

NBER WORKING PAPER SERIES

LONGITUDINAL ANALYSES OF
THE EFFECTS OF TRADE UNIONS

Richard B. Freeman

Working Paper No. 1207

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1983

The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

Longitudinal Analyses
of the Effects of Trade Unions

ABSTRACT

This paper examines how measurement error biases longitudinal estimates of union effects. It develops numerical examples, statistical models, and econometric estimates which indicate that measurement error is a major problem in longitudinal data sets, so that longitudinal analyses do not provide the research panacea for determining the effects of unionism (or other economic forces) some have suggested. There are three major findings:

1) The difference between the cross-section and longitudinal estimates is attributable in large part to random error in the measurement of who changes union status. Given modest errors of measurement, of the magnitudes observed, and a moderate proportion of workers changing union status, also of the magnitudes observed, measurement error biases downward estimated effects of unions by substantial amounts.

2) Longitudinal analysis of the effects of unionism on nonwage and wage outcomes tends to confirm the significant impact of unionism found in cross-section studies, with the longitudinal estimates of both nonwage and wage outcomes lower in the longitudinal analysis than in the cross-section analysis of the same data set.

3) The likely upward bias of cross-section estimates of the effect of unions and the likely downward bias of longitudinal estimates suggests that, under reasonable conditions, the two sets of estimates bound the "true" union impact posited in standard models of what unions do.

Richard B. Freeman
National Bureau of Economic
Research
1050 Massachusetts Avenue
Cambridge, MA 02138
(617) 868-3900

Longitudinal Analyses of the Effects of Trade Unions

But union members are different from nonmembers in unobserved ways, biasing your estimates. You should...make a selectivity bias correction...simultaneously determine union status and economic outcomes...develop an unobservables model...USE LONGITUDINAL DATA.

archetypical comment on virtually any study of the economic effects of unionism, or suitably modified, on any other empirical subject

Longitudinal data, which follow the same worker over time, offer researchers a potentially valuable way to examine often-raised objections to the findings of cross-section studies. Unlike complex "structural model" approaches to cross-section data problems, which often yield unstable and uninformative results (Freeman and Medoff 1981), longitudinal data offer a distinctively different "experiment" for uncovering the effects of changes in economic variables. In the case of unions, what is a more natural way to study what unions do than to compare economic outcomes for workers (firms) before and after they change union status?

This paper presents a critical analysis of the "natural experiment." In contrast to the archetypical comment cited above, it argues that longitudinal analyses do not provide a research panacea for determining the effects of unionism (or other economic forces). The main reason for this is the substantial impact of measurement or misclassification error of the union (other economic) variable on longitudinal work.

The paper is divided into four sections. Section I develops briefly the statistical models used in this (and other) longitudinal investigations of what unions do. Section II examines the effect of measurement error in union status

on estimated effects of unionism in cross-section and longitudinal studies. Section III presents the results of estimating the effect of unionism on outcomes in four longitudinal and cross-section data sets. In contrast to other empirical analyses using longitudinal data, it treats two market outcomes which are at the center of the "voice-response" face of unionism, dispersion of wages and provision of fringe benefits, as well as wages. Section IV considers the argument that cross-section and longitudinal estimates of union effects "bound" the true impact of unionism.

There are three basic findings:

1) The difference between the cross-section and longitudinal estimates is attributable in large part to random error in the measurement of who changes union status. Given modest errors of measurement, of the magnitudes observed, and a moderate proportion of workers changing union status, also of the magnitudes observed, measurement error biases downward estimated effects of unions by substantial amounts.

2) Longitudinal analysis of the effects of unionism on nonwage and wage outcomes tends to confirm the significant impact of unionism found in cross-section studies, with the longitudinal estimates of both nonwage and wage outcomes lower in the longitudinal analysis than in the cross-section analysis of the same data set.

3) The likely upward bias of cross-section estimates of the effect of unions and the likely downward bias of longitudinal estimates suggests that, under reasonable conditions, the two sets of estimates bound the "true" union impact posited in standard models of what unions do.

All told, the paper concludes that because of measurement error and likely selectivity of who changes union status longitudinal analysis is a useful tool

for "checking on" the result the cross-section studies but may very well yield worse estimates of the parameters of interest.

I. Longitudinal Models of What Unions Do

The standard cross-section analysis of the impact of collective bargaining on the economic outcome or behavior of individual workers (or firms) involves a multivariate statistical equation of the form

$$O_i = a + bU_i + cX_i + u_i \quad (1)$$

where O_i = outcome for person i ,

U_i = dichotomous unionization variable (1 = covered, 0 = not covered),

X_i = control variables (education, sex) assumed constant over time, and

u_i = error term.

The recurrent objection to estimates based on (1) is that because of selectivity of union workers u_i is likely to be positively correlated with U_i , leading to an overstatement of the union effect. Since, as Abowd and Farber (1982) have stressed, who gets a union job results from the decisions of both employers and workers, the selectivity argument depends on whose decision dominates the hiring process. In the case of wages it is generally assumed that, given high union wages, firms select more able workers from the queue facing them, producing $E(u_i U_i) > 0$. In the case of nonwage outcomes, it is often claimed that workers sort themselves in such a way that those who have strong desires for union-type work conditions and modes of compensation (and would thus obtain more of those outcomes in nonunion settings than the randomly chosen worker) choose union jobs. In this case firms either are indifferent or prefer those workers as well (since they will be more satisfied).

Longitudinal data provides a way to deal with the correlation between

unionism and the error term. Assuming that the part of u_i that is correlated with U_i is an individual effect constant over time, so that $u_{it} = a_i + \epsilon_{it}$ with $E(\epsilon_{it} U_{it}) = 0$, addition of individual constants (which can be viewed as a form of differencing) will eliminate the correlation between u_{it} and U_{it} . In a two-period linear model one obtains

$$\Delta O_{it} = b\Delta U_{it} + \Delta\epsilon_{it}, \quad (2)$$

where ΔU_{it} takes the values -1, 0, 1. A multivariate analysis of (2) will yield the desired b as long as the change in union status is properly measured and is uncorrelated with the change in the random part of the error term.

Equation (2) can be readily generalized to exploit more fully the longitudinal data by allowing different changes in union status to have different effects on wages. In particular, we can allow changes in outcomes to differ among workers who join unions, leave unions, stay union, and stay nonunion:

$$\Delta O_{it} = \alpha_1 UU + \alpha_2 UN + \alpha_3 NU + \alpha_4 NN + \Delta\epsilon_{it}, \quad (3)$$

where UU , UN , NU , and NN are dummy variables that take the values 1 or 0 depending on the union status in the two periods:

$UU = 1$, if union in both periods;

$UN = 1$, if union in period 1, nonunion in period 2;

$NU = 1$, if nonunion in period 1, union in period 2;

$NN = 1$, if nonunion in each period;

and where the constant term has been suppressed.

Equation (3) shows that the before/after nature of the experiment permits calculation of three different union effects, each answering a somewhat different question:

(1) What happens to nonunion workers who join unions compared to nonunion workers who remain nonunion (obtained as the difference between the coefficients

on NU and NN, NU - NN, for short)?

