THE PRACTICAL IMPORTANCE OF SAMPLE WEIGHTS

David E. Bloom and Todd L. Idson, Columbia University
Todd L. Idson, Columbia University, Department of Economics, New York, NY 10027

KEY WORDS: Regression, Sample weights

Introduction

Few of the large survey data sets that economists regularly rely upon in their empirical work are generated by simple random sampling. Rather, most are created using sampling techniques that select observations randomly within geographic clusters or within strata that are believed to be associated with key variables of interest. These techniques of probability sampling are employed to lower the cost of generating a sample of a given size (cluster sampling) and to increase the expected precision of parameters that a fixed-size sample can be used to estimate (stratified sampling). A cost, though, of cluster sampling methods is that OLS standard errors are biased in the presence of intracluster correlation (see Scott and Holt, 1982, for an empirical analysis of the effect of intracluster correlation on standard errors; see also Glied, 1990). Furthermore, the use of sample weights will reduce efficiency if higher weights are attached to groups with higher variances (e.g., in the estimation of consumption functions with a data set that over-samples the higher income population, see Cramer, 1971).

Failure to use sample weights with stratified samples will not bias regression coefficients if the model is correctly specified since a sufficient condition for OLS estimates to be unbiased is that $Cov(X_{ji},U_i) = 0$ for $i=1,\ldots,n$ observations and all j, where the X_j are the regressors and U is the error term. Yet if the model is incorrectly specified then weights can act to reduce the misspecification bias if, for example, an omitted interaction contains a variable that is correlated with the stratification variable. In fact, DuMouchel and Duncan (1983) have extended this insight to employing sample weights to test for model misspecification.

As a general rule, information on selection probabilities for different sample units needs to be used in constructing unbiased estimates of single population parameters (e.g., means and variances). In addition, it is well known among statisticians that the parameters of multivariate models may be inefficiently and inconsistently estimated if one ignores information on the selection probabilities of different sample observations (i.e., the sample weights). For example, ordinary least squares estimates of the coefficients in a standard multiple regression model will be inconsistent if sample weights are ignored that (1) contain information that is not accounted for (or not correctly accounted for) in the regression specification <u>and</u> (2) are correlated with one or more variables that are included in the specification of independent variables. More generally, coefficient estimates will be inconsistent if sample weights are ignored in cases in which regression parameters vary across sample strata within which sample weights do not vary. The estimated variance-covariance matrix will be inconsistent if the error variance is correlated with information contained in the weights.

On the other hand, least squares estimates that weight each observation by the inverse of its probability of inclusion in the sample will be inefficient (though consistent) if the information contained in the sample weights is already reflected through the regression specification. If regression parameters are constant across population strata then sample design will have no effect, while if these parameters are not homogeneous across subgroups then a regression that treats the data as a random sample will be misspecified. In these cases the use of sample weights will partly correct for heterogeneity across strata. Thus, whether or not researchers use sample weights to estimate the parameters of a multivariate model is ultimately a reflection of the confidence they are placing in their empirical specification.

The starting point for this paper is the observation that the use of sample weights in estimating the parameters of multivariate models is fundamentally an empirical issue, not an analytical one. Our objective is to assess the practical importance of accounting for sample weights by examining the sensitivity of the estimates of several popular models in labor economics to the incorporation of sample weights. We report and compare weighted and unweighted least squares estimates of standard wage and hours equations, and maximum likelihood estimates of a simple probit model of female labor force participation. We experiment with parsimonious specifications that do not control for key labor market influences that are reflected in the sample weights; we also experiment with more elaborate specifications that do control for such information. Estimates are calculated using the March 1989 Current Population

Survey (CPS). (See Bloom and Idson, 1991, for a more extensive empirical analysis of these questions using two additional labor market surveys - the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth.)

A number of researchers have commented on the effects of weighting in particular cases. In their investigation of American marriage patterns, Bloom and Bennett (1990) generally found little effect on either parameter estimates or standard errors when weighting in both the CPS and the National Survey of Family Growth. Estimating total cost functions, Hu and Stromsdorfer (1970) find what they consider large changes in their parameter estimates, yet no changes in sign and only infrequent changes in statistical significance from weighting. Klein and Morgan (1951) find little effect from weighting when they estimate savings equations.

Section I describes the CPS and the sampling designs upon which it is based. This section also reports some descriptive analyses of the sample weights and their covariates. Section II presents estimates of the wage, hours, and labor force participation equations. The main emphasis in this section is on comparing weighted and unweighted estimates of parameters and standard errors computed for identical models using the same data. In this way, we hope to assess the practical importance of accounting for sample weights.

