Testing For Informative Weights And Weights Trimming In Multivariate Modeling With Survey Data

Tihomir Asparouhov[1], Bengt Muthen[2]
Muthen & Muthen[1]
UCLA[2]

Abstract

Analyzing the informativeness of the sampling weights can lead to significant improvement in the precision of model estimation with survey data. A test for weights ignorability was proposed in Pfeffermann's (1993). We propose a modification of this test which improves its performance for small and medium sample size problems. We also generalize the test to a test of equivalence between two different sets of sampling weights, which can be used to test the informativeness of individual weight components. We evaluate the performance of these techniques in simulation studies based on linear regression and multivariate factor analysis models. We also apply the test of equivalence to the problem of finding the optimal level of weight trimming and illustrate this approach with a practical example. We describe the implementation of these techniques in the software package Mplus.

KEY WORDS: Weights Informativeness, Weight Trimming, Test of Weights Ignorability.

1. Overview

Large variation in the sampling weights is a major problem when analyzing survey data. Such variation increases dramatically the variability of the parameter estimates. Bias reductions gained by using the sampling weights can easily be eliminated by the larger variability of the weighted estimates. Using the sampling weights will result in an increase of the mean squared error of the estimates. In addition, sampling weights with large variability increase finite sample size biases because general asymptotic results will then typically require larger samples. Sampling weights could also be non-informative. In such a case using the sampling weights will increase the variability of the estimates without reducing the bias. Therefore its imperative that test for informativeness of the sampling weights is conducted in all practical applications. While many software packages include facilities for model estimation with sampling weights, standardized tools for testing the informativeness of the weights are not included.

In practical applications sampling weights are computed as the product of different components. For example sampling weights obtained by post-stratification will typically be computed as the product of different stratification variables such as SES, race and age. Each of the components could be informative or not-informative. It is imperative to analyze each of these components separately for informativeness. Including non-informative components in the sampling weights will simply decrease the precision of the estimates.

In this article we investigate the performance of the Pfeffermann's (1993) test for informativeness of the sampling weights. The test can be used for univariate as well as multivariate models, single level and multilevel level models. The test applies to the pseudo maximum-likelihood (PML) estimator described by Skinner (1989) and implemented in the software package Mplus (Muthen & Muthen 1998-2007), which we used for all computations in the article. We conduct simulation studies for univariate, multivariate, single level and two-level models using both informative and non-informative sampling weights.

We also provide a modification of Pfeffermann's test for informativeness which improves the test performance for small and medium sample size. This modified test can also be used to determine the informativeness of weight components as well as to compare two different weight variables. The test can be used for determining the optimal level for weights truncation.

2. Testing The Informativeness Of The Sampling Weights

Consider a model with p parameters $\theta = (\theta_1, ..., \theta_p)$ and let $\hat{\theta}$ be the maximum likelihood estimates of θ. Let $\hat{\theta}_w$ be the PML estimates when sampling weight variable w is included in the estimation. Suppose that $V(\hat{\theta})$ and $V(\hat{\theta}_w)$ are the corresponding variance estimates for the θ and θ_w. Pfeffermann (1993) proposed a simple method for testing the informativeness of the sampling weights. Under the null hypothesis of non-informative weights the following test statistic T has a chi-square distribution with p degrees of freedom

$$T = (\hat{\theta}_w - \hat{\theta})[V(\hat{\theta}_w) - V(\hat{\theta})]^{-1}(\hat{\theta}_w - \hat{\theta})^T \sim X^2(p). \quad (1)$$

This is because under the null hypothesis the variables $\hat{\theta}$ and $\hat{\theta}_w - \hat{\theta}$ are asymptotically independent. For finite sample size however this may not be so and the variance of $\hat{\theta}_w - \hat{\theta}$ may be quite different from $V(\hat{\theta}_w) - V(\hat{\theta})$. In fact for small sample size frequently $V(\hat{\theta}_w) - V(\hat{\theta})$ will not be a positive definite matrix and in that case the value of T may be negative. In such cases we interpret the test as accepting the null hypothesis with p-value of 1. We call this test the Pfeffermann's test of ignorability (PTI).