(2) What happens to union workers who leave the union compared to those who remain union (UN - UU)?

(3) Among workers who change, what happens to those who join a union as compared to those who leave a union ((UN - NU)/2 or some such other average)?

It can be readily seen that when union differential is constant over time (UU = NN) and when the effects of joining and leaving unions are the same in absolute value ($|NU - NN| = |UN - UU|$), Equation (3) collapses into Equation (2). Less restrictively, if the only reasons for (3) to differ from (2) are changes in union differentials over time, the estimated parameters will fulfill the equality in absolute values given above, that is, the only difference between the gains of workers who join unions versus those who leave is the changed union differential over time.

Equations (2) and (3) can be readily generalized to analyze data covering more than two periods. The natural extension of (2) is to a fixed effects model with individual constants (differences from mean values) for each person. The natural extension of (3) is to a model with dummy variables for all possible classifications of changes in status. For ease of exposition in this paper I treat only the two-period case.

Interpreting Longitudinal Results

Assuming that $E(\alpha_1 U_i) > 0$, the longitudinal estimates of union impact should be lower than cross-section estimates. In fact, empirical analyses of wages do indeed show a lower impact of unionism in longitudinal than in cross-section data, providing support for the "omitted ability bias" model given above (among the panel studies are Duncan 1977, 1979; Brown 1980; Mellow 1981; Mincer

1981; Chamberlain 1982). In Chamberlain's analysis, for example, addition of individual constants reduces the union coefficient by 32% to 44%, indicating "a substantial heterogeneity (ability) bias." The union wage effect still stands, but its magnitude is smaller than in traditional cross-section analysis. As the archetypical comment at the beginning of the paper indicates, many have interpreted the smaller longitudinal estimates as providing better estimates of the true union effects than the larger cross-section estimates. Indeed, under the fixed-effect assumptions that changes in union status are properly measured and that selectivity of changers does not produce a correlation of the error in the change-in-outcome equation with changes-in-status, the longitudinal estimate is unbiased.

Are these assumptions likely to be valid in empirical work? What does their violation do to longitudinal estimates of union effects?

This paper argues that neither assumption is likely to be valid and that, under reasonable conditions, measurement error and selectivity of changers will bias downward longitudinal estimates of union effects. Because in practice measurement error appears to be the principal econometric problem in analysis of longitudinal data, I focus largely on the measurement error issue.

II. The Problem of Measurement Error

In cross-section studies of unionism, one generally ignores measurement error in the union status variable on the assumption that only a small number of workers are likely to be misclassified and thus that any bias in the estimated union coefficient due to measurement error is modest. Misclassification of a small number of workers will, however, produce a much larger error in longitudinal than in cross-section analysis and thus cannot be readily ignored. The reason for the greater error is twofold. On the one hand, random misclassifica-

Table 1Example of Measurement Error Effect

A. Cross-Section Data Set

Observed	True	Number
U	U	23
U	N	2
N	U	2
N	N	73

B. Longitudinal Data Set

Observed		Consisting of True	With observed means of	
			1	2
UU	13	13UU	1.30	1.30
UN	12	9UN, 1UU, 2NN	1.25	1.03
NU	12	9NU, 1UU, 2NN	1.03	1.25
NN	63	61NN, 1UN, 1NU	1.004	1.004

tion of workers in two periods will produce a larger number of misclassified workers than random misclassification in one period. On the other hand, by obtaining information on union effects from generally small numbers of changers, the longitudinal analysis will contain a smaller number of correct observations. As a result the proportion of observations in error will be much larger in the longitudinal analysis than in the cross-section analysis, producing a larger bias.

A numeric example illustrates the dramatically different effect of modest misclassification on cross-section and longitudinal estimates. Assume we have a sample of 100 workers, of whom 25 are union members and 75 are not. Assume measurement error is such that 2 union workers are misclassified and 2 nonunion workers are misclassified.¹ Then we have the situation shown in Table 1A. If the true value of the outcome variable is 1.00 for nonunion workers and 1.30 for union workers, our estimated means would be 1.28 and 1.01, giving an estimated differential of 27%, a value that is 10% below the true impact of unionism.

Assume that 20 workers switch union status in the period, 10 joining and 10 leaving unions. With 4 workers misclassified in each period, so that 8% of union workers and 2.7% of nonunion workers are incorrectly classified, it can be demonstrated (see Equation 12) that the longitudinal data set will be approximately as shown in Table 1B.

There are three points to note about this data set. First, the longitudinal estimates of the union effect from NU and UN comparisons are the same: 1.25/1.03 or 21%, which is 30% below the true impact of unionism -- an attenuation that is three times as large as that in the cross-section analysis. Second, measurement error produces a pattern of differences in levels of wages between the four sets: for example, workers measured as leaving unions have a

lower wage in period 1 than workers who remain union, workers measured as joining unions have a higher wage in period 1 than workers who remain nonunion, and so on. Third, the best estimate of the difference in wages in the data is the comparison of the mean level of wages for the UU set with the mean level for the NN set, which yields essentially the correct 30% differential. For this to be the best estimate of the union effect, however, workers in the two sets would have to be otherwise identical, contrary to the assumed $E(a_i U_i) > 0$.

More formally, I compare what measurement error in the dichotomous union status variable does to the estimated union coefficient in cross-section Equation (1) to what measurement error in the change in union status variable does to the estimated union coefficient in longitudinal Equation (2). Because of the restricted values of union status or change in status, the measurement error is correlated with the workers' true status, so that the standard measurement error in regression analysis must be modified, along lines set out by Aigner (1973) and by Marquis et. al. (1981).²

Consider first measurement error in a dichotomous variable. Let $M =$ measured union status, $U =$ actual status, and $e =$ error. Then

$$M = U + e \quad (4)$$

where possible errors are: -1 , if a person's true status is union ($U = 1$), producing a nonunion classification ($M = 0$), and 1 , if a person's true status is nonunion ($U = 0$), producing a union classification ($M = 1$).

Now let r_U be the probability that a union worker is misclassified and r_N is the probability that a nonunion worker is misclassified and $1 - r_U$ and $1 - r_N$ be the corresponding probabilities that the workers are correctly classified. Then the relation between the expected error and the true status is:

$$E(e) = r_N + (-r_U - r_N)U \quad (5)$$

so that from (4)

$$E(M) = r_N + (1 - r_U - r_N)U. \quad (6)$$

Hence we can write M as

$$M = r_N + (1 - r_U - r_N)U + v \quad (6')$$

where v is a random variable with mean zero and variance σ_v^2 .

The effect of regressing an outcome O on M rather than on U can be evaluated by substituting (6') into the true equation (1) and treating the random component of measurement error as an omitted variable. Substitution yields:

$$O_i = (b/1 - r_U - r_N)M_i + cX_i - bv_i/(1 - r_U - r_N) + u_i \quad (7)$$

where I have suppressed the constant term.

