Title
Modeling Earnings Measurement Error: A Multiple Imputation Approach

Authors
Brownstone, David
Velletta, Robert G.

Publication Date
1996-06-01

Review of Economics and Statistics, forthcoming

Modeling Earnings Measurement Error :
A Multiple Imputation Approach

David Brownstone*
University of California
Department of Economics
3151 Social Science Plaza
Irvine, CA 92697-5100

and

Robert G. Valletta*
Economic Research Department
Federal Reserve Bank of San Francisco
101 Market Street
San Francisco, CA 94105

July, 1996

* Financial support from the U.C. Irvine Research Unit in Mathematical Behavioral Sciences is gratefully acknowledged. An earlier version of this paper was presented at the NSF-sponsored PSID Event History Conference and the 1992 Census Bureau Annual Research Conference. McKinley Blackburn, John DiNardo, Greg Duncan, Zvi Griliches, Shulamit Kahn, Anders Klevmarken, Kevin Lang, Lee Lillard, and Carole Uhlaner provided useful comments on earlier drafts. We also thank Donald Rubin and four additional anonymous referees. The authors, however, are solely responsible for any errors.

Modeling Earnings Measurement Error :
A Multiple imputation Approach

Abstract

Recent survey validation studies suggest that measurement error in earnings data is pervasive and violates classical measurement error assumptions, and therefore may bias estimation of cross-section and longitudinal earnings models. We model the structure of earnings measurement error using data from the Panel Study of Income Dynamics Validation Study (PSIDVS). We then use Rubin's (1987) multiple imputation techniques to estimate consistent earnings equations under non-classical earnings measurement error in the PSID. Our technique is readily generalized, and the empirical results demonstrate the potential importance of correcting for measurement error in earnings and related data, particularly during recessions.

JEL Classification numbers : C19, J31

Modeling Earnings Measurement Error :
A Multiple imputation Approach

1. Introduction

Earnings equations are used to investigate a wide variety of hypotheses that characterize labor market operations. Although measurement or reporting error in earnings and related data may play an important role in earnings equations, the impact of such error has not been extensively investigated. Few studies acknowledge the presence of measurement error; those that do typically assume that it is "classical" in form and therefore has limited influence on parameter estimates if earnings are used only as the dependent variable.

A major reason for the limited treatment of measurement error is practical: until recently, virtually no data existed that enabled systematic investigation of measurement error in earnings and related variables. An early exception is data from the Gary, Seattle, and Denver income maintenance experiments conducted in the 1970s, in which respondent data were supplemented by employer-reported data available through state employment security agencies (see Greenberg, Moffitt, and Friedmann 1981; Greenberg and Halsey 1983). Because the reporting error was largely a result of experimental design, these data have not been used to examine the form of earnings measurement error more generally. Two validation data sets that have come into recent use, however -- the Panel Study of Income Dynamics Validation Study (see Bound *et. al.* 1990, 1994; Rodgers, Brown, and Duncan 1993; Pischke 1995) and the Current Population Survey-Social Security Earnings Records Exact

Match File (see Bound and Krueger 1991) -- enable systematic investigation of measurement error in earnings and related models. Recent research with these data, as described in Section 2, suggests that measurement error in earnings violates classical assumptions and therefore is likely to bias parameter estimates in earnings equations.

Existing research, however, does not use validated data to correct for the influence of measurement error on coefficients and standard errors estimated using standard data sets. Lee and Sepanski (1995) provide general techniques to combine information from validation data with information from other data sets, but do not apply their techniques. In this paper, we develop and apply an alternative methodology that links the modeling of non-classical measurement error in validated data sets with the estimation of earnings equations using standard unvalidated data (in our case, the Panel Study of Income Dynamics). We model earnings measurement error using the Panel Study of Income Dynamics Validation Study. We then use this model to provide multiple imputations of true earnings in our unvalidated data, based on techniques first developed for handling missing data in surveys (Rubin 1987 and 1996, Rubin and Schenker 1986). Our approach echoes the U.S. Census Bureau's recalibration of 1970 and 1980 census industry and occupation codes. This project produced multiply-imputed 1980 codes based on logistic regression models applied to a 1970 census research sub-sample for which both 1970 and 1980 codes were listed. The imputations are included in the 1970 census public-use samples (see for example Clogg *et. al.* 1991, Rubin and Schenker 1987).

In Section 2 of this paper, we discuss existing research on earnings models and earnings measurement error. In Section 3, we investigate the form of earnings measurement error in our PSIDVS sample and elaborate on results from prior research. Section 4 describes our multiple imputation approach. Section 5 describes our cross-section and longitudinal earnings equations and presents uncorrected regression results. Section 6 applies the multiple imputation technique to these equations. We find important influences of measurement error on the models' statistical properties and on parameter estimates, particularly during recessionary periods. Estimates of the effect of general labor market experience and union and blue-collar status on earnings are particularly sensitive to error. In the conclusion, we summarize the results and discuss applying our approach more broadly.

2. Earnings Equations and Measurement Error

The earnings equation, which models earnings as a log-linear function of personal and job-related characteristics, is one of labor economists' most commonly used tools. Human capital theory provided an early theoretical basis for earnings equations (Becker 1975, Mincer 1974). Subsequent research has focused on estimating the returns to measurable human capital characteristics, such as years of formal education or training, or human capital proxies, such as years of general labor market experience or tenure with a particular firm. A wide variety of other issues, such as the union wage effect and labor market discrimination, are also analyzed using wage equations.

In typical earnings equations, the residual is interpreted as arising from unobservable variables, such as the quality of the worker or job match. Measurement error in earnings, however, may also play a role. Researchers who acknowledge the presence of measurement error typically assume that it is "classical" in form. This requires that measurement error be normally distributed with mean zero and constant variance, uncorrelated with true earnings and the values of explanatory values, and uncorrelated over time for a given individual. Under these assumptions, the estimation problems caused by measurement error in linear models are limited: classical measurement error in an independent variable biases all estimated coefficients toward zero; classical measurement error in the dependent variable increases the model's error variance and standard errors but does not bias parameter estimates or associated statistical tests.

Several recent studies using validated data, however, suggest that the assumption of classical measurement error in earnings data is inappropriate. The two main sources of validated labor market data currently in use are the Panel Study of Income Dynamics Validation Study (the PSIDVS, as described in Bound *et. al.* 1990, 1994) and the 1978 Current Population Survey-Social Security Earnings Records Exact Match File (CPS-SER; see Bound and Krueger 1991). The PSIDVS sample consists of approximately 400 employees surveyed from a large Detroit, Michigan area manufacturing firm. In the PSIDVS, responses to the standard PSID survey instrument are matched with company personnel records on a wide variety of employment information, including earnings, fringe benefits, hours, unemployment spells, and employment tenure. In contrast, the CPS-SER provides a much larger and more representative sample but only validates earnings; it matches responses to the March 1978 and 1979 CPS yearly earnings questions with corresponding Social Security administration payroll tax data as reported by employers.

By assuming that employer records are error-free measures of the variables of interest, researchers can use these data sets to test the properties of errors in survey responses.[1] Results from both data sets suggest that measurement error in survey earnings responses is large -- the share of measurement error variance in total earnings variance ranges from approximately .15 to .82 across all earnings measure -- and is not classical in the sense defined above. In particular, both Bound *et. al.* (1990, 1994) and Bound and Krueger (1991) find that measurement error in earnings is negatively correlated with true earnings and positively autocorrelated over time. These properties have important implications for the use of longitudinal earnings equations. As indicated by Bound and Krueger, they increase the reliability of first-differenced earnings data relative to the classical measurement error case.

Perhaps most important for our purposes, both Bound *et. al.* (1989) and Duncan and Hill (1985) find systematic partial correlations between measurement error in earnings and other observable variables in the PSIDVS. In a cross-section regression of earnings on education, pre-employer experience, and tenure with the current employer, Duncan and Hill find that ignoring these partial correlations cause a downward bias of approximately 30% in the estimated effect of tenure on earnings. Bound *et. al.* report similar results for tenure, and also find that measurement error causes approximately a 33% overstatement of the return to education. The earnings equation specification in both papers is highly restricted, however, and therefore may not reveal the full impact of measurement error in more standard earnings equations that include a wider set of personal and job-related characteristics.