Description of the Data

In order to investigate the importance of sample weights in concrete applications we estimate three popular models in labor economics using a widely analyzed data set. While these models are clearly not exhaustive of the types of economic relationships commonly explored or estimation techniques employed, a comparison of weighted and unweighted regressions for these particular cases will provide some indication of the general order of the effect of weighting on parameter estimates and standard errors in the context of both least squares (LS) and maximum likelihood (ML) estimation. We estimate a human capital (semi-log) wage function using LS, and two labor supply functions describing different dimensions of labor supply behavior, (i) a LS annual working hours function for men and women separately and (ii) a ML probit function for female labor force participation.

The CPS is a national probability sample of households based on a multistage area cluster design where the sample weights primarily incorporate information on age, race, sex, and area cluster. Each unit is initially assigned a baseline weight (which is constant within PSU's, but varies across the 51 state and District of Columbia samples), which is then adjusted to account for noninterviews and for data on the age, sex, and race of the respondents that are interviewed. For the earnings equations we use the basic March CPS weight for the two rotation groups that are asked about their earnings (i.e., about one-fourth of the total sample). While the annual hours information is reported in the March Income Supplement, for which there is a separate weighting variable provided since the March supplement oversamples Hispanics (see Shapiro, 1987), we still use the basic March weight because our sample is restricted to the earnings eligible group in order to have information on wages.

Empirical Examples

Table 1 reports frequency distributions for the various weights employed, and Table 2 reports multivariate estimates of the relationship between the weights and the regressors used in the wage and hours regressions. Table 3 reports LS estimates without and with sample weights for human capital earnings functions, and Table 4 reports LS estimates of annual hours of work regressions for men and women separately, and ML estimates of a simple probit model of female labor force participation. All

Table 1: Distribution of Sample Weights

Categories	Percentage; Count
1	21.6; 2475
2	40.0; 4588
3	22.5; 2578
4	9.6; 1100
5	3.8; 435
6	2.0; 227
7	0.3; 37
8	0.2; 28
9	0.1; 8
10	0.0; 3

Each cell contains first the percentage and second the count of the weights that lies in the particular category. Categories are created by equally dividing the range of the weights into ten equal groups.

Table 2: Sample Weight Regressions		
Independent Variables	Wages R^2=0.061[a]	Labor Supply R^2=0.047[a]
Intercept	7096.662[a] (272.047)	5571.809[a] (214.084)
Education	-124.811[a] (17.518)	-
Experience	19.164 (13.144)	-
Experience^2	-0.623[a] (0.292)	-
ln Wage	-	115.658 (78.559)
Spouse Works	-	-111.170 (114.179)
Spouse Income	-	-0.001 (0.003)
Married	-47.013 (93.123)	-44.645 (112.729)
Kids LT 18	-	-143.201[a] (40.063)
Union Coverage	54.579 (123.875)	85.958 (119.887)
Black	623.666[a] (149.741)	1352.631[a] (145.269)
South	708.020[a] (93.372)	536.161[a] (88.101)
Female	-351.467[a] (87.460)	-328.700[a] (90.666)
Age	-	-3.800 (3.812)
SMSA	2213.010[a] (101.450)	1799.551[a] (96.466)

* Parameter estimates are listed with standard errors reported in parentheses. Superscripts a and b denote marginal significance levels of 1% and 5%, respectively.

of the weighted and unweighted estimates were calculated using SAS. (Each table includes specifications that are clearly misspecified with omitted variables known to be correlated with factors reflected in the weights, and more complete specifications.)

We see from Table 1 that the distribution of weights tends to be highly skewed and fairly disperse. This finding rules out the possibility that the weights might be unimportant because they exhibit little variation. Table 2 provides a description of the multivariate association between the regressors used in the tables and the weights used (significant experience effects no doubt represent the effects of age-based sampling designs).

Our main concern is with the practical significance of weighting in the context of empirical labor economics. Hence, we compare the weighted and unweighted results according to changes in the sign, significance, and size of the parameter estimates, and changes in size for the standard errors. (We also perform the DuMouchel-Duncan test for the desirability of weighting. The results of this test are reported at the bottom of Tables 3 and 4.)

Looking first at the semi-log wage regressions, there are no changes in sign or significance of the estimated parameters, with a pattern of relatively small changes in size of the parameter estimates. As expected, the estimates tend to vary a bit more in the Basic Specification than in the Full Specification. Weighting also appears to have little effect on the estimated standard errors. Table 3 generally indicates that in the context of human capital wage regressions, sample weights have little effect in either simple specifications or more complex specifications that include dummies for factors reflected in the weights. DuMouchel-Duncan test results indicate that sample weights matter for the Basic Specification, yet they do not matter in the Full Specification that controls for a larger number of factors that are ostensibly reflected in the weights.

