Bootstrap versus Analytical Testing of Predictions From Stratified Sample Regressions

Robert G. Valletta*
Economic Research Department
Federal Reserve Bank of San Francisco
101 Market Street
San Francisco, CA 94105
415-974-3271
rob.valletta@sf.frb.org

March 1997 (Revised April 1998)

* I thank John DiNardo for helpful comments; any errors are my own. The opinions expressed in this paper do not necessarily represent the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

Bootstrap versus Analytical Testing of
Predictions From Stratified Sample Regressions

Abstract

Stratified sample regressions are used to form predicted differentials across categorically definable population groups when the regression parameters differ across groups. A common application of this technique is to the analysis of labor market differences in wage and benefit outcomes. I discuss bootstrap and first-order analytic approaches to estimating statistical precision for the predicted differentials obtained from such models. In linear and nonlinear applications, analytic approximations substantially overstate statistical precision for estimated differentials that depend on differences in means.

Bootstrap versus Analytical Testing of
Predictions from Stratified Sample Regressions

I. Introduction

Consider the following two linear regression models for two observable sub-samples of a population:

$$w_g = X_g\beta_g + \epsilon_g \text{ , } g=1,2$$

In this formulation, the w's are row vectors, the X's are sub-sample data matrices that contain observations for identical variable lists (including a constant) of dimension k, the β's are coefficient vectors of dimension k, and the ϵ's are random error terms that are uncorrelated with the elements of X_g. Such models frequently are applied to the analysis of wage differentials between labor market groups. For example, let group 1 be women and group 2 men, assume that the dependent variable is ln(earnings), and define:

$$d_{1,2} = (\hat{\beta}_1 - \hat{\beta}_2)\overline{X}$$
$$\textit{where } \hat{\beta}_{1,2} = \textit{estimates of } \beta_1, \beta_2$$

The statistic "$d_{1,2}$" is the vector cross-product of the difference in female and male coefficient vectors and a vector that contains a weighted combination of corresponding variable means from the male and female samples. This statistic estimates the percentage earnings differential between men and women attributable to differences in returns to measurable characteristics, which is often interpreted as an estimate of market discrimination (see Oaxaca 1973; Blinder 1973; Cain 1986).

This technique is applied in a variety of contexts. Besides race and sex earnings differentials, other applications in a linear setting include analysis of union/nonunion earnings differentials (for example, Lee 1978). Numerous nonlinear analogs also have been estimated, typically using probit models; applications include analysis of declining union membership (Even and Macpherson 1990), male/female and union differences in employee benefit receipt (Robinson 1991, Even and Macpherson 1991, 1994), and racial differences in mortgage lending outcomes (Munnell et. al. 1996). In general, this technique is relevant for analyses that involve decomposition of observed outcome differences across categorically definable groups (or across time periods) into components due to differences in characteristics ($\overline{X}$) and differences due to different model structure (i.e., different coefficients).

For the male/female earnings gap example, consider the question "Is there a statistically significant unexplained wage gap between men and women?" Jackson and Lindley (1989) discuss Chow-type tests of this hypothesis, which are tests of the stratified sample specification. However, such tests of equality of coefficient vectors may over-reject the null hypothesis of no wage gap, since they test coefficient rather than residual differences. For example, in a sample of earners, we may find that men have a low return to education and high return to labor market experience, whereas the opposite holds for women. A Chow test may reject equality of the male and female coefficient vectors even though $d=0$—i.e., men and women face a different earnings structure but earn the same on average (conditional on X). Thus, standard tests for a stratified sample specification are not appropriate for the hypothesis in question.

Furthermore, consider a reasonable alternative question, for the case when predicted differentials have been computed for $t>1$ time periods: "Has the female/male wage gap decreased over the period (T-t, T)?" In order to answer such questions with clearly specified statistical precision, an estimate of the standard error of the statistic "$d_{1,2}$" is required. Provision of such standard errors is rare (notable exceptions are Wellington 1992, Oaxaca and Ransom 1994, Baker et. al. 1995, and Brown and Corcoran 1997), and discussion of the properties of such estimation appears to be absent from the applied econometrics literature.

In this paper, I describe a bootstrap procedure for obtaining standard errors on parameters that are multiplicative combinations of regression coefficients and variable means. This covers both the estimated differential described above, its decomposition into portions accounted for by coefficient and mean differences, and the contribution of individual variables to the latter. I compare this procedure to analytic estimates based on a first-order linear approximation (often referred to as the "delta method"). Although the analytic method imposes a lower computational burden with little added programming complexity, in both linear and nonlinear applications analytic approximations substantially overstate statistical precision for estimated differentials that depend on differences in means across sub-samples.

