THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Response Errors in Labor Surveys: Comparisons of Self and Proxy

No. 28

Daniel Hill

University of Michigan

August 1987

U.S. Department of Commerce U.S. CENSUS BUREAU

Survey of Income and Program Participation

RESPONSE ERRORS IN LABOR SURVEYS:
COMPARISONS OF SELF AND PROXY

by

DANIEL HILL
UNIVERSITY OF MICHIGAN

No. 8706 28

SIPP Working Paper #28

August 1987

ACKNOWLEDGMENTS

This paper is based on work supported by the National Science Foundation under grant SES 84-01460, "On-Site Research to Improve Government-Generated Social Science Data Base". The research was partially conducted at the U.S. Bureau of the Census while the author was a participant in the American Statistical Association/Census Bureau Research Program, which is supported by the Census Bureau and through the NSF grant. This program was prepared for presentationa t the U.S. Bureau of the Census Third Annual Research Conference, held March 24-April 1, 1987.

Any opinions, findings, and conclusions or recommendations expressed here are those of the author and do not necessarily reflect the views of the National Science Foundation or the Census Bureau.

The author would like to thank Heidi Hartmann, Martha Hill, Frank [illegible]berg and Jeff Moore for their helpful comments on earlier versions of the papers. Any errors are the responsibility of the author.

Table of Contents

PREFACE

Simple differences between reports from self and proxy respondents from non-experimental studies can not be taken as evidence of differential reporting behavior. Both groups are to some extent self-selected and mean reports will contain selectivity bias. In this paper, a traditional response error model is modified to incorporate selectivity bias explicitly. The model is estimated on monthly earnings data for prime-age males from SIPP. The results suggest the existence of substantial selectivity and differential reporting biases as well as differential reporting error variance.

Introduction

Requiring all individuals to respond to survey questions for themselves is an expensive practice. Survey costs can be reduced considerably if, for some individuals, some other informed individual in the household is allowed to act as a 'proxy' respondent. The cost savings from allowing proxy reporting may, however, come at the expense of reduced data quality, but the evidence on this possibility from more than thirty years of active research by survey methodologists is inconclusive and contradictory. Given the lack of evidence that proxy reports are worse than self reports, and given the definite nature of the costs savings, it is not surprising that there has been a trend toward more relaxed 'respondent rules' in survey research. Even so, lack of definitive evidence of effects is not the same as definitive evidence of a lack of effect, and as labor economists we should be concerned about the extent to which our knowledge of secular trends in earnings, discrimination, and work hours is being clouded by a concurrent shift toward proxy reporting in the surveys we use.

The lack of progress in determining the relative quality of self and proxy data is the result of the fact that there are two very difficult empirical problems involved. The first of these is that a definitive answer requires some reliable measure of the true value of the behavior in question. We can not assume that simply because individuals should know more about their own behavior that they are likely to provide more accurate answers. Self-presentation bias—the tendency to provide socially desirable answers—is probably stronger for self than for proxy respondents and is probably not limited to alcohol consumption or charitable giving, but may extend to reporting of work hours and income. Obtaining external validating data is quite expensive but without it one can only examine the bias and error variance of proxy reports relative to that of self reports without saying which data provide more accurate estimates.

The second problem is that observed differences in distributions between self and proxy reports can not be taken as evidence of differences in reporting behaviors. There may be a systematic relationship between proxy/self status and the level of the behavior under investigation. If so, then the average reports for the two groups of respondents will each be biased, and not necessarily because of any systematic difference in reporting behaviors. The bias, instead, may simply be the result of differences in the types of people who are likely to respond for themselves versus have some other person respond as a proxy. The failure to recognize and rectify this form of 'selectivity bias' is the single most prominent shortcoming of earlier analysis of the self/proxy data quality question.

The purpose of this paper is to apply recent econometric techniques developed to deal with selectivity bias in a comparison of self and proxy reports. These techniques are also of considerable potential importance even in studies with controlled selection. The reason is that, at least historically, respondents have not been fully co-operative and there is considerable self-selection in the form of nonresponse. Because these techniques require the specification of a behavioral model, and because the results are sensitive to this specification, we confine our attention to a measure for which there is a generally accepted behavioral model—the labor earnings of prime-aged males. We shall not attempt nor claim to provide definitive answers to the question of whether self or proxy reports of earnings are of higher quality. As Moore (1985) observes, to do so would require data from a study combining both random assignment of cases to self/proxy collection modes and outside validating measures. We have neither. What we hope to do in the present paper is provide answers to the more limited question of: Within the context of the human capital model of earnings, are there differences between self and proxy respondents in the distributions of earnings which cannot be attributed to differences in the levels of the determinants of earnings or to selectivity bias? If so, what are these differences, and how sensitive are our estimates of their magnitudes to the details of the specification of the model.

The paper is divided into two main sections. The first section presents a traditional measurement model modified to incorporate selectivity explicitly. The second section applies this model to data on monthly earnings of prime-aged male workers from the 1984 Survey of Income and Program Participation Panel.

Section I
A Model of Self/Proxy Respondent Selection Bias

Following Hansen, Hurwitz and Bershad (1961) we assume that the report provided for individual i's earnings (Y) by respondent type 'r' (r=s,p for self and proxy respondents, respectively) can be expressed as:

$$Y^r_{it} = Y_i + \beta^r_i + \epsilon^r_{it} \tag{1}$$

where: Y_i is the true value of earnings;

β^r_i is the bias of reports provided by the r^{th} respondent type; and

ϵ^r_{it} is random response error of the r^{th} respondent for individual i's earnings recorded in observation or trial t.

For each individual i we observe either Y^s or Y^p depending upon the level [illegible] n unobserved latent index (R) of propensity to self-respond. That is:[1]

[1]The choice of zero as the threshold for self response is done for notational convenienc[illegible] nl[illegible] [illegible]e could just as easily set the threshold at any arbitrary level τ. That is, we could express [illegible] de[illegible] n rule as self respond iff $R^* > \tau$. In this case the above selection rule would still result if $R \equiv$ [illegible] $- \tau$.

$$Y^r_{it} = \begin{cases} Y^s_{it} & \text{iff } R_i > 0 \\ \\ Y^p_{it} & \text{iff } R_i \leq 0. \end{cases} \qquad 2)$$

The practice of assessing relative bias (i.e. $\beta^s - \beta^p$) by simply comparing the average values of reported is appropriate if, and only if, reported income (Y) is uncorrelated with the propensity to self respond (R). In general, this can be assured only if the researcher intervenes in the selection process and, in effect, randomly sets the value of R. Note that the value we would obtain if all respondents were required to report for themselves is $\bar{Y} + \beta(s)$. If R and Y are negatively correlated the estimated average earnings of self-reporters would represent an under estimate of this amount. Furthermore, the corresponding average for proxy respondents would be higher than if all interviews were conducted with proxy respondents. A positive correlation will result in discrepancies of the opposite direction.

It is important to note that both Y and R may be composed of systematic and random components and, therefore, any correlation between them may be due to associations of either component. According to the human capital model, for instance, the actual earnings of individual 'i' are determined by a set (X) of measures of prior investments in human capital and other individual characteristics, and luck. That is:

$$Y_i = X_i\Gamma + \psi_i. \qquad 3)$$

Γ is a vector of structural parameters relating the level of X to earnings and ψ is a random error term uncorrelated with X and with zero expectation and constant variance (Ψ). Similarly, the propensity of the individual to respond for himself is determined by a set Z of characteristics of the individual, of potential proxy respondents, of the interviewer, of the interview situation itself, and, again, luck v. In other words:

$$R_i = Z_i\Lambda + v_i \qquad 4)$$

where Λ is a vector of structural parameters which relate the characteristics Z to the unobserved propensity to self-respond.

