NBER WORKING PAPER SERIES

MEASUREMENT ERROR IN CROSS-SECTIONAL AND LONGITUDINAL LABOR MARKET SURVEYS: RESULTS FROM TWO VALIDATION STUDIES

John Bound

Charles Brown

Greg J. Duncan

Willard L. Rodgers

Working Paper No. 2884

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1989

Prepared for the symposium "Panel Data and Labor Market Studies" held in Amsterdam, December 15-17, 1988. Research reported in this paper was supported by NSF grant SES 86200-86 and reflects the work of the authors, Alan Krueger and Nancy Mathiowetz. Lynn Dielman provided inspired research assistance throughout the data collection of the second wave of the PSID Validation Study and in the analysis of those data represented in this paper. Dawn von Thurn was also instrumental in the data collection phase. It has benefited from comments of participants in the Cornell Labor Workshops. We also owe special thanks to several members of the personnel and payroll offices of the firm that cooperated with us in the PSID Validation Study. We regret that our pledge of confidentiality precludes listing their names to give them some well-deserved recognition. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Working Paper #2884
March 1989

MEASUREMENT ERROR IN CROSS-SECTIONAL AND LONGITUDINAL LABOR MARKET SURVEYS: RESULTS FROM TWO VALIDATION STUDIES

ABSTRACT

This paper reports evidence on the error properties of survey reports of labor market variables such as earnings and work hours. Our primary data source is the PSID Validation Study, a two-wave panel survey of a sample of workers employed by a large firm which also allowed us access to its very detailed records of its workers earnings, etc. The second data source uses individuals' 1977 and 1978 (March Current Population Survey) reports of earnings, matched to Social Security earnings records.

In both data sets, individuals: reports of earnings are fairly accurately reported, and the errors are negatively related to true earnings. The latter property reduces the bias due to measurement error when earnings are used as an independent variable, but (unlike the classical-error case) leads to some bias when earnings are the dependent variable. Measurement-error-induced biases when change in earnings is the variable of interest are larger, but not dramatically so. Various measures of hourly earnings were much less reliable than annual earnings. Retrospective reports of unemployment showed considerable under-reporting, even of long spells.

John Bound
Charles Brown
Greg J. Duncan
Willard L. Rodgers

Institute for Social Research
University of Michigan
Ann Arbor, MI 48106-1248

I. Introduction

In contrast to our knowledge of most other aspects of labor market phenomena, conventional wisdom about the nature and effects of the measurement properties of earnings, work hours, tenure and other key concepts is based on assumption rather than on direct observation. Most empirical studies of labor market behavior ignore measurement error altogether or, at most, view it as a harmless component of the stochastic disturbance of their behavioral models.

Measurement error models presented in econometric and statistical textbooks typically make strong — and exceedingly convenient — assumptions about the properties of error (cf. Fuller, 1987). Most frequently measurement error in a given variable is assumed to be uncorrelated with the true level of that and all other variables in the model, measurement error in other variables and the stochastic disturbance (e.g., Kmenta, 1986; Pindyck and Rubinfeld, 1981). From these assumptions comes the most elementary version of conventional wisdom about the effects of measurement error on estimates of cross-sectional models: (i) error in the dependent variable neither biases nor renders inconsistent the parameter estimates but simply reduces the efficiency of those estimates and (ii) error in the measurement of independent variables produces downward-biased and inconsistent parameter estimates, with the extent of bias and inconsistency dependent upon the extent of the error.

The novelty of panel data has produced a less well-defined body of conventional wisdom about the measurement properties of such data. The most convenient assumption is that measurement errors in a given variable are uncorrelated across time, leading to the conventional wisdom that measures of change produced by panel data are much noisier than corresponding measures of levels. Allowance for autocorrelation in these errors (as in Griliches and Hausman, 1987) but retention of the other assumptions leads to a focus on the relative sizes of the the correlation across time in the error and the true level of the variable.

The research summarized in our paper is based on direct observation of the measurement error properties of interview reports of such labor market measures as earnings, work hours and tenure. Our principal data source, the PSID Validation Study, is a unique two-wave panel survey of a probability sample of workers from a single large manufacturing company. As detailed in the appendix, some 418 workers were interviewed in the first wave, conducted in the summer of 1983, using a sampling scheme that produced approximately equal numbers of hourly and salaried workers and a uniform age distribution. Interviews were conducted four years later with 341 of the original sample and with 151 additional hourly workers drawn at random from company employment lists just prior to the interview.

Access to very detailed company records enabled us to obtain virtually error-free validation of survey responses for a host of interesting labor market measures. However, the advantages of such precise validation are offset to some extent by the restrictive nature of the single-firm sample; all hourly workers in the firm are unionized and the distribution of earnings appears to be much more compressed within the firm than among general population samples of workers even within the same industry. Company records for hourly workers also showed surprising variability in work hours and earnings from one pay period to the next. The extent to which this variability is unique to this company or its industry is an important and, as yet, unanswered issue for considering how our results may apply to data from larger, more representative samples.

A second data source used in the paper makes possible direct observation of measurement error by matching panel data on earnings from the March, 1977 and 1978 waves of a general population survey — the U.S. Current Population Survey — to Social Security earnings records for those same individuals. Sample sizes in this data set are much larger — 2924 men and 465 women. Also described in the appendix, this source provides general population coverage at a cost of some degree of imprecision in the validating information and problems arising from the truncation of earnings in the Social Security records.

These data sources provide much information that challenges the conventional wisdom. We find that the amount of measurement error in cross-sectional reports of annual earnings is rather low in both data sets. Error in reports of annual work hours is higher, while error in reports of hourly earnings, obtained by dividing annual earnings by annual hours, is quite high. An investigation of the error properties of alternative measures of hourly earnings produced the surprising result that reports of either "usual" or last pay period earnings and hours appear to be even less reliable than reports of hourly earnings obtained by dividing annual reports of earnings and hours.

Although not as high as for cross-sectional measures, the reliability of panel survey measures of change in earnings appeared to be surprisingly high. Over 75 percent of the observed variation in first-differenced 1977 and 1978 annual earnings is true variation, while more than half of the observed variation in earnings differenced over the four-year interval between 1982 and 1986 appears to be real.

Some of the surprising reliability in the cross-sectional and panel measures is due to what Bound and Krueger (1988) have called "mean-reverting" measurement error – a pronounced negative correlation between the error and true level of many of the measures we were able to validate. Such correlations are assumed to be zero in classical measurement error models but are clearly pervasive in both of our validation data sets.

Classical measurement error models of error-induced bias on right-hand side variables commonly presume zero covariances between measurement errors in a given measure and the true levels of other variables in the model. As with another validation study using a different data source (Rodgers and Herzog, 1987), we find considerable evidence of nonzero covariances, with earnings reporting errors at times negatively correlated with job tenure and positively correlated with schooling levels.

An investigation of the quality of retrospective reports of unemployment spells in the company sample showed massive underreporting. Scarcely one-third of the unemployment spells that appeared in company records were reported in the interview. Longer and more recent spells were more accurately reported, although the fraction of quite long and quite recent spells still unreported was uncomfortably high. Although consistent with other research on episodic recall, the poor quality with which event-history employment data are reported has received surprisingly little attention.

II. Econometric Issues

If one is likely to have measurement error, assuming it is of the classical variety is obviously convenient. But in most social-science contexts there is no a priori reason to believe that the correlations assumed to be zero in the classical case are in fact zero in one's data. In addition to providing some evidence about the magnitude of measurement errors, validation studies permit one to determine whether measurement errors are uncorrelated with other variables.

Suppose we believe the true model is

$$Y = X\beta + \epsilon$$

where ϵ is uncorrelated with X. Instead of X and Y, we observe

$$\tilde{X} = X + u$$

$$\tilde{Y} = Y + v$$

We will not assume u and v are uncorrelated with X and Y, but we will assume that ϵ is uncorrelated with X, u, and v. The motivation for this last assumption is partly strategic, partly conventional: a validation study in principle allows us to observe u and v but never ϵ (so we have little to say about such correlations) and they are in any case treated in the literature (e.g., correlation between X and ϵ leads to a standard "omitted variable" bias).

Whether we have validation or not, we observe $\tilde{X}$ and $\tilde{Y}$. The least-squares estimator of β is

$$b = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$$

We will present a general approach to dealing with measurement errors in X and Y which are correlated with the true X and/or Y. Before doing so, however, it useful to highlight a few results for the biases due to measurement errors when convenient assumptions hold. To simplify discussion of the various biases, we assume throughout that the Xs have been defined so that $\beta \geq 0$. Consider three special cases:

First, if there is classical measurement error in only one independent variable X_j, $\sigma^2_{\tilde{X}_j} = \sigma^2_{X_j} + \sigma^2_{u_j}$. The proportional bias in estimating β_j (i.e., minus the ratio of the bias to the true β_j) depends on $\sigma^2_{u_j}/\sigma^2_{\tilde{X}_j}$. In particular, with only one independent variable in the regression, the proportional bias is just equal to this ratio. Alternatively, $\sigma^2_{X_j}/(\sigma^2_{X_j} + \sigma^2_{u_j})$, the ratio of true to total variance, gives the ratio of the expected value of the estimate b_j to the true β_j. This is probably the most common textbook result about measurement error.

Second, even if the error u_j is correlated with the true X_j (or other Xs), the proportional downward bias is equal to the regression coefficient for X_j from a hypothetical regression of u_j on the set of measured $\tilde{X}$'s. If there is only one independent variable in the model, this reduces to the simple regression coefficient $b_{u\tilde{X}}$. (In the case where u and X are uncorrelated, $b_{u\tilde{X}}$ is equal to the variance ratio $\sigma^2_u/(\sigma^2_u + \sigma^2_X)$. But if u and X are negatively correlated, $b_{u\tilde{X}}$ can be smaller than in the classical case.)[1]

Third, if the <u>dependent variable</u> Y is measured with error, and that error is correlated with the true Y ($v = \delta Y + v^*$ where v^* is uncorrelated with X, Y, and u) and the Xs are measured without error, then the proportional bias in estimating each β is just equal to δ. To emphasize the similarity to the previous case, note that δ can be thought of as the regression coefficient b_{vY}.