In Section II, I describe the estimation problem and the bootstrap and analytic approximation procedures. Section III presents estimates that illustrate and enable comparison of the procedures. I estimate both linear wage equations and probit models of employer-provided health insurance receipt. In Section IV, I discuss the sensitivity of the standard error estimates to sample size and model fit. I provide concluding comments in Section V.

II. Bootstrap and Analytic Estimates of the Differential Parameter

Basic Linear Case

The first case to be analyzed begins with a set of log-linear regression models for 2 sub-samples of a population:

$$\ln(w_g) = X_g\beta_g + \epsilon_g, \quad g=1,2 \tag{1}$$

w_g is a row vector in each case, which provides observations on an outcome variable (such as wages) for each group. The X's are sub-sample data matrices that contain observations for identical variable lists of dimension K (including a constant). The β's are corresponding coefficient vectors and the ϵ's are row vectors of random error terms that are uncorrelated with the elements of X_g. Each group contributes observations n_g to the full pooled sample, with $\Sigma n_g=N$.

We are interested in estimating the variance in w that is attributable to group status, conditional on the variables contained in X. If $\beta_1=\beta_2$, the model can be estimated by pooling the sample and including a group dummy variable, which produces a common β but separate intercepts for the two groups. In this simple model, the coefficient on the group dummy variable estimates the conditional variation attributable to group status, and the usual finite sample test statistics can be used.

Assume instead that $\beta_1 \neq \beta_2$. This model is of more general interest than the pooled model. For example, in the wage discrimination case it allows for the possibility that employers value characteristics differently across labor market groups—i.e, that workers with different characteristics face different wage differentials (Oaxaca and Ransom 1994). The natural log (approximate percentage) wage gap between the two groups can be expressed several different ways:

$$
\begin{aligned}
\ln(w_1/w_2) = \overline{X}_1\hat{\beta}_1 - \overline{X}_2\hat{\beta}_2 &= \Big((\overline{X}_1 - \overline{X}_2)\hat{\beta}_1\Big) + \Big((\hat{\beta}_1 - \hat{\beta}_2)\overline{X}_2\Big) \\
&= \Big((\overline{X}_1 - \overline{X}_2)\hat{\beta}_2\Big) + \Big((\hat{\beta}_1 - \hat{\beta}_2)\overline{X}_1\Big) \\
&= \Big((\overline{X}_1 - \overline{X}_2)\hat{\beta}^*\Big) + \Big((\hat{\beta}_1 - \hat{\beta}^*)\overline{X}_1 + (\hat{\beta}^* - \hat{\beta}_2)\overline{X}_2\Big) \\
&\equiv c_{1,2} + d_{1,2}
\end{aligned}
\qquad (2)
$$

Each expression on the right-hand side decomposes the full wage gap into portions (enclosed in large parentheses) attributable to differences in mean characteristics and differences in estimated coefficients (returns to characteristics); in the final line, these portions are labeled as $c_{1,2}$ and $d_{1,2}$. The alternative decompositions shown on different lines arise from a standard index number problem, with each decomposition reflecting a different assumption about the wage distribution that would prevail if the two groups received the same return to earnings-related characteristics (Oaxaca and Ransom 1994). The first two lines, respectively, reflect the assumption that the group 1 or the group 2 structure would prevail. In the third line, an intermediate outcome represented by the coefficient β^* is assumed.

Early work on wage gap decompositions typically reported the estimates from the last expression in the first two lines, which were interpreted as bounds on the degree of labor market discrimination. More recently, several authors (Cotton 1988, Neumark 1988) have used weighting criteria implying a non-discriminatory wage structure that would lie between that for the groups in question. For the work reported here, I use the Neumark criterion applied to the third expression. This criterion sets β^* equal to the estimate obtained from a pooled sample regression, thereby implying a non-discriminatory wage structure that is a matrix-weighted average (using the pooled sample data matrix) of the different group wage structures; Oaxaca and Ransom (1994) discuss this approach in more detail.

Although early researchers further decomposed $d_{1,2}$ into portions attributable to differences in specific coefficients, Jones (1983) demonstrated that this necessarily reflects an arbitrary normalization. It is possible, however, to decompose $c_{1,2}$ into portions attributable to specific variables. In particular, for variable "k" in X:

$$c_{1,2}^{k} = (\overline{X}_1^k - \overline{X}_2^k)\hat{\beta}^{*k} \quad (\textit{by definition}, \ \sum_k c_{1,2}^{k} = c_{1,2}) \qquad (3)$$

The purpose of this paper is to illustrate the use of analytical and bootstrap procedures for obtaining the sampling distribution of $d_{1,2}$, $c_{1,2}$, and its components $c_{1,2}^{k}$. I focus on $d_{1,2}$ for illustrative purposes. Let $\delta_{1,2}$ represent the estimate of $d_{1,2}$. Assuming that the regression model is correct, then $E[d_{1,2} - \delta_{1,2}] = 0$; $d_{1,2}$ is an unbiased estimate of the population parameter $\delta_{1,2}$.