In light of 4) equation 2) can be expressed as:

$$Y^r_{it} = \begin{cases} X_i\Gamma + \beta^s + e^s_{it} & \text{iff } v_i > -Z_i\Lambda \\ X_i\Gamma + \beta^p + e^p_{it} & \text{iff } v_i \leq -Z_i\Lambda \end{cases} \tag{2'}$$

where $e^r_{it} \equiv \psi_i + \epsilon^r_{it}$.

The term "selectivity bias" is reserved in the econometrics literature for ases resulting from correlations of the stochastic components only. Any departures of the expectation of observed means from their true values which are due to differences in the distributions of X or Z are termed specification bias. Selectivity bias, in this sense, will result so long as there is any systematic relation between e and v. Suppose, for instance, that they are linearly related. That is:

$$e^r_{it} = bv_i + \mu^r_{it}$$

where μ has zero expectation and is independent of v. The average earnings reported by self-respondents would then be given by:

$$\begin{aligned} E(Y^s_{it}) &= \bar{X}_s\Gamma + \beta^s + bE(v_i | v_i > -Z_i\Lambda) \\ &= \bar{X}_s\Gamma + \beta^s + b\int_{-Z_i\Lambda}^{\infty} v_i\,\phi(v_i)\,dv_i \end{aligned} \tag{5a}$$

where ϕ is the probability density function corresponding to the distribution of v.

The corresponding expression for the mean of proxy reports is:

$$\begin{aligned} E(Y^p_{it}) &= \bar{X}_p\Gamma + \beta^p + bE(v_i | v_i \leq -Z_i\Lambda) \\ &= \bar{X}_p\Gamma + \beta^p + b\int_{-\infty}^{-Z_i\Lambda} v_i\cdot\phi(v_i)\,dv_i \end{aligned} \tag{5b}$$

We should note that even if the expectation of v is zero, equations 5a) a. b) will not correspond to their true population values because the v distribution in each is trunca d. This can be shown graphically as in Figure 1. For graphical simplicity we assume that bot v and e are normally distributed with zero mean and unit variance. Figure 1a illustrates, for the entire population, the situation in which the two error terms are uncorrelated (b=0). The distribution of the selection error term v is plotted on the horizontal axis, while that for the behavioral error e is plotted on the vertical. The limits of the 95% confidence interval for each distribution are at +1.96 and −1.96, and the corresponding confidence area for the *bivariate* distribution is depicted by the circle in the interior of the graph.[2] The centroid (i.e. the point corresponding to the mean of both distributions) of the plot is at the population averages of $<0,0>$. Figure 1b shows the effects of self-selection on both distributions and their conditional means. We observe only cases in which $v > -Z\Lambda$. The distribution of the selection errors is therefore truncated and no longer has a zero mean. Indeed, it can be shown that the mean of such a truncated normal distribution is equal to the probability density at the truncation point divided by the cumulative density from that point to $+\infty$ (i.e. the probability of inclusion in the selected sample).[3] Because the two error terms are uncorrelated, however, this truncation has no effect on the mean of the behavioral and measurement error term—there is a perfect balance in the exclusion of high and low error cases. The centroid of the conditional bivariate distribution is $<\phi(\cdot)/(1-\Phi(\cdot)),0>$ and selection leaves the mean of e and, therefore, the mean of Y unaffected.

The more general situation in which the two errors are correlated is depicted graphically in Figures 1c and 1d. Because the two random variables are, in this illustration positively, correlated, the 95% confidence region is no longer circular (in the parlance of econometrics the errors are non-spherical) but is an upward sloping ellipse. Still, without selection, the unconditional centroid is $<0,0>$ (see Figure 1c). With selection, however, the conditional means of both v and e are altered. This can be seen clearly in Figure 1d. Again, the conditional expectation of the v's is $\phi(\cdot)/(1-\Phi(\cdot))$ but the corresponding expectation of the e's is $b[\phi(\cdot)/(1-\Phi(\cdot))]$. Since the actual population expectation of the e's is zero, this latter amount is the amount by which selectivity biases the average reports of Y.

The relationship between reporting bias and selectivity bias can also be shown graphically, as in Figure 2. Assuming again that there is a positive correlation between the selection and behavioral errors, and that proxy reports are more positively biased than self reports, then the 95% confidence regions we might obtain for the two types of respondents if we had randomized selection would appear as depicted in Figure 2a. Because of the random selection rule (and in spite of the positive correlation reflected by the positively inclined ellipses) the difference in the simple averages of the two reports is an unbiased estimate of the true difference in reporting bias. Figures 2b through 2d illustrate how selectivity affects our ability to draw inferences about relative reporting bias from simple average

[2] If one imagines a mound of probability density projecting upwards from the page, the circle depicted is the projection of all points on the surface corresponding to a density of .05844 ($=\phi(1.96)$). Ninety-five percent of the ***volume*** of the density function is contained within a cylinder above this circle.
[3] See Maddala (1983) pages 365–367.

reports. In Figure 2b differential reporting bias and selectivity bias counteract each other and the difference in observed averages understates the extent of true differences in reported earnings between self and proxy respondents. In Figure 2c, we reverse our earlier assumption that proxy re[illegible]ts are more positively bias than self reports, and, as a result, selectivity bi[illegible] and reporting bias now [illegible]nforce each other, and the obs[illegible]rved averages overstate the true extent of differential reporting bia[illegible] [illegible]inally, in Figure 2d, we see that the simple differences in observed averages will correspor[illegible] [illegible]o true differential reporting bias even with selection if the two errors are uncorrelated.

Remedial Measures

There are a number of techniques available to purge selectivity bias from estimates of average reports self selected groups.[4] They all require some prior knowledge of the form of the probability density function ϕ. The most celebrated of these techniques are those developed by Heckman (1979) and are based on the assumption that ϕ is the normal density function. In this case one can obtain consistent and efficient estimates of the parameters of equations 5a) and/or 5b) by maximizing the likelihood function:

$$L = \Pi\, [\Phi(-Z\Lambda)]^{-1}\, [1/\sigma] \exp\, [-(1/2\sigma^2)(y_i - X_i\Gamma)]$$

$$\text{x}\ \Phi\left[\frac{[-Z_i\Lambda - \rho(y_i - X_i\Gamma)]/\sigma}{(1-\rho^2)^{1/2}}\right]$$

where ϕ and Φ are the standard normal density and cumulative distribution functions, respectively, σ is the standard deviation of the e's, and ρ is the correlation of the e's and υ's.

Alternatively, consistent estimates can be obtained by including an instrumental variable for $E(\upsilon | \upsilon > -Z\Lambda)$ in 5a) and applying Ordinary Least Squares. This is possible because under the assumption of normality of the υ's, equation 5a) reduces to:

$$E(Y^s_{it}) = \bar{X}_s\Gamma + \beta^s + b[\phi(-Z_i\Lambda)/(1-\Phi(-Z_i\Lambda))]$$

and it is possible to obtain consistent estimates of the bracketed terms—known as th[illegible] [illegible]rse Mill's ratio—from an ordinary probit of the response decision.[5] This procedure, known as [illegible] [illegible]eckman Two Step', involves first estimating a probit on whether the case is a self-report. T[illegible] [illegible]rovides consistent estimates of the parameters of the R function (scaled by the standard devi[illegible] [illegible] of υ).

