Oct., 2001	0.1312	0.2355	1.6667	2.4301

Hosmer and Lemeshow (1980) proposed a goodness-of-fit test for binary data in a logistic regression framework. For data from various months, we tested the final logistic model for goodness-of-fit and found 91% of the time the model fit quite well. For example, for October 2001, the Hosmer-Lemeshow test statistic was 4.7362 with 8 degrees of freedom and had a p-value of 0.7854.

The modeling methodology of nonresponse adjustments in LFS seem to work better in explaining the nonresponse pattern of the observed data. The auxiliary variables based on contact information are very significant and improve the model. It was found that the current and modeling methodologies produce identical results in estimating the observed response rates for the categories of geographical and design variables such as province, rotation group, and type of design. On the other hand, as indicated in Figures 1 and 2, the modeling method estimates the observed response rates with high precision for contact variables, while the current method fails to do so. Figure 1 shows these differences for the number of attempts categories, from smallest to largest (for detail see appendix) and Figure 2 displays results for the provinces of Canada from east to west. The observed response rate is what the name suggests the observed response rate for various categories of a given variable based on June 2001 data.

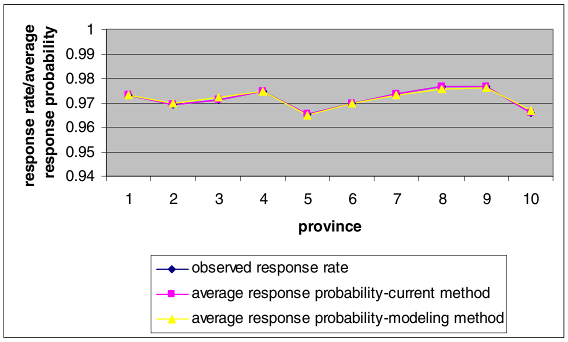
We obtained national and provincial unemployment rates for a series of months using LFS data with weights adjusted for nonresponse using the current and modeling methods. Relative differences in unemployment rates derived from the two methods of nonresponse adjustment were calculated, taking as the base the unemployment rates obtained from the current method. Over the time period considered, the relative differences vary from -1.4% to 1.4%. Hence the two methodologies produce unemployment rates that are very close to each other. Similar comparisons were made when the nonresponse adjusted weights were also calibrated. This is likely due to the low nonresponse rate in the LFS. With a higher nonresponse rate, we would probably observe larger differences.

Figure 1: Observed Response Rate Compared with Estimated Response Probabilities from Current and Modeling Methods for number of attempts variable



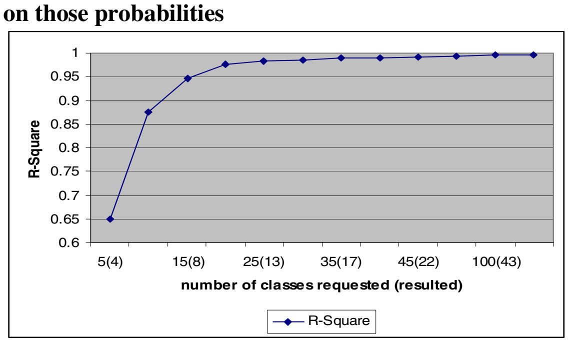
As mentioned earlier, the number of classes resulting from the current method of nonresponse adjustment is around 900 per month. On the other hand, the average number of classes resulting from the modeling method is around 50 per month. Figure 3 shows the effect of increasing the number of nonresponse adjustment classes on $\tilde{R}^2$, the coefficient of determination for the model in which the estimated response probability for each dwelling is the dependent variable and the class corresponds to the independent variable, for one month (June 2001). This graph shows how homogeneous the nonresponse adjustment classes are with respect to the estimated response probabilities. The term 'requested' is used for the number of classes requested in the SAS program, and the term 'resulted' is the number of classes made by SAS. It is clear that 40 to 50 classes achieve sufficient homogeneity.

Figure 2: Observed Response Rate Compared with Estimated Response Probabilities from Current and Modeling Methods for Provinces



We looked at the variability of the nonresponse adjustment factor for reweighting, resulting from the two approaches for various months of the LFS data. For example, for one month it was found that the factor from the current methodology has a smaller variance than the modeling method (0.0064 compared to 0.0085), but the reverse is true for range (3.05 compared to 1.39). The pattern is typical for the range of months considered. It is not surprising that the modeling method has better predictive power and produces the nonresponse adjustment factors that have more variability but are more restricted.