---

[1] In the U.S., there have been occasional proposals to add measurement errors to "strategic" variables in some data files in order to avoid confidentiality problems. The result in the text suggests that if one goes this route, the error (in variables that are expected to be used as independent variables) should be non-classical.

Each of the above results applies to cross-section analysis, and to panel data by substituting ΔX for X, etc. But when one uses $\Delta\tilde{Y}$ and $\Delta\tilde{X}$ as one's dependent and independent variables, another aspect of the data becomes important — the correlation over time in the true values (the correlation between Y at time t and at time t-1, and similarly for X) and in the measurement errors (the correlation between v at time t and at time t-1, and similarly for u). A general result is that, if the variance of a variable (say, X) is the same in both years, the variance of ΔX is equal to $2\sigma_X^2(1-r_{X_t,X_{t-1}})$, which is greater or less than σ_X^2 as $r_{X_t,X_{t-1}}$ is less than or greater than one half. A common concern, usually expressed in the context of classical measurement errors, is that true values of X will be highly correlated over time, while the measurement errors will be more or less uncorrelated. In this case, $\sigma_{\Delta X}^2$ is less than σ_X^2 but $\sigma_{\Delta u}^2$ is greater than σ_u^2, so that moving from "levels" to "changes" intensifies the bias due to errors in measuring the independent variable(s).

This increase in bias does not necessarily mean that using "change" variables is to be avoided. In most cases, differenced models are used when the analyst suspects that the error term ϵ contains a component which is constant over time and, contrary to our assumptions, correlated with $\tilde{X}$ (typically because some variable which doesn't change over time cannot be measured). In this case, regressing $\tilde{Y}$ on $\tilde{X}$ will produce estimates of β which are biased by both measurement error and the omitted variable, while regressing $\Delta\tilde{Y}$ on $\Delta\tilde{X}$ eliminates the latter bias. We have not been able to derive any particularly illuminating results for comparing the gain from eliminating the omitted variable bias with the potential intensification of the bias due to measurement error.[2]

Having highlighted some special cases in which the consequences of measurement error can be succinctly summarized, we turn to a more general model. With u and v potentially correlated with X and Y, the least-squares regression coefficient can be rewritten as

$$b = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(\tilde{X}\beta - u\beta + v + \epsilon)$$

$$= \beta + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(-u\beta + v + \epsilon)$$

Therefore, the bias of the least-squares estimator b is

$$\text{plim } b-\beta = \text{plim}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'(-u\beta + v)$$

It is useful to collect the measurement errors and their coefficients:

$$w \equiv \begin{bmatrix} u \mid v \end{bmatrix}$$

$$\gamma \equiv \begin{bmatrix} -\beta \\ 1 \end{bmatrix}$$

$$\text{plim } b-\beta = \text{plim}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'w\gamma \equiv \text{plim } A\gamma$$

---

[2]To get some sense of how elusive such a result is likely to be, note two results which are easy to see or derive for classical errors in measuring X: (1) If $\tilde{X}$ and ϵ are positively correlated, the omitted variable bias and the bias due to measurement error work in opposite directions, so eliminating the latter may make things worse; (2) Classical error in measuring X tends to reduce the bias due to the omitted variable — in the limit, if $\tilde{X}$ is <u>all</u> classical noise, it can't be correlated with ϵ at all!

If there are k separate variables in the independent-variable matrix X, then A is k by k+1. It can be rewritten in a more intuitive form as

$$A = \left[b_{u\tilde{X}} \mid b_{v\tilde{X}} \right]$$

where the jth column of $b_{u\tilde{X}}$ consists of the coefficients from regressing u_j on $\tilde{X}$, and $b_{v\tilde{X}}$ is the set of coefficients from regressing v on $\tilde{X}$.

If there is measurement error in only one independent variable X_j, only one column of A is non-zero, and $A_{jj} = b_{u\tilde{X}_j}$, as claimed in our discussion of special cases. If $v = \delta Y + v^* = \delta X\beta + \delta\epsilon + v^*$, v^* is uncorrelated with the other variables of the model, and the independent variables are measured without error, then $b_{u\tilde{X}}$ is a matrix of zeros and $b_{v\tilde{X}} = \delta\beta$. Thus, the proportional bias for each coefficient equals δ.

It is also useful to note the analogous expressions among strictly observable variables; i.e., before taking plims. Let b_R be the OLS coefficients from regressing Y on X (the record variables), and b_I be the OLS coefficients from regressing $\tilde{Y}$ on $\tilde{X}$ (the interview variables).

Then

$$b_I = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$$

$$= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(\tilde{X}b_R\text{-}ub_R + v + f)$$

$$= b_R + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'(\text{-}ub_R + v + f)$$

$$= b_R + Ac + (\tilde{X}'\tilde{X})^{-1}\tilde{X}'f$$

where

$$c \equiv \begin{bmatrix} -b_R \\ 1 \end{bmatrix}$$

We can calculate b_I, b_R, A, and c, thus neglecting the last term (which vanishes in the probability limit). The fact that it doesn't necessarily vanish in the actual sample data could in principle tell us whether our assumption is correct that ϵ, the equation error in the model with the correct variables, is uncorrelated with the measurement errors u and v. However, in our data this discrepancy is usually small, so we have not pursued this issue.

III. Measurement Error in Annual Measures of Earnings and Work Hours

Throughout this paper, we treat the "record" value of a variable, either from the company's own records (PSIDVS) or from Social Security records (CPS-SSA), as the "true" value, and treat the difference between the individual's report and this record value as measurement error. We do this for two reasons. First, we have a great deal of confidence in the accuracy and recording of the company records, in part because of the extraordinary cooperation of the company involved. We believe the assumption that the Social Security records are correct is at least defensible, in part because in choosing the sample those who were most likely to present problems (e.g., job changers) were excluded. Second, as a practical matter, there seems to be no way to relax this assumption without making other less plausible ones (e.g., that errors in records are uncorrelated with the true values).

In Table 1, we present simple summary statistics for the errors in measuring earnings in the PSID Validation Study data[3] and in the CPS-SSA data analyzed by Bound and Krueger (1988).[4] In each case, we present five summary statistics described above: the ratio of error to error plus true variance $\sigma_u^2/(\sigma_X^2 + \sigma_u^2)$; the regression coefficient from regressing the earnings error on interview earnings, $b_{u\tilde{X}}$ (which would equal the variance ratio in the first column if the error and true values were uncorrelated); the regression coefficient from regressing the earnings error on its true value, b_{vY}; and, for the change in earnings only, the correlation over time in the measurement errors and in the true values. We switch to v and Y (in column 3) from u and X (in the remaining columns) to highlight the fact that the regression of the

[3]We included only individuals who reported working 520–3500 hours and earning at least $1000.

[4]We used the Bound-Krueger results from Table 4, part B, which gives the relevant variance-covariance matrix for those with earnings below the Social Security earnings ceiling (Social Security earnings, taken as the "true" value, are reported only up to a ceiling ($15,300 in 1976 and $16,500 in 1977). For this sample, the variance-covariance matrix for X and u can be calculated in the usual way. They also report an estimated variance-covariance matrix which uses the full sample, and uses maximum likelihood methods to correct for the fact that X is truncated at the Social Security limit. Not surprisingly the variance of X is larger in this sample (by about one third) but the error variances are very similar.

measurement error on the true value is most interesting when ln-earnings is the dependent variable; but throughout X and Y refer to true ln-earnings while u and v refer to the difference between reported and true values of ln-earnings. All estimates refer to males, who make up an overwhelming fraction of the workers at the plant that cooperated in the PSID Validation Study.

The first two rows of the Table present results for cross-section analyses, based on the PSID Validation Study data. Judged by the variance ratio, the bias due to errors in measuring earnings when earnings is an independent variable is appreciable, but perhaps not alarming; depending on the year in question, the effect of earnings on some other variable would be understated by 15 to 30 percent.[5] However, the variance ratio considerably overstates the likely bias, because the measurement error is negatively correlated with the true value of earnings. As a result, the likely bias is on the order of 8 to 24 percent. This "good news" for using earnings as an explanatory variable is to some extent tempered by the corollary for using it as a dependent variable: the negative correlation between true earnings and the measurement error means that the impact of other variables on earnings could be understated by 10 to 17 percent. So, while the classical assumptions lead to the conclusion that mismeasurement produces bias when earnings are used as an independent variable but not if they are used as the dependent variable, we find that because of violations of the classical assumptions a bias is introduced in the latter case but the bias in the former case is reduced.

---

[5]The 15 percent value for 1982 is considerably lower than the 30 percent for 1986 or the CPS-SSA estimates discussed below. It reflects unusually large true earnings variance in 1982, a year of significant unemployment at the studied firm.

The third row of the table allows us to compare the biases one might expect using change variables with those that arise in cross-section analysis. The variance ratio does go up (to 29 percent, from an average of 23 percent in the two cross sections), and the more appropriate measures $b_{u\hat{X}}$ and b_{vY} also grow by a roughly similar proportion. However, the biases do not increase as sharply as one might have expected, because – while the correlation in the errors is near-zero, as is sometimes assumed – the correlation between 1986 earnings and earnings four years earlier is only .452 in these data.

The second set of three rows of the table present analogous summary statistics from Bound and Krueger's CPS-SSA analysis. The three major conclusions discussed above – that $b_{u\hat{X}}$ is less than the variance ratio, that $b_{u\hat{X}}$ and b_{vY} are of similar magnitude, and that differencing increases the importance of measurement error, but probably not disastrously so for most applications – can be seen in their data as well. The correlations over time for u and X are higher, a point we discuss below.