We now derive a modified Pfeffermann's test of ignorability (MPTI) to achieve two different goals. The first goal is to provide a non-zero estimate for the covariance of $\hat{\theta}$ and $\hat{\theta}_w - \hat{\theta}$ could enable us to improve the finite sample size performance of the PTI. The second goal is to generalize the test to the case when we compare the parameter estimates based on two different sampling weights w_1 and w_2. When one of the sampling weights is $w_1 = 1$, we essentially will give an alternative test for ignorability for the other weight w_2. Denote by $\hat{\theta}_{w_1}$ the PML parameter estimates based on sampling weights w_1 and by $\hat{\theta}_{w_2}$ the PML parameter estimates based on sampling weights w_2. The null hypothesis we want to test is that both sampling weights lead to consistent parameter estimates, i.e., the ratio $f = w_2/w_1$ is ignorable weights component. First we derive the joint distribution of $\hat{\theta}_{w_1}$ and $\hat{\theta}_{w_2}$. Denote by l_i the log-likelihood for the $i-$th unit in the sample and w_{1i} and w_{2i} the two sampling weights for that unit. Let $L_j = \sum_i w_{ji} l_i$ be the weighted log-likelihood, $j = 1, 2$, for the two sampling weights. Let T_j, $j = 1, 2$, be the score equations used to derive the PML estimates

$$T_j = \frac{\partial L_j}{\partial \theta} = \sum_i w_{ji} \frac{\partial l_i}{\partial \theta}. \tag{2}$$

By definition $T_j(\hat{\theta}_{w_j}) = 0$. Let $T = (T_1, T_2)$ and $\hat{\theta}_w = (\hat{\theta}_{w_1}, \hat{\theta}_{w_2})$. Thus $T(\hat{\theta}_w) = 0$. Using the linearizaton method we can obtain the distribution of $\hat{\theta}_w$

$$0 = T_j(\hat{\theta}_{w_j}) \approx T_j(\theta_0) + (\hat{\theta}_{w_j} - \theta_0)\frac{\partial T}{\partial \theta}. \tag{3}$$

$$\hat{\theta}_{w_j} - \theta_0 \approx \left(\frac{\partial T_j}{\partial \theta}\right)^{-1} T_j(\theta_0) \approx \tag{4}$$

$$\left(\frac{\partial^2 L_j(\hat{\theta}_{w_j})}{(\partial \theta)^2}\right)^{-1} \sum_i w_{ji} \frac{\partial l_j(\hat{\theta}_{w_j})}{\partial \theta}. \tag{5}$$

where θ_0 is the true parameter value. Therefore an estimate for the variance of $\hat{\theta}_w$ is given by

$$V(\hat{\theta}_w) = \begin{pmatrix} V(\hat{\theta}_{w_1}) & C \\ C & V(\hat{\theta}_{w_2}) \end{pmatrix} \tag{6}$$

where

$$C = \left(\frac{\partial^2 L_1(\hat{\theta}_{w_1})}{(\partial \theta)^2}\right)^{-1} \cdot M \cdot \left(\frac{\partial^2 L_2(\hat{\theta}_{w_2})}{(\partial \theta)^2}\right)^{-1\ T} \tag{7}$$

$$M = \sum_i w_{1i} w_{2i} \frac{\partial l_i(\hat{\theta}_{w_1})}{\partial \theta} \Big(\frac{\partial l_i(\hat{\theta}_{w_2})}{\partial \theta}\Big)^T \tag{8}$$

Thus the variance estimate for $\hat{\theta}_{w_1} - \hat{\theta}_{w_2}$ is

$$V = V(\hat{\theta}_{w_1}) + V(\hat{\theta}_{w_2}) - 2C. \tag{9}$$

Note that this variance estimate is positive definite even for small sample size. Under the null hypothesis of non-informativeness of $f = w_{2i}/w_{1i}$ the following test statistic T has a chi-square distribution with p degrees of freedom

$$T = (\hat{\theta}_{w_1} - \hat{\theta}_{w_2}) V^{-1} (\hat{\theta}_{w_1} - \hat{\theta}_{w_2})^T \sim X^2(p) \tag{10}$$

We call this test the modified Pfeffermann's test of ignorability (MPTI).