Figure 3: R-square between the estimated probability of response and nonresponse adjustment classes based on those probabilities



We did an analysis of variance with the estimated mean of the individual response probabilities (estimated response probabilities) as the dependent variable and the labour force status of the individuals (employed, unemployed, and not in labour force), which is the variable of most interest in the LFS, as the explanatory variable. We found that the estimated mean of the response probabilities obtained from either the current or the modeling approach, was not very much correlated with the labour force status. We obtained the coefficient of determination $\tilde{R}^2$, which is the square of the coefficient of linear correlation, between the estimated response probabilities and the mean of the estimated response probabilities within each level of the labour force status. For example for October 2001, the $\tilde{R}^2$ values were respectively 0.0000 and 0.0025 for analysis based on the current and modeling methods. The F-test that resulted from the analysis of variance had a p-value of 0.2048 for the current method and 0.0001 for the modeling method, suggesting that response probabilities obtained from the modeling method have different mean values for various labour force classifications.

Another diagnostic to compare the two methods was based on the measure of change in the weights. This test is described in detail in Dufour, Gagnon, Morin, Renaud, and Särndal, 2001. If we define Initial weight as the design weight before the nonresponse adjustment, Intermediate weight as the weight after nonresponse adjustment but before calibration, and the Final weight as the weight after the calibration then the measure of change D is defined as

$$D = R_{01} + R_{12} + R_{int} + G,$$

where R_{01} measures the individual weight changes between the initial and intermediate sets of weights, R_{12} measures the individual weight changes between the intermediate and final weight set, R_{int} measures the interaction between the two types of change and G measures the change in the average weight between the initial and final sets. Table 2 presents the measure of change for three months. According to the empirical results, the bigger the value of D, and more specifically R_{01}, the better the method is in reducing the nonresponse bias. As can be seen from Table 2, differences in measures of change between the current and modeling method seem to be very small, although the modeling method has consistently higher values of D and R_{01}.

Table 2: Measure of Change for Current and Modeling Methods

	Current		Modeling	
	D	R_{01}	D	R_{01}
Mar.,2001	0.0734	0.0059	0.0818	0.0137
Jun., 2001	0.0552	0.0013	0.0613	0.0076
Oct., 2001	0.0609	0.0010	0.0649	0.0050

Finally, we did a simulation study to compare the nonresponse bias and variance based on the current and modeling methods. We took one month respondents data of the LFS, treating the estimated response probabilities from the two-step model as the true values, and generated the nonresponse using Poisson sampling. We repeated the process 100 times to get 100 pseudo samples. Then, we applied the current and modeling nonresponse adjustment techniques separately on each pseudo sample and made nonresponse adjustment classes, did reweighting and obtained the unemployment rates (without calibration, based on the weights that are nonresponse adjusted for nonresponse).

We then obtained the estimates of the Relative Bias (RB) and Relative Root Mean Square Error (RRMSE) for both methods. The estimated RB is defined as

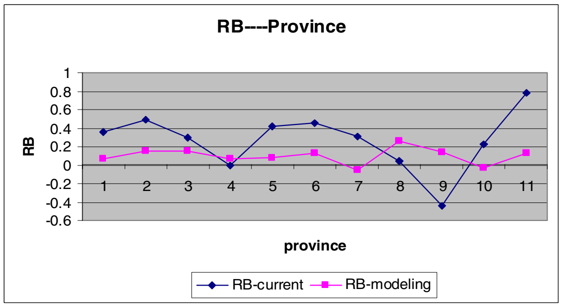
$$est(RB) = \left[\left(\frac{1}{100}\right)\sum_{i=1}^{100}\left(\hat{\theta}_i - \theta\right)\right]\frac{1}{\theta}\times 100\%$$

and the estimated RRMSE as,

$$est(RRMSE) = \sqrt{\frac{\sum_{i=1}^{100}\left(\hat{\theta}_i - \theta\right)^2}{100}}\times\left(\frac{1}{\theta}\right)\times 100\% ,$$

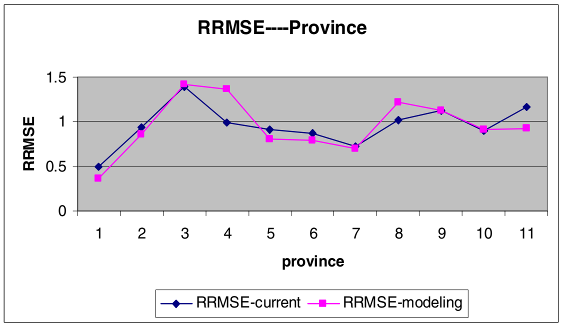
where $\hat{\theta}_i$ is the estimate of unemployment rate for a given domain after the reweighting from i^{th} pseudo sample and θ is the unemployment rate based on the design weight (before nonresponse adjustment) from the original LFS sample.

