Accounting for Complex Sample Designs in Analyses of the Survey of Consumer Finances

Article *in* Journal of Consumer Affairs · December 2016

DOI: 10.1111/joca.12106

2 authors:

Su Shin
University of Utah

Sherman D. Hanna
The Ohio State University

COMMENTARY

SU HYUN SHIN AND SHERMAN D. HANNA

Accounting for Complex Sample Designs in Analyses of the Survey of Consumer Finances

We examined the effects of using bootstrap weights to account for the complex sample design in analyses of Survey of Consumer Finances (SCF) datasets. No article published in this journal that has used the SCF has mentioned the issue of complex sample designs. We compared results obtained without weights and with application of population and bootstrap weights in a logistic regression, and found no substantial differences between the unweighted and the weighted analyses. We also compared results for an ordinary least squares regression, and found few differences between unweighted and weighted models. Unweighted regressions produced more conservative significance tests than the counterpart, and some econometricians have suggested that unweighted analyses are better for hypothesis testing. If estimation of the magnitudes of effects is important, weighted regression may be better because it produces consistent estimators. Researchers should be cautious in drawing conclusions when weighted and unweighted effects are substantially different.

Lindamood, Hanna, and Bi (2007) reviewed articles that used the Survey of Consumer Finances (SCF) datasets and appeared in the *Journal of Consumer Affairs*. They examined the papers for lack of transparency with respect to a number of methodological issues, including weighting of multivariate analyses and the use of the Repeated Implicate Inference (RII) method. However, no articles in this journal that have used SCF datasets have included a discussion on the issue of complex sample designs (Nielsen and Seay 2014; Nielsen et al. 2009). What is the effect on standard error estimates and hypotheses testing when complex sample designs are ignored? Would a consideration of complex sample designs have changed the major conclusions of analyses of SCF datasets?

For our comparisons, we used the logit model of Yuh and Hanna (2010), who analyzed a combination of the 1995–2004 SCF datasets, and

Su Hyun Shin (shin.375@osu.edu) is a Ph.D. candidate and Sherman D. Hanna (hanna.1@osu.edu) is a professor, both at The Ohio State University.

The Journal of Consumer Affairs, Summer 2017: 433–447
DOI: 10.1111/joca.12106

employed a normative economic framework to create their hypotheses. We also used an ordinary least squares (OLS) model similar to the Yuh and Hanna logit analyses. We used the 2010 SCF dataset, and slightly modified some of their specifications. We focused on providing comparisons between an unweighted multivariate analysis of a SCF dataset, one that applies population weights only, and one that uses both population and bootstrap replicate weights. We suggest guidelines for how to choose between unweighted and weighted models.

BRIEF LITERATURE REVIEW

The issue of weighting of multivariate analyses has been controversial. Winship and Radbill (1994) stated that the decision to use sampling weights in regression analysis is complicated, and unweighted regression is preferred if the sampling weights are a function of the independent variables. Deaton (1997, 66) mentioned “the old and still controversial issue of whether survey weights should be used in regression.” He stated that the answer should be based on the purpose of the regression, and that “the strongest argument for weighted regression comes from those who regard regression as descriptive, not structural” (Deaton 1997, 71). He suggested that researchers adopt the approach taken by DuMouchel and Duncan (1983) to compare weighted and unweighted estimates (Deaton 1997, 72).

Lindamood et al. (2007) compared unweighted versus population-weighted analyses of an SCF dataset for three different logistic regressions. They reported that of 99 coefficient estimates for the three logits, nine had a difference between unweighted and weighted in terms of whether the statistical significance level was less than .05 (nine were significant in the weighted model but not significant in the unweighted model). Only one of the 99 coefficients had a statistically significant effect in the unweighted estimate and a nonsignificant effect in the weighted estimate. They cited Deaton’s (1997, 66–73) discussion that weighted regression analysis is suspect for hypothesis testing on datasets with endogenous weights, and recommended that if hypothesis testing is the main research focus, unweighted regression analysis should be used for SCF datasets. In general, the unweighted analysis produced more conservative results in terms of the significance tests than the population-weighted analysis.