In comparing the detailed results, two differences between the studies should be emphasized. First, the PSID Validation data come from a single, unionized firm. This considerably restricts the true variation in earnings (see Appendix Table 1), making measurement error more serious than it might be in a broader sample. It is also possible that measurement errors are smaller, (e.g., because our workers did not change jobs in the year preceeding the interview), though this difference is likely to be less important than the restriction in the variance of true earnings. Second, the PSID Validation Study's earnings change spans a four-year period, while the CPS-SSA data span a pair of adjacent years. Thus, we should expect a higher correlation for u over time and for X over time in the Bound-Krueger data. Indeed, if measurement error had only first-order serial correlation, a one-year correlation of .37 would imply a four-year correlation of .02, so the .073 in Table 1 is higher than the simplest back-of-the envelope calculation would suggest.

The PSID Validation data include interview and record information on hours worked per year for hourly (non-salaried) workers, so one can also explore consequences of errors in earnings per hour. The findings turn out to be quite different for earnings per hour than for earnings per year: the biases due to measurement error are considerably more severe than those in Table 1.

Table 2 has three sections. The first presents results for the annual earnings of all hourly workers. While there are some differences between these numbers and those of the combined sample of salaried and hourly workers shown on Table 1 (the sample of hourly workers has less difference between the two cross sections, larger values of b_{vY} in the cross sections but a smaller value for the changes), these differences are negligible when compared with the impact of moving from annual earnings to earnings per hour for hourly workers.

The biases arising from using earnings per hour as an explanatory variable are, depending on one's perspective, serious or alarming (middle section of Table 2). Using the variance ratio, the downward bias is two-thirds of the true value in cross-section analysis, and (unlike annual earnings) the more general measure $b_{u\tilde{X}}$ gives a very similar estimate. The consequences of using earnings per hour as a dependent variable are less clear, with proportionate biases of essentially zero and 30 percent in the two cross-sections.

If one instead uses the change in earnings per hour, matters are worse still – proportional biases of 82 or 87 percent when the change is the independent variable, and 37 percent as a dependent variable.[6]

To interpret these results, it is important to keep in mind the peculiar features of the data: the variance in true earnings per hour is understated by our focus on a single firm, but the four-year gap spanned by the change in earnings per hour should produce less biased estimates than one would expect from one-year changes.

The last three lines of the table present the summary statistics for hours per year. Measurement error here is severe enough to produce non-negligible biases when hours are either an independent or dependent variable, but they are not so badly measured that the poor showing of earnings per hour can be attributed to the poor measurement of the denominator.[7] Rather, the problems of measuring earnings per hour in the PSID Validation data arise from an unhappy combination of errors in measuring earnings, errors in measuring hours, and the intercorrelations involved. The correlation between true earnings and hours is very high in these data, reducing the variance of true earnings per hour, while the correlation between the errors in earnings and hours is much smaller.[8]

---

[6]Altonji (1986) reaches similar conclusions, using PSID data and a variety of more complex indirect estimation techniques.

[7]It is important to keep in mind that the PSID annual work hours measure is constructed from an elaborate question sequence asking about work lost to sickness, vacation, strike, unemployment and time out of the labor force. Simpler question sequences may produce greater error variance.

[8]The correlation between record ln(earnings) and ln(hours) was .858 in 1986 and .879 in 1982. In contrast, the correlation between the errors in ln(earnings) and ln(hours) was .407 in 1986 and −.169 in 1982.

Implications for earnings functions

As an illustration of the more general methodology, we present an analysis of the consequences of measurement error for estimates of a simple earnings function. As with a similar analysis of the 1983 wave of the PSID Validation Study conducted by Duncan and Hill (1985), we regress the logarithm of annual earnings on education, tenure (years with current employer), and experience prior to starting work with current employer.[9] As is often done in such contexts, our measure of pre-company experience is age minus schooling minus 5 minus tenure. We focus on cross-sectional estimation, because the change in our explanatory variables is either zero (education, pre-company experience) or approximately constant (company tenure).

We have interview and record values of earnings and tenure; as it turns out, errors in the latter are negligible, so our emphasis is on the impact of errors in reporting earnings, and in particular whether they are correlated with the explanatory variables. We have no independent verification of education (as reported in the interview) or pre-company experience.[10] Given that these seem likely a priori to be relatively well-measured, and that we have no independent way of verifying workers' reports, we assume these variables are measured without error.

[9]Mellow and Sider (1983) ran similar regressions using 1977 Current Population Survey data, but had been restricted to information provided by the employers of CPS respondents. Since there was no other attempt to verify the employer information, their data is best thought of as two fallible indicators of the underlying wage and other employment conditions. They find very few differences between coefficients obtained from the interview and employer data.

[10]Pre-company experience is equal to age (based on company records, but these were based on information originally provided by the worker) minus education (reported by the worker) minus 5 minus tenure (from company records).

Table 3 is an elaboration of the algebraic relationship among b_I, b_R, c, and A. The first column gives values of b_I, regression coefficients based on interview data. The second gives values of b_R, based on the record data, and hence (apart from sampling error) gives the "true" coefficients. Since only tenure and ln earnings are assumed to be measured with error, the A matrix has only two non-zero columns, these being the third and fourth columns of the table. The final column is the discrepancy, which occurs because the correlation between the equation error (using the record values of the variables) and u or v is not zero in the sample. The top part of Table 3 refers to the 1986 cross-section, while the bottom part refers to the 1982 data.

The estimated coefficients of the earnings function, based on the interview data (b_I) are similar to what one finds in the literature, with two exceptions. First, they are generally smaller in absolute value, because our data refer to one firm and part of the return to education and experience comes from access to higher-paying firms. Second, the coefficient of pre-company experience is negative in the 1986 data, which is not what one finds in other data sets.

Differences between the coefficients obtained from record data and interview data may be due to three sources: a relationship between the error in tenure and the measured X, $b_{u_{Tenure}\tilde{X}}$; a relationship between the error in earnings and measured X, $b_{v\tilde{X}}$; or the residual discrepancy. In Table 3, the relationship between the error in tenure and the measured variables (including measured tenure) is negligible, because these workers are able to report their tenure very accurately.[11] On the other hand, errors in measuring earnings are significantly related to education in the 1986 data and

[11]The correlation between reported and true tenure exceeded .99 in each year.

to tenure in the 1982 data. As a result, the proportionate difference between b_I and b_R in these cases is not negligible — interview data overstate the record return to education in 1986 by about a third (.025 vs. .018, with a t statistic on the difference of 2.1) and understate the return to tenure in 1982 by almost the same proportion (.011 vs. .014, with a t statistic on the difference of 4.4).[12]

The discrepant pattern of covariances affecting the earnings functions between the 1982 and 1986 cross-sections is disturbing because it implies that biases due to such covariances change, perhaps unpredictably. We searched for factors that might reconcile the discrepancies and discovered that the much more extensive unemployment prior to the 1983 interview seemed largely responsible.

Workers with extensive unemployment tended to overreport earnings, perhaps because they reported "typical" annual earnings, not realizing that 1982 earnings had been reduced by their unemployment. The prevalence of unemployment is negatively related to company tenure, producing the negative covariance between earnings-reporting error and tenure shown in Table 3. Calendar year 1986 produced no unemployment and, perhaps, a more "normal" relationship between reporting error and the earnings function covariates. At any rate, the addition to the 1982 earnings function of the amount of unemployment in 1982 as revealed in company records produced earnings function estimates that were much more similar to those found for 1986.

---

[12]The results for 1982 are quite similar to those of Duncan and Hill (1985), who compare b_I and b_R using a slightly different sample from the first PSID Validation Study wave. Bound and Krueger (1988) report regressions of errors in measuring ln-earnings on measured values of variables like those in Table 3 (plus additional demographic variables such as marital status and region). They find small coefficients which are not very stable across two adjacent years.

Specifically, the regression coefficients and, in parentheses, standard errors for the regression of 1982 earnings error on education, pre-company experience, tenure and dummy variables for actual 1982 unemployment (corresponding to the column labeled $b_{v\tilde{X}}$) in Table 3 were:

$$v = -.0433 + \underset{(.0036)}{.0036}\ \text{Education} - \underset{(.0009)}{.0016}\ \text{Tenure} + \underset{(.0013)}{.0000}\ \text{Pre-company experience}$$

Dummy variables for substantial unemployment had a positive and highly significant effect on earnings error and the addition of the dummy variables increased the R^2 from .054 to .138.

IV. Error Properties of Alternative Measures of Hourly Wages

Evidence presented in Table 2 showed a substantial proportion of measurement error in reports work hours and, especially, earnings per hour. One source of the measurement error may be fluctuation in the conditions about which reports are elicited. Week-to-week variability in hours may be caused by several factors, including holidays, vacations, illness, and overtime. Depending on the type of job, these variations in hours may or may not be reflected in paychecks; and earnings may furthermore be supplemented by bonuses and incentive pay or reduced by disciplinary actions.

Faced with such instability in the target, researchers have adopted various measurement strategies. Perhaps the most straightforward is to ask about a specific (preferably recent) pay period. This is a standard approach in U.S. Bureau of Labor Statistics establishment surveys, but is not common when workers are being interviewed about their earnings. If the concept of interest is a longer range level of compensation, per pay period measurement might be thought to produce a higher proportion of stochastic variability for the sake of a lower proportion of measurement error (although whether this strategy in fact leads to a reduction in measurement error is a matter for empirical verification). A second strategy is to ask questions about a longer period of time, typically a calendar year, on the assumption that the availability of year-end reports from the employer and the preparation of tax returns increases the accuracy of reports of earnings (although there is no reason to make a similar assumption for accuracy of reports on hours worked). This approach is taken by the Current Population Survey in its March Supplement each year and in the PSID. Prior calendar year information has the disadvantage of not reflecting "current" compensation to which current working conditions can be related, an especially critical problem if the respondent worked for a different employer or held a position different from his or her current one for all or part of the preceding calendar year. A third strategy is to ask about "usual" hours and "usual" pay, in effect asking the respondent to do the appropriate integration and trend analysis to arrive at a report that best reflects his or her current conditions. This approach is used by U.S. Current Population Surveys and, e.g., by National Longitudinal Surveys in obtaining weekly earnings and hours.