We are primarily interested in the variance of our estimator, $E[d_{1,2} - \delta_{1,2}]^2$. Assume first that $\overline{X}$ is fixed (non-stochastic). In that case, $d_{1,2}$ is a linear function of the (random) elements of β_1, β_2, and β^*. The linear approximation to such a distribution is exact (Kendall and Stuart 1969, p. 232):

$$\begin{aligned} var(d_{1,2}) &= \overline{X}V_1\overline{X} + \overline{X}V_2\overline{X} + \overline{X}V^*\overline{X} \\ &= \sum_{\hat{\beta}_j \in (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}^*)} \left[\sum_k (\overline{X^k})^2 var\,\hat{\beta}_j^k \ + \ \sum_{k \neq l} \overline{X^k X^l}\, cov(\hat{\beta}_j^k, \hat{\beta}_j^l) \right] \end{aligned} \qquad \textbf{(4)}$$

where the V's represent the variance-covariance matrix of the corresponding β.[1] Similar expressions can be derived for var($c_{1,2}$) and its components, $c_{1,2}^k$.

In practice, however, the X's (and therefore their means) are subject to sampling error. In that case, the expressions for $d_{1,2}$ and $c_{1,2}$ are multiplicative combinations of random parameters, and the linear approximation ignores non-zero higher order terms. Although a higher order expansion could be performed, this can be a complex undertaking.

The bootstrap is a straightforward and accurate alternative to an approach based on higher order expansions. For linear regression models, the bootstrap can be applied either by drawing random samples of the data with replacement, or by drawing random samples of the regression residuals with replacement. I use the former approach, which is less sensitive to specification errors in the regression model (Efron and Tibshirani 1993).

The bootstrap algorithm proceeds in 3 steps:

(1) select B independent bootstrap samples of size N, each drawn with replacement from the full set of (w,X)

(2) for each bootstrap replication, estimate and save $d_{1,2}$, $c_{1,2}$, and the individual variable components $c_{1,2}^k$ (see equations (2) and (3))

(3) The resulting distributions of $d_{1,2}$, $c_{1,2}$, and $c_{1,2}^k$ provide the necessary information to form confidence intervals and test-statistics; for example, the standard errors of these parameters are estimated by the standard deviation of the bootstrap parameter distribution.[2]

---

[1] I omit covariance terms between β_1 and β_2 because the coefficient vectors are estimated from independent samples. Furthermore, to keep the approximation simple, I ignore correlations between these estimates and the estimate of β^*

[2]Jeong and Maddala (1993) discuss alternatives to reporting bootstrapped standard errors, which may be based on skewed distributions and therefore may be misleading. I discuss this issue further in the

In the empirical section, I compare the sampling distributions obtained from the bootstrap and analytical approaches.

Linear Change Model

As noted in the introduction to this paper, obtaining accurate standard errors is likely to be particularly important when making comparisons over time (for example, Wellington 1992, McCrate and Leete 1994). A decomposition of changes over time that is similar to that provided for cross-section data above is:

$$\begin{aligned} \Delta \ln(w_1/w_2) &= \Delta(X_1\hat{\beta}_1 - X_2\hat{\beta}_2) \\ &\simeq \left[(\Delta\overline{X_1})\hat{\beta}_{11} - (\Delta\overline{X_2})\hat{\beta}_{21}\right] + \left[(\Delta\hat{\beta}_1)\overline{X_{10}} - (\Delta\hat{\beta}_2)\overline{X_{20}}\right] \\ &\equiv \Delta c_{1,2} + \Delta d_{1,2} \end{aligned} \qquad \textbf{(5)}$$

where the first or sole subscripts (1 and 2) represent group and the second (0 and 1) represent the two periods being compared. This decomposition is an approximation that sacrifices exactness for simplicity of interpretation (see Wellington 1992) and comparability to its cross-section analogue.[3] The first term on the right-hand side measures the portion of the changing gap explained by changing mean characteristics (weighted by the period 1 coefficients), and the second term measures the portion explained by changing coefficients (weighted by the period 0 means). The expressions needed to obtain the sampling distributions of the characteristic and coefficient portions of this expression are similar to those in the cross-section case.

---

empirical section.

[3]This expression is identical to that evaluated by Wellington (1992; equation 2). I use an approach that exists in the empirical literature rather than formulating a decomposition that is exactly analogous to the pooled cross-section approach described by equation (2).