Nielsen et al. (2009) noted that many large survey datasets had complex sample designs, and that researchers who failed to account for the complex designs would typically increase the probability of a Type I statistical error. We could not find any articles using an SCF dataset with a detailed discussion of the issues raised by Nielsen and Seay (2014) and Nielsen et al. (2009). Nielsen and Seay (2014) listed one journal article that used the SCF and reported accounting for complex sample designs (Yao, Ying, and Micheas 2013). However, that article did not discuss the issue of population weighting, or note the effect of complex sample designs on the estimates of standard errors.

In SCF datasets, bootstrap weights are available to account for complex sample designs but previous methodological discussion of SCF estimation in this journal (Lindamood et al. 2007) has covered only the issue of unweighted analyses versus analyses using the population weight. Therefore, the objective of this research is to provide a detailed comparison of (1) unweighted regression analyses, (2) analyses using the population weight, and (3) analyses using both the population weight and the bootstrap weights. We show that researchers should use either the first approach (unweighted) or the third approach, and discuss ways to decide which approach is more appropriate.

METHODS

Dataset

This study uses 6,482 households in the 2010 SCF dataset. Yuh and Hanna (2010) did not apply population weights to their multivariate analysis, citing Lindamood et al. (2007) as justification that unweighted analysis is better for hypothesis testing. Further, they did not mention the issue of the use of bootstrapping to account for the complex sampling structure (Nielsen and Seay 2014). We present here both unweighted multivariate results and results obtained with the application of both population and the bootstrap replicate weights. In order to demonstrate the separate effect of application of the bootstrap weights, we also present results with the population weights applied but not the bootstrap weights.

Measurement of Variables

For the logit model, we use the same dependent variable as that of Yuh and Hanna (2010). This is a dichotomous measure that represents whether the household reported spending less than its income during the previous year, and is created using a SAS code provided by the Federal Reserve Board (2012). A household is considered to be a saver if the respondent reported that spending was less than income during the previous year (not including spending on investments). Many authors who have used this as a dependent variable refer to spending less than income as saving (e.g., Hogarth and Anguelov 2003). For the OLS model, we used the natural log of liquid assets as a dependent variable and the log of 0.01 for households that do not own any liquid assets.

We followed Yuh and Hanna (2010) in selecting independent variables based on a normative framework. For both the logit and OLS models, the independent variables included employment status, income and wealth-related factors, age, race/ethnicity, education, health, household type, homeownership, health insurance coverage, availability of emergency funds of $3,000 from friends/relatives, and presence of related children under age 18. Yuh and Hanna (2010) listed 14 directional hypotheses for the variables listed above, based on considerations from economic theory.

Yuh and Hanna (2010) included net worth as the natural log of net worth for net worth >0, with Ln (.01) for net worth of 0 or less. However, in 2010, 11.0% of households had negative net worth, with 13.1% having zero or negative net worth, and it might not be reasonable to assume that the effect of net worth was the same for that range as for higher levels of net worth. We tried a specification similar to that used in Hanna (2011), with two spline variables: (1) Ln(net worth) for net worth >0, Ln(.01) for net worth ≤ 0; (2) Ln ($-$net worth) for net worth <0, Ln(.01) for net worth ≥ 0. The results were somewhat similar to those in Hanna (2011), with the likelihood of saving increasing as net worth increased in the positive range, and decreasing as net worth decreased in the negative range. However, we found multicollinearity between the two variables when estimating models with bootstrapped samples, so we instead used the specification as shown in Tables 1 and 2, which allowed for a similar pattern but with the same absolute value of slopes in the positive and negative range, and also added a dummy variable for whether net worth was negative.

Analyses

We used multivariate analyses to test our empirical models, and compare the results. Researchers who use the SCF datasets must contend with two issues: missing values and heteroskedasticity (Pence 2001). For missing data, the SCF provides five imputed implicates for each household based on the conditional distribution of variables. As Lindamood et al. (2007) recommended, we use "repeated-imputation inference" (RII) techniques to estimate both unweighted and weighted models. The RII method is used to combine all five implicates of each household to better represent the

TABLE 1

Logistic Regression Analysis of Likelihood of Saving (Spending Less than Income), 2010