Overall, existing analyses of the form, magnitude, and influence of measurement error in survey earnings data suggest the need for techniques to account for such error. Given the absence of validated data in standard panel data sets (such as the PSID and the NLS), it appears that little can be done about measurement error biases in earnings equations. Rubin (1987), however, proposes a multiple imputation technique for handling missing data in panels that is superior to either ignoring cases with missing data or replacing them with a single set of imputed values. When the measurement error process can be modeled from existing validated data, this technique may be extended to account for measurement error in standard non-validated data (Meng 1994).

In the next section, we explicitly model the earnings measurement error process using the PSIDVS, in order to replicate previous results and also to draw additional implications for our error correction models. The measurement error process identified in previous research and investigated below has implications for both cross-section and longitudinal earnings models; we therefore estimate variants of both. Also, in contrast to Duncan and Hill (1985) and Bound *et. al.* (1989), we estimate an expanded error model that allows for correlations between measurement error and a larger set of covariates.

3. Earnings Measurement Error in the PSID Validation Study

Classical measurement error is fully random. In survey data, classical measurement error might arise from non-systematic misreporting or rounding by respondents. Although such behavior is likely given time constraints and limited incentives for careful responses, more systematic errors are also possible. Among the spate of recent papers on earnings measurement error, only Pischke (1995) attempts to model and test the measurement error process embodied in the PSIDVS data. He proposes a model in which measurement error stems from underreporting of transitory income fluctuations, a person fixed-effect, and a white noise component. He tests the resulting moment conditions and finds that the results support his model. Pischke's approach and results are consistent with prior findings that earnings measurement error in the PSIDVS arises largely due to yearly variations in hours worked, which in turn arise largely from unemployment spells among PSIDVS subjects (Bound *et. al.* 1989). Previous research has identified the importance of event "salience" (memorability) in determining response error, particularly in regard to unemployment spells.[2] Reporting error in hours lost due to unemployment or other events, hence reporting error in earnings, is likely to be affected by the duration and impact of the events.[3]

Alternatively, systematic underreporting of earnings may arise for reasons related to income tax evasion. Clotfelter (1983) finds that tax evasion increases with income and marginal tax rates. Feinstein (1991), however, finds somewhat mixed effects in his more complex models. Existing tax evasion behavior might extend to survey respondents who are dubious about assurances regarding confidentiality. The negative correlation between measurement error and true earnings reported in previous analyses (Bound *et. al.* 1994, Bound and Krueger 1991) is consistent with increases in underreporting as income or marginal tax rates increase.

A key issue for the application of our error correction technique is the extent to which measurement error is systematically related to demographic and economic variables that appear in earnings equations. Both explanations for measurement error given above are consistent with such effects. Response error for reasons related to tax evasion may vary across demographic groups, due to differential tax treatment or differing levels of tax sophistication; Clotfelter (1983) and Feinstein (1991) find such differential tax evasion. Hours fluctuations due to unemployment spells or overtime may also vary systematically with worker characteristics, such as job tenure, union status, and hourly or blue-collar status. The link between unemployment spells and reporting error may in turn link such variables to reporting error. For example, Bound *et. al.* (1989) find that company tenure reduces unemployment incidence in the PSIDVS and therefore is negatively related to earnings response error. Systematic relationships between unemployment and reporting error in earnings will produce different patterns of earnings measurement error across the business cycle, hence over time. In particular, due to unemployment-induced hours fluctuations, it is likely that the extent and impact of earnings measurement error will be more severe in survey data collected during recessions (Pischke 1995). We therefore expect greater measurement error bias in our 1983 data, for which earnings correspond to the recession year 1982, than in our 1987 data, for which earnings correspond to the expansion year 1986.

We now examine these issues empirically, using the PSIDVS, which consists of approximately 400 employees surveyed from a large Detroit, Michigan area manufacturing firm. An initial set of 534 interviews was attempted in 1983, of which 418 were completed. Reinterviews were successfully conducted with 341 individuals in 1987, of whom 275 were respondents in both 1983 and 1987. An additional sample of 151 hourly workers was interviewed in 1987. The resulting data set matches standard PSID survey responses with company personnel records on a variety of employment variables, including earnings, fringe benefits, hours, unemployment spells, and employment tenure. The company records are highly accurate and are interpreted as error-free variables in our analysis. Although company record data was collected for the entire period 1982-87, the survey was only administered in 1983 and 1987. Therefore, for practical reasons, our error correction and earnings models are restricted to the 1983 and 1987 cross-sections, and to a longitudinal model with 1983 and 1987 data only.[4] Restriction to observations with non-missing values of key variables produced the sample sizes in Tables 1 and 2. Summary statistics for the 1983 cross-section are provided in the appendix table.

Use of these data entails several drawbacks. The primary weakness is that the PSIDVS was administered to a single company, not to a representative labor market sample. For example, the company work force was older and had higher tenure than the national work force. Our PSID sample restrictions (see Section 5), however, produce a sample that is more comparable to the PSIDVS sample than a representative work force sample would be. Also, the distribution of earnings in the PSIDVS has a higher mean and lower variance than do national earnings distributions, even when the two samples are restricted to be roughly comparable (Bound *et. al.* 1990).[5] In this paper, we assume that the PSIDVS allows for construction and estimation of an error correction model that is appropriate for our PSID sample. Support for this assumption is provided by the similar pattern of earnings errors across the PSIDVS and CPS-SER samples (as reported in Bound *et. al.* 1990), the latter of which is more representative of the national work force.

A key finding in previous research is that the earnings measurement error process is non-classical. Consider the following equation:

$$(1) \qquad y - y^{*} = f(y^{*}) + pZ + m$$

where y represents ln(interview earnings), y^* is ln(validated earnings), $f(y^*)$ is an unknown function, Z is a matrix of covariates, p is a corresponding vector of coefficients, and m is an error term. The function $f(y^*)$ allows for non-linearities in the relationship between measurement error and true earnings; Lee and Sepanski (1995) allow for similar non-linearities (in y) in their validation-based error correction models.

In this framework, classical measurement error implies the restrictions $f(y^*)=0$, $p=0$, and m has mean zero with constant variance and no autocorrelation.

Although our model above is specified with true earnings on the right-hand side, our multiple imputation technique requires that interview earnings be used as an explanatory variable. We therefore report results from both types of model. The dependent variable in the cross-sections is [ln(interview earnings)-ln(true earnings]; for the longitudinal model covering changes between 1983 and 1987, it is [Δ(ln(interview earnings)-Δ(ln(true earnings))]. In all cases, the earnings period is the year immediately prior to survey administration.

Table 1 presents the results of regressing measurement error in the natural log of earnings on validated earnings and interview earnings, and a set of covariates typically found in earnings models, for the 1983 and 1987 cross-sections. In these models, measurement error is significantly and negatively related to true earnings.[6] Thus, the "mean reversion" property found by previous investigators (Bound and Krueger 1991, Bound *et. al.* 1990, 1994, Pischke 1995), in analyses that excluded our additional covariates, is robust to their inclusion.[7] Conversely, measurement error is positively related to interview earnings or its change.

Several of the additional covariates have significant effects on measurement error in Table 1. In the models that include record earnings, women tend to underreport earnings in both 1983 and 1987. F-tests applied separately to the potential experience and tenure terms indicate small but significant effects of these variables on measurement error in 1983. Marital and hourly status are significantly related to measurement error in 1987. In the models that include interview earnings, blue-collar status causes a significant understatement of earnings in both years. Potential experience has a significant negative effect, and schooling a marginally significant negative effect, in 1983.

Table 2 presents the results for the longitudinal error models. Quadratic functions of true and interview earnings provide the best fit in these models. We also control for both the change and the base level of several independent variables, in order to account for time-varying effects (as described in Section 5). Hourly earner status is not in change form because no changes in this variable occurred in our PSIDVS data.