The bias on the coefficient on M_i from omitting v_i is the coefficient of v_i in (7) times the regression coefficient of v_i on M_i , holding the X s fixed. Assuming, for ease of presentation, that M is uncorrelated with the X s in (7), we obtain the coefficient of v_i on M from (6') as σ_v^2/σ_M^2 , the random measurement error component of the measured variance. Then the regression for (7) yields for the coefficient on M_i (\hat{b}).

$$\begin{aligned} E(\hat{b}) &= [b/(1 - r_U - r_N)] (1 - \sigma_v^2/\sigma_M^2) \\ &= [b/(1 - r_U - r_N)] (1 - r_U - r_N)^2 \sigma_U^2/\sigma_M^2 \\ &= b(1 - r_U - r_N) \sigma_U^2/\sigma_M^2. \end{aligned} \quad (8)$$

Since union status is binomial, $\sigma_U^2 = \bar{U}(1 - \bar{U})$, where \bar{U} = mean proportion union. If as in our numeric example, we assume that $\bar{M} = \bar{U}$, which holds whenever $r_U U = r_N(1 - U)$, Equation (8) simplifies to

$$E(\hat{b}) = b(1 - r_U - r_N) \quad (9)$$

When M is correlated with $X(r_{MX})$ the comparable equation is:

$$E(\hat{b}) = \frac{b(1-r_U - r_N) - b_{UX}b_{XM}}{1 - r_{MX}^2} \quad (10)$$

where b_{ux} and b_{xm} are the simple regression coefficients. Here the bias depends on the relation between the X s and both observed and true union status. If we assume that the random component of the measurement error is independent of $x[\sigma_{vx} = 0]$ then, noting that $b_{ux}b_{xm} = r_{mx}^2$, (10) becomes

$$E(\hat{b}) = \frac{b(1 - r_U - r_N) - r_{MX}^2}{1 - r_{MX}^2} \quad (10')$$

Since the bias in (10) is greater than the bias in (9), we conclude that as long as the random component of measurement error is uncorrelated with the X s the cross-section estimate of the union effect is biased downward by at least $1 - r_U - r_N$ percent.³

Turning to the effect of measurement error on longitudinal estimates, we proceed in a similar manner to the preceding analysis. In this case, the equation relating measured and true changes in union status is

$$\Delta M = \Delta U + e, \quad (11)$$

where ΔM = measured change in union status (= 1, 0, -1),

ΔU = true change (=1, 0, -1), and

e = error (2, 1, 0, -1, -2).

When r_U and r_N are independent over time, the relationship between the true changes and the measured changes can be written as functions of r_U and r_N and of the true changes from one state to the other T_{ij} ($i, j = U$ or N) as follows:

$$M_{NN} = (1 - r_N)(1 - r_N)T_{NN} + (1 - r_N)r_U T_{NU} + (1 - r_U)r_N T_{UN} + r_U r_U T_{UU} + v_{NN}$$

$$M_{NU} = (1 - r_N)r_U T_{NN} + (1 - r_U)(1 - r_N)T_{NU} + r_N r_U T_{UN} + (1 - r_U)r_U T_{UU} + v_{NU}$$

$$M_{UN} = (1 - r_N)r_N^T r_N^T r_N^T + r_N r_U^T r_N^T + (1 - r_U)(1 - r_N)^T r_N^T + (1 - r_U)r_U^T r_U^T + v_{UN} \quad (12)$$

$$M_{UU} = r_N r_N^T r_N^T + (1 - r_U)r_N^T r_N^T + (1 - r_U)r_N^T r_N^T + (1 - r_U)(1 - r_U)^T r_U^T + v_{UU},$$

where v_{ij} is a random error.

Equation (12) is the critical equation in our analysis. The three terms in each equation in which an r_i or r_j is multiplied by a $(1 - r_j)$ represent misclassification errors. The terms in which $(1 - r_i)$ is multiplied by $(1 - r_i)$ represent true changes in the measured observations. As before, the error term can take on only a limited set of values, dependent on the value of the true change. The relation between the true values of ΔU and the possible error is defined as in Table 2 below.

But from this array it can be seen that

$$E(e) = -(r_U + r_N)\Delta U \quad (13)$$

so that

$$e = -(r_U + r_N)\Delta U + v \quad (13')$$

and

$$\Delta M = -(r_U + r_N)\Delta U + v, \quad (14)$$

where v is a random measurement error. Substituting (14) into (2) and applying the omitted variable bias formula for omission of v yields for the expected value of the estimated longitudinal impact of unionism

(\hat{b}_L)

$$E(\hat{b}_L) = b/(1 - r_U - r_N) (1 - \lambda) \quad (15)$$

where λ is the ratio of random variance (σ_v^2) to measured variance ($\sigma_{\Delta M}^2$).

From (14) $\sigma_{\Delta M}^2 = (1 - r_U - r_N)^2 \sigma_{\Delta U}^2 + v^2$, yielding

$$E(\hat{b}_L) = b(1 - r_U - r_N) \sigma_{\Delta U}^2 / \sigma_{\Delta M}^2. \quad (16)$$

According to (16) the downward bias in the longitudinal analysis will

exceed the downward bias in the cross-section analysis as long as $\sigma_{\Delta U}^2 < \sigma_{\Delta M}^2$.

Table 2Relation Between True Value of ΔU and Possible ErrorFrequency of Error Assuming True Value of ΔU

<u>Error</u>	<u>1</u>	<u>0</u>	<u>-1</u>
2	0	0	$r_U r_N$
1	0	$(1-r_N)r_N + (1-r_U)r_U$	$r_N(1-r_U) + r_U(1-r_N)$
0	$1-(r_N+r_U-r_U r_N)$	$1-2[(1-r_N)r_N-(1-r_U)r_U]$	$1-(r_N+r_U-r_U r_N)$
-1	$r_N(1-r_U) + r_U(1-r_N)$	$(1-r_N)r_N + (1-r_U)r_U$	0
-2	$r_U r_N$	0	0

Calculating variances we find that

$$\sigma_{\Delta U}^2 = (T_{UN} + T_{NU}) + (T_{UN} - T_{NU})^2 \quad (17)$$

$$\sigma_{\Delta M}^2 = (M_{UN} + M_{NU}) + (M_{UN} - M_{NU})^2. \quad (17')$$

For ease of analysis, assume that the true mean of unionism, (\bar{U}) , is constant over time and that there is no constant response bias, $E(\bar{M}) = \bar{U}$. But it can be shown (Marquis, et. al. 1981, p. 101) that M_{UN} depends on T_{UN} ,

$$\begin{aligned} E(M_{UN}) &= (1 - r_U - r_N)^2 T_{UN} + \delta_M^2 \\ E(M_{NU}) &= (1 - r_U - r_N)^2 T_{NU} + \delta_M^2 \end{aligned} \quad (18)$$

where $\delta_M^2 = (1 - r_U)r_U\bar{U} + (1 - r_N)r_N(1 - \bar{U})$, the average variance of the measurement error.

With constant \bar{U} , $T_{UN} = T_{NU}$. Now let T_C be the proportion of workers changing union status in the sample ($T_C = T_{UN} + T_{NU}$). Then (17) simplifies to

$$\sigma_{\Delta M}^2 = T_C, \quad (19)$$

while substitution of (18) into (17') yields

$$\sigma_{\Delta M}^2 = (1 - r_U - r_N)^2 T_C + 2\delta_M^2. \quad (19')$$

Equation (19') is an approximation due to the absence of terms reflecting the equation-specific error terms (v_{UN} , v_{NU} of [12]).