[4]Olsen (1980) provides what is perhaps the clearest and most [illegible] [illegible]ments of these [illegible]thods. Maddala (1983) provides a more detailed development.

[5]Few statistical package programs provide options for computing the Mill's ratio from [illegible] probit coefficients. Appendix B provides source code which will accomplish it, however.

These coefficients are then used to construct an instrument for the inverse-Mill's ratio which is included as a predictor in the behavioral equation.

Because a change in any one of the predictors of earnings can exert its effect either directly or via its effect on the probability of self-response, care must be taken that the combined model is identified. Technically, differences in the functional form of the behavioral and selection models is sufficient to guarantee identification. This type of identification, however, inspires less confidence than what we might call rank identification—i.e. identification through the inclusion of unique variables in each structural relationship. As we shall see below, because of the 'respondent rules' of the SIPP it is relatively easy to find variables which should affect the probability of self-response which should not, at least directly, affect earnings. Variables which should affect earnings and not selection, however, are far more difficult to isolate. In our analysis, therefore, we will rely on both functional form and structural identification.

Several alternative procedures are based on alternative assumptions regarding the distribution of the v's. These include Olsen's OLS correction procedure based on the assumption that the v's follow a uniform distribution, and Lee's procedures for a variety of other distributions.[6]

An especially convenient estimating equation can be obtained by combining the modified versions of equations 5a and 5b into a single model. This yields:

$$Y_i^r = (\alpha + \beta_i^p) + (\beta_i^s - \beta_i^p)\, S_i + X_i \Gamma + b\, \lambda_i + e_i^r \qquad 6)$$

where α is the intercept, S is a dummy variable taking on the value one if the case is a self report, and λ is the expected value of v which will depend on whether the case is a self or proxy report. In the case of the Heckman two-step procedure λ, for each individual i, will equal $\phi(\text{-}Z\Lambda)/(1-\Phi(\text{-}Z\Lambda))$ for self respondents, and $-\ \phi(\cdot)/\Phi(\cdot)$ for proxy respondents. The coefficient on the dummy variable S is directly interpretable as the net reporting bias, and that on λ as the effect of selectivity on the overall average report.

[6]One might think that once the probabilities of self-response are estimated it should be possible to develop inverse probability weights which would eliminate the selectivity bias. The problem is that the conditional expectation of the v's in equation 5a (5b) are positive (negative) for all self (proxy) respondents, and the only weights which would make the expectation across individuals of this term zero are zero weights. The only other possibility for weights to correct the problem is if the weighted expectations of v are orthogonal to not only all the X variables but to Y itself. Clearly weights based on the predicted probabilities from a probit of self response will not guarantee this result.

Section II
Empirical Analysis

Model Specification and Sample Restrictions

The behavioral model we employ in our analysis is the so-called human capital model of earnings determination. According to this model, earnings are determined by the individual's marginal productivity which, in turn, is determined by the individual's past investments in 'human capital'. The most important forms of investment in human capital are formal education and on-the-job learning. The amount of on-the-job investment is thought to increase at a decreasing rate with the amount of time spent in the labor market—i.e. with 'experience'. In addition to these human capital variables, social factors such as race are also thought to affect earnings. Non-whites will earn less than whites either because of direct wage discrimination, or as a result of racial differences in the quality of education and access to on-the-job learning opportunities.

The simple human capital model which this implies has been most frequently restricted to, and most generally agreed upon for, prime-age male workers. For this reason, we restrict our sample to males aged 25 to 55 who worked for pay in at least two months of the first twelve months of the SIPP. Because age and experience for these men are so highly correlated, it is common to use age less education (less six years) as the measure of experience. One very attractive aspect of this model for the purposes at hand is that each of the above determinants of earnings is subject to very little reporting error. When we look for discrepancies between self and proxy reports of these characteristics for cases where the respondent changed between waves one and two, we find perfect agreement on race, only one percent of the sample disagreeing on age (year of birth), and only about three percent disagreeing on educational attainment.

In addition to this age-gender-work status restriction, we restrict our sample to those individuals who provide complete data. The reason is that we do not want to confound the effects of the SIPP imputation procedures with the effects of self versus proxy response status. While the existence of imputed amounts is an important dimension of data quality, and while it is more prevalent for proxy than for self respondents,[7] it is not part of what we generally mean by either pr[illegible] bias or proxy error variance. Finally, because there is no chance for primary individuals to have a [illegible]oxy response, they are eliminated from our analysis.

We will confine our attention in the subsequent analysis to the earnings of men [illegible]or[illegible]n Wave II of the 1984 SIPP Panel. Indeed, most of our attention will center on their earnings [illegible]n a [illegible]e month—December 1983. The reason for focusing on Wave II rather than Wave I is that we wa[illegible]ed

[7]Approximately twice as many wage-salary amounts in Wave II of the 1984 SIPP Panel were imputed for proxy- (5.5%) as for self-(2.7%) prime-age male respondents.

the interviewers and respondents to have some time to become accustomed to the reporting task and respondent rules. December is chosen because it is common to all three rotation groups in our sample.[8] Table 1 presents the simple (weighted) average reports of December 1983 earnings for our sample of men by response status. The average earnings for the two groups are virtually identical, although the variability of proxy reports is greater than that of self reports. This result is somewhat remarkable, because, as we shall see below, there are systematic differences in the characteristics of the men being reported on by the two groups of respondents. One such difference is apparent from the means presented in the Table. Self reporters are more highly educated than proxy. They are also slightly more likely to be non-white. While these differences may not appear substantial, the effect of the education difference on earnings would amount to about one hundred dollars a month. That it does not, suggests that there may be some differential reporting bias, with self reporters giving lower reports than proxies.

Another remarkable aspect of the figures presented in Table 1 is that somewhat more than half of the reports for prime-age male (non-primary individual) workers in Wave II are provided by proxy respondents. The implication of this is that if there are differences in the quality of reports then these differences are going to have a major impact on the overall quality of SIPP earnings data for prime age men. If proxy (self) response were a rare phenomenon, then there would be little overall effect on data quality even if the reports were of very poor quality. As it stands, the distribution of report status is such that it will maximally affect overall quality.

Determinants of Self/Proxy Response Status

As noted above, understanding the self/proxy selection mechanism is crucial to any attempt to correct for selectivity bias in making self/proxy data quality comparisons. In this section self/proxy status is viewed as the result of a complicated interaction of family time allocation decisions, survey respondent rules, and interviewer judgements. Since the interview takes place in the home, the probability of finding the individual there and securing his co-operation should be inversely related to the amount of time he spends away from the home at work. Similarly, the chances of finding an acceptable proxy respondent should be inversely proportional to the work hours of other potential respondents. For this reason, our empirical specification of the self/proxy model includes the labor hours of the man in question, as well as an indicator of the amount of market work of other family members—their *per capita* labor earnings.[9] Furthermore, since the amount of time spent in the

[8]Rotation Group IV is eliminated from our sample because we use a three wave merged data file, and the third wave is missing for it.

[9]This is an imperfect proxy for what we would like to include which is the 'home-time' and wage rate of each potential proxy respondent. Unfortunately, the data management task of constructing such measures from the SIPP data files is beyond the scope of this investigation.