In column (1), which includes true earnings on the right-hand side, the explanatory power of the model is very small. Only record earnings (and its square) and the change in potential experience significantly affect longitudinal measurement error. For purposes of constructing an imputation equation, however, the column (2) regression of measurement error on reported earnings is more relevant. In addition to interview earnings and its square, measurement error is significantly affected by potential experience and its change, and by changes in blue-collar status. Because the negative coefficient on the change in potential experience is identified by changes in schooling, this coefficient suggests that workers who acquire additional schooling overstate their earnings changes. This may occur because surveyed workers do not fully account for the effects of additional schooling on hours or productivity. Marital status and the level of schooling in 1983 may also affect measurement error in this model.[8]

The models presented in this section generally have low explanatory power, as indicated by the adjusted r-squares. Our imputation models, however, incorporate the information embodied in interview earnings, so that models of true earnings (rather than measurement error) as a function of interview earnings are more indicative of the explanatory power of our imputation equations. These models have substantially more explanatory power than the corresponding models reported in Tables 1 and 2, with adjusted r-squares in the range of .80 for all models.

Overall, the results in Tables 1 and 2 indicate the presence of systematic measurement error in earnings that is correlated with true earnings, reported earnings, and several explanatory variables from our earnings equations. This pattern of measurement error is likely to bias estimation of earnings equations. In particular, as described by Bound *et. al.* (1994), the negative correlation between measurement error and true earnings will tend to bias downward the coefficients from earnings equations. This simple bias, however, is complicated by the estimated relationship in Tables 1 and 2 between earnings measurement error and several common covariates. The technique described in the next section uses our measurement error models to generate multiple imputations of true earnings, which are then used to adjust the earnings equations for the presence of measurement error.

4. Measurement Error Models

There is a large literature on measurement error models (see Fuller 1987), but most of that literature is developed in a setting where true values of variables are unobserved. The methods described in that literature typically require strong identifying assumptions and complex estimation procedures. This section shows that if an independent validation study with observed true values is available (such as the PSIDVS), then consistent estimates of coefficients and corresponding sampling distributions can be obtained using simple modifications of standard linear regression techniques.

The basic modeling assumption is that the true value, y^*, and the reported value, y, follow a bivariate normal distribution conditional on exogenous variables X and Z. Because y^* is unobserved in the main sample but is observed in the validation sample, while y is available in both the main and validation samples, it is convenient to write the model as:

(2a) $y^* = Xb + e,\ e \sim N(0,\ s_e^2 I)$

(2b) $y^* = yg_0 + Zg_1 + h,\ h \sim N(0,\ s_h^2 I)$,

where X has N rows and *K* columns and Z also has N rows. The first equation (2a) represents the conditional distribution of y^* given X, which is the relationship we want to estimate. The second equation represents the conditional distribution of y^* given y and Z.[9] Note that we have temporarily simplified the measurement error model in equation (1) by assuming that f(•) is linear.

The model in equations (2) is identified if X and Z are not identical. If they are identical, then the model can be identified by assuming that the structural errors (e and h) are independent. Rubin (1987) shows that models allowing correlation between the structural errors without exclusion restrictions on X and/or Z can only be identified by strong functional form assumptions. This means that there is always an observationally equivalent model with no correlation between the structural errors. Bound *et. al.* (1994) also assume uncorrelated errors across the equation of interest and the measurement error equation. As they note, validation studies provide no information about e beyond that in standard data sets, so that analysts using validated data have little to say about such correlation.

The model in equations (2) could be estimated by maximum likelihood or method of moments, but these would require custom programming because the reduced form covariance terms depend nonlinearly on g_0 (see Lee and Sepanski, 1995, and Imbens and Hellerstein, 1993). Multiple imputation, however, provides a general and computationally simple method for inference in missing data problems (see Rubin, 1987; Schenker and Welsh, 1988; and Brownstone, 1991). The measurement error problem is related to missing data models by considering the true values, y^*, as missing for all observations in the main data sample. The only difference is that in our model interview earnings (y) provides information about true earnings (y^*) beyond that available in the usual missing data case. Rubin (1987, chapter 4) provides general proofs of the validity of multiple imputation procedures for many imputation methods. Brownstone (1991) provides direct proofs of the consistency of the multiple imputation estimators used in this study. Brownstone also provides results from some Monte Carlo studies which show that the asymptotic approximations are valid for sample sizes smaller than those encountered in this study.

The remainder of this section describes the multiple imputation methods used in this study and explains heuristically why they are consistent. As we will show in equation (5), the consistency of the multiple imputation approach requires that conditioning on Z implies conditioning on X in (2a). We will therefore assume this to be the case for the remainder of this paper.

The first, and key, step is to use the validation sample to calculate imputed values which match the first two moments (conditional on X) of the unobserved y^*. We will use the "normal imputation" procedure given in Rubin (1987, chapter 5, example 5.1) and Schenker and Welsh (1988), which creates one set of imputed values according to:

(3) $\quad \hat{y}^* = E(y^*|y,Z,g^*) + h^*$ for the main sample

$\qquad = y^*$ for the validation sample,

where g^* are drawn from the sampling distribution of a consistent estimator of g in equation (2b), and h^* are independent draws from the distribution of h. For the model in equations (2), g and s_h^2 can be estimated by $\hat{g}$ and s_h^2 , the least squares estimates from regressing y^* on y and Z using the validation study. Therefore, $\hat{g}$ follows a $N(g,\ s_h^2\Delta^{-1})$ distribution (where $\Delta = [y\ Z]'[y\ Z]$) and $ds_h^2/\ s_h^2$ follows an independent Chi-squared distribution with degrees of freedom d equal to the number of observations in the validation data minus the number of columns in [y Z]. The following steps create one set of valid imputations for the model in equations (2) according to (3):

a) set $s_h^{2*} = s_h^2 c^*/d$, where c^* is drawn from a c_d^2 distribution.

b) draw g^* from a $N(\hat{g}\ , s_h^{2*}\Delta^{-1})$ distribution.

(4) c) draw h^* independently from a $N(0, s_h^{2*})$ distribution.

d) set $\hat{y}^* = yg_0^* + Zg_1^* + h^*$.

Note that $E(\hat{y}^*) = E(E(y^*|y,Z,g^*)) = E(y^*|X)$ (since we have assumed that conditioning on Z implies conditioning on X), so that the imputed values match the first conditional moment of the unobserved y^*. The second moments also match since:

$$
\begin{aligned}
E(\hat{y}_i^{*2}|Z_i,g^*) &= E(E(y_i^*|y_i,Z_i,g^*)^2) + E(E(h_i^{*2})) \\
&= V(E(y_i^*|y_i,Z_i,g^*)) + (E(y_i^*|Z_i,g^*))^2 + E(V(y_i^*|y_i,Z_i,g^*)) \\
&= V(y_i^*|Z_i,g^*) + (E(y_i^*|Z_i,g^*))^2 \qquad (5) \\
&= E(y_i^{*2}|X_i,g^*) .
\end{aligned}
$$

Therefore the least squares estimators from regressing $\hat{y}^*$ on X, $\hat{b}^*$ $(= (X'X)^{-1}X'\hat{y}^*)$ and s_e^{2*} $(= \hat{y}^{*\prime}(I - X(X'X)^{-1}X')\hat{y}^*/(N-K))$, are unbiased estimates of the least squares estimates of y^* on X, $\hat{b}$ and s_e^2, and are therefore unbiased for b and s_e^2.

However, the least squares covariance estimator, $\hat{\Omega}^* = s_e^{2*}(X'X)^{-1}$, is inconsistent and downward biased for the covariance of $\hat{b}^*$ since it does not account for the uncertainty in g^*. This covariance can be decomposed according to:

$$
\text{(6)} \qquad Cov(\hat{b}^*) = E(Cov(\hat{b}^*|g^*)) + Cov(E(\hat{b}^*|g^*)).
$$

$\hat{\Omega}^*$ is clearly an unbiased estimator of the first term on the right-hand side of equation (6), but we need some way of estimating the other term. The multiple imputation technique solves this problem by drawing multiple sets of independent imputed values according to (4). For each of these M draws, compute $\hat{b}_j^*$ and $\hat{\Omega}_j^*$ (where the subscript j denotes the jth set of imputations). The final estimate of b is the average of the M completed data point estimates:

(7) $$\tilde{b}_M = \sum_{j=1}^{M} \hat{b}_j^* \Big/ M.$$

If $\tilde{\Omega}_M$ is the corresponding average of the covariance estimates, define

(8) $$B_M = \sum_{j=1}^{M} \left(\tilde{b}_M - \hat{b}_j^*\right)\left(\tilde{b}_M - \hat{b}_j^*\right)' \Big/ (M-1).$$

Then

(9) $$T_M = \tilde{\Omega}_M + (1 + M^{-1})B_M$$

is the multiple imputation estimate of the covariance of $\tilde{b}_M$. The term T_M can be interpreted as the sum of the average covariance within a set of imputed values and the covariance across independent sets of imputed values.

