Latent Variable Models for Nonresponse Adjustment and Calibration October 2015

Brian Meekins
Bureau of Labor Statistics, 2 Massachusetts Ave, NE, Washington, DC 20212

Abstract
Traditional weight adjustments for survey sampling error are often constructed through multiple stages, where design weights are based on the inverse of the probability of selection, and in a separate stage nonresponse adjustments are derived from weighting cells or classes, or based on model-deduced response propensities. More recent efforts by Little and Vartivarian (2003) have advocated the use of propensity models that incorporate both design information, as well as variables that are, ideally, related to both nonresponse and the survey outcome. There is often a third stage of adjustment that involves calibration to known or reliable population totals. It would be useful to incorporate this calibration stage into a propensity model containing the design information and variables related to response behavior. This can be accomplished via a latent constructs that are constrained (by totals or proportions) to the external information being used. By simultaneously estimating the response propensity under calibration and incorporating design variables, additional variance due to adjustment would be minimized.

Key Words: survey nonresponse, propensity model, Latent Class Analysis

1. Motivation

Unit nonresponse is an ever increasing problem in sample surveys. Response rates have continued to fall, putting more strain on adjustment procedures. When unit nonresponse is not completely at random (MCAR) we can expect some amount of bias to be introduced in our survey estimates. Corrections for this bias are usually made by weighting responding sample units to match the sampling frame or population as closely as possible.

One methodology that is recently gaining purchase in the survey sampling community is propensity weighting. This technique regresses the log odds of the response on sample frame information that is available for both respondents and nonrespondents. It then utilizes the inverse of the predicted outcomes (the "propensity" to respond) from this logistic model as weights. One advantage of these propensity models over classic weighting adjustment using adjustment cells, is that they can easily incorporate auxiliary information with continuous distributions. However, it is often the case that these models do not fit very well, and succeed in decreasing bias only modestly at the expense of increasing weight variability to an unacceptable degree. Nonresponse adjustment weights based on response propensity are therefore frequently trimmed or attenuated in some way.

Calibration weighting is also frequently used to adjust for nonresponse. With calibration the responding sample units are adjusted to known population totals. For example, sample units may be weighted to reflect the race distribution in the sample region based on U.S.

Census estimates. If more than one group of population totals is used like employment, and the intersection of these two groups is unknown in the population (for example, one does not know the population total for unemployed whites) than iterative proportional fitting, or generalized raking, is typically utilized to calculate the nonresponse weights. In this case of calibrating to known population totals the weights typically do not utilize any information from respondents only.

Many surveys use both propensity weighting and calibration to adjust for nonresponse. This is typically done in two steps where the two independently calculated weights are multiplied together. While this does not affect the ability of the weights to adjust for bias, it can increase the variability of the weights dramatically and, in some cases, increase the mean square error to levels larger than the contribution of the bias one seeks to correct.

2. Model

A possible solution to the potential increase in the variability of the overall weights due to estimating the calibration weights in a separate step from the propensity weights is to estimate a propensity model under the constraints of the calibration variables population totals using latent variable models. In this way, the proposed latent variable model extends the basic logistic propensity model to incorporate population information for calibration.

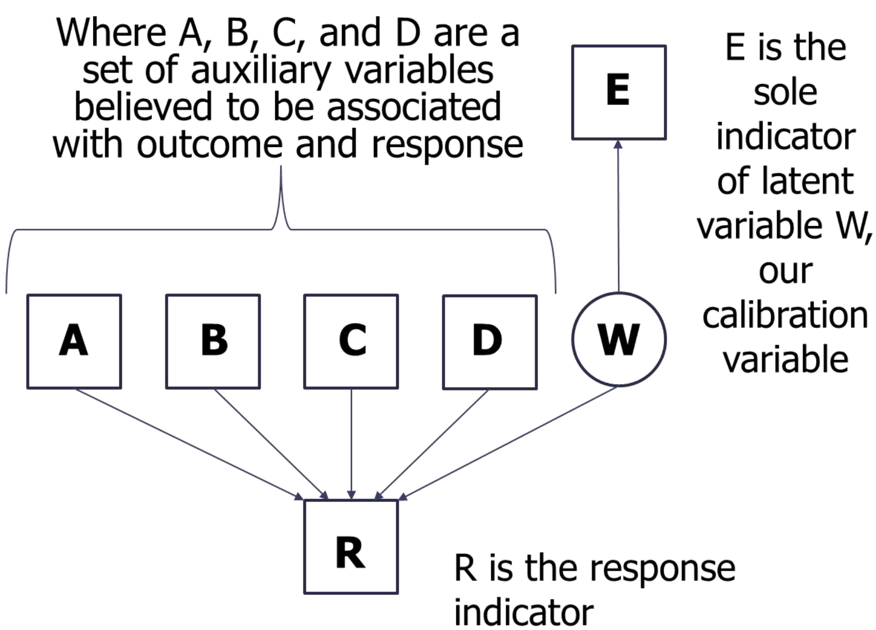

Figure 1: Latent variable model incorporating a single calibration variable and four auxiliary variables from sample frame

Figure 1 illustrates the proposed latent variable model. In this diagram *R* is the response indicator variable; *A, B, C,* and *D,* are auxiliary variables available on the sampling frame; and *E* is an indicator of the latent variable *W*. The latent variable *W* is constrained to match known population totals, where *E* is the corresponding sample totals. In cases where calibration is conducted with information from only respondents, an extra category is simply added to E to include the unit nonresponders. There is no extra category for W in this case. This effectively serves to reclassify missing sample units to categories based on population values. Unlike usual calibration procedures, this parameterization takes into account measurement error.