	(i) Unweighted RII Analysis			(ii) RII Analysis with Population Weights			(iii) RII Analysis with Population and Bootstrap Weights		
Variable	Coefficient	Standard Error	2-tail p Value	Coefficient	Standard Error	2-tail p Value	Coefficient	Standard Error	2-tail p Value
Log (net worth) {Ln(−net worth) if net worth < 0} & Ln(0.01) if net worth = 0}	**0.1343**	**0.0143**	**.000**	**0.1055**	**0.0175**	**.000**	**0.1055**	**0.0119**	**.000**
Negative net worth (1 if net worth < 0)	**−0.8556**	**0.1075**	**.000**	**−0.8709**	**0.1147**	**.000**	**−0.8709**	**0.0919**	**.000**
Log (income) {Ln(.01) if income ≤ 0}	**0.0819**	**0.0141**	**.000**	**0.1276**	**0.0434**	**.003**	**0.1276**	**0.0315**	**.000**
Own home	**−0.1616**	**0.0783**	**.039**	**−0.1530**	**0.0842**	**.069**	**−0.1530**	**0.0640**	**.017**
Racial ethnic status of the respondent (reference category = White)									
Black	−0.0987	0.0926	.286	−0.0914	0.0978	.350	−0.0914	0.0712	.199
Hispanic	−0.0748	0.1014	.461	−0.1175	0.1080	.277	−0.1175	0.0754	.119
Asian/other	**0.2882**	**0.1405**	**.040**	**0.2943**	**0.1475**	**.046**	**0.2943**	**0.1080**	**.006**
Employment status of head (reference category = employed by someone else, not working)									
Head self-employed	−0.0429	0.0787	.586	−0.1173	0.0922	.203	−0.1173	0.0615	.056
Head retired	**−0.2796**	**0.1005**	**.005**	**−0.2608**	**0.1128**	**.021**	**−0.2608**	**0.0792**	**.001**
Education of household (reference category = less than high school)									
High school degree	0.0564	0.1251	.652	0.0105	0.1334	.937	0.0105	0.0907	.908
Some college	0.0811	0.1295	.531	0.0494	0.1393	.723	0.0494	0.0996	.619
Bachelor degree	0.1699	0.1366	.214	0.1136	0.1479	.442	0.1136	0.0951	.232
Post-bachelor degree	**0.2861**	**0.1454**	**.049**	0.2444	0.1607	.128	**0.2444**	**0.1169**	**.037**
Perceived health status (reference category = poor health)									
Excellent health	**0.6595**	**0.1325**	**.000**	**0.6196**	**0.1450**	**.000**	**0.6196**	**0.1145**	**.000**

TABLE 1
continued

Variable	(i) Unweighted RII Analysis			(ii) RII Analysis with Population Weights			(iii) RII Analysis with Population and Bootstrap Weights		
	Coefficient	Standard Error	2-tail p Value	Coefficient	Standard Error	2-tail p Value	Coefficient	Standard Error	2-tail p Value
Good health	**0.3897**	**0.1206**	**.001**	**0.4002**	**0.1311**	**.002**	**0.4002**	**0.1014**	**.000**
Fair health	**0.3052**	**0.1255**	**.015**	**0.3507**	**0.1361**	**.010**	**0.3507**	**0.1061**	**.001**
All in household covered by health insurance	**0.2958**	**0.0757**	**.000**	**0.2497**	**0.0818**	**.002**	**0.2497**	**0.0483**	**.000**
Age of head	**−0.0623**	**0.0108**	**.000**	**−0.0647**	**0.0116**	**.000**	**−0.0647**	**0.0089**	**.000**
Age of head squared	**0.0005**	**0.0001**	**.000**	**0.0005**	**0.0001**	**.000**	**0.0005**	**0.0001**	**.000**
Household type (reference category = married couple)									
Partner couple	−0.1362	0.1167	.243	−0.1070	0.1251	.392	−0.1070	0.0861	.214
Single male	−0.1504	0.0902	.095	−0.1163	0.0992	.241	−0.1163	0.0903	.197
Single female	**−0.4929**	**0.0774**	**.000**	**−0.4567**	**0.0857**	**.000**	**−0.4567**	**0.0607**	**.000**
Have dependent child under 18	**−0.2961**	**0.0705**	**.000**	**−0.3359**	**0.0773**	**.000**	**−0.3359**	**0.0624**	**.000**
Expectation of household income (reference category = sure it will increase faster than prices)									
Sure same	−0.0973	0.1030	.345	0.0564	0.1134	.619	−0.0564	0.0699	.420
Sure less	**−0.4497**	**0.1079**	**.000**	**−0.2958**	**0.1188**	**.013**	**−0.2958**	**0.0734**	**.000**
Unsure	**−0.3855**	**0.1010**	**.000**	**−0.2775**	**0.1123**	**.013**	**−0.2775**	**0.0772**	**.000**
Income relative to normal income (reference category = about the same)									
Higher than normal	**0.2688**	**0.1182**	**.023**	0.2361	0.1374	.086	**0.2361**	**0.0947**	**.013**
Lower than normal	**−0.4017**	**0.0691**	**.000**	**−0.3523**	**0.0788**	**.000**	**−0.3523**	**0.0577**	**.000**
Could get $3,000 from friends or relatives in emergency	**0.3708**	**0.0639**	**.000**	**0.4193**	**0.0692**	**.000**	**0.4193**	**0.0447**	**.000**
Intercept	−0.6281	0.3652	.085	−0.8809	0.5159	.088	−0.8809	0.3910	.024