Brownstone (1991) extends Rubin's (1987) and Schenker and Welsh's (1988) results to show that the multiple imputation estimator, $\tilde{b}_M$, is consistent and asymptotically normally distributed for fixed $M \geq 2$, and T_M is a consistent estimator of its asymptotic covariance. Although these results support consistent inference for small M, increasing the number of multiple imputations clearly reduces the covariance of B_M and, as Brownstone's (1991) results show, the asymptotic covariance of $\tilde{b}_M$. Rubin (1987) shows that as *M* gets large, then the Wald test statistic for the null hypothesis that $b = b^0$,

$$(10) \qquad \left(b^0 - \tilde{b}_M\right)' T_M^{-1}\left(b^0 - \tilde{b}_M\right) \; /K ,$$

is asymptotically distributed according to an F distribution with *K* and *n* degrees of freedom, where

$$(11) \qquad n = (M\text{-}1)(1+ r_M^{-1})^2 \text{ and}$$

$$r_M = (1+M^{-1}) \, \mathrm{Tr}(B_M \, \Omega_M^{-1})/K.$$

This suggests increasing *M* until *n* is large (greater than 100), which is the stopping rule used in the empirical work described in the next section. This stopping rule requires between 50 and 80 multiple imputations for these applications. Note that this is much larger than the number of multiple imputations reported in previous applications, and it is probably due to the large amount of missing data in our particular application. If it is not computationally feasible to use this stopping rule, Li *et. al.* (1991) and Meng and Rubin (1992) can be consulted for more accurate alternative approximations and test statistics for small values of *M*.

Obtaining consistent and accurate results from these multiple imputation methods requires that the validation study contain sufficient information to specify and estimate an imputation model with small prediction errors. This suggests that validation studies should be designed to insure sufficient variation in the important explanatory variables entering the measurement error process. Rubin (1987, pp. 81-87) shows that the validation study also needs to be designed so that the data can be pooled with the main study data and used to estimate equation (2a). If the validation study data are not used to estimate (2a), then the covariance estimator, T_M, is no longer consistent since the completed data estimators, $\hat{b}^*_j$, are no longer conditioned on all of the observed data. In this case the completed data estimators, $\hat{\Omega}^*_j$, will have a probability limit larger than the true $\mathrm{Cov}(\hat{b}^*_j|g^*_j)$, which in turn means that the standard errors of the parameter estimates will be overstated.

If the validation and main samples cannot be pooled to estimate equation (2a), then consistent standard error estimates can be obtained by including interactions of all of the variables (X) with a dummy for inclusion in the validation sample. Note that this dummy variable and its interactions do not cause any identification problems because conditioning on Z includes conditioning on membership in the validation sample. One can then test for the possibility of "partial pooling" where some subset of the parameters are constant across both samples. This approach leads to an efficient parametrization of the differences between the main and validation samples. If important differences are found, then this might be cause to suspect the key assumption that the imputation equation (2b) is valid for the main sample.[10] Unfortunately there is no way to test this assumption since the correct values are only observed in the validation sample. This discussion suggests that an ideal validation sample will be a random subsample of the main sample. Under these circumstances complete pooling would be trivially satisfied and there would be no need to worry about the imputation model's validity for the main sample.

Multiple imputation estimators are more efficient than single imputation estimators, but they are not guaranteed to be fully efficient. However, they are easier to implement with standard statistical software packages than maximum likelihood or direct computation of the covariance of the completed data estimator. For example, the estimates provided in this paper were obtained using STATA Version 3.1 (1993). Beyond standard OLS regression techniques, implementation of our multiple imputation procedure required only random draws from the appropriate distributions and simple manipulation of the covariance matrices from each imputation.[11] No optimization routine or other custom programming is required.

Our approach also provides a general consistent alternative to the unfortunately common practice of ignoring uncertainty in the imputation process and basing inference on least squares estimates from a single imputation. This occurs most frequently because standard data sets such as the Current Population Survey do not provide the information necessary for applied researchers to account for the imputation procedures. Although estimation based on a single imputation produces consistent coefficient estimates in our setting, the corresponding standard errors are severely underestimated. For example, the least squares standard error estimates for our longitudinal model earnings model (the square root of the diagonal elements in $\tilde{\Omega}_M$ from equation 9) are often less than 30 percent of their true values (the corresponding elements of T_M).

An alternative approach is discussed by Lee and Sepanski (1995), who propose instrumental variables estimators to combine validated and unvalidated data. Their framework is highly general, and allows for measurement error in independent and dependent variables. Our specific model is a special case within their framework,[12] but the multiple imputation approach is more general because it can easily be applied to discrete and limited variables.

5. PSID Data and Uncorrected Results

We apply the techniques discussed in Section 4 to a sample of male and female household heads from the PSID for the years 1981-87, combined with a corresponding sample from the PSIDVS. As noted in Section 4, our multiple imputation approach requires pooling of the two samples. Our PSID sample was designed to approximate the sample available in the PSIDVS. We restricted our PSID sample to individuals aged 16-64 in 1981 who were heads of respondent households continuously during 1981-1987 and who were never self-employed on their main job during that period. These restrictions yielded a sample of 3409 male and female heads. Further restriction to individuals with non-zero and non-allocated earnings, and non-missing independent variable values, produced a 1983 PSID cross-section sample of 2504, a 1987 cross-section sample of 2293, and a longitudinal sample (for changes between 1983 and 1987) of 1905. Weighted sample means for the 1983 cross-section are provided in the appendix table.[13]

In our cross-section earnings equation, we model the natural log of an individual's total wage and salary earnings in the previous year as a linear function of years of formal education (top-coded at 17 for receipt of a graduate degree), potential labor market experience (defined as (age - education - 6)), employment tenure (years with current firm), potential experience squared (divided by 100), tenure squared (divided by 100), and a vector of personal and job characteristic dummy variables that includes race (black), sex, marital status, coverage by a union contract, blue-collar occupation, whether paid hourly, residence in the South, and an interaction between residence in the South and race.

The latter three variables are included primarily to facilitate pooling of the PSID and validation study in our cross-section earnings equations. As described in Section 3, the PSIDVS sample is taken from a single firm and therefore is not a random population sample. Due to extensive union coverage among hourly employees in the PSIDVS, inclusion of the hourly dummy was necessary to produce a similar union effect across the two samples. Similarly, the South-black interaction insures that the corrected black/white wage gap is similar in the two samples (the PSIDVS sample firm is in the North). Similar adjustments could not be made, however, to account for the lower return to schooling, higher earnings in blue-collar jobs, and smaller male/female earnings gap in the PSIDVS sample.[14] Also, average earnings in the PSIDVS sample are higher than in our PSID sample. The regression models therefore include a dummy variable indicating that the observation is from the PSIDVS sample, and interactions between this dummy and the schooling, blue-collar, and female variables. As discussed in Section 4, including such interaction terms enables application of our technique in settings for which the main and validation samples do not fully pool.

The estimated coefficients on labor market experience and tenure are of substantive interest in our study. Potential experience was chosen over a more detailed actual experience measure because it is the standard variable used in earnings equations to proxy for general human capital accumulation. We use years with current firm as our tenure measure, to account for the accumulation of firm-specific skills or other shared investments.[15]

Columns (1) and (3) in Table 3 present uncorrected results from our 1983 and 1987 cross-section equations. The error-corrected results in the adjoining column for each regression are discussed in Section 6. The cross-section coefficients are precisely estimated and broadly agree with results from other studies. The estimated return to tenure is large relative to the return to potential experience. The estimated return to schooling is also large. Blacks (particularly in the South) and women earn substantially less than do white males. Unmarried, blue-collar, and hourly workers also earn less than their counterparts. The estimated union wage gap is similar to that produced by other studies that apply OLS to a cross-section of individuals. Finally, the interaction terms indicate that blue-collar and female workers earn substantially more in the PSIDVS than in our PSID sample, but the return to schooling may be lower in the PSIDVS. Restricting the analysis to a male sample produced very similar results.