If in addition to sampling weights, the sampling design could also include stratification and cluster sampling. Suppose that w_{1sci} and w_{2sci} are the sampling weight for individual i in cluster c in stratum s. All of the above formulas still apply with the exception of formula (8) which is modified as follows

$$M = \sum_s \frac{n_s}{n_s - 1} \sum_c (z_{1sc} - \bar{z}_{1s})(z_{2sc} - \bar{z}_{2s})^T \tag{11}$$

where

$$z_{ksc} = \sum_i w_{ksci} \frac{\partial l_{sci}(\hat{\theta}_{w_k})}{\partial \theta} \tag{12}$$

$$\bar{z}_{ks} = \frac{1}{n_s} \sum_c z_{ksc} \tag{13}$$

and n_s is the number of clusters in stratum s.

In this article we consider only single level models, however the PTI and MPTI tests apply also for multilevel models with sampling weights and the MPML estimation method (Asparouhov, 2006).

Alternative way to derive the distribution of $\hat{\theta}_w$ is to use the linearization method on the total score equation T, however the formulas become somewhat more complicated. This approach will yield yet another finite sample size approximation to the PTI test that will asymptotically be equivalent to the PTI test.

In the next section we compare the PTI and the MPTI test applied to the case when $w_1 = 1$ for testing the ignorability of the sampling weights.

3. Simulation Study For Testing Ignorability Of Sampling Weights

In this simple simulation study we compare the performance of the PTI and the MPTI tests for three simple models and various sample sizes. The first model is a univariate mean and variance estimation

$$\text{Model 1} : Y_i = \mu + \varepsilon_i \tag{14}$$

where ε_i are zero mean normally distributed residuals. We estimate two parameters μ and the variance σ of ε_i (and Y_i). We generate the data from the standard normal distribution, i.e., the true parameters values are $\mu = 0$ and $\sigma = 1$. We also generate an ignorable set of weights

$$w_i = Exp(\xi_i) \tag{15}$$

Table 1: Rejection rates for PTI and MPTI tests when weights are ignorable.

Test	Model	p	n=200	n=500	n=2000	n=10000
PTI	1	2	12%	15%	6%	3%
MPTI	1	2	10%	5%	4%	2%
PTI	2	3	19%	20%	10%	5%
MPTI	2	3	9%	10%	7%	4%
PTI	3	15	20%	40%	31%	9%
MPTI	3	15	30%	20%	9%	6%

Table 2: Rejection rates for PTI and MPTI tests when weights are not ignorable.

Test	Model	p	n=200	n=500	n=2000	n=10000
PTI	1	2	76%	85%	100%	100%
MPTI	1	2	100%	100%	100%	100%
PTI	2	3	77%	85%	99%	100%
MPTI	2	3	100%	100%	100%	100%
PTI	3	15	46%	63%	97%	100%
MPTI	3	15	100%	100%	100%	100%

where ξ_i are independent standard normal deviates. The unequal weight effect (UWE) measures the variability of the weights and is computed as follows

$$UWE = \frac{n\sum_i w_i^2}{(\sum_i w_i)^2} \approx \frac{E(w_i^2)}{(E(w_i))^2}. \quad (16)$$

Thus if w_i are generated as in equation (15) where ξ_i is a standard normal variable with variance θ the UWE effect is approximately $Exp(\theta)$, i.e., in case our case the UWE effect is approximately 2.71. This approximation is however for large samples. For small samples the UWE effect will very substantially from sample to sample.

The second model is a univariate linear regression model with one predictor variable

$$\text{Model 2}: Y_i = \mu + \beta X_i + \varepsilon_i. \quad (17)$$

The predictor variable X_i is a standard normal deviate. We estimate three parameters μ, the residual variance σ and the slope β. The true parameter values for μ and σ are as in Model 1, and $\beta = 0.5$. The third model is a multivariate factor analysis model with five dependent variables and one factor variable

$$\text{Model 3}: Y_{ji} = \mu_j + \lambda_j \eta_i + \varepsilon_{ji}. \quad (18)$$

where $j = 1, ..., 5$ and η is a standard normal unobserved factor variable. There are 15 parameters in this model μ_j, λ_j and the residual variances σ_j. The true parameter values are $\lambda_j = \sigma_j = 1$ and $m_j = 0$.