Notes: Effects significantly different from 0 at $p < .05$ for 2-tail test are in bold. If both estimates are significant, the variable name is also in bold.

TABLE 2
Regression Analysis of Log of Liquid Assets, 2010

	(i) Unweighted RII Analysis			(ii) RII Analysis with Population Weights			(iii) RII Analysis with Population and Bootstrap Weights		
Variable	Coefficient	Standard Error	2-tail *p* Value	Coefficient	Standard Error	2-tail *p* Value	Coefficient	Standard Error	2-tail *p* Value
Log (net worth) {Ln(−net worth) if net worth < 0} & Ln(0.01) if net worth = 0}	**0.6216**	**0.0152**	**.000**	**0.6010**	**0.0149**	**.000**	**0.6010**	**0.0110**	**.000**
Negative net worth (1 if net worth < 0)	**−0.9568**	**0.1337**	**.000**	**−0.7656**	**0.1657**	**.000**	**−0.7656**	**0.1307**	**.000**
Log (income) {Ln(.01) if income ≤ 0}	**0.0928**	**0.0171**	**.000**	**0.1881**	**0.0409**	**.000**	**0.1881**	**0.0323**	**.000**
Own home	0.0447	0.0993	.652	−0.0256	0.1197	0.831	0.0447	0.1033	.402
Racial ethnic status of the respondent (reference category = White)									
Black	**−0.7352**	**0.1164**	**.000**	**−0.7973**	**0.1588**	**.000**	**−0.7973**	**0.1203**	**.000**
Hispanic	**−0.3805**	**0.1278**	**.003**	**−0.3593**	**0.1591**	**.024**	**−0.3593**	**0.0953**	**.000**
Asian/other	−0.1861	0.1699	.273	−0.1311	0.1919	.495	−0.1311	0.1373	.170
Employment status of head (reference category = employed by someone else, not working)									
Head self-employed	−0.0628	0.0950	.509	−0.1231	0.1111	.268	−0.1231	0.0747	.050
Head retired	**−0.2832**	**0.1235**	**.022**	−0.2933	0.1520	.054	**−0.2933**	**0.1221**	**.008**
Education of household (reference category = less than high school)									
High school degree	**0.9606**	**0.1548**	**.000**	**0.9056**	**0.2210**	**.000**	**0.9056**	**0.1572**	**.000**
Some college	**1.5948**	**0.1606**	**.000**	**1.5284**	**0.2225**	**.000**	**1.5284**	**0.1585**	**.000**
Bachelor degree	**2.1096**	**0.1700**	**.000**	**2.0113**	**0.2238**	**.000**	**2.0113**	**0.1716**	**.000**
Post-bachelor degree	**2.2381**	**0.1798**	**.000**	**2.0610**	**0.2309**	**.000**	**2.0610**	**0.1745**	**.000**
Perceived health status (reference category = poor health)									
Excellent health	**0.9334**	**0.1628**	**.000**	**0.9789**	**0.2016**	**.000**	**0.9789**	**0.1699**	**.000**