Given distinct patterns of labor force behavior for men and women we have separately estimated OLS annual hours regressions for men in Table 4. Using a linear specification of annual hours regressions reveals greater effects of sample weights than those for the semi-log wage regressions, though the effect is still not particularly dramatic. A change in sign is observed only in the female hours regression for the variables Married and SMSA (although they are both statistically insignificant), yet changes in statistical significance (measured at conventional levels) do occur for a few other variables. While we do observe some changes in standard errors, weighting seems to have its major effect through changes in the parameter estimates. Similarly, we observe no sign changes or changes in significance for the ML probit estimates. In all cases the DuMouchel-Duncan test fails to reject the hypothesis that sample weights do not affect the estimated parameters (at conventional significance levels).

Table 3: Human Capital (ln) Wage Regressions (n=9,527)		
Independent Variables	Without Weights	With Weights
1. Basic Specification	R^2=0.241	R^2=0.250
Intercept	0.648[a]	0.634[a]
	(0.028)	(0.028)
Education×10	0.940[a]	0.952[a]
	(0.019)	(0.019)
Experience×10	0.306[a]	0.313[a]
	(0.014)	(0.015)
$\text{Experience}^2\times 10^3$	-0.438[a]	-0.451[a]
	(0.032)	(0.033)
2. Full Specification	R^2=0.356	R^2=0.359
Intercept	0.689[a]	0.671[a]
	(0.027)	(0.028)
Education×10	0.902[a]	0.912[a]
	(0.017)	(0.018)
Experience×10	0.259[a]	0.262[a]
	(0.014)	(0.014)
$\text{Experience}^2\times 10^3$	-0.373[a]	-0.375[a]
	(0.030)	(0.031)
Married	0.092[a]	0.090[a]
	(0.009)	(0.010)
Union Coverage	0.194[a]	0.192[a]
	(0.013)	(0.013)
Black	-0.130[a]	-0.136[a]
	(0.016)	(0.015)
South	-0.090[a]	-0.091[a]
	(0.010)	(0.010)
SMSA	0.155[a]	0.158[a]
	(0.010)	(0.012)
Female	-0.245[a]	-0.237[a]
	(0.009)	(0.009)

Parameter estimates are listed with standard errors reported in parentheses. Superscripts a, b and c denote marginal significance levels of 1%, 5% and 10%, respectively. The F-statistic for the DuMouchel-Duncan test is 2.459 (P=0.043) for the basic specification and 0.721 for the full specification (P=0.706).

Conclusions

This paper has addressed the practical influence of using sample weights by estimating a number of widely used models in labor economics. Overall we find weak effects associated with weighting both on the estimated coefficients and the standard errors. Empirical applications in which the use of sample weights have an important, substantive effect on the sign, size, or significance of parameter estimates appears to be an infrequent occurrence in labor economics.

REFERENCES

Bloom David E. and Neil G. Bennett, "Modeling American Marriage Patterns," Journal of the American Statistical Association, December 1990: 1009-17.

Bloom, David E. and Todd L. Idson, "The Practical Importance of Sample Weights," Columbia University Working Paper #563, October 1991.

Cramer, J.S., Empirical Econometrics, North-Holland Publishing Co., 1971.

DuMouchel, William H. and Greg J. Duncan, "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples," Journal of the American Statistical Association, September 1983: 535-43.

Glied, Sherry, The Economics of Health Status and Labor Market Outcomes, Unpublished Ph.D Dissertation, Harvard University, July 1990.

Klein, Lawrence R. and J.N. Morgan, "Results of Alternative Statistical Treatments of Sample Survey Data," Journal of The American Statistical Association, December 1951: 442-60.

Hu, Teh-wei and Ernst W. Stromsdorfer, "A Problem of Weighting Bias in Estimating the Weighted Regression Model," Proceedings of the Business and Economics Section of the American Statistical Association, 1970: 513-16.

Shapiro, G.M., "CPS Supplemental Weighting: Specifications for Weighting the Combined March CPS and Hispanic Supplements Beginning in March 1988," Bureau of the Census, 1987.

Scott, A.J. and D. Holt, "The Effect of Two-Stage Sampling on Ordinary Least Squares Methods," Journal of the American Statistical Association, December 1982: 848-54.