Nonlinear Models

A further application is to nonlinear models. For example, Even and Macpherson (1994) outline a method for decomposing predicted probabilities obtained from stratified sample probit equations; they use this method to examine male/female differences in pension receipt. In addition to examining pension receipt, Robinson (1991) examines male/female differences in employer-provided health insurance coverage. Other applications include analysis of declining union membership (Even and Macpherson 1990), union/nonunion differences in employee benefit receipt (Even and Macpherson 1991), the contribution of declining unionism to the male/female wage gap (Even and Macpherson 1993, Doiron and Riddell 1994), and racial differences in mortgage lending outcomes (Munnell et. al. 1996).

The analytic and bootstrap procedures in this setting proceed as follows. Letting I denote a 0-1 indicator variable, Z and γ be a vector of covariates and corresponding coefficients, and "1" and "2" denote two different groups, the probit equations estimated are:

$$\begin{aligned} Pr(I{=}1 \mid Z_1) &= \Phi(Z_1\gamma_1) \\ Pr(I{=}1 \mid Z_2) &= \Phi(Z_2\gamma_2) \end{aligned} \tag{6}$$

In this case, we are interested in decomposing the difference in predicted probabilities across the two groups:

$$
\begin{aligned}
Pr(I{=}1)_1 - Pr(I{=}1)_2 &= \frac{1}{n_1}\sum_{i\in 1}\Phi(Z_i\gamma_1) - \frac{1}{n_2}\sum_{i\in 2}\Phi(Z_i\gamma_2) \\
&\simeq \left[\frac{1}{n_1}\sum_{i\in 1}\Phi(Z_i\gamma_1) - \frac{1}{n_2}\sum_{i\in 2}\Phi(Z_i\gamma_1)\right] + \\
&\quad \left[\frac{1}{n_1}\sum_{i\in 1}\Big(\Phi(Z_i\gamma_1) - \Phi(Z_i\gamma_2)\Big)\right] \qquad \textbf{(7)} \\
&\equiv c_{1,2} + d_{1,2}
\end{aligned}
$$

where Z_i=a vector of independent variable values for a particular individual and $n_{1,2}$=the number of individuals in each group. The second line decomposes the full probability difference into explained and unexplained components, labeled (as before) $c_{1,2}$ and $d_{1,2}$ on the third line. Due to the nonlinearity of these equations, the second line represents an approximation. The explained component is defined as the average difference in probabilities if both groups' characteristics are rewarded by the group 1 structure of returns. The unexplained component is defined as the average across all group 1 sample members of the difference between their actual Pr(I=1) and Pr(I=1) if their characteristics are rewarded by the group 2 structure of returns. Due to the nonlinearity of the probit model, these probabilities must be averaged over observations (as opposed to weighting the coefficient difference by sample means and then evaluating the probit expression).

Furthermore, nonlinearity of the probit model precludes exact decomposition of the explained differential into individual variable contributions. However, in order to satisfy the constraint that the individual effects sum to the total effect, a linear weighting approximation can be applied (Even and Macpherson 1990, 1994).[4] Then the contribution of an individual variable *k* is measured by:

$$c_{1,2}^{k} = \left[\frac{\left(\overline{Z_1^k} - \overline{Z_2^k}\right)\hat{\gamma}_1^k}{\left(\overline{Z_1} - \overline{Z_2}\right)\hat{\gamma}_1} \right] c_{1,2} \qquad \textbf{(8)}$$

This expression provides a single variable analog to the full explained gap estimate; it does so by scaling $c_{1,2}$ by the share of the total explained difference across groups in the index function $Z\gamma$ that is accounted for by differences in variable *k*.

I provide both linearized and bootstrapped estimates of the standard errors of $d_{1,2}$, $c_{1,2}$, and the individual variable effects, $c_{1,2}^{k}$. The linearized estimates are obtained by taking a first-order approximation to the sampling variance of the two probit estimates; the expression is provided in Greene (1993, pp. 645-646). For example, the approximation for the asymptotic variance of $d_{1,2}$ in (7) is:

$$\begin{gathered} Var[d_{1,2}] = (\hat{\phi}_{11})^2 Z_1 V_1 Z_1 + (\hat{\phi}_{12})^2 Z_1 V_2 Z_1 \\ where\ \ \hat{\phi}_{1j} = \frac{1}{n_1}\sum_{i\in 1} \phi(Z_i\gamma_j),\ \ V_j = Asymptotic\ \ Variance\ \ of[\gamma_j] \quad (j=1,2) \end{gathered} \qquad \textbf{(9)}$$

where Z_i=a vector of independent variable values for a particular individual. Evaluating (9) requires choosing a particular vector of values for Z_1 in the first line; I use the sample means for group 1. The expressions for the individual contributions in (9) are similar.

---

[4]Alternatively, Doiron and Riddell (1994) propose a method based on a first order Taylor series approximation to the probability function. In their analysis of male/female differences in union membership, their technique yields results similar to those based on Even and Macpherson’s technique.