The key question is will $\sigma_{\Delta U}^2$ always be less than $\sigma_{\Delta M}^2$?

Examination of (19) - (19') shows the answer to be negative. When measurement error is large so that $(1 - r_U - r_N)^2$ is close to zero and when T_C is large $\sigma_{\Delta U}^2 > \sigma_{\Delta M}^2$. For example, let $r_U = r_N = .40$ and $\bar{U} (= r_U/(r_U + r_N)) = .50$. Then $\sigma_{\Delta M}^2 = .04T_C + .48$, so that for $T_C > .5$, $\sigma_{\Delta U}^2 > \sigma_{\Delta M}^2$. In this case, measurement error biases the cross-section estimate more than the longitudinal estimate.

On the other hand, when measurement error is modest, -- as in our earlier numeric example -- $\sigma_{\Delta U}^2$ will be less than $\sigma_{\Delta M}^2$ for moderate values of T_C , producing a greater downward bias in the longitudinal calculation.

Since the effect of measurement error on longitudinal as opposed to cross-section analysis thus depends on the magnitudes of the various parameters in the measurement error formula, I turn next to estimates of the critical magnitudes.

Evidence on Measurement Error

The first parameters needed to evaluate the importance of measurement error are the actual errors themselves -- r_U and r_N . I have identified two surveys which provide the type of information needed to estimate r_U and r_N : separate measures of the union status of the same workers at essentially the same time. The first survey is a special supplement to the January 1977 Current Population Survey, which asked workers whether or not they were covered by collective bargaining and then asked their employers the same question. The second is the May 1979 Current Population Survey which asked workers about their collective bargaining status on the "dual job" supplement and on the "pension" supplement. While there are differences in the timing of the questions in both surveys, the time differences are sufficiently slight so that differences in answers provides us with a reasonable first-order approximation to random measurement error in union status.

Table 3 tabulates the responses to these two surveys. It shows that while r_U and r_N are, as stated, modest in value, they are sufficiently non negligible to produce potentially large response error bias in longitudinal data. In the 1979 CPS sample 6.4 to 8.1% of workers in the union category are misclassified, and 1.9 to 2.3% of those in the nonunion category giving a value of 8.7 to 10.0%

Table 3Misclassification of Union Status on Two Surveys

A. Current Population Survey, May 1979

Covered by Collective Bargaining on Main Survey	Covered by Collective Bargaining on Pension Supplement		
	yes	no	total
yes	3,976	272	4,248
row (%)	93.6	6.4	
column (%)	91.9	1.9	23.2
no	321	13,688	14,009
row (%)	2.3	97.7	
column (%)	8.1	98.1	76.8
<u>Total</u>	4,297	13,950	18,257
row (%)	23.5	76.5	100

B. Employer - Employee Matched Survey, January 1977

Covered by Collective Bargaining, by Employers	Covered by Collective Bargaining by Employees or Household Respondent		
	yes	no	total
yes	707	57	764
row (%)	92.5	7.5	
column (%)	92.5	2.3	23.2
no	57	2,476	2,533
row (%)	2.3	47.8	76.8
column (%)	7.5	97.8	
<u>Total</u>	764	2,533	3,297
row (%)	23.2	76.8	100

Source: A, tabulated from May 1979 CPS.

B, tabulated from January 1977 Employee Employer Matched Sample.

for the critical $r_U + r_N$ figure. In the 1977 matched employer - employee sample, 7.5% of workers in the union category are misclassified and 2.3% of those in the nonunion category giving a 9.8% value to $r_U + r_N$.

To check whether the differences in classification on the samples can, in fact, be interpreted as resulting from random measurement error, I have estimated union wage equations for the sample of workers for whom there are conflicting estimates of union status, and for the sample for whom there are no such conflicts. If the conflict in responses is due to random misclassification, one would expect no significant union wage effect for persons in the sample in which estimates conflict, compared to a sizeable union effect in the sample for which there are no conflicts in whether a person is union or not. As can be seen in the unnumbered table below, estimates of standard log wage equations (with the usual demographic and human capital controls) for the samples yield the expected results:

	agreement	Union Status disagreement
January 1977 sample estimated union coefficient (standard error)	.26 (.02)	<u>+</u> .05 (.07)
May - June 1979 sample estimated union coefficient (standard error)	.21 (.01)	+ .06 (.07)

where the + before the coefficient reflects the change in sign depending on which estimate of unionism is used as the independent variable.⁴

Finally, taking the magnitudes of the estimated misclassification errors in Table 3 as valid, we can apply the formulas given earlier to evaluate the impact of measurement error on regression estimates of union impacts, given different proportions of workers truly changing union status. As can be seen in Table 4, when only 5% of workers change status the longitudinal estimate is less than half of cross-section estimate and just 40% of the true b , whereas if 15% -

Table 4Potential Impact of Measurement Error on Estimates of Union Effects

Proportion of Workers Truly Changing Union Status (%)	Estimated Bias in Cross-Section Estimate (a)	Estimated Bias in Longitudinal Estimate (b)	Relative Bias (a)/(b)
5	.90	.40	.45
10	.90	.59	.66
15	.90	.70	.78
20	.90	.77	.86
25	.90	.82	.91
30	.90	.86	.95
80	.90	1.00	1.12

Source: Calculated using formulas (9) and (19') assuming $r_U = 7.5\%$, $r_N = 2.5\%$, and $\bar{U} = .25$, so $\delta_M^2 = .036$.

20% change status the estimates are closer together. Consistent with the preceding analysis, when the proportion changing union status rises to relatively high levels, the longitudinal estimates exceed the cross-section estimates.

Table 5 turns to the next obvious issue: the proportion of workers who actually change union status in a longitudinal data set. It examines the proportions measured as changing status in four major longitudinal surveys: the May 1974-75 Current Population Survey (CPS), the National Longitudinal Survey of Men Aged 14-24 in 1966 (NLS) for the period 1970-78, and the Quality of Employment Panel Survey (QES), 1973-77, and also records estimates of the true proportion changing. The estimated true proportion changing are obtained by summing the expected values of M_{UN} and M_{NU} from (18), which yields

$$E(M_C) = (1 - r_U - r_N)^2 T_C + 2\delta_M^2 \quad (20)$$

where T_C is the proportion of true changers, and solving for T_C . In the cases where \bar{U} changes over time, Equation (20) is still applicable because the impact of changes in U has offsetting effects on $E(M_{UN})$ and $E(M_{NU})$.⁵

In three of the samples, the calculations yielded reasonable estimates of the true proportion changing, and those figures are reported in the table. In the May CPS sample, however, the formulas yielded no estimate, because under the assumptions, measurement error by itself should have produced virtually the proportion of changers observed.

The key finding in Table 5 is that whether one looks at the measured proportion of changers or at the estimated true proportion the values are on the low side of the figures in Table 4. The measured changes (M_C) range from 6.2% (CPS) to 24.2% (NLS) while the "true" proportion changing vary from 9.1% (QES) to 21% (NLS). With these changes, measurement error biases downward the longitudinal estimates by 14% (NLS) to 29% (PSID) to 34% (QES) and by even larger

amounts in the CPS, according to the estimates in Table 2.