Table 1

Simple Average Reports by Type of Respondent
(Prime-age male workers)

	Self	Proxy
December 1983 earnings	$1,899.77 (1,355.38)	$1,896.04 (1,551.62)
Education	13.35 (2.96)	12.92 (2.90)
Experience	18.16 (9.30)	18.26 (9.36)
% Non-white	11.58 (32.00)	10.16 (30.21)
n	1800	2048

home is likely to vary across the life-cycle, and for cultural reasons by race, we include measures of the individual's age, race, and marital status. Finally, since one's motivation to complete the interview oneself probably increases with one's exposure to the types of research performed with survey data, we would expect self-response to increase with education.

The respondent rules for the SIPP restrict the interviewer to choosing only certain household members as proxy respondents when the interview can not be readily taken with the designated respondent. In particular, these must be adults, and within the group of adults the interviewer is required to take first spouses, and then other relatives.[10] For this reason we include measures of whether the reference individual is the spouse or child of the householder and of the number of potential respondents in the estimating equation.

Finally, the interviewer him or herself makes judgements as to the ability of potential respondents to provide the information required to complete the interview. If, from prior experience with the individual, or from comments of the potential proxy respondents, the interviewer believes that the individual's situation is too complex for a proxy respondent to accurately report it, she or he has the option of rescheduling the interview for some other time when the individual will be at home and able to be interviewed. While most interviews are taken during the work week, he or she may choose to return to the house during the weekend to do the interview with the designated person. For these reasons we include, as a proxy for how complicated the individual's situation is, the actual amount of time required to complete the questionnaire, and as an indicator of how important the interviewer thought obtaining self-reports for this individual, a dummy variable for whether the interview was taken on a weekend day.

Table 2 presents the estimated effects of these variables on the probability of self-response during Wave 2 of the SIPP under two assumptions regarding the distribution of the selection error term—uniformity, and normality. The first column presents Ordinary Least Squares results obtained when the dependent variable is specified as 1 if the individual responded for himself, and 0 if a proxy response was obtained. This specification is appropriate if the v's are uniformly distributed. Column two presents the corresponding probit parameter estimates which are appropriate under the assumption that the v are distributed normally. In both cases the most important predictor, in terms or explanatory power as measured by the t-ratio, of self response is the labor earnings of other household members. The OLS coefficient suggests that each thousand dollars of other's labor income increases the probability that the individual will respond for himself by seventeen percent. The corresponding probit coefficient of .00043 is interpretable as saying that each thousand dollars of other's labor earnings increases the propensity for self response index by .43 standard normal deviates,

[10] In the second and subsequent waves, the proxy respondent from the previous wave is to be given higher priority than the spouse of the designated respondent.

which at the mean corresponds to an increase of, again, 17% (16.6%) in the probability of self response.[11] Thus, as hypothesized, the more constrained, or valuable, the time of other household members, the less likely they are to act as proxy respondents for the man in question.

Also as hypothesized, the more limited the home time of the designated respondent lower is his probability of responding for himself. The OLS estimates indicate that a man workin standard forty-hour week would be approximately 6.5% less likely to self-respond than an otherwi milar man who was not working at all. Again the magnitude of this effect is very close to that i ated by the probit coefficient.

While education has an effect in the direction expected, and while this effect significant in both specifications, neither age nor race have any discernable effect on self/proxy reporting behavior. Marital status and the individual's relationship to the reference person are, on the other hand, quite important. Married men, *ceteris paribus*, are far less likely (roughly 30% less likely in both specifications) to report for themselves than are unmarried men.

The effects of the respondent rules are also clearly important determinants of self-reporting in both specifications of the model. Adult children of the householder are thirty percent and spouses thirteen percent less likely than the householder himself to be self-reporters. Furthermore, for each additional potential respondent, the probability that a self-report will be given falls by approximately three percentage points.

Finally, the results presented in Table 2 suggest that interviewer judgement and behaviors may play an important role in determining response status. The more complicated the interview, as measured by the amount of time it takes the respondent to complete it, the more likely it is that the designated individual will be the respondent. Each ten minutes increase in interview length is associated with nearly a seven percent increase in the probability of self-response. Furthermore, interviews taken on weekends are nearly fifteen percent more likely to be conducted with the designated respondent than those taken during the week. Of course, for each of these latter measures it is possible that causation is running in the opposite direction from that hypothesized. It is possible, for instance, that self-reporters spend more time thinking about and formulatin e answers to questions, and therefore the interview takes longer. To the extent that this is the cas e coefficients estimated in Table 2 will suffer from simultaneity bias. Since our purpose here is to o as good an instrument for the self-reporting obabilities as possible, however, rather than to pred the e ts on self reporting of changes in e independent variables, such simultaneity bias is a serious problem.

[11] This is the area under the standard normal density function from zero to .43.

Table 2

Self/Proxy Respondent Selection Model Estimates Under Two Distributional Assumptions
(All working males 25–54 years of age living with a potential
proxy: Dependent variable whether self-report in Wave 2)

	Uniform	Normal
Constant	.521**	.131
	(.073)	(.204)
Other's per capita labor earnings ($1000)	.174**	.429**
	(.020)	(.053)
Work hours (100's)	−.163**	−.451**
	(.050)	(.133)
Age (100's)	.186	.477
	(.099)	(.265)
Non-white	.019	.051
	(.026)	(.069)
Married	−.297**	−.877**
	(.045)	(.135)
Education (100's)	.982**	2.757**
	(.266)	(.715)
# Eligible proxies	−.029**	0.075**
	(.009)	(.024)
Spouse of reference person	−.131**	−.356**
	(.034)	(.094)
child of reference person	−.286**	−.853**
	(.049)	(.145)
Whether weekend	.147**	.385**
	(.023)	(.243)
Length of interview	.675**	1.905**
	(.085)	(.243)
$\bar{R}^2$	.0697	–
F, χ^2	27.42**	292.39**
	(11,3870)	(11)

*Significant at the 5% level.
**Significant at the 1% level.

Overall, both specifications of the model do a credible job of 'explaining' self/proxy response status. Tests of the significance of the goodness of fit are hugely significant. Furthermore, well over sixty percent of the cases in the sample are correctly classified into self and proxy categories using the estimated coefficients. Nevertheless, there is considerable room for improvement since less than ten percent of the variance in the dependent variable of the OLS specification is explained by the predictors. A better fit would result in more precise estimates of the effects of selectivity bias in the subsequent analysis.

Self/Proxy Reporting Bias and Selectivity Bias for Monthly Earnings

Table 3 presents labor earnings generating equations estimated according to equation 6. The equations include terms to capture the effects of differential reporting bias (the self-report variable) and selectivity bias (the inverse-Mill's ratio), and are estimated on the full sample of 3848 prime-age males. The most powerful predictors of earnings in December 1983 for these sample men are experience, education and race. As in almost all human capital models, experience increases earnings in a non-linear fashion—initial levels of experience having a stronger impact than subsequent ones. The estimates suggest that the first year of experience increases earnings about seventy-two dollars per month (= (1070–2*176)/10), while the incremental effect of a year's experience at the sample mean of 1.8 decades is only forty-three dollars per month.