The bootstrapped estimates are obtained using the 3-step bootstrap algorithm described earlier.

III. Estimates

Basic Linear Model

I illustrate the bootstrap and analytic procedures for obtaining standard errors using the outgoing rotation group samples from the 1983 and 1993 Current Population Surveys (CPS-ORG), and also the April 1993 CPS Benefits Supplement. These data sets provide basic information on earnings and worker characteristics for a large, representative sample of the U.S. labor force. The Benefits Supplement provides health insurance information, which I use to estimate probit models below.

I focus on the male/female wage gap because the estimated wage gaps are large and the corresponding literature is well developed. I estimate the separate regression equations needed to evaluate equation (2) from Section II using my CPS data set, restricting the analysis to full-time workers (defined as those who work at least 35 hours per week). To maintain this paper's focus on the general econometric technique, I use a relatively restricted set of covariates (education, potential experience and its square, union membership, marital status, and a race dummy) and ignore issues related to appropriate specification of wage equations for estimation of unexplained differences in wages and benefit provision. The estimates provided here therefore are likely to overestimate the degree of labor market discrimination against women.

Table 1 provides sample means and estimates for the 1994 sample. I provide the pooled coefficient (β^* from equation 2) in the first column, group means in the next two, and the contribution to the explained gap in the final column. The total explained and unexplained gaps are listed at the bottom of the table. The coefficient estimates generally are unsurprising and not the focus of my analysis, so I do not discuss them.

As indicated by the sum of the explained and unexplained gap estimates listed at the bottom of the table, among full-time workers the average logarithmic wage gap between men and women was 21% in 1993.[5] Very little of this gap (3.6 percentage points) is explained by differences in mean characteristics for men and women. Among these characteristics, a substantial negative effect of lower union membership for women is offset by somewhat higher female educational attainment. The unexplained wage gap is large (17.7 percentage points).

Our focus here is on the standard errors for the explained and unexplained gap estimates. These are shown in the final column of the table, below the coefficient estimates. The analytic standard error estimate is provided immediately below each parameter estimate, followed by the bootstrapped standard error. The latter was obtained using 500 bootstrap replications, which was adequate to produce bootstrap distributions for the column (4) estimates for which normality could not be rejected.[6] The bootstrap confidence intervals are well-approximated by these standard errors, so in discussion of testing based on the bootstrapped distributions I focus on test statistics formed by treating the standard errors as normal statistics.

The results indicate that for the explained gap estimates, the analytical standard errors underestimate the bootstrapped standard errors by as much as a factor of 5. For the full

[5]I leave these figures in natural log terms for simplicity. The exact numerical gap corresponding to each estimated gap g is ln(1+g).

[6]Stability of the bootstrapped standard errors occurred after as few as 50 replications. Far more were required to produce approximately normal distributions.

explained gap, the bootstrapped standard error is approximately 3 times as large as the analytical standard error. In contrast, the difference between the two standard errors on the unexplained gap estimate is small. This pattern is unsurprising. The analytical technique treats the sample means as fixed, whereas the bootstrap estimate accounts for their sampling variability. The resulting difference in the standard error estimates is larger for the explained gap, which incorporates the difference in means across the two groups, than it is for the unexplained gap, which does not rely on the difference in means. When the bootstrap method is applied with the means treated as fixed, the resulting standard error estimates are virtually identical to those produced by the analytical technique. Thus, the difference between the bootstrapped and analytical estimates is explained by the bootstrap's incorporation of stochastic means.

Linear Change Model

A recent set of papers examine changes over time in the explained and unexplained portions of the male/female and black/white earnings gaps (for example, Wellington 1992, McCrate and Leete 1994). Table 2 reports the results of this exercise for the male/female earnings gap using the 1983 and 1993 CPS data.[7] The table lists the net coefficient changes in the first column, the change in the female and male means in the next two columns, the contribution of each variable to the change in the explained gap in the final column, and the change in the total gaps at the bottom.[8]

---

[7]Because the yearly CPS-ORG is very large, I used a 50% random sample for each year.

[8]Following other studies, Wellington (1992) estimates male and female wage equations corrected for selectivity in employment status. Because she does not find evidence of significant selectivity in the wage equations, I do not estimate similar equations here. Researchers should note, however, that the inclusion of selectivity terms introduces additional complexity for the estimation of sampling errors; the

The total gap figures at the bottom of Table 2 reveal a large decrease in the full-time earnings gap between men and women over this period. Of the approximately 13 logarithmic percentage point decline, 3.9 points are explained (primarily by a smaller decline in union membership for women than for men) and the remainder is unexplained. The gap parameters are precisely estimated in general. The bootstrapped standard errors are substantially larger than their analytical counterparts for the explained gap components, although the difference is not as large as in the basic linear model. The analytical and bootstrap estimates are essentially identical for the unexplained gap estimate.