2.1 Adjustment Weights

Adjustment weights are the inverse of the probability of response given the auxiliary information and the latent variable. For the previous example shown in *Figure 1* the weights (w_i) are calculated by:

$$w_i = 1 / P(R=1|ABCDW)$$

Calculating propensity and calibration weights in a single step should decrease, in general, the variance of the weights – if the estimation of the model itself does not add a significant amount. These models are easy to estimate using existing software such as, Lem, Mplus, LatentGold, LISREL, and R.

3. Simulations

Simulations were conducted in order to compare the properties of the weights calculated using the latent variable model to those of a two-step procedure where a logistic propensity model is estimated separately from the calibration step. In the two-step procedure the weights are then combined by simply multiplying the two. This procedure is fairly common in large surveys.

3.1 Simulation Details

The sample size for the simulated data is 8,000 with a 60% response rate. Because the latent variable models took a fairly long time to estimate (~5 minutes), only 100 iterations for each condition was completed. Lem, a free latent variable software written by Jeroen Vermunt, was used to estimate the latent variable models. To avoid local maxima each model was run 10 times with the best fitting run selected for each of the 100 iterations. A number of conditions were varied. Here we summarize the set of conditions where we vary the association of the auxiliary and calibration variables with response and a hypothetical variable of interest, or outcome variable.

3.2 Results

Table 1 through *Table 5* summarize the findings from the simulation studies. Due to space constraints only selected results are included here. Listed in each table are the mean (across the iterations) of the bias, the variance of y, the Mean Square Error (MSE), and the mean variance of the weight variable. For each of these, a lower value is preferable. The "true" distribution of the five class calibration variable, as well as the distribution for the weighted sample for the two-step estimation procedure and the latent variable model is included at the bottom of each table.

Table 1: Auxiliary Information Strongly Related to Both Response and Outcome

Estimates

	Two-Step	*Latent*
Mean bias	-349.83	-233.90
Mean variance y	236287.44	222105.81
Mean MSE	358734.48	287147.89
Mean variance w_i	0.329	20.592

Distribution of Calibration Variable

	1	*2*	*3*	*4*	*5*
True	.15	.15	.25	.30	.15
Two-Stage	.15	.15	.25	.30	.15
Latent	.14	.14	.36	.22	.14

Table 1 shows the results of the comparison of the two approaches when the auxiliary information and calibration variables are strongly related to both response and the outcome variable. There is some modest improvement in the mean bias and a slight improvement in the variance of the weighted outcome variable using the latent variable model. These improvements are quite small and it should be noted that distribution of the calibration variable is somewhat different for the latent variable model.

Table 2: Auxiliary Information Strongly Related to Response Only

Estimates

	Two-Step	*Latent*
Mean bias	-476.66	-478.47
Mean variance y	159800.63	172433.27
Mean MSE	387048.08	401920.35
Mean variance w_i	0.331	20.518

Distribution of Calibration Variable

	1	*2*	*3*	*4*	*5*
True	.15	.15	.25	.30	.15
Two-Stage	.15	.15	.25	.30	.15
Latent	.14	.09	.33	.30	.15

Table 2 shows the results of the comparison when the auxiliary information and calibration variables are strongly related to response but not related to the outcome variable. There is almost no difference in the bias of the estimate and only a modest difference in the MSE, due to the latent variable model producing weights that increase the variability of the outcome variable. Once again, the distribution of the calibration variable is incorrect.

Table 3: Auxiliary Information Strongly to Outcome Only

Estimates

	Two-Step	*Latent*
Mean bias	-591.63	-602.50
Mean variance y	229303.97	348480.91
Mean MSE	579364.57	734456.59
Mean variance w_i	0.001	97.457

Distribution of Calibration Variable

	1	*2*	*3*	*4*	*5*
True	.15	.15	.25	.30	.15
Two-Stage	.15	.15	.25	.30	.15
Latent	.20	.24	.11	.21	.25

Table 3 shows the results of the comparison when the auxiliary information and calibration variables are strongly related to the outcome variable only but not related to response. While the bias is only modestly larger for the latent variable model, the variance of the outcome variable is quite a bit larger for the latent variable approach. The distribution of the calibration variable is also very different than the true distribution.