We conduct the simulation studies with four different sample sizes $n = 200$, 500, 2000 and 10000. We generate 100 samples for each sample size and conduct the PTI and MPTI test of ignorability of the sampling weights (15). We reject the true null hypothesis when the test statistic value exceeds the 95% quantile of the corresponding chi-square distribution and thus we expect that the test rejects no more than the nominal 5% of the time for sufficiently large sample size. Table 1 contains the rejection rates for this simulation study. The results suggest that both PTI and MPTI perform correctly for large sample sizes for all three models, however for small sample size both tests tend to reject incorrectly the null hypothesis more often than the nominal 5% rate. This is especially the case for the model with larger number of parameters p. The results also suggest that MPTI outperforms the PTI test for small and medium sample size.

Note that the inflated rejection rates in Table 1, while being undesirable, should not be a deterrent for utilizing the tests. The consequences from not being able to establish the non-informativeness of the weights is relatively minor. Essentially the less precise weighted estimates will be used even when the unweighed estimates are better. Not utilizing the ignorability test makes this exact error 100% certain.

To evaluate the power of the two tests we conduct a simulation study with informative weights. We generate sampling weights according to (15) again however ξ_i and ε_i are generated from a bivariate normal distribution with correlation 0.5, for Model 1 and 2. For Model 3 we generate ξ_i and η_i from a bivariate standard normal distribution with correlation 0.5. The correlation between the data in the model and the sampling weights causes the unweighed parameter estimates to be biased. Since the sampling weights are informative we expect the PTI and MPTI tests to reject the null hypothesis of ignorable weights 100% of the times. The higher the rejection rate the more powerful the test is. Table 2 contains the

Table 3: Rejection rates for MPTI tests when the two weights are equivalent.

Test	Model	p	n=200	n=500	n=2000	n=10000
MPTI	1	2	9%	9%	6%	5%
MPTI	2	3	15%	13%	13%	12%
MPTI	3	15	41%	27%	17%	11%

Table 4: Rejection rates for MPTI tests when the two weights are not equivalent.

Test	Model	p	n=200	n=500	n=2000	n=10000
MPTI	1	2	100%	100%	100%	100%
MPTI	2	3	96%	100%	100%	100%
MPTI	3	15	99%	100%	100%	100%

results of this simulation study. Clearly the MPTI test outperformed the PTI test here as well. The low power of the PTI test for smaller sample size suggest that this test could lead incorrectly to the conclusion that the sampling weights are non-informative. This could be a major drawback for using the PTI test because it could lead to less accurate estimates.

4. Simulation Study For Testing Equivalence of Two Weighted Estimates

In this section we conduct a simulation study to evaluate the performance of the MPTI for testing significant differences between two sets of weighted parameter estimates. Suppose that w_1 and $w_2 = w_1 \cdot f$ are two sets of sampling weights and we want to test the equivalence of the corresponding sets of weighted parameter estimates. This situation can arise for example when f is constructed to reduce the variability of the weights, such as a truncation factor or another weights shrinkage factor. Alternatively, the sampling weights can be obtained from a multistage sampling scheme where unequal probability of selection has been used at each stage of the sampling process. In that case the sampling weight is the product of the inverse probability of selection for each sampling stag. Another example would be post-stratification sampling weights where the sampling weight is composed of multiplicative factors, one factor for each stratification variable.

Using the same models as in the previous section we conduct simulation study with different sample sizes to evaluate the performance of the MPTI. The PTI test is not applicable for comparison of two weighted estimates. We generate the sampling weights for Models 1 and Model 2 as

$$w_{1i} = 1 + Exp(Y_i) \qquad (19)$$

and for Model 3 as

$$w_{1i} = 1 + Exp(\eta_i) \qquad (20)$$

The UWE effects for Models 1 and 3 are approximately 2.26 and for Model 2 it is 2.93. Model generated data is included in the samples with probability proportional to $1/w_{1i}$. We also generate a uninformative weight factor f according to (15) with an independent standard normal ξ_i. The second weight is computed as $w_{2i} = w_{1i}f$. We test the ignorability of f by the MPTI test. The rejection rates are given in Table 3. The results suggest that the MPTI test works correctly for sufficiently large sample size, however for models with larger number of parameters and smaller sample size a substantial deviation from the nominal rejection rate is found, which implies that in some cases informative weight factor may not be detected by this test.