The effects of education on earnings in our estimates are also consistent with the human capital model and those reported elsewhere in the literature. Unlike experience, the human capital model suggests that the *rate of return* to education should be constant. In the more traditional semi-log specification, this would result in a linear relationship between education and the natural log of earnings. Since, in the present specification, we are employing an earnings generating function similar to that used in evaluating manpower programs, we should and do observe education affecting earnings exponentially.[12] The rate of return to education implied by our estimates of slightly more than eleven percent (=(19.69+2*7.39*13)/1899) at the sample mean of 13 years of education is higher than that typically found, however. The reason is that, unlike most analysts, we include the unemployed in our sample and unemployment is negatively associated with education. This is also the probable cause of our estimate of the effect of being non-white being more strongly negative than that found

[12]The reason we choose to examine levels of earnings rather than the natural logarithm of earnings is that, as in the case in the training evaluation problem, cases with zero earnings are important to us. A major source of differential reporting error could be differences in the tendencies of proxy- and self-respondents to report zero earnings when in fact there were earnings during the month in question, and *vice versa*. The natural logarithm of zero is undefined. While we could add some arbitrary amount to everyone's earnings before taking logs, past experience shows that the precise quantity chosen can make a substantial difference in the estimates.

Table 3

Earnings Generating Functions with Selectivity Correction and Self/Proxy Treatment Terms

	OLS	Empirical Bayes Ridge Regression	Partitioned Self	Sample Proxy
Constant	−666.53*	−738.34	−1074.25	−665.16
	(282.79)		(298.78)	(283.00)
Whether self report	−408.86*	−329.89	[−409.09]	
	(172.98)	(153.78)		
Inverse mills ratio	217.72*	168.38	216.87†	216.87†
	(111.14)	(98.92)	(110.94)	(110.97)
Education	19.69	58.77	19.76	19.76
	(41.49)	(33.53)	(4.145)	(41.45)
Education squared	7.39**	5.80**	7.38**	7.38**
	(1.60)	(1.29)	(1.60)	91.60)
Experience decades	1070.00**	840.00**	1071.00**	1071.00**
	(102.00)	(88.00)	(102.00)	(102.00
Experience squared	−176.00**	−122.00**	−176.00**	−176.00**
	(24.00)	(32.00)	(24.00)	(24.00)
Whether non white	−463.62**	−465.15**	−463.08**	−463.08**
	(69.28)	(88.35)	(69.16)	(69.16)
$\bar{R}^2$	.1777	.1735	.1797	.1797
F,χ^2	119.73**	115.43**	55.30**	55.30**
d.f.	(7,3840)	(7,3840)	7	7

†Significant at the 10% level.
*Significant at the 5% level.
**Significant at the 1% level.

elsewhere. For December 1983 our estimates indicate that whites enjoyed a monthly wage premium of more than $460 over nonwhites. That the unemployment rate for whites at that time was less than half that for nonwhites, is, no doubt, a substantial part of the reason.

Of course, our primary interest at present is not in the structure of labor markets but in the relative quality of self and proxy reports of earnings. The coefficient of −408.86 on the self-report dummy variable indicates substantial and significant relative reporting bias.[13] Once one controls for differences in education, experience, and race, and for selectivity, self respondents provide reports which are more than four hundred dollars per month lower than those provided by proxy [illegible]ondents. This could happen if, for instance, self reporters are basing t[illegible] on the "take-ho[illegible]" amount of their paychecks while pro[illegible]porters are forming the answer by dividing annual labo[illegible]rnings by twelve. If so, then the self reports would be seriously biased downward (from what w[illegible]tended by the study designers) because take-home pay excludes taxes, social security, insuranc[illegible]d a host of other deductions from gross wages.[14] The answer provided by proxy's from the [illegible]-cut method might well be closer to the gross monthly pay than is the actual take-home [illegible]. The SIPP questionnaire actually encourages the respondent to recall paychecks, but the procedures allow the respondent to formulate the answer in other ways if the paychecks can not be recalled. Furthermore, it seems quite plausible that, especially when considering pay checks, the net pay is more salient than gross pay to many respondents.

Selectivity somewhat more than offsets the negative relative reporting bias. The large positive coefficient on the inverse Mill's ratio means that there is a positive association between the stochastic components in both the earnings and the selection equations. Thus, men who are doing better (in terms of earnings) than we would expect given their education, experience and race are more likely to respond for themselves than we would expect given their characteristics and the characteristics of the interview situation. Conversely, those doing less well than they 'should' are less likely to talk about it than we would otherwise expect.

If one accepts this finding then one must conclude that the practice of allowing interviews to be taken with proxy respondents results in estimated average monthly earnings which are some $200 per month higher than would have been obtained under a 'self-report only' respondent rule. This is the meaning of the coefficient on the inverse-Mill's ratio of 217.72, which is significant, although barely, at the 5% level of confidence.[15]

[13]Throughout our discussion of the empirical results we will base our tests on variance estimates calculated under the assumption of simple random sampling. The SIPP is, of course, a complex multi-stage probability sample, and proper variance estimates would, in all likelihood, b[illegible]mewhat larger than those presented. Conservative tests can be constructed by assuming a desi[illegible]ct of 1.3 and dividing F and Chi-square statistics by this amount, or t-ratios by its square root.

[14]This is the reason suggested by some of the most knowledgeable SIPP analysts [illegible]PP earnings estimates being lower than those of the CPS. If this is the case then it argues strong[illegible]r changing the SIPP instrument.

[15]We should note, at least in passing, that the same general result of negative relative [illegible]ing bias and positive selectivity bias holds under a wide range of alternative specifications. In [illegible]cular, it holds under the assumption of uniformly distributed sele[illegible] errors, as well as in most sp[illegible]ications which include additional predictors of selection in the behavioral model. The result does not hold, however, if one includes work hours as a predictor of earnings. It is not clear, however, what the meaning of such a model is.

Of course, there are many reasons why one might not accept these estimates. Even if one accepts the basic human capital model and the self-selection model, there is always the possibility that the results are being driven by a purely technical problem such as multi-colinearity. The inverse-Mill's ratio and the self report 'treatment' dummy variable are highly colinear with a correlation coefficient of .9637. Part of the reason for this is purely mechanical—by construction, each self-respondent's Mill's ratio is positive while each proxy-respondent's is negative. While the resulting multi-colinearity is not sufficiently high to cause serious problems in and of itself,[16] these measures are included with other predictors such as education and its square which are correlated with them and with each other, and it may be that the coefficients are unstable as a result. At the expense of imparting some bias toward zero in the coefficients, we can attempt a correction for multi-colinearity by means of the so-called 'Empirical Bayes Ridge Regression' (EBRR).[17] As with all Ridge Regressions, the basic idea is to stabilize the estimates by adding a constant to the diagonal elements of the sums of squares and cross-products matrix of the predictors before inverting it. The EBRR procedure is somewhat less *ad hoc* than some other procedures in that the size of the constant to be added is determined by the variance of the prior distribution of the parameters, and this is estimated from the data.

The second column of figures in Table 3 presents the results of the EBRR procedure applied to our selectivity bias/self/proxy treatment model. The result is an approximate twenty percent reduction in the magnitude of both the estimated selectivity term and self-report relative bias term. The standard errors for these coefficients are reduced only slightly more than ten percent. Never the less, the negative coefficient on the self-report dummy remains significant and both coefficients remain substantial, even though we know they are biased, by the procedure, toward zero. In fact, our attempt to improve the stability of the estimates seems to have been most effective for education and its square—perhaps because these are the variables which are most seriously affected by multi-colinearity in the first place. Thus, the EBRR estimates, while suggesting that there may be some problems with multi-colinearity for the education variables, do not differ sufficiently from the OLS results to indicate that colinearity is the major factor leading to our finding significant selectivity and relative reporting bias.