Weekly earnings

All three of these measurement strategies were employed for hourly workers in the PSID Validation Study.[13] Because information was available from company records

---

[13] A measurement strategy for wage rates that is used in some surveys is to ask respondents to report their hourly earnings as CPS does for workers paid by the hour, rather than (or in addition to) dividing their reports of earnings for a week or other period by the hours worked in that period as was done exclusively in the present study. This strategy has the advantage that some employees (particularly hourly, as opposed to salaried, workers) may be more aware of their wage rates than of their gross pay for a

about hours and pay on a weekly basis, it is possible to assess the accuracy of answers to each type of question. The findings are summarized in Tables 4 through 6. Table 4 shows that the correlation between (the log of) company record earnings for the previous calendar year (1986) and (the log of) the respondents' reports of their earnings for that year is 0.81. This is a high value, compared with others that will be examined, but nevertheless indicates only two-thirds of the variance in the survey reports reflects valid variance. (Note that this corresponds to one minus the ratio of the error to the total variance, the statistic reported in the first column of Table 1.) The correlation between records and survey reports for the pay period immediately preceding the interviews is only 0.46, indicating that less that a quarter of the variance in this measure is valid.[14] The third measure of earnings, "usual" earnings, has about the same level of validity as the reports on the preceding pay period. The correlation between the reports of usual earnings and the average value in the records for the preceding 12 "normal" weeks[15] is 0.46.

---

pay period. A disadvantage is that many workers do not have a constant hourly rate, especially if they work overtime. We were unable to validate responses to such a question for hourly workers in our firm sample.

[14]As noted, the analysis reported in this section is restricted to hourly employees because of the absence of records on hours worked by salaried employees. We did, however, examine the correlations of records and survey reports on annual earnings and earnings for a recent pay period for salaried workers. The correlation of the annual earnings reported by salaried employees and the record value is 0.715, compared to the correlation of 0.806 observed for hourly employees. The correlation of earnings reported for the most recent pay period, and the recorded earnings for the last payperiod in June, is 0.667, compared to the 0.456 observed for hourly employees.

[15]Cases were excluded from this analysis if fewer than 12 of the most recent 22 weeks were "normal" weeks, as defined by having worked at least 30 hours (according to the records) and earned at least $100. That is, weeks during which the respondent did not work close to full time (because of illness or vacation, primarily) are not included in

Table 4 also provides information about the stability of earnings. If the mean value of earnings for the preceding 12 "normal" weeks (labelled "REMN" in the table) is taken as the best indicator of "current" earnings, the correlations of the other two measures abstracted from the records (labelled "RES86" AND "RELST" in the table) provide information about stability of earnings, and in this case indicate a rather low level of stability. The correlation between the current mean and the 1986 level is 0.67, while the correlation between the current mean and earnings in the preceding pay period is only 0.52.

Mean earnings per week over the preceding 12 normal weeks were about 15 percent higher than mean earnings per week during 1986, and about 4 percent higher than earnings in the preceding pay period. These differences in means reflect several sources of variation, including the fact that there are "abnormal" weeks when the individual works few or no hours because of vacation or illness, or on the other hand receives incentive pay, bonuses, or other types of income beyond that based on hours worked during that week. The average correlation between the earnings in each pair of weeks among the 12 "normal" included in the mean is 0.49. From the covariances among these items, the reliability of the mean of these 12 weeks as an indicator of current "average earnings" is estimated to be quite high – 0.92.[16] That is, it is

---

the calculation of average earnings. Cases with average earnings per week across the most recent "normal" weeks of less than $200 were also excluded. As in all analyses reported in this section, only males were included (there were not enough females in the sample to support analysis of sex differences), and only hourly workers were included (because records about hours worked were not available for salaried workers).

[16]This is the alpha coefficient – a measure of reliability developed by psychometricians (see, for example, Nunnally, 1967, p. 196).

appropriate to think that this records-based indicator itself has error. If account is taken of this error, by dividing the observed correlation between the record mean and the reports on usual earnings by the square root of the reliability coefficient, the corrected correlation rises slightly from 0.45 to 0.47.

The rather low week-to-week stability in earnings, as assessed from the company records for these workers, raises two questions. First, to what extent does the instability reflect "abnormal" weeks — weeks with a lot of overtime hours, or receipt of overtime pay, for example? Second, how typical are the earnings patterns observed for the employees in this particular plant of hourly workers in general? Further analysis provides some insights relevant to both of these issues, although direct assessment of the second question is not possible without information from the employment records of other companies.

Examination of the distributions of company-recorded weekly earnings reveals that there are indeed a small proportion of "abnormal" weeks, during which the respondent worked very few or very many hours or received large bonuses or other types of pay beyond those based on hours worked during that particular week. In using the mean of the preceding 12 "normal" weeks to define "true" usual earnings, the previously reported analyses took account of one extreme — weeks during which a respondent worked fewer than 30 hours or received less than $100 were ignored in calculating the average earnings. If such weeks are not ignored, the average between-week correlation is 0.38 instead of 0.49. If, however, weeks at the other extreme (specifically, those in which the respondent worked more than 80 hours, received more than $1,800, or had an hourly pay rate of more than 30 dollars) are also ignored, the between-week correlation does not improve (in fact, it drops slightly, to 0.48).

With respect to how typical these workers are of hourly workers in general, we are not aware of data based on company records of earnings for a broad sample that could be used for comparison with the PSID Validation Study sample, so we cannot address this issue directly. It is possible, however, to ask how sensitive the correlation between reports of usual earnings and the record values is to stability in earnings for individual workers. Across the whole sample, the average value of the variance in the logarithms of weekly earnings (ignoring "abnormal" weeks of both types described in the previous paragraph) is 0.085, with a range from 0.001 to 0.171. If the workers with above-average standard deviations are ignored, the correlation of a worker's record mean with his report of usual earnings does not increase at all. Specifically, for all workers the correlation is 0.445; if those with the highest ten percent of values on the standard deviation are ignored, the correlation is 0.439; if the highest twenty-five percent are ignored, the correlation is 0.438; and if the highest fifty percent are ignored, the correlation is 0.460. Thus, our finding that most of the variation in interview reports of usual earnings is "noise" rather than "news" appears quite insensitive to the variability of record earnings.

Weekly hours worked

Data about agreement between survey reports and company records on hours worked per week are shown in Table 5. Unlike the case for weekly earnings, there is little evidence that any one of the three survey measures is superior to the other two. The correlations between the survey reports and the corresponding records information are all in the range of 0.60 to 0.64. There is considerable week-to-week variation in number of hours worked: the average correlation between hours worked in each of the twelve preceding normal weeks is just 0.42.[17] This variability in hours is reflected in the rather low correlations among the three measures derived from company records. These correlations of variables based on records are, in fact, at best only slightly higher than the correlations among the corresponding survey reports. The answers to the survey question about hours worked in the preceding pay period actually correlate somewhat more highly with the mean record value for recent weeks (r=.65) than do the answers to the question about number of hours usually worked (r=.61).

Unlike the pattern observed for weekly earnings, eliminating respondents with high week-to-week variability in hours worked does at least slightly improve the correlation between reports of usual working hours and the mean of the record hours. Across all respondents (and ignoring both unusually "low" and unusually "high" weeks), the correlation between these two measures is 0.607. Eliminating the ten percent with the greatest variance in weekly hours improves this correlation to 0.639; eliminating the highest fifty percent improves it to 0.664. Part of this improvement is probably best interpreted as reflecting improved reliability of the records-based measure, but the rest of the improvement (if statistically significant, a hypothesis that is not readily testable)

---

[17]This is the correlation if weeks are ignored during which, according to the records, the respondent worked less than 30 hours or earned less than $100. If such weeks are not ignored, the correlation is only 0.27. If weeks at the other extreme (as defined earlier, those during which the respondent worked more than 80 hours, earned more than $1800, or more than $30 per hour worked) are also ignored, however, this does not improve the week-to-week stability (the correlation actually drops slightly, from 0.42 to 0.41).

may reflect greater accuracy on the part of those respondents who have a more stable work pattern.[18]

Hourly wage rates

The respondents were not asked to report their hourly wage rates directly, but these rates were calculated by dividing the various reports on weekly earnings by the corresponding reports on number of hours worked per week. These ratios were also calculated from the records. The correlations between the wage rates as calculated from the respondent reports and those calculated from the records are shown in Table 6. The calendar year-based survey and record measures of wage rates correlate somewhat less closely with one another (at r=.56) than do the corresponding survey and record measures of hours worked (at r=.64) and considerably less than do the corresponding survey and record measures of weekly earnings (at r=.81). The other pairs of measures of wage rates, however, correlate even less with company records: for the reports based on the most recent pay period, the correlation of the survey measure of hourly earnings with the records measure is 0.35, while the hourly earnings measure based on reports of "usual" hours and earnings correlates with the average wage rate for recent weeks at only 0.25. The latter correlation suggests that the proportion of valid variance in a

[18]Correcting the correlation between the two types of measure for the assessed reliability of the records-based measure, the correlation for all respondents is 0.642, while that for respondents with standard deviations on the weekly hours below the median is 0.683.

commonly used indicator of wage rates is only six or seven percent of its total variance.[19]

The apparently extremely low validity of the survey-based measure of hourly wage rate raises the issue of how respondents arrived at their answers to questions about their earnings and work hours. The rather low correlations of these reports with the mean values from company records for recent weeks, and the very low correlation of the ratio of these reports to the corresponding ratio based on the records, suggests that respondents may assess their "usual" earnings and work hours in a manner other than simply taking an average across recent weeks. Several possibilities can be imagined; for example, they may give more weight to recent pay periods than to less recent ones; they may give more weight to "above average" or to "below average" weeks; or they may report a "typical" week. If the latter, the question arises as to whether "typical" is operationalized more closely by the median or the mode, rather than the mean.