Table 4: Annual Hours and Labor Force Participation Functions

	ANNUAL HOURS-MALES (n=6,327)		ANNUAL HOURS-FEMALES (n=5,474)		FEMALE LABOR FORCE PARTICIPATION (n=9,668)	
Independent Variables	Without Weights	With Weights	Without Weights	With Weights	Without Weights	With Weights
1. Basic Specification	R^2=0.104	R^2=0.103	R^2=0.138	R^2=0.144	L=-3369	L=-3344
Intercept	1254.967^{a}	1276.250^{a}	1073.722^{a}	1069.317^{a}	-1.309^{a}	-1.321
	(29.815)	(29.661)	(31.657)	(35.510)	(0.030)	(0.030)
ln Wage	324.441^{a}	314.886^{a}	406.031^{a}	412.347^{a}	-	-
	(12.778)	(12.715)	(16.285)	(16.687)		
Spouse Works	71.811^{a}	77.265^{a}	-6.771	15.142	-0.177^{a}	-0.157^{a}
	(19.676)	(19.927)	(22.547)	(23.348)	(0.045)	(0.045)
Spouse Income × 10^2	-0.159^{c}	-0.191^{b}	-0.525^{a}	-0.624^{a}	0.000546^{a}	0.000561^{a}
	(0.094)	(0.095)	(0.054)	(0.057)	(0.000095)	(0.000094)
Kids LT 18	-32.069^{a}	-30.131^{a}	-72.254^{a}	-75.037^{a}	0.114^{a}	0.107^{a}
	(6.561)	(6.663)	(8.339)	(8.700)	(0.015)	(0.016)
2. Full Specification	R^2=0.141	R^2=0.137	R^2=0.146	R^2=0.151	L=-3302	L=-3280
Intercept	1282.443^{a}	1304.100^{a}	906.795^{a}	911.496^{a}	-0.867^{a}	-0.870^{a}
	(36.001)	(36.953)	(44.134)	(45.997)	(0.069)	(0.072)
ln Wage	270.779^{a}	268.832^{a}	402.228^{a}	410.063^{a}	-	-
	(13.856)	(13.873)	(16.896)	(17.392)		
Spouse Works	-38.815^{c}	-32.478	-2.239	7.971	-0.170^{a}	-0.132^{a}
	(20.837)	(21.267)	(25.517)	(26.518)	(0.052)	(0.053)
Spouse Income × 10^2	-0.194^{b}	-0.222^{b}	-0.509^{a}	-0.617^{a}	0.000604^{a}	0.000592^{a}
	(0.093)	(0.093)	(0.055)	(0.058)	(0.000098)	(0.000097)
Married	254.177^{a}	240.127^{a}	-15.581	2.115	-0.054	-0.059
	(20.696)	(20.930)	(23.346)	(24.174)	(0.047)	(0.047)
Kids LT 18	0.554	0.176	-67.665^{a}	-71.493^{a}	0.107^{a}	0.100^{a}
	(7.108)	(7.248)	(8.565)	(8.946)	(0.015)	(0.016)
Union Coverage	-92.595^{a}	-78.257^{a}	29.808	24.547	-5.431	-5.411
	(18.971)	(18.962)	(29.785)	(30.360)	(1670.962)	(1731.603)
Black	-81.788^{a}	-66.258^{a}	33.121	13.649	-0.015^{b}	-0.103^{a}
	(26.597)	(24.115)	(29.346)	(28.372)	(0.064)	(0.057)
South	26.023^{c}	38.914^{b}	80.107^{a}	77.469^{a}	0.054	0.038
	(15.544)	(15.546)	(18.438)	(18.562)	(0.037)	(0.039)
Age	1.954^{a}	1.406^{a}	3.901^{a}	3.711^{a}	-0.008^{a}	-0.010^{a}
	(0.718)	(0.733)	(0.754)	(0.765)	(0.002)	(0.002)
SMSA	-44.948^{a}	-38.611^{c}	3.359	-1.381	-0.149^{a}	-0.101^{a}
	(16.861)	(18.517)	(20.428)	(22.682)	(0.039)	(0.043)

* Parameter estimates are listed with standard errors reported in parenthesis. Superscripts a, b and c denote marginal significance levels of 1%, 5% and 10%, respectively. For males hours regressions, the DuMouchel-Duncan test statistic is .468 and 1.128 for the Basic and Full specifications (P-values of .8 and .334), respectively. For females the corresponding values are 1.744 and 1.228 (.121 and .262). Likelihood ratio tests for the labor force participation regressions yield test statistic values of 1.752 and 10.691, both of which are insignificant at the 10% level (the critical chi-squared values are 7.779 and 15.987, respectively).