Figure 4: Comparison of RB of Unemployment Rates for Current and Modeling Methods, for June 2001



The domains considered were provinces and employment insurance economic regions. Figures 4 and 5 present the comparisons for provinces.

Figure 5: Comparison of RRMSE of Unemployment Rates for Current and Modeling Methods, for June 2001



It is clear that for most provinces the modeling method reduces the nonresponse bias. In addition the RRMSE from the two methods is almost the same except for a few provinces with a higher modeling RRMSE. This pattern points to the fact that variance of nonresponse adjustment factor is higher for the modeling method as we have more variability of weights. We also compared the RB and RRMSE for other months and the same pattern shown in Figures 4 and 5 emerged.

5. Conclusions

The current methodology of nonresponse adjustment of weights for unit nonresponse in the LFS has been compared with a new methodology based on modeling of response using logistic regression.

It was found that the current and modeling methods of creating nonresponse adjustment classes estimate the response rate of dwellings in different domains with equal precision when domains are based on geographical or design variables. On the other hand, the modeling method estimates the observed response rates with higher accuracy in domains based on design based or contact information.

Various diagnostics have shown the general superiority of modeling method over the current method. The next step would be to further improve the logistic regression model, in terms of adding more contact information and refining the variables already in the model, used in creating nonresponse adjustment classes. Also, a plan to write specifications of the modeling method is underway.

Appendix

Following is the detailed information on categories of the five main effects used in one-stage modeling.

Province = 1, Newfoundland
Province = 2, Prince Edward Island
Province = 3, Nova Scotia
Province = 4, New Brunswick
Province = 5, Quebec
Province = 6, Ontario
Province = 7, Manitoba
Province = 8, Saskatchewan
Province = 9, Alberta
Province = 10, British Columbia

1 = January or July rotation
2 = February or August rotation
3 = March or September rotation
4 = April or October rotation
5 = May or November rotation
6 = June or December rotation

if type of stratum = 0 then type of design = 1;
if 1 <= type of stratum <= 9 then type of design = 2;
if 11 <= type of stratum <= 19 then type of design = 3;
if 21 <= type of stratum <= 29 then type of design = 4;
if 31 <= type of stratum <= 39 then type of design = 5;
if 41<= type of stratum <= 59 then type of design = 6;
if type of stratum = 61 then type of design = 7;
if 65<= type of stratum <= 98 then type of design = 8;
if type of stratum = 99 then type of design = 9;

if # of attempt = 1 then # of attempt category = 1;
if # of attempt = 2 then # of attempt category = 2;

if 3 <= # of attempt <= 5 then # of attempt category = 3;
if 6 <= # of attempt <= 10 then # of attempt category = 4;
if 10 < # of attempt then # of attempt category = 5;

if midnight <= start time of the last attempt < 11:00 am then time=1;
if 11:00 am <= start time of the last attempt < 2:00 pm then time=2;
if 2:00 pm <= start time of the last attempt < 5:00 pm then time=3;
if 5:00 pm <= start time of the last attempt < 7:00 pm then time=4;
if 7:00 pm <= start time of the last attempt < midnight then time=5

References

Cox, D.R. and Snell, E.J. (1989), "*The Analysis of Binary Data*", Second Edition, London: Chapman and Hall.

Dufour, J., Gagnon, F., Morin, Y., Renaud, M., and Särndal, C.-E. (2001), "A Better Understanding of Weight Transformation through a Measure of Change", *Survey Methodology*, 27, 97-108.

Eltinge, J. L., Yansaneh, I. S. (1997), "Diagnostics for Formation of Nonresponse Adjustment Cells, With an Application to Income Nonresponse in the U.S. Consumer Expenditure Survey", *Survey Methodology*, 23, 33-40.

Haziza, D., Chow, O., Charbonnier C., Beaumont, J.-F. (2001), "Construction of Imputation Cells for the Canadian Labour Force Survey", *2001 Proceedings of the Symposium*, Statistics Canada.

Hosmer, D. W., Lemeshow, S. (1980), "A Goodness-of-Fit Test for the Multiple Logistic regression Model", *Communications in Statistics*, A10, 1043-1069.

Hosmer, D. W., Lemeshow, S. (1989), *"Applied Logistics Regression",* Wiley Series in Probability and Mathematics.

Lafortune, Y. (2002), "*Diagnostic Analysis of Logistic Regression Models With survey Data*", SSMD-2002-006E, Methodology Branch Working Paper, Statistics Canada

Little, R. J. A. (1986), "Survey Nonresponse Adjustment for Estimate of Means", *International Statistical Review*, 54, 139-157.

Statistics Canada. (1998), "*Methodology of the Canadian Labour Force Survey",* Catalogue no. 71-526-XPB