TABLE 2
continued

Variable	(i) Unweighted RII Analysis			(ii) RII Analysis with Population Weights			(iii) RII Analysis with Population and Bootstrap Weights		
	Coefficient	Standard Error	2-tail *p* Value	Coefficient	Standard Error	2-tail *p* Value	Coefficient	Standard Error	2-tail *p* Value
Good health	**0.7255**	**0.1489**	**.000**	**0.7046**	**0.1898**	**.000**	**0.7046**	**0.1520**	**.000**
Fair health	**0.3348**	**0.1538**	**.029**	0.3723	0.1969	.059	**0.3723**	**0.1492**	**.006**
All in household covered by health insurance	**0.8601**	**0.1000**	**.000**	**0.8059**	**0.1245**	**.000**	**0.8059**	**0.0863**	**.000**
Age of head	**−0.0722**	**0.0132**	**.000**	**−0.0855**	**0.015**	**.000**	**−0.0855**	**0.0117**	**.000**
Age of head squared	**0.0008**	**0.0001**	**.000**	**0.0009**	**0.0001**	**.000**	**0.0009**	**0.0001**	**.000**
Household type (reference category = married couple)									
Partner couple	−0.1594	0.1468	.278	−0.1738	0.1886	.357	−0.1738	0.1251	.082
Single male	−0.1825	0.1128	.106	−0.2675	0.1410	.058	**−0.2675**	**0.1022**	**.004**
Single female	**−0.2608**	**0.0987**	**.008**	**−0.2672**	**0.1092**	**.014**	**−0.2672**	**0.0914**	**.002**
Have dependent child under 18	**−0.2257**	**0.0882**	**.010**	**−0.3183**	**0.0997**	**.001**	**−0.3183**	**0.0721**	**.000**
Expectation of household income (reference category = sure it will increase faster than prices)									
Sure same	0.1227	0.1198	.306	0.2056	0.1339	.125	**0.2056**	**0.0849**	**.008**
Sure less	−0.0398	0.1284	.756	0.0248	0.1394	.859	0.0248	0.0950	.397
Unsure	**−0.4168**	**0.1203**	**.001**	**−0.3761**	**0.1495**	**.012**	**−0.3761**	**0.1022**	**.000**
Income relative to normal income (reference category = about the same)									
Higher than normal	**0.3494**	**0.1394**	**.012**	**0.3107**	**0.1343**	**.021**	**0.3107**	**0.0948**	**.001**
Lower than normal	−0.1301	0.0864	.132	−0.1139	0.1079	.291	−0.1139	0.0712	.055
Could get $3,000 from friends or relatives in emergency	**0.7050**	**0.0819**	**.000**	**0.7596**	**0.0966**	**.000**	**0.7696**	**0.0633**	**.000**
Intercept	−1.7960	0.4501	.000	−2.2332	0.6111	.000	−2.2332	0.4893	.000

Notes: Effects significantly different from 0 at $p < .05$ for 2-tail test are in bold. If all three estimates are significant, the variable name is also in bold.

covariance of estimators than when one uses only one implicate or does not use the method (Montalto and Sung 1996).

Heteroskedasticity may exist when individuals in various strata have different probabilities of being selected for the sample because of differences in sampling and response rates (DuMouchel and Duncan 1983). This might be the case in the SCF datasets because the survey employs a complex dual sampling design based on multi-stage geographic regions and a list of tax returns provided by the Internal Revenue Service (Kennickell 2012). Owing to this sampling procedure, the probability that a household will be included in the SCF sample depends on its location and wealth. If the population is heterogeneous depending on each stratum, i.e., if estimators produced by empirical models greatly differ by strata even after controlling for various factors, both weighted and unweighted models may be biased (Deaton 1997, 70). This issue may be crucial if the parameters of interest (e.g., wealth) differ substantially depending on strata.

In order to adjust for heteroskedasticity, Deaton (1997, 71) suggested using bootstrap estimates to correct the standard errors that are essential for inference. The bootstrap method enables us to mitigate this issue by obtaining the averages of the bootstrapped estimators and the asymptotic covariance matrix using numerous replications of the data (Greene 2012, 612). The Federal Reserve provides replicate weights (WT1B1-WT1B999) for 999 replicate samples and multiplicity factors (the number of times the case was selected in the replicate: MM1-MM999) for the first replicate only (Kennickell 2012). We calculated bootstrapped standard errors from the first replicate using 999 bootstrap replications, and adjust them by using the RII procedure. For weighted analyses, we adjusted the population weight (X42001) to the U.S. population in order to match the sample size of each implicate to the actual sample size, as Lindamood et al. (2007) did in their weighted multivariate analyses.