Another possibility is that our results are merely an artifact of a violation of some of the other assumptions underlying the model presented in equation 6).[18] By placing both self and proxy

[16]The eigen values of the moments matrix of these two predictorsare .015 and .94.
[17]See Amemiya (1986), p 60–61.
[18]Two additional assumptions which might be violated are the assumption regarding the normality of the v's, and the assumption that education, age, and race belong in the selection equation. Appendix A presents the results obtained when we estimate the model under the alternative assumptions that the v's are uniformly distributed and the education, age and race do not belong in the selection equation. As it turns out, the results are quite robust to these assumptions.

reporters in the same estimating equation, we were implicitly assuming that, aside from differences related to selectivity, they share a common systematic and stochastic structure. The onsistency of this assumption with the data can be tested by reformulating our model slightly. ather than estimate it as a single equation, we can partition the sample int groups f and p v reporters, and estimate a selection corrected model for each simultaneously. The assumption a mmon structure can then be imposed in the form of cross-group constraints on the coeffici Each such constraint will reduce the overall goodness of fit of the model to the data and likeli ratio tests of the appropriateness of these constraints can be conducted. The details of h ese tests are constructed are laid out in detail in Appendix C.

Viewed in this way the model presented in equation 6) has seven over-identifying constraints—one each for the inverse Mill's ratio, education, its square, experience, its square, race, and one for the variance of the composite-error term (e). When we estimate the model using Jöreskog and Sörbom's LISREL.IV (1978) algorithm we obtain the results presented in columns three and four of Table 3. The differential self-proxy bias, or treatment effect, implicit in this formulation is obtained from the difference in the intercept terms. With all seven of the over-identifying constraints in place, the parameter estimates are virtually identical to those obtained by the OLS procedure and presented in column 1.[19] The chi-square statistic of 55.3 with 7 degrees of freedom, which is defined as twice the value of the likelihood function, is strongly significant. Since this is a measure of the harm done to the goodness-of-fit by assuming proxy and self reports share a common structure, we must conclude that they do not.

In order to see which of the assumptions does most harm to the goodness of fit, we first relax them all. The estimates obtained from this just-identified, or 'saturated', model are presented in columns one and two of Table 4a. While all of the formerly constrained coefficients now differ between the two sub-samples, only one differs by more than a single standard error. The error variance for proxy-reports of earnings is nearly forty-percent larger than that for self-reports. This difference is hugely significant and is the primary reason for the poor performance of the constrained model.[20] This result could reflect either of two things, or some of both. First, xy reports could have higher measurement error variances. This seems the most plausible e nation. Proxy respondents probably do not have as good access to the information needed to answ e questions as do self respondents. Furthermore, as noted in the section on the self/proxy select odel results, they seem to spend less time formulating their answers. The second possibility, howe can not be ruled out. This is that the people for whom proxy reports are obtained represent a g inely more heterogenous group in terms of their earnings experiences than of self responde s.

[19]Indeed, the only reason they are not identical is that the convergence criterion employed is based on the value of the likelihood function and is set at a somewhat generous level.
[20]This result holds even if the inverse-Mill's ratio is removed from the model entirely.

Table 4a

Partitioned Sample Earnings Generating Function Under Various Cross-group Constraints

	Free		$\beta^s = B^p$	
	Self	Proxy	Self	Proxy
Constant	−996.90 (410.96)	−576.74 (394.01)	−777.84* (284.43)	−777.84* (284.43)
Implied self treatment	[−420.16] (−)		0 (−)	
Inverse Mill's ratio	120.40 (141.01)	329.36* (171.07)	98.21 (137.78)	313.41† (169.71)
Education	26.79 (58.82)	11.63 (56.23)	−.51 (45.76)	37.35 (46.63)
Education squared	6.55** (2.24)	8.27** (2.27)	7.51** (1.83)	7.33!** (1.89)
Experience (decades)	1123.35** (133.53)	1016.00** (152.22)	1113.19** (132.83)	1031.02** (150.86)
Experience squared	−186.43** (32.31)	−165.04** (35.88)	−185.52** (32.29)	−166.91** (35.79)
Non white	−398.07** (89.71)	−532.58** (104.15)	−401.20** (89.62)	0531.13** (104.14)
$\hat{\sigma}^2$[E+4]	14.58** (.49)	20.13** (.64)	14.58** (.49)	20.13** (.64)
χ^2	0	0	.54	.54
d.f.	0	0	1	1

†Significant at the 10% level.
*Significant at the 5% level.
**Significant at the 1% level.

The remainder of Tables 4a and 4b present the results obtained when various cross-group constraints are imposed on the model. Columns three and four of Table 4a refer to the situation when the intercepts are constrained to equal each other, but all other parameters are allowed to vary across groups. This is one test of the hypothesis of no differential reporting bias. The harm done to the goodness of fit is trivial and insignificant, and we can not reject the hypothesis of no differential reporting bias. But what does it mean for the structural parameters relating education, experience and race to differ between the two subsamples, and is this plausible? There are four possible explanations. First, men who are self respondents and men who are proxy respondents may actually face different labor market conditions, and the differences in the coefficients may be reflecting this. This seems hard to believe. Second, the coefficients may differ because of some omitted variable which differs between the two groups and is not captured in the self selection model. Such omitted variables are always possible and we can not rule this possibility out. Third, the coefficients may appear to differ because of differential measurement error which is correlated with the predictors. The PSID validity study (Duncan and Hill, 1985), yielded evidence of a significant negative correlation between measurement error in earnings and the level of experience. If there is more measurement error in the proxy reports of earnings, and if the correlation of this error with experience is at least as strong for proxy as self reports then we would expect the estimated effects of experience to be smaller for proxy respondents than for self respondents. The point estimates of these effects in Table 4a are consistent with this hypothesis in that the age/earnings profile suggested by the proxy reports is both flatter and less convex than that suggest by the self reports.

The fourth possible explanation for coefficients on the human capital variables to differ between self and proxy respondents is, of course, that these differences are the result of pure chance. We can test this hypothesis by imposing equality constraints on these coefficients and examining the goodness of fit of the model. Columns one and two of Table 4b do just this. When these five restrictions are imposed the Chi-square increases from zero to 5.88. This is far from significant and we can not, therefore, rule out the possibility that the apparent differences between self and proxy parameters are merely a reflection of the luck of the draw.

In other words, while we can rule out the possibility that self and proxy reports of earnings share a common error structure, we can not rule out their sharing a common behavioral structure. If so, then the results presented in columns three and four of Table 4b provide strong evidence of significant differential reporting bias. The incremental damage done to the fit of the data from adding the constraint that the two groups share a common intercept is 5.3, which is highly significant with one degree of freedom. By this second test we can soundly reject the hypothesis of no differential reporting bias. We should note that this result is somewhat stronger than that from our earlier single-equation specification because the present model does not rely on the automatic cross-group correlation between the inverse-Mill's ratio and the self-report dummy to estimate the effects of selectivity. It uses only the within-group correlation of earnings and the Mill's ratio to estimate selectivity effects separately for each group. Thus, unlike the earlier specification, the reporting bias result can not be being driven merely by the colinearity between these variables.