Nonlinear Model

To demonstrate the application of these procedures to nonlinear models, I analyze differential receipt of health insurance by men and women. Such results may be informative in regard to sex discrimination in benefits provision, or differences across men and women in the prevalence or insurer treatment of pre-existing medical conditions. The data come from the April 1993 CPS Benefits Supplement, which provided detailed information on health insurance and other benefits in the current job. I restricted the sample to married men and women who work full-time. Relative to the earnings models, I use an expanded set of independent variables to account for determinants of benefit provision that differ from the determinants of wages; the additional variables are two child measures, an MSA dummy, and 4 firm size dummies. The differential parameters are estimated as described in equations 7 and 8 in Section II. The columns of Table 3 list the female sample coefficient, the female and male sample means, and

---

bootstrap is well suited to the estimation of test statistics for such models.

each variable's contribution to the total explained gap.[9]

The explained gap estimates at the bottom of Table 3 indicate that female individual and job characteristics yield a probability of receiving health insurance on the job that is approximately 5.9 percentage points larger than that based on male characteristics. A substantially lower female incidence of working in the smallest firm size category (firms with fewer than 10 employees) accounts for virtually all of this explained differential. In contrast, women have an 8.9 percent lower probability of receiving health insurance for unexplained reasons, as captured by differences in the probability effects of the independent variables.

As in the linear models, the bootstrapped standard errors are substantially larger than their analytical counterparts for the explained gap estimates, which rely on differences in sample values of the independent variables. In contrast, the bootstrapped standard error for the unexplained gap is approximately 40 percent larger than its analytical counterpart, a substantial amount but one that is much smaller than the typical difference in standard errors for the explained gap estimates.

IV. Sensitivity to Sample Size and Regression Fit

The high degrees of precision for the estimated standard errors of the differential parameters reported in Section III arise in part due to the large sample sizes available in the CPS. The estimated standard errors on the fitted differentials vary nearly in direct proportion to the inverse of the square root of the number of observations. This simple relationship exists because

[9]The sample means are informative but are not used to calculate the explained effects. Instead, as described in Section II, the explained effects are derived from the average difference in fitted probabilities between the female and male samples.

the fitted differentials combine multiplicative terms of coefficients and sample means; the sampling variation in each is inversely related to the square root of the sample size or degrees of freedom. The resulting increase in standard errors as the sample size falls may be salient for wage differential models estimated using data sets that are smaller than the CPS—for example, panel data sets such as the National Longitudinal Survey or the Panel Study of Income Dynamics, or data sets based on personnel information for a single large firm. Reduction to a 1% sub-sample increases the estimated standard errors by a factor of 10. This reduction in sample size to 561 for the Table 1 model causes the estimate of the total explained gap to become insignificantly different from zero.

The estimated standard errors also are sensitive to the fit of the underlying regression model. In particular, the standard error on the coefficient differential ($d_{1,2}$) varies in direct proportion to the square root of the residual sum of squares in the regression model, for the same reason that it varies inversely with the square root of the sample size. In contrast, the standard error on the characteristics differential ($c_{1,2}$) does not have a simple dependence on the residual sum of squares from the regression model, because the characteristics differential relies more heavily on differences in variable means than on differences in estimated coefficients.

V. Conclusion

I presented analytical and bootstrap methods for estimating the sampling properties of fitted differentials obtained from stratified sample regressions, in both linear and nonlinear settings. The estimated standard errors indicate very precise estimates of the fitted differentials in wage and benefit models. However, the bootstrapped estimates of these standard errors are in general substantially larger than their analytical counterparts for estimated gaps that incorporate differences in means across samples. This arises because the first-order analytical approximations treat the samples (and their means) as fixed, whereas the bootstrapped estimates account for stochastic variation in sample characteristics.

In the illustrations provided, the difference between the analytical and bootstrapped standard errors did not generally determine whether the estimates attained conventional levels of statistical significance. However, the high degree of precision of these estimates primarily is due to the large sample sizes available in the Current Population Survey. As discussed in Section IV, the standard errors vary in direct proportion to the inverse of the square root of the sample size. In smaller data sets, the substantial differences between the bootstrap and analytical standard errors may determine whether or not the corresponding test statistics attain conventional critical values. Although the bootstrap technique is somewhat time-consuming, particularly for models estimated by likelihood techniques, it would be prudent for applied econometricians to check the bootstrapped confidence intervals at an intermediate stage and to provide them for the final manuscript or published estimates of fitted differential parameters.