The dependent variable for the longitudinal model in Table 4 is the difference between the log of reported yearly earnings in the 1987 and 1983 surveys. Difference models of this form are commonly used to account for omitted variables that are fixed over time; "fixed effects" that are correlated with the independent variables will bias OLS cross-section regression results. To derive the longitudinal specification, consider the following cross-section earnings equations for a fixed sample of individuals, where the subscripts 0 and 1 refer to the beginning and end periods of the analysis.

(12) $\ln(\mathrm{W}_0) = a_0 + b_0\mathrm{X} + g_0\mathrm{Z}_0 + Y + m_0$

(13) $\ln(\mathrm{W}_1) = a_1 + b_1\mathrm{X} + g_1\mathrm{Z}_1 + Y + m_1$

In these equations, W is our earnings variable, X is a vector of variables whose values are fixed over time (such as race and sex), Z is a vector of variables whose values can change over time in the data (for example, potential experience, tenure, and union status), and the ms are random disturbances that are uncorrelated with the independent variables. The Ys are unobserved determinants of earnings which do not vary over time. If Y is correlated with the observed variables X and Z, then OLS applied to the cross-sections yields inconsistent estimates of the parameters b and g. However, subtracting (12) from (13) yields:

$$\ln(W_1) - \ln(W_0) = (a_1 - a_0) + (b_1 - b_0)X + (g_1 Z_1 - g_0 Z_0) + (m_1 - m_0) \tag{14}$$

This eliminates the influence of the fixed effect Y (which could, for example, be an unobserved measure of worker quality that is correlated with employment tenure). Note that our model allows the effects of the independent variables to vary over time, which obviates the need for more complex modeling of the omitted effect in our setting.[16] For ease of interpretation, our empirical specification below includes both the change in the Z's and their base year (1983) levels (their respective coefficients equal g_1 and $g_1 - g_0$).

Table 4 presents results for this model. Separate effects of changes in tenure and changes in potential experience are identified by job changers and by increases in schooling attainment. We are unable to estimate a coefficient on schooling increases, however, because such increases are perfectly collinear with increases in potential experience; we are only able to estimate schooling level effects on earnings changes.[17] We also estimate the effects of changes in marital status (marriage and divorce), union coverage, blue-collar and hourly employment, and employer, along with race, sex, region and validation sample effects.

Compared to the cross-section results in Table 3, longitudinal estimation increases the size (but reduces the precision) of the return to potential experience and substantially reduces the return to tenure. The reduced tenure effect is consistent with fixed-effects results from Brown and Light (1992). Although we did not specify a possible correlation between components of the error terms and accumulated tenure, it is likely that our fixed-effects approach reduces the estimated return to tenure by accounting for unobservables that are positively correlated with the decision to remain with a firm. The coefficient on unionism in Table 4 is reduced relative to the cross-section coefficient in Table 3, consistent with other analyses that use longitudinal data (see Jakubson 1991). Similar to the cross-section results, we find that changes to hourly and blue-collar status substantially reduce earnings. In contrast to the cross-section results, however, the validation sample dummy and interactions are insignificant here. Finally, note that the coefficients on several base-level variables are significant in our longitudinal model, which suggests that simple first-differencing of explanatory variables in fixed-effect wage equations causes misspecification.

Overall, the cross-section and longitudinal results are reasonably consistent with results from previous analyses using similar data. We now turn to earnings equation results that apply our multiple imputation error correction technique.

6. Results Corrected for Measurement Error

We now apply our multiple imputation technique to the correction of our earnings equations. In Tables 3-4, each column of results from the uncorrected earnings models is followed immediately by the corresponding column of error-corrected results. In the corrected regressions, log interview earnings (or its change) for each observation in the main PSID sample is replaced by multiple imputed values of log true earnings, which are obtained through repeated application of algorithm (4) in Section 4. Conditional on the use of a correctly specified imputation model and variables that account for the different earnings structures in the PSID and PSIDVS, the error-corrected coefficients and standard errors are consistent. The models in this section are identified by assuming no correlation between the structural errors in equation (2). The longitudinal models are further identified by including quadratic earnings terms in the imputation equation.

The results reveal substantial changes relative to the corresponding uncorrected results in Tables 3-4. Hausman tests (Hausman 1978) of the error corrected results against the uncorrected OLS equations reject the null hypothesis of no error correction at the 1% level in all specifications except the 1987 cross-section. Several of the changes in specific coefficients follow directly from the corresponding coefficients from the measurement error models in Tables 1 and 2, but other changes appear to arise from a more complex covariance pattern across the earnings and measurement error equations.

In the 1983 cross-section in Table 3, error correction increases the coefficient on potential experience and the absolute value of the coefficient on its square by about 33%. The negative effect of blue-collar status on earnings is almost halved due to error correction. The return to union status in 1983 is substantially reduced in size and becomes insignificant, the coefficients on tenure and tenure squared are reduced by 20% and 33%, and the negative effects of being female or black are substantially reduced. The reduction in the tenure effect due to error correction is surprising because Duncan and Hill (1985) find that measurement error biases the tenure coefficient in a 1983 PSIDVS cross-section earnings equation downward. The difference between our results and Duncan and Hill's is due to our less restrictive earnings and measurement error model specifications. Duncan and Hill only include schooling, experience, and tenure in their equations. They find a negative partial correlation between tenure and measurement error while we find a positive one using our more general specification. Finally, although the magnitude and statistical significance for most coefficients is reduced by error correction, this shrinkage is not uniform: the coefficients on marital status and hourly earner status remain relatively constant across the corrected and uncorrected specifications.

In contrast, the Hausman test does not reject the absence of measurement error bias in the 1987 cross-section earnings model in Table 3. This follows directly from the relative absence of large and significant coefficients in the 1987 cross-section measurement error equation (Table 1). Only the coefficient on blue-collar status appears to be substantially affected by error correction; it is approximately halved (from -.20 to -.11). Otherwise, the coefficients are mostly shrunk and their standard errors inflated somewhat, as in the 1983 cross-section. As noted in Section 3, we expected more severe measurement error bias in the 1983 sample than in the 1987 sample, due to greater transitory hours fluctuations in the recession year 1982 than in the expansion year 1986.

In the full longitudinal specification in Table 4, measurement error bias is again severe; the Hausman test statistic is significant at better than the 1% level. Error correction increases the size and significance of the coefficient on potential experience but decreases the negative effects of blue-collar and hourly status, and their changes. The joint effect of union status and its change is reduced in size and significance. The return to tenure increases slightly in size but its significance is substantially decreased.

We also estimated our error-corrected longitudinal earnings models on separate subsamples of hourly workers and workers who did not change jobs. We find greater effects of error correction when hourly wages are used, presumably due to more severe measurement error in hourly than in yearly earnings (Bound *et. al.* 1990). For the sample of individuals who do not change jobs the effects of error correction in this sample are reasonably comparable to those based on the full longitudinal sample in Table 4.

Although our technique corrects for measurement error in the dependent earnings variable, measurement error in the covariates may also be important. Among the variables in our earnings model, the PSIDVS provides validated values for tenure (1983 and 1987) and union status (1987 only). Measurement error in such variables is a general problem for obtaining consistent estimates of earnings equations. Errors in these variables, however, are very rare in the PSIDVS (Duncan and Mathiowetz 1985), which limits our ability to construct useful imputation models for these variables. We ran regressions of true and interview earnings on validated and unvalidated tenure and union status (plus the other covariates) to assess the relative roles of measurement error in earnings versus measurement error in tenure and union status. There is essentially no effect of reporting errors in tenure on the results. The effect of reporting errors in union status also appears limited, but it is difficult to assess due to a close relationship between union and blue-collar status in the PSIDVS.