Table 4: Auxiliary Information Not Related to Response or Outcome

Estimates

	Two-Step	*Latent*
Mean bias	505.51	506.57
Mean variance y	159843.47	264519.70
Mean MSE	415411.46	522210.28
Mean variance w_i	0.001	85.563

Distribution of Calibration Variable

	1	*2*	*3*	*4*	*5*
True	.15	.15	.25	.30	.15
Two-Stage	.15	.15	.25	.30	.15
Latent	.27	.28	.10	.17	.17

Table 4 shows the results of the comparison when the auxiliary information and calibration variables are not related to either response or the outcome variable. As in previous models, the difference in the mean bias between the two weighting procedures is negligible. However, in this case the MSE and the variance of the outcome variable is significantly different - the latent variable model producing weights that greatly increase the variance. In addition, the distribution of the calibration variable is unacceptably different from the true distribution.

Table 5: Auxiliary Information Strongly Related to Both Response and Outcome (Calibration Variable Present for Respondents Only)

Estimates

	Two-Step	*Latent*
Mean bias	169.24	165.96
Mean variance y	235808.40	496202.70
Mean MSE	264520.96	525533.30
Mean variance w_i	0.33266	344002

Distribution of Calibration Variable

	1	*2*	*3*	*4*	*5*
True	.15	.15	.25	.30	.15
Two-Stage	.15	.15	.25	.30	.15
Latent	.14	.18	.52	.03	.13

Table 5 shows the results of the comparison when the auxiliary information and calibration variables are strongly related to both the response and the outcome variable. However, in this case, the calibration variable is only available for respondents rather than the entire sample. The parameterization of *E* in *Figure 1* has an additional class which includes the unit nonrespondents. This seemed to affect the estimation of *W* (the latent construct), such that the distribution is quite different from the true distribution. The variance of the outcome variable is quite large, as well, leading to a very high MSE. However, like many of the models the bias is very similar between the two weighting methods.

4. Summary and Discussion

Decreasing response rates has led most survey practitioners to incorporate a variety of weighting techniques to adjust for possible bias introduced from nonresponse. Propensity model based weighting and calibration are two popular techniques that utilize auxiliary information from the sample frame and known population totals. Combining both of these procedures into one step should lead to a decrease in the variance of the adjustment weights and the weighted response variable.

However, the proposed method, fitting the data with a latent variable model, produced mixed results. Only in the case where auxiliary and calibration variables are strongly related to both response and the outcome variable did the latent variable model reduce both the bias and variance of the weighted outcome compared to that of the two-step method. Indeed, while the bias was comparable regardless of the relationship of the model information to response and the outcome variable, the variance of the outcome variable (and therefore the MSE) was significantly increased in most cases.

It is not particularly surprising that when the model information is not related to response that latent variable models produce weights that are not optimal in terms of their variance. Like propensity models – but even more so – poorly fitting or misspecified latent variable models produce highly variable estimates. Thus, the same criticisms levied at propensity models apply to latent variable models, with consequences that appear to be more severe.

Much more testing needs to be conducted. In particular research should be aimed at estimating latent variable models in such a way that they can be as robust to misspecification as logistic propensity models – if not more so. It would be interesting to include more than one calibration variable and compare that to a generalized raking procedure. In addition, we could parameterize the latent variable model to assume no measurement error in the calibration variable. In this case, we would specify that the probability of the calibration variable indicator being a certain value “x” given the calibration latent variable is the same value “x” is one, or

$$P(E=x|W=x) = 1$$

It would also be interesting to incorporate the outcome variable into the model as an additional latent construct. This should help to reduce bias, and, to the extent the auxiliary information is related to outcome, reduce variability in the weighted outcome variable.

Acknowledgements

Thanks to Jeroen Vermunt for his assistance with specifying the model in Lem.

References

Deville, C., and Särndal, C-E. 1992. "Calibration Estimators in Survey Sampling." *Journal of the American Statistical Association*, **87**, 376 – 382.

Kott, P. S., and Chang, T. 2010. Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse. *Journal of the American Statistical Association*, **105**(491), 1265-1275.

Kott, P. S. and Liao, D. 2015. "One Step or Two? Calibration weighting from a Complete List Frame with Nonresponse." Survey Methodology, **41,** 165-181.

Little, R. and Vartivarian, S. 2003. "On Weighting the Rates in Non-Response Weights." *Statistics in Medicine*, **22**, 1589 – 1599.

Lundström, S., and Särndal, C-E. 1999. Calibration as a Standard Method for Treatment of Nonresponse. *Journal of Official Statistics*, **15**, 305-327.

Vermunt, J.K 1997. *LEM 1.0: A General Program for the Analysis of Categorical Data*. Tilburg: Tilburg University.

Vermunt, J.K. 1997. *Log-linear Models for Event Histories*. Advanced Quantitative Techniques in the Social Sciences Series, Volume 8, 348 pages, Thousand Oaks: Sage Publications.