Coefficients

The coefficients of multivariate models were obtained from the RII procedure by averaging five coefficients from each implicate separately with population weights (Pence 2001; Wenzlow et al. 2004).

Standard Errors

We calculated standard errors using the following formula (Pence 2001; Wenzlow et al. 2004):

$$\sqrt{\left(\frac{6}{5}\right) * \text{Imputation variance} + \text{Sampling variance}}.$$

The imputation variance is the between-imputation variance with five replicates, which is equivalent to the variance obtained from the RII procedure. We calculated the variance by applying STATA's "*scfimp*" command with population weights. The sampling variance is the average variance from the bootstrapped sample with population weights. We used STATA's "*scfboot*" command with 999 bootstrap replications. The Stata ado files for "*scfimp*" and "*scfboot*" are available from the Federal Reserve Board staff upon request. The newly released "*scfcombo*" Stata ado file simplifies the calculation of standard errors of SCF data adjusting for bootstrap weights. The ado file is built-in, so researchers can easily install the program with the command of "*ssc install scfcombo*."[1]

RESULTS

Comparison of Multivariate Analyses with and without Weighting

Logit Model

Table 1 shows the effects of independent variables on the likelihood of saving for three logistic regressions (logits), each using RII. The first three columns (Section i) are the results without weighting. The second three columns (Section ii) are the results with population weights applied. The last three columns (Section iii) are the results with both population weights and bootstrapping.

None of the results from the unweighted logistic regression (i) differ substantially from the results in the regression with both population weights and bootstrapping applied (iii). Most of the results in all three estimates are similar to those reported by Yuh and Hanna (2010), who analyzed a combination of the 1995–2004 SCF datasets. It is plausible that some of the differences from the earlier study are due to our different survey year, 2010. For instance, Yuh and Hanna found Black households were less likely to save than White households, while we find no difference; but we found that Asian/other households were more likely to save than White households. Our specification of age was different from the previous study, but our results were similar, in that the likelihood of saving decreases with age for all values of age in the sample, with very similar patterns in the unweighted and weighted logits. Our specification for net worth is more complicated than the previous study, but the combined effects of the log of net worth and of the dummy variable for negative net worth imply that the likelihood of saving increases with net worth as net worth increases above

1. A help file is also available. The ado file has been validated by the Federal Reserve Board (more information is available at http://cfs.wisc.edu/presentations/scf_combo_brief.pdf).

zero; for slightly negative levels of net worth the likelihood of saving is low, but increases as net worth decreases in the negative range. The pattern based on the weighted logit is very similar to the pattern based on the unweighted logit. The only major difference between our results and those in the earlier study was in the effect of perceived health status, as in both our unweighted and weighted results, better health was associated with an increased likelihood of saving.[2]

Coefficient estimates in the unweighted model (i) were different from those in two weighted models, and the differences were due to the application of the population weights. In the model with population weights only (ii), variables for households with post-bachelor degree and the expectation that income definitely would increase the same as prices did not have significant effects ($p < .05$) on whether or not households save, but these two variables had significant effects in the unweighted model (i) and the model with both population weights and bootstrap weights (iii). The absolute value of the z statistic was higher for the unweighted estimate (i) than for the estimate based on population weighting only (ii) for 79% of the variables.

In comparing the results for the population weighted model without bootstrap weights (ii) to the results for the population weighted model with bootstrap weights (iii), the coefficient estimates were identical, but z values of all estimates were higher with the bootstrap weights than the z values without the bootstrap weights.

Finally, for the logit results in Table 1, in comparing the results for the population weighted model with bootstrap weights (iii) to the unweighted model (i), many of the significance levels were similar between the unweighted model and the population weighted model with bootstrapping. There were no variables for which the significance levels changed from being insignificant ($p \geq .05$) to being significant ($p < .05$), or vice versa. Of the 14 directional hypotheses listed by Yuh and Hanna (2010), conclusions regarding the hypothesis tests obtained from the unweighted model and from the population weighted model with bootstrapping were identical.

OLS Model

For a comparison of the effect of population and bootstrap weights in an OLS regression, we regressed the natural log of the level of liquid assets on the set of independent variables used for the logistic regression. Directional hypotheses based on the normative reasoning presented in Yuh and Hanna (2010) are similar to those presented by those authors. Table 2 shows results

2. This may be due to a coding error in Yuh and Hanna (2010).

of three OLS regressions on household liquid assets, each using RII. The first three columns are the results without weighting (i), the second three columns are the results with population weights only (ii), and the last three columns are the results with both population weights and bootstrapping (iii). As with the logit results, the differences in coefficient estimates were due to the application of population weights.