In sum, given measurement errors in union status that produce values of $r_U + r_N$ of about .10, and true proportions of workers changing status below .20, the analysis in this section suggests that longitudinal estimates of the effect of unionism on economic outcomes will be below cross-section estimates and, more important, below the true effect of unionism as well.

Comparisons of Longitudinal and Cross-Section Estimates of Union Effects

As noted in Section I, there have been several studies of union wage effects using longitudinal data. These studies have found lower union effects than are found in comparable cross-section studies. By contrast, while there is a large and growing cross-section literature on the effects of unions on outcomes other than level of wages, such as dispersion of wages, labor turnover (notably quit behavior), fringe benefits, and the like (see Freeman and Medoff [1981] for a summary), there has been little longitudinal evidence regarding the effect of union membership on these outcomes. This section provides evidence that for two important "nonwage" outcomes, the dispersion of wages and fringe benefits, and for wages, longitudinal analysis yields smaller estimated union effects than does cross-section analysis, but that the estimated effects are still fairly sizable and economically significant. This finding leads us to reject criticisms that the results of cross-section studies of the nonwage outcomes are more subject to "heterogeneity" or fixed effects bias than are the results of wage studies. As measurement error should reduce the estimated impact of unionism on all outcomes, this is consistent with the models given in Section II.

The analysis treats the four data sets set out in Table 5. In each case I

Table 5
Proportion of Workers Measured as Changing Union Status in
Diverse Surveys

Status	Survey (Sample Size)			
	May 1974 -75 CPS (7,887)	Michigan Panel Survey of Income Dynamics 1970-79 (635)	National Longitudinal Survey 1970-78 (1,905)	Quality of Employment Survey 1973-77 (543)
NN	.714	.400	.609	.595
NU	.028	.098	.160	.057
UN	.034	.094	.087	.101
UU	.225	.408	.149	.247
M _C (UN or NU)	.062	.192	.242	.158
U ₁ (UU or UN)	.259	.502	.231	.348
U ₂ (UU or NU)	.253	.506	.309	.304
Estimated T _C	---	.117	.210	.091

Source: Tabulated from relevant survey with estimates of true UN or NU as described in the text with $r_U + r_N = .10$ and that $r_N/(r_U + r_N)$ equal the average rate of unionization in the period. Thus, for the NLS, I set $r_N/(r_U + r_N) = 1/2 (.231 + .309) = .27$ and obtain $r_N = .027$, $r_U = .073$. The same procedure is used for the other data sets. Note the Michigan PSID includes all of the "poverty" sample, producing a large proportion of union workers.

sought the largest possible sample for which the outcome variables and the union variable were reported. In the Michigan PSID sample, in which one has a number of possible years to examine, I report the results from a relatively long time span, 1970-79, though I examined shorter spans as well. In contrast to some studies, I include all of the special "poverty" sample as well as the random sample in the survey. In the NLS sample I also chose a relatively long time span to examine. As the May CPS sample covers one year and the QES covers three years, the result is significant variation in the time span covered and, as seen in Table 5, significant variation in the proportion of persons changing union status as well.

Wages

Table 6 presents the results of my longitudinal analysis of union wage effects in the four data sets.⁶ It records the log wages for the four union-change groups before and after the change, the change in log wages, and the implied union effects and, for comparison, the cross-section estimate of the union wage effect in the same data. While there is some variation among the three types of longitudinal estimates, the general pattern of results is clear: the longitudinal calculations yield lower estimates of the union effects than do cross-section calculations. As many longitudinal analyses focus on the difference in changes in wages between those joining and those leaving unions, the most significant comparison is between the $(NU - UN)/2$ estimates and the cross-section estimates.⁷ Consistent with the results of Mellow (1981) they show a great reduction in the estimated union effect in the May 1974-75 CPS. As this is the group with the smallest measured proportion of changers, this is to be expected from measurement error. There is, however, one aberrant case in the

Table 6

Log Wages, Changes in Log Wages Associated with Changing
Union Status, and Estimated Union Effects

Group and Survey	Log Wage Before	Log Wage After	Δ	Group	Estimated Union Effects
A. May CPS, 1974-75					
NN	1.24	1.34	.10	NU - NN	.09
NU	1.28	1.47	.19	UU - UN	.08
UU	1.58	1.67	.09	(NU - UN)/2	.09
UN	1.46	1.47	.01	cross-section	.19
B. National Longitudinal Survey of Young Men, 1970-78					
NN	.97	1.84	.87	NU - NN	.12
NU	.94	1.93	.99	UU - UN	.09
UU	1.34	2.05	.71	(NU - UN)/2	.19
UN	1.22	1.84	.62	cross-section	.28
C. Michigan PSID, 1970-79					
NN	.95	1.61	.67	NU - NN	.08
NU	1.06	1.81	.75	UU - UN	.26
UU	1.29	2.02	.73	(NU - UN)/2	.14
UN	1.16	1.63	.47	cross-section	.23
D. QES, 1973-77					
NN	1.38	1.85	.48	NU - NN	.19
NU	1.24	1.91	.67	UU - UN	.11
UU	1.55	2.00	.45	(NU - UN)/2	.16
UN	1.35	1.70	.34	cross-section	.14

Source: Calculated from the surveys. Cross-section estimates based on multivariate regression model with standard set of controls for demographic and human capital variables.

table: in the QES, the $(NU - UN)/2$ comparison yields a larger rather than smaller estimated union effect than does the cross-section analysis. In this case, the cross-section difference in wages was only moderately above the longitudinal difference (UU and NN differ by .17 and .15) so that inclusion of regression controls reduced the cross-section estimate to the lower level. Note also that the pattern of differences in the log wages themselves, before and after the change, are also generally, although not always, in line with the impact of measurement error. The before-change log wages show that union leavers have lower wages than union stayers, which agrees with the Section II numerical example. The after-change log wages also show that union joiners have lower wages than union stayers in all cases. By contrast, the before and after comparisons of changers with nonunion stayers show a less consistent pattern.

Finally, if we assume that the estimates of measurement error used in Table 5 apply to these data, we can calculate the proportion of the difference between cross-section/longitudinal coefficients due to measurement error. To do this we estimate the relative bias of longitudinal to cross-section estimates from Table 4, using the estimated true proportion of changers from Table 5, and multiply the resulting statistic by the cross-section estimate in Table 6. This yields .24 for the NLS and .16 for the PSID as the expected estimates from the longitudinal analyses, if measurement error were the only factor operating. Comparing these figures to the actual longitudinal estimates in Table 6, we see that measurement error explains 44% (NLS) to 77% (PSID) of the cross-section/longitudinal differences. While further analysis is required to pin down the specifics of the misclassification effects in each data set, our analysis suggests that measurement error can explain much of the difference between cross-section and longitudinal estimates of union wage effects.