Evidence on these issues is shown in Table 7. The correlation between the interview report of usual earnings per pay period with recorded earnings for the most recent pay period is 0.32, but this rises to 0.40 if earnings during the most recent "normal" pay period are considered (that is, substituting for the most recent pay period if the employee happened to work few hours during that two-week period). The correlation rises again, to 0.45, if an average is taken of the recorded earnings for the twelve most recent "normal" weeks. A similar pattern is observed with respect to the measures of hours worked per week, but no such pattern exists for the ratio of these measures (i.e.,

---

[19]Correcting the correlations between the two measures for the assessed reliability of the records-based measures, the correlation for all respondents is 0.266, while that for respondents with standard deviations on hourly wage rates below the median is 0.297.

for the dollars earned per hour worked). The pattern for weekly earnings and hours worked suggests the importance of variability in the actual pay and hours worked by these employees. Other entries in Table 7 indicate that the mean value of recent normal weeks is probably the closest counterpart to the reports by the respondents of their usual pay and hours. For weekly earnings, the correlation of usual pay with the mean of the record values, at 0.45, is somewhat higher than for the median or mode of the record values, while for hours worked the median is correlated at approximately the same level as the mean (0.62) and both correlate somewhat more strongly than does the mode (0.60). The data also indicate that the minimum earnings and hours worked correlate less strongly with the reports on usual levels than do the maximum values, but that both extremes correlate less strongly than do the measures of central values. This pattern is also seen when the weekly values are ranked and means taken of each quartile.

Additional analysis showed little evidence that any more complex combinations of data from records would have a higher correlation with the reports. Very little explanatory power is gained by using earnings (or hours or hourly wage rates) of each individual week, rather than their mean, as predictors in a multiple regression analysis of the survey reports of usual pay. It appears, then, that these respondents arrived at their answers to the questions about usual pay and usual hours by a fairly straightforward process of finding a central value (corresponding most closely to the mean of recent weeks), but did so with considerable error. It also appears to be the case that their answers to the two questions, about usual earnings and usual hours, were arrived at independently rather than, for example, using estimates of their hourly wage rate and of their hours to calculate their weekly earnings. The consequence is that the measures of hourly wage rate, as calculated by dividing the reports of earnings by the reports of hours, contain measurement error from both sources and therefore are (with the exception of the measure based on annual earnings and work hours) only barely associated with the corresponding ratios based on records.[20]

Bias in survey reports

As shown in the first row of Table 8, the average biases in the various survey measures of weekly earnings (where bias is operationalized as the discrepancy between the survey measure and the corresponding measure derived from company records) are generally small. Also shown in this row are the average values of the absolute values of these discrepancies. The lowest bias, and the lowest average measurement error, is for the report of total earnings in 1986, while the largest bias (an underestimation by about six percent – the only bias in Table 8 that differs significantly from zero) is for the reports of usual earnings as compared with the mean recorded earnings for the twelve preceding normal weeks. However, the standard deviation of the errors in the reports for the preceding pay period (shown in the second row of Table 8) is considerably higher than the standard deviation of the errors in reports on usual earnings, so the root mean

---

[20]The low validity of the survey measure of hourly wage rate relative to the validities of the survey measures of weekly earnings and hours is not due to greater instability in the actual hourly wage rate. Across the 12 most recent normal weeks, the average correlation of the recorded weekly earnings is .492; that for the recorded hours worked is .420; and that for the hourly wage rate (derived as the ratio of the preceding values) is .476. It is perhaps surprising that the stability of the hourly wage rate is not higher than that for weekly earnings or for hours. At least in the company used for this study, hourly wage rates depend on the particular job performed at any given time and increase for overtime hours.

square error (shown in the third row) is greater for the reports on the preceding pay period than for the reports on usual earnings.

The lower half of Table 8 shows the correlations of the discrepancy scores with other variables. The first covariate shown is the variability in weekly earnings, which is seen to be associated both with bias and with total error in the reports on usual earnings, but not significantly related to errors in the other two survey measures. More specifically, individuals with greater week-to-week variations in their earnings underestimate their usual earnings to a greater extent than do those with smaller weekly variations. The remaining rows of Table 8 indicate greater underreporting of usual earnings, and larger absolute discrepancies with the records, by those with less education, by older workers, and by those who have worked longer for the company.[21] The correlations of the discrepancies on the other survey measure based on annual earnings generally follow the same pattern, but tend to be weaker and are less likely to be statistically significant. None of the correlations of the discrepancies based on earnings in the previous pay period is statistically significant.

Table 9 shows similar data for the various survey measures of hours worked per week. The relative errors in the reports on usual hours per week are considerably smaller than for reports on usual earnings per pay period, so that the total error (as assessed by the root mean square error) in this measure of hours is only slightly larger than that in the measure of annual hours. Moreover, there is little bias in this measure,

---

[21]Correlations in the simple discrepancy columns are the bivariate counterpart to the regression-based partial correlations shown in columns of Table 3 labelled "$b_{v\tilde{X}}$". Table 3 information is based on annual earnings reports for both salaried and hourly workers.

and the amount of bias is not related significantly to week-to-week variations in hours worked. The pattern of correlations with education, age, and tenure are quite similar to that observed in Table 8.

Errors in the derived measures of hourly wage rates are summarized in Table 10. The measure based on reports of usual hours and usual earnings is underreported, on average, by about six percent, while the hourly wage rate measure based on reports about the preceding pay period has a positive bias of about four percent. Moreover, the total error in the measure based on a usual weekly earnings and hours is actually greater than for the measure based on the preceding pay period, as well as being much greater than for the measure based on the preceding calendar year. Errors in the measure of usual wage rate are not related to weekly variability in the wage rate, nor are there (with one exception) significant correlations with education, age, or tenure.

The data presented in this section lead quite clearly to the conclusion that the most valid measure of earnings, at least among the three evaluated in this study, is one based on the preceding calendar year. This measure is much more strongly related to the records than is a measure either of usual earnings or of earnings in the preceding pay period. Moreover, unlike the reports on usual earnings, the reports on annual earnings are essentially unbiased. For measuring work hours, the choice between a measure based on reports of annual hours and one based on reports of usual hours is less clear cut. The annual reports are somewhat more valid, but the difference is much smaller than for reports on earnings, and there is little bias in either of these survey measures. For measuring hourly wage rates, which is generally the variable of most conceptual interest to labor economists, the choice suggested by the present data is a measure obtained by dividing annual earnings by annual hours. While the validity of this measure is lower than might be wished, it is considerably more valid than measures based on reports of usual weeks or the preceding pay period. Moreover, the annual-based measure is only weakly biased compared with the other two measures.

The evidence displayed here supporting the superiority of annual measures of earnings and wage rates is limited to the amount of measurement error in the various measures. The conclusion that the annual measures have less measurement error than do other measures must be weighed together with considerations about the correspondence of each type of measure to the theoretical concept of interest.

V. The Quality of Retrospective Reports of Unemployment[22]

Event-history models of labor market phenomena such as unemployment typically rely heavily on retrospective information provided in interviews. There have been very few validation-based studies of the quality of such data, and those studies that have been conducted on episodic recall suggest the potential for massive measurement error, the implications of which have been almost universally ignored.

Typical of the findings on the quality of episodic recall in survey settings, although unusually thorough in methodology, is the research conducted by Cannell and his colleagues on the quality of retrospective reports of hospitalizations (Cannell and Fowler, 1963). Overnight stays in hospitals might be expected to be at least as salient as labor market events such as unemployment or job or position changes. When asked to recall

[22]This section draws heavily from work reported in Mathiowetz and Duncan (1988) and Mathiowetz (1985).

hospitalizations within ten weeks of the interview, respondents failed in only three percent of the cases. If the elapsed time increased to one year, however, the failure rate increased to well over 25 percent. Not only length of recall period, but also interviewer behavior, question wording, the social desirability of the response, the salience of the event and the number of related events have been linked to a respondent's ability to report accurately (e.g., Bradburn, Sudman and Associates, 1979; Cannell, Fisher and Baker, 1965; Jabine, Straf, Tanur and Tourangeau, 1984; Lansing, Ginsburg and Braaten, 1961; Tulving and Thompson, 1973).

The 1983 wave of the PSID validation study gathered retrospective reports of unemployment episodes that had occurred between January 1, 1981 and the date of the interview.[23] Detailed employee records covering the same period provided precise information on periods of time when an individual was not working for the given company. (Four respondents who reported employment with other firms were eliminated from the analysis, since it was impossible to validate their secondary employment.)

---

[23]The actual questions were as follows: "Were there any periods since the beginning of the year before last, January, 1981, when you were unemployed and looking for work or temporarily laid off for a week or more?" "What months(s) and year(s) (was that/were those)?" "Any other such periods?" "Were there any periods since the beginning of the year before last, January, 1981, when you were completely out of the labor force, that is, neither unemployed nor temporarily laid off nor looking for work for a week or more?" Since these questions followed a sequence of questions that asked the respondent to account for weeks of work, vacation, sick time, and other reasons for nonemployment during calendar years 1982 and 1981, they might be expected to stimulate recall of the nonemployment episodes. On the other hand, there was no attempt to ask the episodic unemployment questions in the context of an event calendar that associated employment history with other domains of life events. Such calendars appear to improve the accuracy of reporting of event-history information (Freedman et al. 1988). Similar questions were incorporated into the 1987 questionnaire, although there was so little unemployment during the 1985–87 period that it was impossible to replicate the first wave analysis on the second wave sample.