As in the previous section, by introducing a correlation between ξ_i and ε_i of 0.5, for Model 1 and Model 2, and between ξ_i and η for Model 3, we obtain informative weight factor f. In this case we expect the MPTI test to reject 100% of the time. The results of this simulation study are presented in Table 4. The rejection rate is nearly 100% in all cases.

5. Application to Weights Trimming

In this section we illustrate how the MPTI can be used to select proper levels for weights trimming when the sampling weights are not ignorable but still too variable to include in the estimation. We use the following example originally presented in Chantala and Suchindran (2006). The data comes from the National Longitudinal Study of Adolescents (Add Health), a longitudinal study of adolescents in grades 7-12 during the 1994-1995 academic year. A sample of 130 schools (PSU) were chosen with unequal probability of selection. Let p_j be the probability of selection for school j. Within each school students were also selected with unequal probability. Let p_{ij} be the probability of selection for student i in school j. The total number of students in the sample are 18087. The school and individual sampling weights are available and are computed as follows

$$w_{1ij} = 1/p_{ij} \quad (21)$$

$$w_{2j} = 1/p_j \quad (22)$$

The combined weight w_{ij}

$$w_{ij} = w_{1ij} w_{2j} \quad (23)$$

can be used with the PML method to estimate population average models. In this illustration we estimate the a regression model of the body mass index of the students (B variable) on the hours spent watching TV or using computers (W variable) and the availability of a school recreation center (R variable) as well as the interaction of the two predictor variables

$$B = \mu + \beta_1 W + \beta_2 R + \beta_3 WR + \varepsilon \quad (24)$$

The model has 5 parameters: the intercept, the 3 slopes and the residual variance parameter θ thus the PMTI test has 5 degrees of freedom. The sample design includes cluster sampling. The schools represent the PSUs and therefore we facilitate formula (11) in the computation of the PMTI to account for the cluster sampling. Chantala and Suchindran (2006) consider a two level model where random intercept and random slope for W are estimated. For simplicity however in our illustration we use a single level, population average model.

The UWE for w_1 is 2.4, for w_2 it is 3.9 and for the combined weight it is 2.0. These levels of variability indicate that non-informativeness of the sampling weights can lead to poor estimates and therefore it is imperative to analyze the weights. The first step of our weights analysis is to evaluate the informativeness of the total weight variable $w_1 w_2$ as well as each of the two weights components w_1 and w_2. Table 5 shows the results of these 3 PMTI tests. Both components when tested separately appear to be non-informative, however when both components are tested simultaneously they appear to be marginally informative. It is in general unclear how to proceed in such a situation. If only one of the components was non-informative, we would have just dropped that component, but in our case it appears that both weights are slightly informative and in combination they can not be treated as non-informative. One approach to resolve this situation is to simply drop the component which is less informative, in our case the w_1 weight variable. Another approach would be to drop the w_1 component and to trim the w_2 component to the maximum level that gives non-informative reduction. A third approach is to simply trim both weights simultaneously to the maximum level that gives non-informative reduction. Here we will illustrate the last approach.

Denote w_q the $q-$th quantile of the weight variable w. Let $0 < l < u < 1$. Let the weight variable $w(l, u)$ be the trimmed weight variable at quantiles l and u

$$w(l, u) = min(max(w, w_l), w_u), \quad (25)$$

Table 5: MPTI test for informativeness of weight components

Test	w_1	w_2	$w_1 w_2$
MPTI value	4.2	7.1	12.4
p-value	0.522	0.215	0.030

Table 6: PMTI test for informativeness for weight trimming at different levels of trimming

l	u	p-value
0	0.5	0.027
0	0.7	0.050
0	0.75	0.092
0.1	0.75	0.104
0.2	0.75	0.095
0.3	0.75	0.058
0.35	0.75	0.027
0.4	0.75	0.015