Table 4b

Partitioned Sample Earnings Generating Function Under Various Cross-group Constraints

	$\Gamma^s = \Gamma^p$		$\beta^s = \beta^p, \Gamma^p = \Gamma^s$	
	Self	Proxy	Self	Proxy
Constant	−1037.39** (300.50)	−644.72	−804.57 (283.17)	−804.57 (283.17)
Implied self treatment	[−392.67] (−)		0 (−)	
Inverse Mill's ratio	160.60 (138.28)	259.07 (166.49)	−57.00 (101.0)	−15.59 (116.34)
Education	22.33 (41.36)	22.33 (41.36)	21.71 (41.39)	21.71 (41.39)
Education squared	7.19** (1.59)	7.19** (1.59)	7.00** (1.59)	7.00** (1.59)
Experience	1076.50** (100.31)	1078.50** (100.31)	1082.01** (100.36)	1082.01** (100.36)
Experience squared	−177.87** (23.99)	−177.87** (23.99)	−179.81** (23.99)	−179.81** (23.99)
Non white	−453.20* (68.01)	−453.20** (68.01)	−467.62** (67.76)	−467.62** (67.76)
$\hat{\sigma}^2$	14.60** (.49)	20.16** (.64)	14.61** (.49)	20.20** (.64)
χ^2	5.88	5.88	11.18	11.18
d.f.	5	5	6	6

†Significant at the 10% level.
*Significant at the 5% level.
**Significant at the 1% level.

In sum, there is evidence of strong differential reporting bias on earnings for December 1983 between self and proxy reporters in the SIPP. Once control is made for education, experience, race and self selectivity, self respondents provide reports which are some four hundred dollars lower than the reports provided by proxy respondents. Whether or not this differential bias is statistically significant, however, is matter of interpretation. If one believes that the men being reported on by themselves and by proxy respondents are drawn from the same labor market and therefore share a common behavioral structure, then the differential bias is significant. If, on the other hand, one believes, that for any one of a number of reasons, the behavioral parameter estimates should differ between the two groups, then the differential bias is not significant. One can reject neither of these hypotheses with the data. *In either event, the evidence of larger composite error-variances for proxy respondents is strong and compelling.* This result holds with or without cross-group constraints in other parameter estimates and with or without control for selectivity.

There are a number of ways for the estimated composite error variance to be larger for proxy than for self reports. It could be that a few extreme proxy reports of earnings are driving the result. In order to test for this possibility we re-estimate the partitioned sample model excluding cases with reported monthly earnings in excess of eight thousand dollars. The results of this analysis are presented in Table 5. The elimination of these out-lying cases reduces the sample size by roughly two and one-half percent for each subsample. It also results in an estimated composite error-variance which is lower for proxy than for self reporters.[21] This implies that our previous result that proxy reports were more variable than self reports was the result of a relatively small number of cases with extremely high reported earnings. Once these cases are removed, the estimated variance is lower. This does not, however, necessarily mean that proxy reports contain less measurement error. It is quite possible that the strategy employed by proxy respondents to formulate their answers to the earnings questions are based on what the respondent believes is a typical level of earnings for someone with the subject's level of education, experience, and his race. If so it may fit quite well with the human capital model and still be quite wrong.

Conclusions

The wisdom of lenient respondent rules is not at all clear. The reduced survey costs and nonresponse rates they enable may be coming at a considerable price in terms of data quality. Any comfort one might derive from the fact that the average earnings reported for prime-age males by proxy respondents is virtually identical to that reported by self respondents may be purely illusory. Once controls are attempted for self-selectivity, the data suggest substantial differential reporting bias and error variance between self and proxy reports of earnings. While the net differential bias is very imprecisely estimated in our models, the point estimates indicate that the earnings reported by self respondents is more than twenty percent less than that reported by proxy respondents. If the differential bias is this large, and especially since there is also strong evidence of differential error

[21] We should note that the relative reporting and selection bias results continue to hold when outliers are removed. If anything, their effects are even stronger than with the full sample.

Table 5

Partitioned Sample Earnings Generating Function Estimates: Outliers Removed

	Self	Proxy
Constant	−1095.79** (239.64)	−456.91
Implied self treatment	[−638.88]**	
Inverse Mill's ratio	395.01** (123.41)	160.63 (123.41)
Education	25.80 (32.32)	25.80** (32.32)
Education squared	6.44** (1.25)	6.44** (1.25)
Experience	992.23** (79.97)	992.23** (79.87)
Experience squared	−163.91** (19.02)	−163.91** (19.02)
Non white	−437.36** (54.13)	−437.36** (54.13)
$\hat{\sigma}^2$	11.36** (3.83)	9.88** (4.40)
χ^2	7.45	7.45
d.f.	5	5

variance, then lenient respondent rules are probably not optimal in a mean-square-error sense. It would take substantial costs savings, indeed, to justify this level of contamination. We must not forget, however, that our estimates are model-based and violations of any of the assumptions of this model may have major impacts on our results. Whether or not the estimated differential reporting bias is significant, for instance, depends crucially on whether or not proxy and self respondents share a common behavioral structure. If they do, then the differential bias is significant. If they do not, then it is not. Current data can not provide a definitive answer.

Even if we can say definitively that differential reporting bias is substantial and important, without some external validating data we are not able to say which type of report is of higher quality. We would only be able to say that one report is better than the other and therefore we should collect data from only, or primarily, one type of respondent—without being able to say which. This unhappy state of affairs is depicted graphically in Figure 3 which plots the mean-squared error (E) implied by the estimates presented in Table 4b, as a *function* of the unknown true level of mo   y earnings. The relationship between MSE and true average earnings is parabolic for both   and proxy respondent reports. Because our estimates indicate that self reporting bias is mo   egative than proxy reporting bias, the minimum point for the self-MSE curve lies to the left of that for the proxy-MSE curve. Furthermore, since self reports had lower residual variance the minimum of the self-MSE curve is lower than that of proxys. Never the less, if true average earnings is greater (less) than about \$1900 per month, then proxy reports will be unambiguously better (worse) than self reports. If the true average were sufficiently above (below) this level we would say that all reports should be obtained from proxy (self) respondents.

Given the potential but still uncertain importance of differential data quality between self and proxy reports, and given the increasing prevalence of lenient respondent rules, it seems only prudent that the data necessary to obtain definitive answers be collected. Ideally, as Moore (1985) suggests, these data could be produced from a study combining controlled selection to self and proxy collection modes with highly reliable validating information.

References

Amemiya, Takeshi, (1985) *Advanced Econometrics* (Cambridge, Mass: Harvard University Press)

Barnow, B. S., G. G. Cain, and A. S. Goldberger (1981) "Issues in the Analysis of Selectivity Bias", in W. E. Stromsdorfer and G. Farkas (eds.) *Evaluation Studies Review Annual, Vol 5.* (Beverly Hills: Sage), pp 43–59.

Duncan, Greg J. and Daniel H. Hill (1985) "The Extent and Consequences of Measurement Error in Labor Economic Surveys", *Journal of Labor Economics*, Vol. 3, No. 4.

Hansen, M. H., W. Hurwitz and M. Bershad (1961) "Measurement Errors in Censuses and Surveys", *Bulletin of the International Statistical Institute*, 38, pp 359–374.

Heckman, James J. (1979) "Sample Bias as a Specification Error", *Econometrica* Vol 47 No. 1, pp 153–162.

Jöreskog, Karl and Dag Sörbom (1976) *LISREL IV: Estimation of Structural Equation Systems by Maximum Likelihood Methods* (Chicago: International Educational Resources).

Lee, L. F. (1983) "Generalized Econometric Models with Selectivity", *Econometrica*, 51 pp 507–512.

Maddala, G. S. (1983) *Limited-dependent and Qualitative Variables in Econometrics* (London: Cambridge University Press).