There was no attempt to distinguish between the states of "unemployment" and "out of the labor force", since validation of the distinction was impossible. Technically, then, the validation check was against episodic reports of nonemployment from the given company.

Company records gave the precise dates of the beginnings and endings of all spells of nonemployment. The information obtained in the interview was less precise, dating nonemployment to the month in which it occurred. A case-by-case examination of records and interviews, incorporating a rather generous allowance for what constituted a "correct" interview report, produced data on which unemployment spells appearing in the company record were accurately recalled by the respondent.[24]

Table 11 shows the performance of respondents in reporting unemployment spells of various lengths and at various times since the previous interview. It is obvious that respondents have great difficulty in recalling unemployment spells, especially short and distant ones. Only one-third of all unemployment spells that appeared in the company records were reported in the interview.[25] Even very long spells (more than 29 weeks) were seriously underreported; the fraction not reported was more than one-third. Similarly, although spells occurring close to the interview were recalled more accurately, more than half of such spells were not reported in the interview.

---

[24]If company records showed that a respondent was unemployed in a given month and if that respondent reported any unemployment in either that or an adjacent month, then the given unemployment spells was considered to be accurately reported in the interview.

[25]Underreporting of unemployment spells was far more prevalent than overreporting. There were some 45 unemployment spells reported in the interviews that did not occur within one month of spells appearing in company records, as compared with 321 company record spells that were not reported in the interviews.

Mathiowetz and Duncan (1988) take a closer look at the nature and correlates of unemployment reporting error. Interestingly, they find that while unemployment spells elicited through episodic recall are seriously underreported, estimates of calendar year amounts of unemployment in the two calendar years prior to the 1983 interview appeared relatively unbiased. They speculate that episodic recall and estimation place different demands on memory and that respondents might be able to provide reasonably unbiased estimates of total amounts of time spent in given states without being able to recall the precise timing of the episodes.

Mathiowetz and Duncan (1988) also estimate a model of response error in which the probability that employment status in a given month is reported erroneously is related to a set of demographic factors and measures of the likely salience of the events of the given month. Consistent with some past research, they find a number of simple associations between reporting error and demographic measures, with younger and less educated workers more likely to provide erroneous reports of employment status in a given month. Measures of the likely salience of employment status in a given month (e.g., the total length of the spell of unemployment in which a given month's unemployment was imbedded) were also found to be important predictors of reporting error that account for the simple associations between demographic factors and reporting error.[26]

---

[26]Duncan and Mathiowetz (1985) conduct an analogous investigation of response error in retrospective reports of positions held in the company between January 1, 1981 and the 1983 interview. They find that the chance that a given month's occupation is misreported rises from about 12 percent at the time of the interview to over 20 percent for more distant months. (The 12 percent error at the time of the interview appeared to be accounted for largely by hourly workers who had been demoted to lower-status positions and continued to report holding their previous higher-status positions).

It is difficult to draw clear implications from this evidence for the estimation of event-history models that rely on retrospective data. Since the spells of nonemployment from the company are entirely a function of company policy rather than worker behavior, one cannot use interview and validation record data to compare estimates of "error-ridden" and "true" unemployment duration models as was done for earnings functions in Table 3. At this point we can only note that the ingredients for grave concern about the quality of parameter estimates are certainly present: massive underreporting of spells and correlations between reporting error and demographic measures typically included in such models. Clearly more work is needed in this area – on the econometrics of the effects of measurement error on estimates from duration models, on the validity of actual reports of retrospective data and on ways in which the quality of survey-based event-history data can be improved.

VI. Conclusions

The two validation data sets used in this paper produce a number of facts that contradict assumptions made in and implications drawn from traditional measurement error models. Since the validation data sets themselves have features that might limit their relevance for other situations, we order our discussion of implications according to the confidence we have in their generality.

Both data sets showed that annual earnings are fairly reliably reported. The tendency for workers with lower-than-average earnings to overreport and high-wage

---

Estimates of a reporting error model showed associations with demographic factors and measures of salience that were quite similar to those found in models of reporting error in unemployment status.

workers to underreport their earnings — a covariance almost always assumed to be zero in measurement error models — increased the reliability of annual earnings reports considerably. The implied biases due to errors in measuring earnings when earnings is a right-hand independent variable ranged from 18 to 24 percent. Mean-reverting error also produced biases to right-hand side variable coefficients when annual earnings is a dependent variable that ranged from 10 to 17 percent. The restricted variability of true earnings from the single-company sample probably leads to an overstatement of these biases.

Furthermore, each data set also showed a surprisingly small decrement to reliability when going from cross-sectional measures of earnings level to panel measures of annual earnings change —there was more "news" than "noise" when earnings were differenced over either one- or four-year intervals. Reliability was also fairly high in panel reports of change in annual work hours. Indeed, apparently turbulent employment conditions produced cross-sectional reports of earnings and hours in one of the survey waves that were less reliable than the corresponding change measures.

Covariance between earnings error and right-hand side measures such as education, age and job tenure also appeared in both validation data sets. These covariances are also typically assumed to be zero in measurement error models, helping to produce the conventional wisdom that error can only bias right-hand side coefficients toward zero. However, depending on the pattern of covariances between the error in measuring an independent variable and true levels of independent variables, measurement error can readily lead to either downward or upward biases in right-hand side variable coefficients.

We found that the size and statistical significance of these covariances varied across data sets and within waves of each data set; none was consistently large. In one wave of our data a positive covariance between earnings error and schooling produced an upward bias in the payoff to schooling when a cross-sectional earnings regression was estimated with interview data, while a negative covariance between earnings error and job tenure produced a downward bias on the tenure coefficient. However, variability in the estimated pattern of these covariances leaves us unable to assert with confidence what these covariances and likely biases will be in general population data sets.

As shown by other papers at this conference, longitudinal studies of labor market phenomena are increasingly turning to event-history models and data. In the company sample we were able to validate reports over a two-and-a-half year period of spells of nonemployment from the firm as well as changes in positions held within the company. We concentrated on quality of retrospective reports of unemployment and found that only one-third of the spells of nonemployment appearing in company records were reported in the interviews. Shorter and more distant spells were more likely to be unreported, although the fraction of presumably salient longer and more recent spells unreported still exceeded one-third. Furthermore, the incidence of reporting error appeared to be correlated with typical right-hand measures such as age and schooling. Thus, all of the ingredients for coefficient bias due to measurement errors would appear to be present in unemployment event-history data. But despite longstanding evidence that reports of episodic events are quite faulty, there seems to have been virtually no attention paid in the econometrics literature to the possible biases caused by measurement error in retrospective event-history reports.

Perhaps our most surprising — and tenuous — findings came for the variable most dear to the hearts of labor economists: hourly earnings. Here we find that all of the cross-sectional reporting of hourly earnings measures we could validate appeared to be quite unreliable. In this case the characteristics of our validation study sample lead us to be rather cautious. Reports of hourly earnings could be validated only for hourly workers and the pay period-to-pay period variation in work hours and earnings in company records were larger than most readers would anticipate. Whether the conventional wisdom is in need of revision here too is less clear.

At any rate, we found that only about one-quarter of the variation in a cross-sectional hourly earnings measure obtained by dividing interview reports of annual earnings by annual work hours was valid, while only about one-tenth of the variance of hourly earnings measures based on "usual" or last pay period hours and earnings was valid. The implied bias in using any of the hourly earnings measures as right-hand side variables is very large. Measurement properties of the one measure of change in hourly earnings available to us (four-year change in the ratio of annual earnings to annual hours) showed it to be even less reliable than its cross-sectional counterpart.

Taken together, the results from the two validation studies show a clear need to recognize the potential importance of measurement error and incorporate more realistic assumptions about the properties of measurement error into measurement error models. Mean-reverting negative covariances between the error and true level of a given measure were pervasive in our data. Covariances between error and the true levels of other measures of interest were also widespread although not as consistent across waves and between the two data sets. Positive autocorrelation in measurement errors was also apparent in our panel data, constituting a third type of error covariance that need to be built into error models of panel data.

Building realistic (i.e., nonzero) assumptions about these covariances into measurement error models will complicate the conventional wisdom regarding measurement errors, but also force researchers to consider whether conventional assumptions are warranted in their models. A useful by-product of these considerations is the recognition of the need for direct measurement of the covariances through validation studies.

The case for additional survey-based validation studies of labor market panel data is compelling. Here the importance of the quality of the validation data leads us to recommend that additional firm samples be drawn. Such samples will shed light on the question of how representative the employment practices of the firm that cooperated in providing data for our study are as well as opportunities for testing new methods for motivating respondents to provide high quality survey information. The disturbingly high error in retrospective reports of employment events and cross-sectional reports of hourly earnings makes these topics the highest priority for future studies.

Table 1

Summary Statistics for Earnings Errors for Salaried and Hourly Workers

Data Source	Earnings Variable	$\frac{\sigma_u^2}{\sigma_X^2+\sigma_u^2}$	$b_{u\tilde{X}}$	b_{vY}	$r_{u_t,u_{t-1}}$	$r_{X_t,X_{t-1}}$
PSID Validation Study	ln (1986 Earnings) N=422	.303	.239 (.028)	−.172 (.031)	—	—
	ln (1982 Earnings) N=320	.150	.076 (.024)	−.104 (.023)	—	—
	4-yr. Δ ln Earnings N=206	.294	.213 (.043)	−.214 (.043)	.073	.452
CPS-SSA Data	ln (1977 Earnings) N=1575	.221	.158 (.014)	−.138 (.014)	—	—
	ln (1976 Earnings) N=1575	.210	.108 (.015)	−.190 (.013)	—	—
	1-year Δ ln Earnings N=1575	.322	.231 (.017)	−.238 (.017)	.372	.635

Note: Standard errors are given in parentheses beneath coefficients.