This conclusion, while at odds with the widely used fixed-effects interpretation of the difference between longitudinal and cross-section analysis, is consistent with recent evaluations by other researchers. Chowdhury and Nickell (1982) who correct for measurement error bias in standard covariance estimates by instrumenting unionization on lagged unionization (on the grounds that serial correlation in the U variable is strong but is absent from measurement error), found that a longitudinal estimate of the union effect of .10 increased to .30 in the instrumental analysis. Their conclusion was that "omitted quality variables bias the union effect upwards and the 'old-style' cross section estimates are of the right order of magnitude after all." H. Gregg Lewis, (1983) in an evaluation of the effect of measurement error on union wages estimates, has also reached a conclusion similar to mine.

Dispersion of Wages

The proposition that trade union wage policies are designed to reduce inequality of wages within firms and across firms for workers doing similar work has a long history in labor economics, stretching back to the Webbs. Numerous cross-section comparisons of wage inequality have found that inequality is less in union than in nonunion settings (see, e.g., Hyclak 1977, 1979; Freeman 1980, 1982; Hirsch 1982; Plotnick 1982). Standard wage regressions provide corroborating evidence, showing that for the most part the impact of most wage-determining variables is smaller on the wages of union than on the wages of nonunion workers. The magnitude of the estimated union impact is sufficiently sizable to suggest that, despite the increase in dispersion due to union monopoly wage effects, unionism reduces overall inequality of wages.

Do comparisons of dispersions of wages in a longitudinal framework con-

firm the cross-section results? How much smaller, if at all, is the estimated union effect on dispersion?

To answer these questions I have tabulated the standard deviation of the log of earnings for workers by their change in union status in the four data sets referred to earlier. The resulting calculations are given in Table 7, which follows the same format as Table 6. As can be seen, the longitudinal calculations confirm the cross-section finding of lower wage dispersion under unionism. Dispersion tends to fall when workers join unions and decline when they leave, confirming the reduction in dispersion under unionism. There are, however, notable differences in the magnitude and consistency of the effects by groups, the NU - NN and $(NU - UN)/2$ comparisons showing larger union effects than UU - UN comparisons and with the PSID and QES showing more variable results than the other samples. To compare the longitudinal estimates to cross-section estimates, I have made some crude calculations of what a full cross-section analysis (which involves correcting observed differences in variances by observed differences in characteristics) might yield by reducing the difference in standard deviations between UU and NN workers in the before and after data by 30%, a figure consistent with a full analysis of May 1973-75 CPS data (Freeman 1980, Table 4). Without the adjustment the impact of unionism on dispersion estimated with the longitudinal data is much smaller than the impact estimated with the cross-section data. With the adjustment, the longitudinal estimate is still noticeably smaller, by magnitudes comparable to those obtained in Table 6 for wages.

Finally, note that comparisons of the levels of the standard deviations among groups tell a stronger story than did the comparison of the levels of wages.⁸ In the before data, workers who leave unions have larger dispersions

Table 7

Standard Deviations of Log Wages, Changes in Standard
Deviations Associated with Changing Union Status, and
Estimated Union Effects

Group and Survey	Standard Deviation in Log Wages			Group	Estimated Union Effects
	Before	After	Δ		
A. May CPS, 1974-75					
NN	.59	.58	-.01	NU - NN	-.08
NU	.52	.43	-.09	UU - UN	-.05
UU	.38	.35	-.03	(NU - UN)/2	-.06
UN	.46	.48	.02	cross-section	-.15
B. National Longitudinal Survey of Young Men, 1970-78					
NN	.47	.53	.06	NU - NN	-.10
NU	.39	.35	-.04	UU - UN	-.13
UU	.29	.30	.01	(NU - UN)/2	-.09
UN	.32	.46	.14	cross-section	-.14
C. Quality of Employment Survey, 1973-77					
NN	.55	.55	.00	NU - NN	-.23
NU	.52	.32	-.20	UU - UN	.03
UU	.38	.36	-.02	(NU - UN)/2	-.07
UN	.54	.49	-.05	cross-section	-.13
D. Michigan PSID, 1970-79					
NN	.46	.53	.07	NU - NN	-.15
NU	.45	.37	-.08	UU - UN	.01
UU	.31	.30	-.01	(NU - UN)/2	-.03
UN	.40	.38	-.02	cross-section	-.13

Source: Tabulated from the various surveys. The cross-section effect is estimated by taking 70% of the difference in standard deviations between UU and NN (averaged for before and after). This is an approximate correction for differing characteristics of union and nonunion workers.

than those who stay, and workers who join unions have larger dispersions than nonunion workers who remain nonunion. In the after data, workers joining unions have greater dispersion than workers who were always union members while workers leaving unions have less dispersion than workers who remain nonunion. While these patterns could be due to factors other than error in measuring union membership, they are consistent with a pure measurement error interpretation.

I conclude that, as with wages, the impact of unions on dispersion found in cross-section studies is confirmed in a longitudinal analysis and that the magnitude of the effect is commensurably lower, at least partly as a result of error in measuring union status.

Fringe Benefits

The third cross-section finding which I examine with longitudinal data in this paper is the finding that unionism increases the fringe component of compensation, particularly those fringe benefits that are most desired by older workers, such as pensions (for studies of fringe benefits, see Duncan 1976; Goldstein and Pauly 1976; Donsimoni 1978; Solnick 1978; Leigh 1980; Viscusi 1980; Freeman 1981, 1983). As the QES is the only data set which provides fringe benefit figures over time, my longitudinal analysis is limited to that data set. I consider two measures of fringes, the number of fringes reported by workers and the proportion with pensions.

Table 8 presents the results of a longitudinal analysis for these two variables, again following the Table 6 format. While changes in the list of fringes in the surveys causes the number of fringes reported for the majority of workers to fall, the evidence shows that workers who went from nonunion to union gained fringes, while those going from union to nonunion lost relative to those

Table 8

Numbers of Fringes and Presence of Pensions, Changes in
Numbers of Fringes and Presence of Pensions Associated
with Changing Union Status, and Estimated Union Effects,
QES, 1973-77

Group	Before	After	%Δ	Group	Estimated Union Effects
Number of fringes:					
NN	3.01	2.56	-15	NU - NN	32%
NU	2.59	3.02	17	UU - UN	0%
UU	3.64	3.28	-10	(NU - UN)/2	13%
UN	3.16	2.55	-10	cross-section	13%
Proportion of Workers with Pensions					
NN	.65	.70	.05	NU - NN	.34
NU	.55	.90	.35	UU - UN	.02
UU	.95	.96	.01	(NU - UN)/2	.18
UN	.78	.77	-.01	cross-section	.25

Source: Tabulated from Quality of Employment Panel, 1973-77. Pension figures based on 429 NNs, 185 UUs, 66UNs, and 48NUs.

who remained union. The implied union effects are all positive, with, however, considerable difference in magnitude. The NU - NN estimate, in particular, greatly exceeds UN - UU. The pension coverage figures show a similar pattern, with a sizeable increase in the proportion with pensions for workers joining unions but no real change for those leaving unions. Comparisons of the longitudinal with the cross-section estimates show no difference for number of fringes but the usual diminution of the union effect for provision of pensions.⁹ Finally, note that the pattern of differences in levels of fringes is similar to that found in dispersion for comparisons of UN's or NU's or UU's but is mixed in comparisons of changes with NN's.