Although not reported, I also estimated the differential parameters presented here using the single equation dummy variable alternative to the stratified sample approach. The resulting estimates were similar to those obtained from the stratified sample approach in all cases, and the associated standard errors were similar in size to those estimated by the bootstrap approach. For investigators primarily interested in a point estimate of the differential, this might be a reasonable approach. However, if the model suggests stratified samples, or if the contribution of individual variables are of interest, the stratified sample approach is preferred.

References

Baker, Michael, Benjamin, Dwayne, Desaulniers, Andree, and Grant, Mary. 1995. "The Distribution of the Male/Female Earnings Differential: 1970-1990." *Canadian Journal of Economics* 28(3), pp. 479-501.

Blinder, Alan. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4), pp. 436-55.

Brown, Charles, and Corcoran, Mary. 1997. "Sex-Based Differences in School Content and the Male-Female Wage Gap." *Journal of Labor Economics* 15(3, pt.1), pp. 431-465.

Cain, Glen G. 1986. "The Economic Analysis of Labor Market Discrimination: A Survey." In O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics, Vol. 1*. Amsterdam: Elsevier Science Publishers.

Cotton, Jeremiah. 1988. "On the Decomposition of Wage Differentials." *Review of Economics and Statistics* 70(2, May), pp. 236-43.

Doiron, Denise J., and Riddell, W. Craig. 1994. "The Impact of Unionization on Male-Female Earnings differences in Canada." *Journal of Human Resources* 29(2, Spring), pp. 504-534.

Efron, Bradley, and Tibshirani, Robert J. 1993. *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Even, William E., and Macpherson, David A. 1990. "Plant Size and the Decline of Unionism." *Economics Letters* 32(4, April), pp. 393-398.

Even, William E., and Macpherson, David A. 1991. "The Impact of Unionism on Fringe Benefit Coverage." *Economics Letters* 36(1, May), pp. 87-91.

Even, William E., and Macpherson, David A. 1993. "The Decline of Private-Sector Unionism and the Gender Wage Gap." *Journal of Human Resources* 28(2, Spring), pp. 279-296.

Even, William E., and Macpherson, David A. 1994. "Gender Differences in Pensions." *Journal of Human Resources* 29(2, Spring), pp. 555-587.

Greene, William H. 1993. *Econometric Analysis*. Englewood Cliffs, NJ: Prentice Hall.

Jackson, John D., and Lindley, James T. 1989. "Measuring the Extent of Wage Discrimination: A Statistical Test and a Caveat." *Applied Economics* 21 (4, April), pp. 515-40.

Jeong, Jinook, and Maddala, G.S. 1993. "A Perspective on Application of Bootstrap Methods in Econometrics." In G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics, Vol. 11*. Amsterdam: Elsevier Science Publishers.

Jones, F.L. 1983. “On Decomposing the Wage Gap: A Critical Comment on Blinder’s Method.” *Journal of Human Resources* 18(1, Winter), pp. 126-130.

Kendall, Maurice G., and Stuart, Alan. 1969. *The Advanced Theory of Statistics*, Volume 1. New York: Hafner Publishing Co.

Lee, Lung-Fei. 1978. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19 (2, June), pp. 415-433.

McCrate, Elaine, and Leete, Laura. 1994. "Black-White Wage Differences among Young Women, 1977-86." *Industrial Relations* 33(2, April), pp. 168-183.

Munnell, Alicia H., Tootell, Geoffrey M.B., Browne, Lynn E., and McEneaney, James. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86(1, March), pp. 25-53.

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *Journal of Human Resources* 23, pp. 279-295.

Oaxaca, Ronald L. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3), pp. 693-709.

Oaxaca, Ronald L., and Ransom, Michael R. 1994. "On Discrimination and the Decomposition of Wage Differentials." *Journal of Econometrics* 61, pp. 5-21.

Robinson, Michael D. 1991. "Sex Discrimination in Non-Wage Compensation: Pension and Health Insurance Participation." *Eastern Economic Journal* 17 (4), pp. 463-468.

Wellington, Alison J. 1992. "Changes in the Male/Female Wage Gap, 1976-85." *Journal of Human Resources* 28(2), pp. 383-411.

Table 1 -- Wage Equation and Gap Estimates,
Full-Time Workers, 1993 CPS[1] (Standard errors in parentheses)