We performed several tests of robustness. We estimated the cross-section models with an earnings quadratic included in the measurement error equation, and the longitudinal model with the earnings quadratic excluded. We also dropped highly insignificant variables (t-stat < 1) from the measurement error equations. These changes did not appreciably affect our results. We also estimated our models with interactions between interview earnings and two important covariates -- potential experience and blue-collar status -- in the measurement error equations. Although both interactions are significant in the longitudinal model, and the earnings/experience interaction is significant in the 1983 cross-section model, the error-corrected earnings results change very little (except for slight strengthening of the effects of measurement error). Our results appear relatively robust to small specification changes.

An alternative to our approach is to estimate the earnings equation using the validation sample only. The simplicity gains of this alternative approach must be weighed against two main benefits of our approach: (1) combining information from both samples increases estimation efficiency; (2) our approach accounts for possible differences across the two samples in the parameters of the earnings equation. In our data, the efficiency gains are limited in the cross-section equations but large in the longitudinal equation: the standard errors for the cross-section earnings equation parameters are slightly smaller using our approach, but for the longitudinal equation they are approximately 25%-50% lower using our approach. In the cross-section, however, our technique also accounts for significantly different earnings equation parameters across the PSIDVS and PSID (as reflected in the PSIDVS interaction coefficients). Under the assumption of similar error structure across the two samples, our technique provides a flexible method for combining the two samples to estimate the equation of interest.

7. Conclusion

This paper complements the growing body of literature on the implications of measurement error for estimation of earnings and related models. We have proposed and demonstrated a multiple imputation methodology used to correct for measurement error in earnings models estimated from standard data sets where true earnings are unobservable. Using the PSID Validation Study, we find evidence for systematic earnings measurement error that is correlated with individual variables commonly used in earnings models. Accounting for this measurement error changes cross-section and longitudinal earnings equation results in substantive ways. Most importantly, we find that accounting for measurement error in a PSID sample of approximately 2000 household heads for the survey years 1983 and 1987 increases the estimated return to general labor market experience, reduces the negative effect of blue-collar status, reduces the return to union status, and may affect the returns to tenure and other variables, in both a cross-section and longitudinal setting.

Because theories about the incidence of earnings measurement error are sparse, and because the effects of error correction vary over time and across different samples, caution should be exercised in the interpretation of our specific results. Our technique, however, should prove increasingly useful and reliable as additional validation samples come into use and more extensive modeling of the measurement error process is performed. One conclusion for now is that earnings measurement error is potentially related to numerous independent variables that commonly appear in earnings and related equations; future research should account for this. Also, we find greater influence of earnings measurement error in survey year 1983, which corresponds to earnings in the recession year 1982, than in the later expansionary period (survey year 1987). In conjunction with prior work on earnings measurement error, this suggests that labor market events that affect hours worked -- unemployment, work sharing, retraining and additional formal education -- are likely to also generate survey response errors in earnings. Systematic measurement error of this nature may cause difficulties for labor market accounting under changing labor market conditions.

One potentially important methodological extension is to permit heteroskedasticity in the imputation model (equation 2b). We tested one model in which the residual variances in the imputation depended on the same covariates as in the 1983 cross-section imputation model. Although the results are virtually identical to the homoskedastic results given in the tables, heteroskedasticity may be important when using larger validation studies.

Our general methodology can be usefully extended and applied to a diverse set of models from labor economics and other fields. In this paper, we account only for measurement error in earnings, which appears as the dependent variable in our earnings equations. A natural extension is to simultaneously apply error correction to earnings and independent variables in earnings equations. For example, errors in reported fringe benefit values may be large and potentially correlated with measurement error in wages and salaries. Accounting for such measurement error may substantially affect the estimation of wage/fringe benefit tradeoffs. Also, measurement error is particularly severe in both hourly earnings and reported hours. As noted by Heckman (1993), correction of such measurement error may substantially improve estimation of labor supply parameters, particularly in a panel data setting. Another fruitful extension would be to the analysis of unemployment events. The

PSIDVS has extensive information on true and interview reports of unemployment spells, and existing research suggests the importance of recall bias in reporting of unemployment incidence and duration. Application of our approach to duration and other nonlinear models requires further refinement of our techniques, which is left for future research. Given the strong linkages between earnings measurement error, hours variation, and unemployment spells, perhaps future attempts to account for measurement error should more explicitly model the relationships among these outcomes.

References

Akerlof, George A. and Janet L. Yellen, "Unemployment Through the Filter of Memory," *Quarterly Journal of Economics* 100 (August 1985), 747-773.

Angrist, Joshua D. and Whitney K. Newey, "Over-Identification Tests in Earnings Functions With Fixed Effects," *Journal of Business and Economic Statistics* 9 (July 1991), 317-324.

Becker, Gary S., *Human Capital* (New York: Columbia University Press, 1975).

Brown, James N. and Audrey Light, "Interpreting Panel Data on Job Tenure," *Journal of Labor Economics* 10 (July 1992), 219-257.

Bound, John and Alan B. Krueger, "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics* 9 (Jan. 1991), 1-24.

Bound, John, Charles Brown, Gregory J. Duncan, and Willard L. Rodgers, "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Results from Two Validation Studies," NBER Working Paper #2884 (March 1989).

Bound, John, Charles Brown, Gregory J. Duncan, and Willard L. Rodgers, "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Validation Study Evidence," in J. Hartog, G. Ridder, and J. Theeuwes (eds.), *Panel Data and Labor Market Studies*, (Amsterdam: North Holland, 1990).

Bound, John, Charles Brown, Gregory J. Duncan, and Willard L. Rodgers,"Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12 (July 1994), 345-368.

Brownstone, David, "Multiple Imputations for Linear Regression Models," mimeo, UC Irvine (September 1991).

Card, David and Thomas Lemieux, "Changing Wage Structure and Black-White Wage Differentials," *American Economic Review*, 84 (May 1994), 29-33.

Clogg, Clifford C., Donald B. Rubin, Nathaniel Schenker, Bradley Schultz, and Lynn Weidman, "Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression," *Journal of the American Statistical Association* 86 (March 1991), 68-78.

Clotfelter, Charles, "Tax Evasion and Tax Rates: An Analysis of Individual Returns," *Review of Economics and Statistics* 65 (Aug. 1983), 363-373.

DuMouchel, William H. and Gregory J. Duncan, "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples," *Journal of the American Statistical Association* 78 (June 1983), 535-543.

Duncan, Gregory J. and Daniel H. Hill, "An Investigation of the Extent and Consequences of Measurement Error in Labor-economic Survey Data," *Journal of Labor Economics* 3 (Oct. 1985), 508-532.

Duncan, Gregory J. and Nancy A. Mathiowetz, *A Validation Study of Economic Survey Data*, (Ann Arbor: Institute for Social Research, the University of Michigan, 1985).

Feinstein, Jonathan S., "An Econometric Analysis of Income Tax Evasion and Its Detection," *Rand Journal of Economics* 22 (Spring 1991), 14-35.

Fuller, Wayne A., *Measurement Error Models* (New York: John Wiley, 1987).

Greenberg, David and Harlan Halsey, "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment," *Journal of Labor Economics* 1 (Oct. 1983), 380-407.

Greenberg, David, Robert Moffitt, and John Friedmann, "Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment," *Review of Economics and Statistics* 63 (Nov. 1981), 581-589.

Hausman, Jerry A., "Specification Tests in Econometrics," *Econometrica* 46 (Nov. 1978), 1251-1271.

Heckman, James J., "What Has Been Learned About Labor Supply in the Past Twenty Years," *American Economic Review* 83 (May 1993), 116-121.

Imbens, Guido W., and Judith K. Hellerstein, "Raking and Regression," mimeo, Dept. of Economics, Harvard University (September 1993).

Jakubson, George, "Estimation and Testing of the Union Wage Effect Using Panel Data," *Review of Economic Studies* 58 (Oct. 1991), 971-992.

Lee, Lung-fei and Jungsywan H. Sepanski, "Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data," *Journal of the American Statistical Association* 90 (March 1995), 130-140.

Li, K-H., T.E. Raghunanthan, and Donald B. Rubin, "Large-Sample Significance Levels From Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution," *Journal of the American Statistical Association* 86 (Dec. 1991), 1065-1073.

Mathiowetz, Nancy A. and Gregory J. Duncan, "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment," *Journal of Business and Economic Statistics* 6 (April 1988), 221-229.