We found some differences in the significance levels between the unweighted model and the population weighted model without bootstrapping. In the model with population weights only, variables for having a retired household head and for fair health did not have significant effects ($p < .05$) on the level of liquid assets, but these two variables had significant effects in the unweighted model.

The coefficient estimates in model (ii) and model (iii) were identical, but standard error estimates were smaller in model (iii), so that the z values of all estimates were greater with model (iii) than with the model (ii). Although the application of population weights but not bootstrap weights produced the most conservative results with regard to hypotheses testing, researchers should be aware of the potential existence of endogeneity between survey weights (X42001) and control variables that are included in the model (e.g., income). Thus, the application of population weights without bootstrapping may produce biased estimates of variances. As it is not reasonable to apply the population weights but not the bootstrap weights, we now focus on the choice of the unweighted model versus the model with population and replication weights.

In comparing Table 2 results for the population weighted model with bootstrap weights (iii) to the unweighted model (i), many of the significance levels were similar. For all but two variables, the significance levels did not change from being insignificant ($p \geq .05$) to being significant ($p < .05$), or vice versa. In the population and bootstrap weighted model (iii), variables for single male households, and the expectation that income definitely would increase the same as prices are significant factors related to household liquid asset holdings, but these variables are not significant in the unweighted model. As Lindamood et al. (2007) concluded in discussing differences in significance between weighted and unweighted analyses, one should be cautious in drawing conclusions.

Simple Diagnostic Test for Model Selection

The results of our logit and OLS analyses do not provide any obvious pattern that can help us choose whether application of both the population and bootstrap weights will be better than unweighted analyses. However, DuMouchel and Duncan (1983) suggested using a Hausman test (1978) to differentiate between unweighted and weighted estimators. Under the null hypothesis, both unweighted and weighted estimators are consistent, but unweighted estimators are better because they are efficient, i.e., they have the smallest asymptotic variance. Under the alternative hypothesis, weighted estimators are consistent, while unweighted estimators are not. The test statistic is calculated as follows, and the critical value is $\chi^2{}_{(28)} = 41.34$:

$$H = \left(b_{\text{unweighted}} - b_{\text{weighted}}\right)' \left(\text{Var}\left(b_{\text{weighted}}\right) - \text{Var}\left(b_{\text{unweighted}}\right)\right)^{-1} \left(b_{\text{unweighted}} - b_{\text{weighted}}\right)$$

Logit Model

The test statistic calculated from the logit model is 29.59, which is smaller than the critical value. Therefore, we cannot reject the null hypothesis, so both unweighted and weighted estimators are consistent, but nonetheless, it is better to use the unweighted model because unweighted estimators are efficient.

OLS Model

Comparing unweighted and population weighted with bootstrapping estimators, we obtain the test statistic of 86.444, which is greater than the critical value, and thus reject the null hypothesis: under the alternative hypothesis, only weighted estimators are consistent. Therefore, for the OLS model, the Hausman test implies that the model with population weights and bootstrapping should be used.

DISCUSSION AND IMPLICATIONS

In analyses of the 2010 SCF comparing analyses with population weights to analyses with both population and bootstrap weights, standard errors were lower when bootstrap weights were applied. We recommend that researchers who apply population weights with SCF datasets should also apply bootstrap weights to account for complex sample design. We also discussed the issue raised by Deaton (1997, 66) on the use of endogenous weights for hypothesis testing.

In our analyses of the 2010 SCF dataset, the similarity between the estimates for the unweighted logistic regression and the logistic regression with both population and bootstrap weights applied suggests that, at least for similar types of dependent variables (dichotomous attitude variables), application of the bootstrap weights will make little difference. The Hausman test also supports the use of the unweighted over weighted RII estimators for this particular example. All of the differences in coefficients are due to application of the population weights, which Lindamood et al. (2007) explored with comparisons for three different logistic regressions.