Table 2

Errors in Earnings, Hourly Earnings, and Hours for Hourly Workers

Earnings Variable	$\frac{\sigma_u^2}{\sigma_X^2 + \sigma_u^2}$	$b_{u\tilde{X}}$	b_{vY}	$r_{u_t,u_{t-1}}$	$r_{X_t,X_{t-1}}$
ln (1986 Earnings) N=277	.276	.190 (.035)	−.192 (.035)	—	—
ln (1982 Earnings) N=141	.241	.120 (.049)	−.208 (.044)	—	—
4-yr. Δ ln Earnings N=86	.260	.217 (.056)	−.104 (.064)	.146	.330
ln (1986 Earnings/Hour) N=277	.670	.667 (.028)	.021 (.086)	—	—
ln (1982 Earnings/Hour) N=141	.693	.737 (.047)	−.301 (.125)	—	—
4-yr. Δ ln (Earnings/Hour) N=86	.821	.870 (.047)	−.369 (.230)	.084	.542
ln (1986 Hours) N=277	.366	.306 (.040)	−.244 (.043)	—	—
ln (1982 Hours) N=141	.284	.184 (.053)	−.226 (.050)	—	—
4-yr. Δ ln (Hours) N=86	.312	.267 (.061)	−.140 (.072)	.050	.126

Source: PSID Validation Study

Note: Standard errors are given in parentheses beneath coefficients.

Table 3

Impact of Errors on Cross-Section Earnings Function for Salaried and Hourly Workers

Independent Variable	b_I	b_R	$b_{u_{Tenure}\tilde{X}}$	$b_{v\tilde{X}}$	Discrepancy
1986 Earnings N=417					
Education	.025 (.006)	.018 (.006)	−.005 (.020)	.008 (.003)	−.00012
Pre-Company Experience	−.008 (.003)	−.007 (.003)	−.005 (.009)	−.001 (.002)	−.00021
Tenure	.003 (.001)	.004 (.001)	.002 (.005)	−.001 (.001)	−.00002
1982 Earnings N=317					
Education	.035 (.008)	.038 (.008)	.032 (.026)	−.003 (.004)	−.00032
Pre-Company Experience	.004 (.003)	.005 (.003)	−.000 (.011)	−.001 (.001)	−.00018
Tenure	.011 (.002)	.014 (.002)	−.007 (.006)	−.004 (.001)	.00048

Source: PSID Validation Study

Note: Standard errors are given in parenthesis beneath coefficients.

Table 4

Correlations of Interview Reports and Records Data about Weekly Earnings for Hourly Workers

	RE86	IE86	RELST	IELST	REMN	IEUSU
RE86	1.000					
IE86	0.806	1.000				
RELST	0.299	0.270	1.000			
IELST	0.368	0.384	0.456	1.000		
REMN	0.672	0.616	0.523	0.549	1.000	
IEUSU	0.299	0.389	0.321	0.576	0.461	1.000
Mean	6.485	6.469	6.578	6.539	6.624	6.437
St.Dev.	0.181	0.194	0.350	0.310	0.221	0.240

Source: PSID Validation Study

Note: Variables are defined as follows (all measures are logarithmic transformations of earnings in dollars):

RE86: Records data on total earnings in 1986, divided by 52.
IE86: Survey report on total earnings in 1986, divided by 52.
RELST: Records data on earnings in last two week pay period, divided by 2.
IELST: Survey report on earnings in last two week pay period, divided by 2.
REMN: Records data on earnings in preceding twelve "normal" weeks, divided by 12.
IEUSU: Survey report on usual earnings per pay period, divided by 2.

Table 5

Correlations of Interview Reports and Records Data about Weekly Hours for Hourly Workers

	RHR86	IHR86	RHRLST	IHRLST	RHRMN	IHRUSU
RHR86	1.000					
IHR86	0.640	1.000				
RHRLST	0.374	0.237	1.000			
IHRLST	0.313	0.374	0.603	1.000		
RHRMN	0.539	0.380	0.682	0.649	1.000	
IHRUSU	0.364	0.462	0.390	0.610	0.613	1.000
Mean	3.668	3.683	3.840	3.865	3.844	3.813
St.Dev.	0.136	0.150	0.194	0.189	0.138	0.147

Source: PSID Validation Study

Note: Variables are defined as follows (all measures are logarithmic transformations of hours):

RHR86: Records data on total hours worked in 1986, divided by 52.
IHR86: Survey report on total hours worked in 1986, divided by 52.
RHRLST: Records data on hours worked in last two week pay period, divided by 2.
IHRLST: Survey report on hours worked per week in last two week pay period.
RHRMN: Records data on hours worked in preceding twelve "normal" weeks, divided by 12.
IHRUSU: Survey report on usual hours per week.

Table 6

Correlations of Interview Reports and Records Data about Hourly Wage Rate for Hourly Workers

	RWG86	IWG86	RWGLST	IWGLST	RWGMN	IWGUS
RWG86	1.000					
IWG86	0.558	1.000				
RWGLST	0.610	0.310	1.000			
IWGLST	0.261	0.230	0.351	1.000		
RWGMN	0.758	0.420	0.817	0.334	1.000	
IWGUS	0.213	0.212	0.288	0.483	0.254	1.000
Mean	2.812	2.775	2.764	2.679	2.766	2.624
St.Dev.	0.089	0.154	0.117	0.212	0.104	0.216

Source: PSID Validation Study

Note: Variables are defined as follows (all measures are logarithmic transformations of earnings in dollars per hour):

RWG86: Records data on total earnings in 1986, divided by total hours worked in 1986.
IWG86: Survey report on total earnings in 1986, divided by hours worked in 1986.
RWGLST: Records data on earnings in last two week pay period, divided by hours worked in that pay period.
IWGLST: Survey report on hours worked per week in last two week pay period.
RWGMN: Records data on earnings in preceding twelve "normal" weeks, divided by hours worked in those pay periods.
IWGUS: Survey report on usual earnings per pay period, divided by twice the number of usual hours per week.

Table 7

Correlations of Interview Reports on Usual Earnings and Hours for Hourly Workers with Summary Measures from Records Data

	Pay/week	Hours/week	Dollars/hour
Most recent pay period	.321	.380	.258
Most recent "normal" pay period	.395	.452	.242
Across given worker's 12 most recent "normal" weeks:			
Mean	.449	.616	.227
Median	.434	.620	.205
Mode	.397	.595	.196
Minimum	.373	.356	.198
Maximum	.392	.482	.209
Quartiles of given worker's 12 most recent "normal" weeks			
Lowest	.428	.484	.219
2nd	.433	.614	.215
3rd	.428	.617	.201
Highest	.411	.549	.233
Number of cases	275	287	275

Source: PSID Validation Study

Table 8

Average Magnitude and Bias in Survey Reports of Weekly Earnings for Hourly Workers

	1986 Totals		Usual		Last Pay Period	
	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.
Mean	−.006	.033	−.066*	.096	−.019	.075
Standard Deviation	.051	.039	.109	.084	.156	.138
Root mean square error	.051		.127		.157	
Correlation with:						
Record variability	.037	.055	−.187*	.176*	−.076	.085
Education	.081	−.161*	.178*	−.144*	.112	.029
Age	−.184*	.098	−.151*	.109	.068	−.046
Tenure	−.165*	.087	−.197*	.180*	.035	−.046

Notes:

All reports and record entries have been transformed by taking logarithms.

The discrepancies labelled "1986 totals" are defined as the difference between the annual measures (survey and records), after dividing both by 52.

The discrepancies labelled "usual" are defined as the difference between the respondent's report of usual earnings (divided by 2) and the mean of the records for the 12 most recent "normal" weeks.

The discrepancies labelled "last pay period" are defined as the difference between the respondent's report and the record of earnings for the most recent two-week pay period (both divided by 2).

"Record variability" is defined as the standard deviation of the records data (for each individual) for the 12 most recent "normal" weeks.

* Discrepancy or correlation differs significantly from zero ($p < .05$). (Such tests are inappropriate for the absolute discrepancies.)

Table 9

Average Magnitude and Bias in Survey Reports on Hours Worked per Week for Hourly Workers

	1986 Totals		Usual		Last Pay Period	
	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.
Mean	.005	.033	−.009*	.042	.011*	.042
Standard Deviation	.053	.042	.056	.039	.072	.060
Root mean square error	.053		.057		.073	
Correlation with:						
Record variability	.032	.089	−.051	.180	−.021	.094
Education	.131*	−.065	.207*	−.114	.100	.068
Age	−.135*	−.069	−.199*	.086	−.019	−.045
Tenure	−.099	−.051	−.243*	.121*	−.021	−.020

Source: PSID Validation Study

Notes: See notes to Table 8.

Table 10

Average Magnitude and Bias in Survey Reports on
Hourly Wage Rate for Hourly Workers

	1986 Totals		Usual		Last Pay Period	
	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.	Discrp.	ABS Discrp.
Mean	−.013*	.039	−.059*	.086	.038*	.061
Standard Deviation	.056	.043	.101	.079	.089	.076
Root mean square error	.057		.117		.097	
Correlation with:						
Record variability	.024	−.002	.036	−.057	−.165	−.081
Education	−.029	−.163*	.081	−.039	−.040	.031
Age	−.041	.064	−.057	.037	−.071	−.105
Tenure	−.069	.090	−.080	.090	−.030	−.059

Source: PSID Validation Study

Notes: See notes to Table 8.

Table 11

Fraction of Actual Unemployment Spells Reported in Interview by Length of Spell and Recall Period

	Percent of Spells Reported in Interview	Number of Spells
Length of Spell in Weeks		
1	25%	243
2	34	117
3–4	39	31
5–12	43	14
13–20	56	34
21–28	51	23
29 or more	63	19
Length of Recall Period in Months		
8 or less	49%	47
9–12	44	169
13–18	26	131
19 or more	25	140
Total	34%	487

Source: Calculated from Mathiowetz (1985), Table 1, based on data from the PSID Validation study.