Meng, Xiao-li, "Multiple Imputation with Uncongenial Sources of Input" (with discussion), *Statistical Science* 9, (Spring 1994) 583-574.

Meng, Xiao-li and Donald B. Rubin , "Performing Likelihood-Ratio Tests with Multiply-Imputed Data Sets," *Biometrika* 79 (Jan. 1992), 103-111.

Mincer, Jacob, *Schooling, Earnings, and Experience* (New York: Columbia University Press, 1974).

Pischke, Jan-Steffan, "Measurement Error and Earnings Dynamics: Some Estimates From the PSID Validation Study," *Journal of Business and Economic Statistics* 13(3) (July 1995), 305-314.

Rodgers, Willard L., Charles Brown, and Gregory J. Duncan, "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages," *Journal of the American Statistical Association* 88 (Dec. 1993), 1208-1218.

Rubin, Donald B., *Multiple Imputation for Nonresponse in Surveys* (New York: John Wiley, 1987).

Rubin, Donald B., "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association* 91 (June 1996), 473-489.

Rubin, Donald B. and Nathaniel Schenker, "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse," *Journal of the American Statistical Association* 81 (June 1986), 366-374.

Rubin, Donald B. and Nathaniel Schenker, "Interval Estimation from Multiply-Imputed Data: A Case Study Using Census Agriculture Industry Codes," *Journal of Official Statistics* 3 (Winter 1987), 375-387.

Stata Corporation, *Stata Release 3.1 Reference Manual* (College Station, Texas: Stata Corporation, 1993).

Schenker, Nathaniel and A.H. Welsh, "Asymptotic Results for Multiple Imputation," *Annals of Statistics* 16 (Dec. 1988), 1550-1566.

Sudman, Seymore and Norman M. Bradburn, *Response Effects in Surveys* (Chicago: Aldine Publishing Company, 1974).

Table 1 - Measurement Error Models, PSIDVS, 1983 and 1987 Cross-Sections
dependent variable = ln(interview earnings)-ln(company record earnings)
(coefficients; standard errors in parentheses)

	(1) 1983	(2) 1983	(3) 1987	(4) 1987
ln(company record earnings)	-.0673* (.0284)	--	-.130* (.0286)	--
ln(interview earnings)	--	.179* (.0250)	--	.220* (.0256)
Schooling	-.00128 (.00372)	-.00570 (.00348)	.00102 (.00316)	-.00411 (.00299)
Potential Experience	-.00593 (.00370)	-.00867* (.00345)	.000421 (.00242)	.000968 (.00230)
(Potential Experience)^2/100	.0111 (.00759)	.0146* (.00710)	-.000687 (.00295)	-.00114 (.00279)
Company tenure	.00631 (.00449)	.000512 (.00422)	.00158 (.00420)	-.00179 (.00398)
(Company tenure)^2/100	-.00189 (.00106)	-.000846 (.00100)	-.000472 (.000804)	.000010 (.000762)
Black	-.0196 (.0219)	-.0219 (.0206)	.0165 (.0164)	.00325 (.0155)
Female	-.0691* (.0293)	-.0382 (.0276)	-.0848* (.0227)	-.0180 (.0218)
Union coverage	.0829 (.0842)	.0216 (.0791)	.0474 (.0367)	-.00617 (.0348)
Married	-.0198 (.0168)	-.0256 (.0157)	.0262* (.0130)	.00100 (.0123)
Blue-collar	-.00399 (.0168)	-.0386* (.0157)	-.00236 (.0164)	-.0469* (.0154)
Hourly	-.0765 (.0857)	.0485 (.0805)	-.103* (.0382)	.0144 (.0365)
Constant	.750* (.281)	-1.64* (.250)	1.36* (.297)	-2.20* (.268)
Adjusted R^2	.0713	.183	.0796	.174
Number of Observations	344	344	451	451

* - indicates significance at 5% level, two-tailed test

Table 2 - Measurement Error Models, PSIDVS, Longitudinal (change between 1983 and 1987)
dependent variable = Δln(interview earnings)-Δln(company record earnings)
(coefficients; standard errors in parentheses)

	(1)	(2)
Δ**ln(company record earnings)**	-.140	--
	(.0731)	
[Δ**ln(company rec. earnings)]^2**	.112	--
	(.0620)	
Δ**ln(interview earnings)**	--	.378*
		(.0456)
[Δ**ln(interview earnings)]^2**	--	-.133*
		(.0393)
Δ**Potential Experience**	-.0258*	-.0244*
	(.0129)	(.0109)
Potential Experience 1983	.000563	.00398*
	(.00219)	(.00185)
Δ**Tenure**	-.0108	-.0100
	(.0162)	(.0138)
Tenure 1983	.00232	.00202
	(.00247)	(.00212)
Black	.0121	-.0120
	(.0369)	(.0316)
Female	.0205	.0436
	(.0414)	(.0353)
Δ**Union coverage**	-.0222	-.0296
	(.0619)	(.0528)
Union coverage 1983	-.0352	.0301
	(.171)	(.146)
Married (between 1983 and 1987)	.0408	.0520
	(.0384)	(.0327)
Divorced (between 1983 and 1987)	-.00566	-.0624
	(.0586)	(.0498)
Schooling 1983	.00802	.00813
	(.00613)	(.00522)

DBlue-collar	-.0481 (.0343)	-.0656* (.0293)
Blue-collar 1983	.00582 (.0287)	-.0108 (.0244)
Hourly	-.0154 (.171)	-.0814 (.146)
Constant	.0228 (.130)	-.107 (.110)
Adjusted R^2	.060	.313
Number of Observations	229	229

* - indicates significance at 5% level, two-tailed test

[1] All PSIDVS hourly workers were union members in 1983.

Table 3 - Earnings Models, pooled PSID and PSIDVS sample, 1983 and 1987 Cross-Sections
dependent variable = ln(interview earnings)
(coefficients; standard errors in parentheses)

	(1) 1983	(2) 1983 Corrected[1]	(3) 1987	(4) 1987 Corrected[1]
Schooling	.0646*	.0584*	.0770*	.0630*
	(.0067)	(.0108)	(.0072)	(.0112)
Potential Experience	.0172*	.0230*	.0060	.0039
	(.0047)	(.0061)	(.0040)	(.0039)
(Potential Experience)^2/100	-.0320*	-.0413*	-.0160*	-.0116*
	(.0092)	(.0121)	(.0060)	(.0057)
Company tenure	.0379*	.0300*	.0389*	.0314*
	(.0043)	(.0071)	(.0039)	(.0068)
(Company tenure)^2/100	-.0666*	-.0452*	-.0733*	-.0563*
	(.0129)	(.0167)	(.0110)	(.0141)
Black	-.114*	-.0707	-.189*	-.148*
	(.052)	(.0496)	(.051)	(.048)
Female	-.476*	-.346*	-.422*	-.306*
	(.046)	(.069)	(.048)	(.063)
Union coverage	.184*	.122	.226*	.180*
	(.029)	(.090)	(.032)	(.050)
Married	.112*	.116*	.129*	.101*
	(.034)	(.036)	(.035)	(.033)
Blue-collar	-.166*	-.0975*	-.202*	-.107*
	(.032)	(.0366)	(.036)	(.038)
Hourly	-.145*	-.160	-.179*	-.153*
	(.030)	(.092)	(.032)	(.048)
South	.0411	.0337	-.0012	-.0005
	(.0304)	(.0269)	(.0323)	(.0269)
South*Black	-.174*	-.141*	-.0882	-.0674
	(.069)	(.063)	(.0712)	(.0595)
PSIDVS dummy	.448*	.367	.797*	.614*
	(.214)	(.193)	(.214)	(.197)
PSIDVS*schooling	-.0253	-.0208	-.0613*	-.0471*
	(.0148)	(.0132)	(.0147)	(.0139)

PSIDVS*blue-collar	.163*	.132*	.241*	.184*
	(.069)	(.063)	(.071)	(.065)
PSIDVS*female	.457*	.367*	.280*	.214*
	(.115)	(.113)	(.104)	(.092)
Constant	8.69*	8.68*	8.91*	9.07*
	(.113)	(1.24)	(.132)	(1.30)
Number of Observations	2848	2848	2744	2744
Hausman Test Statistic	241.6*		5.81	

* - indicates significance at 5% level, two-tailed test

[1] Corrected for earnings measurement error using multiple imputation technique described in text.