Taking the results of Tables 6-8 as a whole, a reasonable generalization is that longitudinal analyses confirm the qualitative findings of cross-sectional analyses, with, however, smaller estimated union effects, possibly due in large part to the greater impact of errors of measurement on longitudinal than on cross-section statistics.

IV. Bounding the True Impact?

If, as researchers usually assume, there is a substantial selectivity problem in cross-section analysis, which dominates any problems of measurement error, then cross-section estimates of union effects overstate true union effects. The preceding sections show that if there is a substantial measurement error problem in longitudinal analysis, and if there is no countervailing problem of selectivity of changers, then longitudinal estimates of union effects understate true union effects. When both of these statements are true, we have an important "bounding" result:

Theorem: Under reasonable assumptions about the impact of measurement

error and of selectivity of persons into unions, cross-section estimates of union effects provide an upper bound and longitudinal estimates provide a lower bound on the "true" union impact in the model under study.

To prove the theorem, it is necessary to show that (a) measurement error biases longitudinal estimates downward to a greater extent than it does cross-section estimates, which is done in Section II; (b) selectivity of unionists in a cross-section biases cross-section estimates upward more than measurement error biases those estimates downward, which I shall assume on the basis of the modest estimated effect of measurement error in the cross-section; and (c) selectivity of who changes union status in longitudinal data either biases longitudinal estimates downward or biases them upward by less than measurement error biases them downward.

In this section I consider proposition (c). I examine the likely impact of selectivity in who changes union status on longitudinal estimates of union effects. I shall argue that under plausible models of the economics of unionism, selectivity of changers biases longitudinal estimates of union effects downward, reinforcing rather than weakening or offsetting the effects of measurement error. Hence, as long as (b) holds, the bounding theorem will be valid.

Modelling Selectivity

There are two types of selectivity involved in who becomes union or nonunion: workers' choice of working union (nonunion) jobs and employers' choice of workers. I model selectivity on the part of workers, then examine how the analysis changes when employers select workers from the queue desiring union jobs.

Consider the workers' decision to switch from union to nonunion status when the outcomes are determined by

$$\begin{aligned} O_{Uij} &= \bar{d} + d_j + a_j + \epsilon_{Uij} \\ O_{Nij} &= a_j + \epsilon_{Nij} \end{aligned} \quad (21)$$

where O_{Uij} = outcome for j th worker in i th period ($i = 1, 0$) when

\bar{d} = average union differential,

d_j = differential for j th worker relative to average differential with $E(d_j) = 0$,

a_j = individual "ability" effect, and

$\epsilon_{Uij}(\epsilon_{Nij})$ = error when j works union (nonunion) with expected values 0 and variances σ_U^2 and σ_N^2 .

A worker will choose to accept a union job when

$$O_{U1j} - O_{N1j} > K, \quad (22)$$

where K measures cost of mobility. Assume a bivariate normal distribution of the outcome variables. Then the truncated mean gain from working union is

$$E(O_{U1j} - O_{N0j} \mid O_{U1j} - O_{N1j} > K) = \bar{d} + \frac{\sigma_d^2 + \sigma_U^2}{\sigma^*} \frac{f[(K - \bar{d})/\sigma_*]}{1 - F[(K - \bar{d})/\sigma_*]} \quad (23)$$

where $\sigma_*^2 = \sigma_d^2 + \sigma_U^2 + \sigma_N^2$ and where $f/(1 - F)$ is the "inverse Mills" ratio correction for truncation. Equation (23) overstates the union differential because it averages only over workers with especially high gains.

Similarly, for workers leaving unions, we obtain

$$E(O_{U1j} - O_{U0j} \mid O_{N1j} - O_{U0j} > K) = -\bar{d} + \frac{\sigma_d^2 + \sigma_N^2}{\sigma^*} \frac{f[(K + \bar{d})/\sigma_*]}{1 - F[(K + \bar{d})/\sigma_*]} \quad (24)$$

as the expected mean change.

As our estimate of the union effect we take $(1/2)(NU - UN)$, which in the present context is $1/2$ of (23) minus $1/2$ (24). This yields

$$\bar{d} + \left[\frac{\sigma_U^2 f[(K - \bar{d})/\sigma^*]}{\sigma^* [1 - F[(K - \bar{d})/\sigma^*]} - \frac{\sigma_N^2 f[(K + \bar{d})/\sigma^*]}{\sigma^* [1 - F[(K + \bar{d})/\sigma^*]} \right] / 2 \quad (25)$$

where d is the union effect and the remaining components reflect selectivity of changers. Assume, for simplicity, that $\sigma_U^2 = \sigma_N^2$ and that there is a true union effect $\bar{d} > 0$. Then the selectivity bias is negative since

$$f[(K - \bar{d})/\sigma^*] / [1 - F(\cdot)] < f[(K + \bar{d})/\sigma^*] / [1 - F(\cdot)] \text{ because } k + \bar{d} > k - d.$$

If, as is plausible given our findings on dispersion, $\sigma_U^2 < \sigma_N^2$, the negative bias is enhanced. If, by contrast, $\bar{d} = 0$, and $\sigma_U^2 = \sigma_N^2$, selectivity has -- logically enough -- no such bias effect.¹⁰

In this model if there is a union effect, the selectivity of changers biases longitudinal estimates of that effect downward. Even if there is not, we have established that selectivity on the part of workers does not bias upward the longitudinal estimate and thus cannot offset the predicted downward bias from measurement error.

What about selectivity by employers?

Rather than providing a detailed analysis of this question (which involves complex double integrals), let us simply evaluate the qualitative impact of such selectivity on our previous results. Since only union firms have a queue of workers outside their plants, I assume that the only firm selectivity is selection of workers into union jobs. Firms will choose to hire workers with low $d_{j,s}$ -- that is, those for whom the true union effect is smallest (with a fixed union wage effect, this involves picking workers with the highest productivity¹¹ -- and try to displace those with high $d_{j,s}$.

With respect to workers who join unions, employer selectivity will augment the downward bias in the longitudinal estimate. This is because firms will be

selecting lower values of d_{NUj} from the sample of workers for whom $d_{NUj} > K + \epsilon_{Nij} - \epsilon_{Uij}$. This will reduce the inverse Mills ratio component of (23).

With respect to workers who leave union jobs, the easiest assumption is that because of seniority rules, firms have no selectivity, leaving (24) as is. If firms are able to select who leaves, however, there is an additional negative bias component to (24), so that we can no longer sign the net effect of selectivity in (25). For the bias in (25) to remain negative, it is necessary that the effect of firm selectivity on who joins a union dominate the effect of firm selectivity on who leaves. This is plausible given that firms are free to hire whom they want but not to fire or lay off.

All told, our analysis of selectivity in who changes union status suggests that, under reasonable selection criteria but simplified statistical assumptions, the longitudinal estimates of union effects will be biased downward, establishing the bounding theorem.