Appendix

The validation study providing data for this paper gathered interview and company record data from a single large manufacturing firm with several thousand employees. The firm's hourly work force is completely unionized and virtually all workers, both hourly and salaried, work full time. At the time of the initial interviewing in the summer of 1983, the company work force was considerably older (and with more job tenure) than was true of a national sample of workers, in part resulting from layoffs and relatively few new hires in the two years prior to the initial interview. These deviations were offset by a sampling procedure that stratified the employee list by age and type of worker (hourly vs. salaried) and selected a larger proportion of younger and salaried workers.

The resulting 1983 sample was evenly divided between salaried and hourly workers and had a fairly uniform age distribution. Interviews were conducted by telephone with a questionnaire similar to that used in the Panel Study of Income Dynamics, with 418, or 78.3 percent of the 534 potential respondents (or proxies) completing interviews. A more detailed explanation of study procedures involved in the initial round of interviewing is given in Duncan and Mathiowetz (1985).

A second round of interviewing was conducted in the summer of 1987 with the 1983 respondents and a fresh sample of hourly workers. Some 122 of the individuals with whom interviews were attempted in 1983 had retired, left the company for other reasons, had unlisted telephone numbers or for some other reason were not appropriate potential 1987 respondents. Reinterviews were successfully conducted with 341, or 82.4 percent of the remaining 1983 sample; 275 individuals were respondents in both 1983 and 1987. Given the much richer company record information available for hourly workers, it was decided that additional interviews should be conducted only on hourly workers. An additional random sample of 202 hourly workers was drawn and interviews were successfully conducted with 151, or 75 percent of them. Thus, the 1987 data collection had a total of 492 interviews with a total response rate of 79.9 percent.

A comparison of the interview reports of earnings and work hours from the 1983 and 1987 validation study samples to data from the Panel Study of Income Dynamics and Current Population Surveys is given in Appendix Table 1. The distributions of annual and hourly earnings for the validation study sample have considerably higher means and much lower variance than for the national samples, even when those national samples are restricted to unionized hourly workers in durable manufacturing industries and the validation sample is restricted to hourly workers. The reduced variance has implications for the generalizability of findings about the extent of measurement error found for the validation study sample.

Validation sources were company payroll records, company "activity records" showing for hourly workers a daily accounting of work or unemployment, and general company policies on various fringe benefits. Our analysis of the interview data is restricted to information for which there was a very close match between the information sought in the questionnaire and the information available in the company records and for which validating information was judged likely to be highly accurate. Interview and company records were compared to identify cases with substantial apparent reporting error. Both interview and record information in all such cases were rechecked to ensure that differences were not the result of coding errors.

Appendix Table 1.
Means and Standard Deviations of Interview Reports of Earnings and Work Hours in the PSID Validation Study, PSID and Current Population Survey

	Salary and Hourly					Hourly Only		
		Panel Study of Income Dynamics		Current Population Survey			Panel Study of Income Dynamics	Current Population Survey
	PSID Validation Study	All Industries	Durable Manu-facturing Only	All Industries	Durable Manu-facturing only	PSID Validation Study	Unionized Workers in Durable Manufacturing Industries	Unionized Workers in Durable Manufacturing Industries
<u>ln Annual Earnings</u>								
1987 report of 1986	10.50 (.224)	-	-	-	-	10.44 (.193)	-	
1885 report of 1984	-	9.94 (.715)	10.10 (.579)	-	-	-	10.15 (.339)	-
1983 report of 1982	10.30 (.284)	9.82 (.728)	9.97 (.627)	-	-	10.20 (.256)	9.87 (.451)	-
<u>ln "Usual" Weekly Earnings</u>								
1987 Report	6.55 (.298)	-	-	6.04 (.589)	6.17 (.461)	6.44 (.256)	-	6.15 (.300)
<u>ln Hourly Earnings</u>								
1983 report of straight time hourly wage rate	-	-	-	-	-	2.44 (.170)	2.28 (.262)	2.25 (.300)
1987 report of 1986 annual earnings/annual hours	2.83 (.188)	-	-	-	-	2.79 (.161)	-	-
1985 report of 1984 annual earnings/annual hours	-	2.33 (.625)	2.46 (.498)	-	-	-	2.50 (.318)	-
1983 Report of 1982 annual earnings/annual hours	2.78 (.197)	2.24 (.631)	2.40 (.534)	-	-	2.77 (.208)	2.37 (.342)	-
<u>ln Annual Hours</u>								
1987 report of 1986	7.65 (.148)	-	-	-	-	7.63 (.152)	-	-
1985 report of 1984	-	7.61 (.296)	7.64 (.254)	-	-	-	7.65 (.184)	-
1983 report of 1982	7.54 (.217)	7.58 (.313)	7.56 (.247)	-	-	7.43 (.223)	7.49 (.251)	-

NOTE: All samples used for annual measures are restricted to male respondents who work 520 - 3500 annual hours in the given calendar year, had more than $1000 in earnings and had unimputed data. The CPS and PSID samples are restricted to males age 25-65, and the PSID is further restricted to household heads.

Matched CPS-Social Security data come from the 1978 CPS-SER Exact Match File, which was created jointly by the U.S. Census Bureau and Social Security Administration. The Exact Match file matches respondents to the March, 1978 Current Population Survey with Social Security Administration earnings histories from 1950 to 1978. CPS earnings questions and Social Security earnings records closely approximate the same concept. As described in Bound and Krueger (1988), cases on this file were then matched to the March, 1977 Current Population Survey Demographic File to obtain the subset of individuals who responded to both the March 1977 and March 1978 CPS and for whom 1976 and 1977 Social Security earnings information was available.

The file was further restricted to private, nonagricultural workers with positive Social Security earnings in both years, positive, non-imputed CPS earnings in both years and who reported neither self-employment income nor income from tips in either year and who reported that their longest job held during the year was in a covered occupation and industry. The latter restrictions should ensure that the earnings concept used in the Social Security records match with the earnings concept sought in the CPS interview, although some residual matching error no doubt remains. These various matching and sample restrictions resulted in a sample of 2924 males and 465 females of a total of 27,485 potentially matchable households, although an examination of the characteristics of the remaining cases, described in Bound and Krueger (1988), suggests that their mean ages, schooling, weeks worked and self-response rate are similar to the full 1978 March CPS.

A prominent problem with the Social Security earnings information is that they are censored at the maximum taxable amount — $15,300 in 1976 and $16,500 in 1977.

Although only about five percent of the women had earnings at or above the maximum, nearly half of the men had censored earnings. This paper restricts its analysis to the nontruncated sample; Bound and Krueger (1988) use maximum likelihood procedures on the whole sample and obtain similar results.

References

Altonji, Joseph G. (1986). "Intertemporal substitution in labor supply: Evidence from Micro Data". Journal of Political Economy, 94, No. 3, Part 2, S176-S215.

Bound, John and Krueger, Alan B. (1988). "The exhort of measurement error in longitudinal earnings data: Do two wrongs make a right?" Working Paper. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Bradburn, Norman, Sudman, Seymore, and Associates (1979). Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.

Cannell, Charles C. and Fowler, Floyd J. (1963). A Study of the Reporting Visits to Doctors in the National Health Survey. Research Report. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Cannell, Charles, Fisher, Gordon, and Bakker, Thomas (1965). "Reporting of Hospitalization in the Health Interview Survey." (Vital and Health Statistics. Ser. 2, Report 6). Washington, DC: Public Health Service.

Cannell, Charles F., Miller, Peter V. and Oksenberg, Lois (1981). "Research on Interviewing Techniques." In Samuel Leinhardt (ed.), Sociological Methodology 1981. San Francisco: Jossey-Bass.

Duncan, Greg J. and Hill, Daniel H. (1985). "An investigation of the extent and consequences of measurement error in labor economic survey data." Journal of Labor Economics, 3, No. 4, 508–522.

Duncan, Greg J. and Mathiowetz, Nancy A. (1985). A Validation Study of Economic Survey Data, monograph. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Freedman, D., Thornton, A., Camburn, D., Alwin, D. and Young-Demarco, L. (1988). "The Life History Calendar: A technique for collecting retrospective data." In N. Tuma (ed.), Sociological Methodology. San Francisco:Jossey Bass.

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

Griliches, Z. and Hausman, J.A. (1986). "Errors in Variables in Panel Data", Journal of Econometrics 31, 93–118.

Jabine, Thomas, Straf, Miron, Tanur, Judith , and Tourangeau, Roger (eds.) (1984). Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington DC: National Academy Press.

Kmenta, J. (1986). Elements of Econometrics 2nd ed. New York: MacMillan.

Lansing, John B., Ginsburg, Gerald P. and Braaten, Kaisa (1961). "An Investigation of Response Error" ("Studies in Consumer Savings," No. 2), University of Illinois, Bureau of Economic and Business Research.

Mathiowetz, Nancy A. and Duncan, Greg J. (1988). "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment," Journal of Business and Economic Statics, 6, 2. 221–229.

Mathiowetz, Nancy A. (1985). "The Problem of Omissions and Telecopying Error: New Evidence from a Study of Unemployment." Proceedings of the Social Statistics Sector, American Statistical Association.

Mellow, Wesley and Sider, Hal (1983). "Accuracy of response in labor market surveys: Evidence and implications." Journal of Labor Economics, 1, 331-344.

Nunnally, J.C. (1967). Psychometric Theory. New York:McGraw-Hill.

Pindyck, R.S. and Rubinfeld, D.L. (1981). Econometric Models and Economic Forecasts 2nd ed. New York: McGraw-Hill.

Rodgers, W.L. and Herzog, A.R. (1988). "Covariances of measurement errors in survey responses." Journal of Official Statistics Vol. 3, No. 4, 403–418.

Tulving, Endel, and Thomson, D. (1973). "Encoding Specificity and Retrieval Processes in Episodic Memory." Psychological Review, 80, 352–372.

</ref_section>