Table 4 - Longitudinal Earnings Models (1987-1983), pooled PSID and PSIDVS sample
dependent variable = Δln(interview earnings)
(coefficients; standard errors in parentheses)

	(1) - Uncorrected	**(2) - Corrected**[1]
ΔPotential	.0361	.0491*
Experience	(.0208)	(.0202)
Potential	-.00916*	-.00886*
Experience 1983	(.00141)	(.00257)
ΔTenure	.0145*	.0192
	(.00328)	(.0147)
Tenure 1983	.00128	-.00213
	(.00204)	(.00275)
Black	-.0373	.0209
	(.0536)	(.0519)
Female	.0166	-.0254
	(.0360)	(.0464)
ΔUnion coverage	.133*	.0943
	(.0416)	(.0564)
Union coverage	.0511	.00084
1983	(.0303)	(.136)
Married (between	.0816	-.00443
1983 and 1987)	(.0513)	(.0445)
Divorced (between	.0169	.0579
1983 and 1987)	(.0601)	(.0613)
Schooling 1983	-.00006	-.0100
	(.00670)	(.00740)
ΔBlue-collar	-.0832*	.0249
	(.0369)	(.0385)
Blue-collar 1983	-.0478	-.0255
	(.0342)	(.0362)
ΔHourly	-.147*	-.0851*
	(.0371)	(.0327)
Hourly 1983	-.0277	.0663
	(.0320)	(.133)

Changed firms	-.0153	.00597
	(.0387)	(.0304)
South	-.0491	-.0380
	(.0283)	(.0224)
South*Black	.0639	.0105
	(.0695)	(.0540)
PSIDVS dummy	-.150	-.140
	(.236)	(.180)
PSIDVS*schooling	-.00123	.00258
	(.0164)	(.0125)
PSIDVS*blue-collar	.0303	.0225
	(.0692)	(.0534)
PSIDVS*female	-.116	-.0925
	(.115)	(.0905)
Constant	.310*	.332*
	(.122)	(.148)
Number of Observations	2134	2134
Hausman Test Statistic	79.7*	

* - indicates significance at 5% level, two-tailed test

[1] Corrected for earnings measurement error using multiple imputation technique described in text.

Appendix Table - Means for the 1983 Cross-Section PSIDVS and PSID Samples (standard deviations in parentheses)

	PSIDVS	PSID[1]
Interview Earnings	30,380	22,187
	(7,659)	(12,681)
Company Record Earnings	30,426	--
	(7,735)	
Years Schooling	12.83	12.88
	(2.16)	(2.57)
Potential Experience	18.74	22.33
	(10.46)	(12.35)
Company Tenure	16.04	9.51
	(8.81)	(9.17)
Black	.099	.106
	(.299)	(.308)
Female	.073	.186
	(.260)	(.389)
Union coverage	.494	.297
	(.500)	(.457)
Married	.744	.662
	(.437)	(.473)
Blue-collar	.637	.486
	(.482)	(.500)
Hourly	.488	.455
	(.500)	(.498)
South	0	.298
		(.458)
South*Black	0	.061
		(.240)
Number of Observations	344	2504

[1] Calculated using 1983 weights.

Endnotes

[1] Articles that test earnings measurement error properties include Duncan and Hill (1985), Bound et. al. (1989, 1990, 1994), Rodgers, Brown, and Duncan (1993), and Pischke (1995), using the PSIDVS, and Bound and Krueger (1991) using the CPS-SER.

[2] Sudman and Bradburn (1974) provide general discussion of the role of salience in survey response error. The salience of an event or events depends on recency, importance, complexity, and potential for repressed memory. Papers that focus on the role of salience in reported unemployment include Duncan and Mathiowetz (1985), Akerlof and Yellen (1985), and Mathiowetz and Duncan (1988).

[3] Other factors -- such as elapsed time between the period surveyed and the date when the survey is administered, and the task difficulty of the survey -- may also affect reporting error; see Duncan and Mathiowetz (1985, chapter 6) for discussion and additional references. We do not discuss or incorporate these factors because they either do not vary across individuals in the PSIDVS or are unobservable to us.

[4] To investigate the costs and benefits of administering the PSID every other year, the PSIDVS asked respondents about their earnings for each of the past 2 years. To maintain comparability across the PSID and PSIDVS survey instruments for our sample, we only use PSIDVS interview earnings from the latest year.

[5] See Duncan and Mathiowetz (1985) and Bound et. al. (1994) for additional description of the samples and procedures.

[6] Higher order terms in record and interview earnings were not included because they do not improve the fit of the cross-section models.

[7] The inclusion of our additional covariates reduces the degree of mean reversion in the 1983 cross-section, and in the longitudinal model reported below, but increases mean reversion in the 1987 cross-section.

[8] We also estimated these models for hourly workers only, using yearly earnings divided by yearly hours as the earnings measure. Despite greater measurement error (Bound *et. al.* 1990) and greater mean reversion in hourly than in yearly earnings, the explanatory power of these models is higher than for the yearly earnings models.

[9] The measurement error models from Section 3 were estimated with measurement error as the dependent variable, for ease of interpretation. Our imputation equation, however, is easier to interpret written in the form (2b). Although the imputation equation could be estimated with either true earnings or measurement error as the dependent variable -- the results are computationally equivalent -- we use (2b) because it eliminates one step from the estimation.

[10] A related problem may occur if the imputations required in the main sample are based on Z values outside the range of the Z values in the validation sample, since in this case there is no way to empirically test the validity of the imputation model for these Z values. For our applications it turns out that the range of the key continuous variables in the PSID (income, schooling, and tenure) are reasonably well covered in the PSIDVS.

[11] Our STATA 3.1 programs will be available on request to the authors after publication of this paper.

[12] The Lee and Sepanski (1995) approach to our problem is implied in Bound *et. al.* (1994), but in a more restrictive form than our model, and without any suggestions regarding the statistical properties of such estimation.

[13] Our inclusion of Survey of Economic Opportunity sub-sample observations makes the PSID sample non-representative of our target population. In the empirical work described below, we use the PSID sampling weights and apply the weighting procedures discussed in DuMouchel and Duncan (1983) to correct for this. Because the PSIDVS does not provide population weights, we assigned the average PSID sample weight to each PSIDVS observation. Comparison to weighted least squares regression results revealed that standard weighted regression procedures generally underestimate the true standard errors produced by the correct weighting technique by at least 20 percent.

[14] The PSIDVS contains a high proportion of skilled craft workers, which explains the high blue-collar earnings in the PSIDVS. The low return to schooling in the PSIDVS is also partially explained by its occupational distribution. The lower male/female wage gap may be explained by the large size of the PSIDVS sample firm, which may make it a likely target for federal Equal Employment Opportunity enforcement.

[15] Following Brown and Light (1992), we impose longitudinal consistency on our tenure variable. We assume that reported tenure in 1981 is correct. For individuals who do not report changing firms between adjacent survey dates, tenure is increased by one year. For individuals who report changing firms between survey dates and also indicate that their current tenure exceeds one year, tenure is recoded to 0.5 years.

[16] We could also allow the omitted effect Ψ to vary by a proportionality factor over time, as in Jakubson (1991) and Card and Lemieux (1994). This model requires more complex estimation techniques, because it leads to inclusion of an endogenous variable -- lagged earnings -- on the right-hand side of equation (14). In earlier versions of this paper we estimated the more complex model, using multiply-imputed values for both the dependent change variable and lagged earnings as an explanatory variable. We were unable to reject the time invariant omitted effect assumption embodied in (14), however, due largely to the inclusion of the change and base level of potential experience and tenure. Also, an F-test rejects (at better than the 1% level) a difference specification that excludes the level variables. We therefore only report results for the model described in the text.

[17] The PSID schooling questions are not asked every year. Updating of the schooling information in 1985, however, increased schooling attainment for 11.1% of our longitudinal sample. As discussed in Angrist and Newey (1991), although direct estimation of a coefficient on schooling increases is not possible in fixed effect earnings equations that include changes in potential experience, a reduced-form schooling effect can be identified through inclusion of a quadratic term in potential experience. We do not investigate this issue.