For the OLS regression model, we found that two coefficients changed from insignificant ($p > .05$) in the unweighted estimates (i) to significant ($p < .05$) in the weighted estimates (iii). The Hausman test results suggest that researchers should use the weighted model for OLS regression analyses. The Hausman test is a quick and simple way to differentiate the unweighted and the weighted models, and the test could be a useful tool in choosing the appropriate model.

In summary, ignoring the issue of complex sample designs in SCF datasets can affect conclusions, though it is likely that for many models it will not make much difference. The discussion by Deaton (1997) of the issue of unweighted versus the population weighted multivariate analysis is relevant. If the only research focus is hypothesis testing, unweighted analysis is preferable because it is likely to be more conservative in terms of inferring significance than would be the application of both population weights and replicate weights. If estimation of the magnitudes of effects is a focus, it is important to apply population and replicate weights because estimators obtained from this model are consistent in general. Also, applying population and replicate weights may be preferable in order to suggest policy implications. The main purpose of utilizing the population and bootstrapped sample is to avoid bias from a dual sampling frame, and if researchers can mitigate this bias by applying both weights, these results can be generalized to the U.S. population. Would a consideration of complex sample designs have changed the major conclusions of previous analyses of SCF datasets? In the examples we present in this article, accounting for the complex sample design would not have changed any major conclusions, but researchers using SCF datasets should consider comparing results from the unweighted and the weighted model carefully in case they find differences in significance levels or effects of their primary variables in question.

REFERENCES

Deaton, August. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore, MD: Johns Hopkins University Press.

DuMouchel, William H. and Greg J. Duncan. 1983. Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association*, 78 (383): 535–543.

Federal Reserve Board. 2012. Macro SAS Code Designed to Be Used with SCF Datasets from 1989 to 2010. http://www.federalreserve.gov/econresdata/scf/files/bulletin.macro.txt.

Greene, William H. 2012. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hanna, Sherman D. 2011. The Demand for Financial Planning Services. *Journal of Personal Finance*, 10 (1): 36–62.

Hausman, Jerry A. 1978. Specification Tests in Econometrics. *Econometrica*, 46 (6): 1251–1271.

Hogarth, Jeanne M. and Chris E. Anguelov. 2003. Can the Poor Save? *Financial Counseling and Planning*, 14 (1): 1–18.

Kennickell, Arthur B. 2012. *Codebook for 2010 Survey of Consumer Finances*. Washington, DC: Board of Governors of the Federal Reserve System. http://www.federalreserve.gov/econresdata/scf/files/codebk2010.txt.

Lindamood, Suzanne, Sherman D. Hanna, and Lan Bi. 2007. Using the Survey of Consumer Finances: Methodological Considerations and Issues. *Journal of Consumer Affairs*, 41 (2): 195–214.

Montalto, Catherine P. and Jamie Sung. 1996. Multiple Imputation in the 1992 Survey of Consumer Finances. *Financial Counseling and Planning*, 7: 133–146.

Nielsen, Robert B. and Martin C. Seay. 2014. Complex Samples and Regression-Based Inference: Considerations for Consumer Research. *Journal of Consumer Affairs*, 48 (3): 603–619.

Nielsen, Robert B., Michael Davern, Arthur Jones, Jr., and John L. Boies. 2009. Complex Sample Design Effects and Health Insurance Variance Estimation. *Journal of Consumer Affairs*, 43 (2): 346–366.

Pence, Karen M. 2001. 401(k)s and Household Saving: New Evidence from the Survey of Consumer Finances. FEDS Working Paper No. 2002-06. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=287453.

Wenzlow, Audra T., John Mullahy, Stephanie A. Robert, and Barbara L. Wolfe. 2004. An Empirical Investigation of the Relationship between Wealth and Health Using the Survey of Consumer Finances. Institute for Research on Poverty Discussion Paper No. 1287-04. http://www.russellsage.org/research/reports/empirical-investigation-relationship-between-wealth-health.

Winship, Christopher and Larry Radbill. 1994. Sampling Weights and Regression Analysis. *Sociological Methods and Research*, 23 (2): 230–257.

Yao, Rui, Jie Ying, and Lada Micheas. 2013. Determinants of Defined Contribution Plan Deferral. *Family and Consumer Sciences Research Journal*, 42 (1): 55–76.

Yuh, Yoonkyung and Sherman D. Hanna. 2010. Which Households Think They Save? *Journal of Consumer Affairs*, 44 (1): 70–97.