Variable	(1) Pooled Sample Coefficient	(2) Female Mean	(3) Male Mean	(4) Contribution to total explained gap[2]
Education	.065 (.0008)	12.7 (2.06)	12.3 (2.21)	.026 (.0003) (.001)
Potential Experience	.014 (.0009)	19.3 (8.54)	19.1 (8.44)	.003 (.0002) (.001)
(Potential Experience)^2/100	-.019 (.002)	4.46 (3.53)	4.35 (3.54)	-.002 (.0002) (.0007)
Union member	.335 (.004)	.156 (.362)	.298 (.457)	-.048 (.0006) (.001)
Married	.079 (.004)	.599 (.490)	.678 (.467)	-.006 (.0003) (.0005)
Black	-.127 (.005)	.145 (.352)	.102 (.302)	-.005 (.0002) (.0005)
Constant	1.16 (.014)	1.0 (0.0)	1.0 (0.0)	0.0 (0.0) (0.0)
Adjusted R^2	.297	--	--	--
Total Explained Gap ($c_{f,m}$) $(\overline{X}_f - \overline{X}_m)\beta^*$	--	--	--	-.036 (.0007) (.002)
Total Unexplained Gap ($d_{f,m}$) $((\beta_f - \beta^*)\overline{X}_f + (\beta^* - \beta_m)\overline{X}_m)$	-.177 (.004) (.003)	--	--	--

[1] The sample includes 25,437 women and 30,684 men.
[2] The delta method standard error is listed first; the bootstrapped standard error (with means treated as stochastic) is listed second.

Table 2 -- Gap Estimates: Change from 1983 to 1993[1]
(Standard errors in parentheses)

Variable	(1) $\Delta\beta_f - \Delta\beta_m$	(2) Change in Female Mean	(3) Change in Male Mean	(4) Contribution to change in total explained gap[2]
Education	.010	.383	.385	.008 (.001) (.003)
Potential Experience	.008	1.00	.520	-.000 (.002) (.003)
(Potential Experience)^2/100	-.016	.267	.053	-.002 (.001) (.002)
Married	.039	-.072	-.108	.008 (.001) (.001)
Black	-.025	.010	.003	-.000 (.001) (.001)
Union	.018	-.041	-.115	.024 (.001) (.002)
Constant	-.145	0.0	0.0	0.0
ΔTotal Explained Gap ($\Delta c_{f,m}$) ($(\Delta\overline{X}_f)\beta_{f1} - (\Delta\overline{X}_m)\beta_{m1}$)		--	--	.039 (.002) (.005)
ΔTotal Unexplained Gap ($\Delta d_{f,m}$) ($(\Delta\beta_f)\overline{X}_{f0} - (\Delta\beta_m)\overline{X}_{m0}$)	.091 (.007) (.007)	--	--	--

[1] The pooled 1983/1993 (50% random) sample includes 20,662 women and 27,751 men.
[2] Calculated as described in the text. The delta method standard error is listed first; the bootstrapped standard error (with means treated as stochastic) is listed second.

Table 3 -- Male/Female Differences in Health Insurance Receipt, Probit Equations (Asymptotic standard errors in parentheses)

Variable	(1) Female Coefficient	(2) Female Mean	(3) Male Mean	(4) Contribution to total explained gap[2]
Education	.033 (.010)	13.6 (2.43)	13.6 (2.66)	-.0004 (.00002) (.0007)
Potential Experience	-.008 (.014)	19.1 (8.10)	21.3 (8.94)	.006 (.0003) (.0109)
(Potential Experience)^2/100	.000 (.000)	4.31 (3.25)	5.35 (4.16)	-.002 (.0001) (.01260)
Black	.006 (.089)	.070 (.256)	.048 (.214)	.000 (.000003) (.0008)
# kids < 18	-.118 (.026)	1.02 (1.06)	1.27 (1.19)	.011 (.0006) (.003)
presence of kids <6	.065 (.065)	.231 (.421)	.312 (.463)	-.002 (.0001) (.0020)
MSA dummy	.020 (.046)	.519 (.500)	.529 (.499)	-.0001 (.000004) (.0002)
Firm size <10	-1.19 (.080)	.091 (.287)	.201 (.401)	.048 (.003) (.004)
Firm size btwn. 10 and 25	-.917 (.083)	.077 (.267)	.078 (.268)	.0003 (.00002) (.0018)
Firm size btwn. 26 and 99	-.446 (.068)	.127 (.333)	.127 (.333)	-.00003 (.000002) (.0011)

(continued)

Table 3 (continued)

Variable	(1) Female Coefficient	(2) Female Mean	(3) Male Mean	(4) Contribution to total explained gap[2]
Firm size btwn. 100 and 499	-.055 (.062)	.184 (.388)	.137 (.344)	-.001 (.00005) (.0011)
Constant	.844 (.249)	1.0 (0.0)	1.0 (0.0)	0.0 (0.0) (0.0)
Pseudo-R^2	.090	--	--	--
Total Explained Gap ($c_{f,m}$) (see eq. 7 in text)	--	--	--	.059 (.003) (.007)
Total Unexplained Gap ($d_{f,m}$) (see eq. 7 in text)	-.089 (.007) (.010)	--	--	--

[1] The sample includes 3612 women and 6762 men.
[2] The delta method standard error is listed first; the bootstrapped standard error (with means treated as stochastic) is listed second.