Moore, Jeffrey C. (1985) "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature", (Washington: Center for Survey Methods Research, Bureau of the Census).

Olsen, Randall J. (1980) "A Least Squares Correction for Selectivity Bias", *Econometrica* Vol 48 No. 7 pp 1815–1820.

APPENDIX A

This appendix presents estimates for the selectivity corrected self/proxy reporting bias model obtained under two alternative specifications of the selection model. The first (Table A1) is that obtained under the assumption that the selection errors are uniformly distributed. This is the model developed by Olsen (1980). The second alternative is that obtained under the assumption that neither education, race, nor age affect the probability of self reporting. As can be seen by comparing these estimates with the corresponding estimates in the main text (Table 4b), the results are quite similar under all these alternative specifications.

Table A1

Partitioned Sample Earnings Generating Function Under Two Cross-group Constraints

Olsen's OLS Correction Technique

	$\Gamma^s = \Gamma^p$		$\beta^s = \beta^p, \Gamma^p = \Gamma^s$	
	Self	Proxy	Self	Proxy
Constant	−1018.45** (299.48)	−653.06	−802.71 (282.36)	−802.71 (282.36)
Implied self treatment	[−365.39] (−)		0 (−)	
$\hat{P} - 1, \hat{P}$	−220.44 (220.21)	−401.69 (271.65)	99.08 (162.64)	20.91 (187.88)
Education	23.07 (41.41)	23.07 (41.41)	21.84 (41.43)	21.84 (41.43)
Education squared	7.15** (1.59)	7.15** (1.59)	6.99** (1.59)	6.99** (1.59)
Experience	1077.80** (100.33)	1077.80** (100.33)	1082.06** (100.37)	1082.06** (100.37)
Experience squared	−177.77** (24.00)	−177.77** (24.00)	−179.82** (23.99)	−179.82** (23.99)
Non white	−454.06* (68.04)	−454.06** (68.04)	−467.88** (67.77)	−467.88** (67.77)
$\hat{\sigma}^2$	14.60** (.49)	20.16** (.64)	14.61** (.49)	20.20** (.64)
χ^2	5.92	5.92	10.54	10.54
d.f.	5	5	6	6

†Significant at the 10% level.
*Significant at the 5% level.
**Significant at the 1% level.

Table A2

Partitioned Sample Earnings Generating Function Under Two Cross-group Constraints

Alternative Probit Model

	$\Gamma^s = \Gamma^p$		$\beta^s = \beta^p, \Gamma$ Γ^s	
	Self	Proxy	Self	Proxy
Constant	−897.13** (290.33)	−605.81 (299.83)	−781.88 (283.11)	−781.88 (283.11)
Implied self treatment	0 (−)		0 (−)	
Mill's Ratio	43.64 (134.23)	248.06 (164.95)	−112.25 (101.63)	39.13 (115.72)
Education	20.83 (41.36)	20.83 (41.36)	23.47 (41.35)	23.47 (41.35)
Education squared	7.10** (1.59)	7.10** (1.59)	6.96** (1.59)	6.96** (1.59)
Experience	1075.44** (100.40)	1075.44** (100.40)	1082.90** (100.36)	1082.90** (100.36)
Experience squared	−178.37** (23.99)	−178.37** (23.99)	−179.74** (23.99)	−179.74** (23.99)
Non white	−462.13* (67.80)	−462.13** (67.80)	−467.69** (67.76)	−467.69** (67.76)
$\hat{\sigma}^2$	14.60** (.49)	20.16** (.64)	14.61** (.49)	20.19** (.64)
χ^2	5.44	5.44	8.60	8.60
d.f.	5	5	6	6

†Significant at the 10% level.
*Significant at the 5% level.
**Significant at the 1% level.

Appendix B
Source Code for Computing the Inverse-Mill's Ratio

The following code can be modified and used to compute the inverse-Mill's ratio in computer packages which do not have Cumulative Density Functions routines for the normal distribution. The code is written in the OSIRIS.IV Recode dialect of FORTRAN. Translation to other dialects and languages should be straight forward.

```
&COMMENT CALCULATING THE RESPONSE INDEX FOR EACH CASE USING THE PROBIT RESULTS
      R101=(.131071+.0004289*R414+.027567*R1109+.00477062*V1054+.050977* -
       R1107-.877341*V2068-.07526*V2299+.019052*V2305-.0045068*V4223+.384756* -
       R201-.35642*R2056-.85254*R2057)
      R102=ABS(R101)
      R110=.9999999
&COMMENT ROUTING EXTREME CASES AROUND THE COMPUTATIONS
      IF R101 LT -5.2 THEN R111=.00001 AND R112=-.00001 AND GO TO L1
      IF R101 GT 5.2 THEN R111=9 AND R112=-9 AND GO TO L1
&COMMENT CALCULATING THE PROBABILITY DENSITY
      R103=.39894228*( 2.7182818 EXP (-.5*R102*R102))
      NAME R103'PROBABILITY DENSITY'
&COMMENT CALCULATING THE CDF USING THE POLYNOMIAL APPROXIMATION DEVELOVOPED BY
 DUNLAP & DUFFY BEHAVIORAL RESEARCH METHODS & INSTRUMENTATION 7 (1) (1971).
      R104=1/(1+.2316419*R102)
      R105=.31938153*R104
      R106=-.356563782*R104*R104
      R107=1.781477937*R104*R104*R104
      R108=-1.821255978*R104*R104*R104*R104
      R109=1.330274429*R104*R104*R104*R104*R104
      R110=1-R103*(R105+R106+R107+R108+R109)
      IF R101 LT 0 THEN R110=1-R110
      NAME R110'ONE - PHI'
&COMMENT CALCULATION THE INVERSE MILL'S RATIO
      IF R110 GT 0 THEN R111=R103/R110 ELSE R111=9
      IF R110 LT 1 THEN R112=-R103/(1-R110) ELSE R112=-9
L1    CONTINUE
      NAME R111'INVRS MILLS RATIO-SELF',R112'INVRS MILLS RATIO-PROXY'
```

APPENDIX C

Estimation was performed by comparing the product moment matrices implied by the model presented in equations 5a and 5b (Π) with the actual product moment matrices (P) calculated from the sample. For each respondent class r, the product moment matrix implied by the model is:

$$\Pi_r = n^r \begin{bmatrix} \Omega^r W^{r\prime} W^r \Omega^r + (\Psi^r + \sigma \phi) & W^{r\prime} W^r \Omega^r \\ \cdots & \cdots \\ \Omega^r W^{r\prime} W^r & W^{r\prime} W^r \end{bmatrix}$$

where $\Omega^{r\prime} = [(\alpha + \beta^r) | b^r | \Gamma^r]$ and $W^{r\prime} = [1 | \lambda^r | X^r]$.

The concentrated log-likelihood of the model given the sample for both self and proxy respondents combined is:

$$L = \sum_{r=s}^{p} [\log|\Pi_r| + tr(P_r \Pi_r^{-1}) - \log|\Pi_r| - C]$$

where C is the rank of P and Π. The various hypotheses are tested by constraining the various parameters to equality in the self and proxy sub-models (i.e. setting $\Omega^s_{jk} = \Omega^p_{jk}$) and then comparing the resulting value of the likelihood function L. Under the assumption the the combined error terms e^r are normally distributed, twice the value of the likelihood function is distributed χ^2 (asymptotically) with degrees of freedom equal to the number of constraints imposed.