i.e., the weight variable is trimmed at the upper level at quantile u and at the lower level at quantile l. Trimming the weights is the simplest way to reduce the weights variability, however other methods can be used here as well, for example the power-shrinkage method, see Chihnan (2006) et alt. We now illustrate how to select proper level of trimming. For every set of quantiles l and q we can test the informativeness of the reduction factor $f = w/w(l, u)$, by the MPTI test using $w_1 = w$ and $w_2 = w(l, u)$. If the MPTI test does not reject the hypothesis that f is non-informativeness we conclude that the weight trimming at levels l and q is appropriate. In general the p-value of the MPTI test of w v.s. $w(l, u)$ will be decreasing function of l and increasing function of u, i.e, as we trim more and more the p-value decreases. This statement is only approximately so, small deviations on full monotonicity is expected. The p-value for $w(0, 1)$ is 1. The p-value for $w(l, u)$ when $u = l$ is the p-value for informativeness of w, in our case 0.03.

Our strategy for choosing the optimal trimming level is to first trim the upper part of the weights to the lowest level with p-value above the nominal 5% value simply using a line search with step 5%. The upper trimming is followed by a similar lower trimming of the weights. In our illustration we trim both weight components at the same quantile levels however other approaches are possible too. Table 6 contains the p-values we obtained in the search for the optimal u and l values. We conclude that trimming the weights at the 0.30 and 0.75 quantiles is the optimal. The trimmed portion of the weights is not informative with p-value 5.8%. The UWE for the total weight variable is reduced from 2.03 for the original weight variable to 1.52 for the trimmed weights. This reduction indicates that the trimmed weight estimates could be

Table 7: Parameter estimates and standard errors for weighted trimmed and unweighted estimation.

parameter	weighted	trimmed	unweighted
μ	58.539(0.778)	-0.32 (0.82)	-0.50 (0.79)
β_1	0.054(0.019)	-0.11 (0.89)	0.58 (0.68)
β_2	-2.891(1.071)	0.88 (0.86)	1.39(0.81)
β_3	0.110(0.025)	-0.54 (0.88)	-1.72(0.88)
θ	813.762(11.694)	-0.41 (0.73)	-0.85 (0.69)

substantially more precise than the fully weighted estimates. Table 7 shows the parameter estimates and their standard errors for the fully weighted, trimmed, and unweighted estimation. In the fully weighted column we report the parameter estimates and the standard errors in the parenthesis. In the trimmed and unweighted column we report the standardized change in the parameter estimates and the efficiency ratio. The standardized parameter change is the parameter change divided by the weighted standard error. The efficiency ration is the ratio between the standard error for the trimmed/unweighted estimates and the weighted estimates. The results show that the trimmed estimates are quite different from the weighted estimates and the change is almost always in the direct of the unweighted estimates. The results also show that the standard errors of the trimmed and the unweighted are very similar.

6. Conclusion

In this article we investigated the performance of Pfeffermann's test of informativeness for the sampling weights and proposed a modification of this test which improves the performance for small and medium sample sizes. We also generalized the test to the situation when we need to test one set of sampling weights against another set. This generalized test can be used to test separately the informativeness of different weights components which can be useful in eliminating uninformative weight components to improve the precision of the estimates. The generalized test can also be used to select proper level for weights reduction techniques such as weight trimming or power shrinkage.

REFERENCES

Asparouhov, T. (2006), "General Multilevel Modeling with Sampling Weights," *Communications in Statistics: Theory and Methods*, **35**, 439-460.

Chantala, K. and Suchindran, C. (2006), "Adjusting for Unequal Selection Probability in Multilevel Models: A Comparison of Software Packages," *Proceedings of the Survey Research Methods Section, ASA*, 2815-2824.

Chihnan C., Nanhua D., Xiao-Li M., & Margarita A. (2006), "Power-Shrinkage and Trimming: Two Ways to Mitigate Excessive Weights," *Proceedings of the Survey Research Methods Section, ASA*, 2839-2846.

Muthen, L.K.; Muthen, B.O. (2006), *Mplus User's Guide, Fourth Edition*, Los Angeles, CA: Muthen & Muthen.

Pfeffermann, D (1993), "The Role of Sampling Weights when Modeling Survey Data," *International Statistical Review*, **61**, 317-337.

Skinner, C. J. (1989), "Domain Means, Regression and Multivariate Analysis," *Analysis of Complex Surveys (eds. C.J.Skinner, D.Holt and T.M.F. Smith)*, 59-87, Wiley.