Conclusion

In this paper I have tried to show that measurement error is a significant problem in analysis of longitudinal data. I have developed some models of measurement error, examined numerical examples, and estimated the impact of measurement error in four data sets. My analysis has not been complete. I gave only cursory treatment to issues of the correlation between the random component of measurement error and control variables and ignored completely the potential impact of standard exclusion rules (such as requiring positive wages and sensible values of explanatory variables) on longitudinal as opposed to cross-section analyses. These errors of omission aside, the analysis suggests that longitudinal analysis is not the research panacea it is sometimes seen to be.

While omitted fixed effects bias upward cross-section estimates of union effects, measurement error and possibly selectivity of changers bias downward longitudinal estimates. Under reasonable conditions, the two sets of estimates bound the true impact of unionism, and thus should be viewed as complementary research tools. While neither is likely to yield the true parameter, together they enable us to estimate the magnitude of the effects of unionism, which appear to be quite substantial in empirical work.

Footnotes

I have benefited from comments of seminar participants at the University of Chicago, Caltech, Australia National University, and the University of California, Irvine, and the suggestions of John Abowd, Gary Chamberlain, and H. Gregg Lewis.

¹The assumption that equal numbers of workers are misclassified implies that the observed proportion union is an unbiased estimate of the true proportion. It is a useful simplifying assumption that appears consistent with actual measurement error (see Table 3) but is not critical to the numeric example or to the ensuing statistical analysis.

²Much of what follows is based on Marquis et al. (1981). I have also benefited from Aigner (1973).

³We ask if the following inequality holds

$$1 - r_U - r_N > \frac{1 - r_U - r_N - r_{MX}^2}{1 - r_{MX}^2}$$

Multiply by $(1 - r_{MX}^2)$ to obtain $(1 - r_U - r_N)(1 - r_{MX}^2) > 1 - r_U - r_N - r_{MX}^2$.

But simplifying we obtain $(r_U + r_N)r_{MX}^2 > 0$, which proves the inequality.

⁴In the January survey there are two reported wages: one from the individuals, the other from employers. I have used the wage reported by the individuals in this analysis.

⁵Specifically, the formulas with changes in the value of U between the periods are (Marquis et al. 1981, p. 101):

$$E(M_{NU}) = r_U(-\Delta U) + (1 - r_U - r_N)^2 T_{NU} + \delta_U^2 \quad \text{and}$$

$$E(M_{UN}) = r_U(\Delta U) + (1 - r_U - r_N)^2 T_{UN} + \delta_U^2$$

so that the sum becomes

$$E(M_C) = E(M_{NU}) + E(M_{UN}) = (1 - r_U - r_N)^2 T_C + 2\delta_U^2$$

⁶The measurement of wages varies across the data sets. In the CPS I measure wages by the ratio of usual weekly earnings to usual weekly hours; in

the PSID, I use average hourly wages; in the NLS, I use the reported hourly rate; while in the QES wages are annual earnings from work divided by hours worked times 52.

⁷In regression analyses which impose $NU = UN$, the coefficient is a weighted average dependent on relative numbers changing status. The reader can readily calculate weighted averages for contrast, if desired.

⁸A full analysis of the effect of measurement error or dispersion differs somewhat from that of analysis of measurement error in the regression format, but the qualitative effects of error are the same.

⁹The cross-section regression for number of fringes is based on regressions using 635 persons with 10 occupation, 6 industry, tenure, tenure squared, education, race, sex, years of schooling, and marital status controls. The regression for proportion with pensions is based on the same sample and model.

¹⁰I have benefited immensely from the comments of John Abowd in this section. The statistical analysis which follows relies extensively on John Abowd (1983).

¹¹We can also compare the bias in the $NU - NN$ and $UU - UN$ estimates. Following the analysis in the text, we find that the mean for NN is

$$\alpha_N^2 / \sigma^* \cdot (f[(K - \bar{d}) / \sigma^*] / \{1 - F[(K - \bar{d}) / \sigma^*]\}),$$

so that the mean for $NU - NN = \bar{d} + \{[(\sigma_d^2 + \alpha_U^2 - \alpha_N^2) / \sigma^*] f\} / (1-F)$,

which is less than \bar{d} when $\sigma_d^2 + \alpha_U^2 < \alpha_N^2$, which is likely since the dispersion of wages is less than the dispersion of nonunion wages. Hence, here too we have an underestimate.

¹²That is, a reasonable specification is $d_j = -\lambda \alpha_j$, where α_j is our ability indicator with $E(\alpha) = 0$.

References

- Abowd, John M. "Comments on the Effects of Selectivity Biases on Estimates of Union/Nonunion Effects in a Panel Setting." Prepared as a review of "Longitudinal Analyses of the Effects of Trade Unions" by Richard Freeman, 1983. University of Chicago, mimeo.
- Abowd, John M. and Farber, Henry S. "Job Queues and the Union Status of Workers." Industrial and Labor Relations Review 35(3) (April 1982): 354-367.
- Aigner, Dennis J. "Regression with a Binary Independent Variable Subject to Errors of Observation." Journal of Econometrics 1 (1973): 49-60.
- Brown, Charles. "Equalizing Differences in the Labor Market." Quarterly Journal of Economics 94(1) (February 1980): 113-134.
- Chamberlain, Gary. "Multivariate Regression Models for Panel Data." Journal of Econometrics 18 (1982): 5-46.
- Chowdhury, G., and Nickell, S. "Individual Earnings in the U.S.: Another Look at Unionization, Schooling, Sickness and Unemployment Using Panel Data." Discussion Paper No. 141, London: Centre for Labour Economics, London School of Economics. November, 1982.
- Donsimoni, Marie-Paule Joseph. An Analysis of Trade Union Power: Structure and Conduct of the American Labor Movement. Ph.D. thesis, Harvard University, 1978.
- Duncan, Greg J. "Earnings Functions and Nonpecuniary Benefits." Journal of Human Resources Vol 10 (Fall 1976), 462-83.
- _____. "Paths to Economic Well-Being." In Five Thousand American Families -- Patterns of Economic Progress, edited by Greg J. Duncan and James N. Morgan, Vol 5. Ann Arbor: Institute for Social Research, 1977.
- _____. "An Empirical Model of Wage Growth." In Five Thousand American Families -- Patterns of Economic Progress, edited by Greg Duncan and James N. Morgan, Vol 7. Ann Arbor: Institute for Social Research, 1979.
- Freeman, Richard B. "Unionism and the Dispersion of Wages." Industrial and Labor Relations Review, Vol 34, No 1 (October 1980): 3-23.
- _____. "The Effect of Trade Unionism on Fringe Benefits." Industrial and Labor Relations Review, Vol 35 (July 1981): 489-509.
- _____. "Union Wage Practices and Wage Dispersion within Establishments." Industrial and Labor Relations Review 36(1) (October 1982): 3-21.
- _____. "Unionism, Pensions, and Union Pension Funds." National Bureau of Economic Research, Conference on Pensions, Labor and Individual Choice (March 23-26, 1983).

- Freeman, Richard B., and Medoff, James L. "The Impact of Collective Bargaining: Illusion or Reality?" In U.S. Industrial Relations 1950-1980: A Critical Assessment, edited by Jack Steiber, Robert McKersie, and Quinn Mills. Madison: Industrial Relations Research Association, 1981.
- Goldstein, Gerald, and Pauly, Mark. "Group Health Insurance as a Local Public Good." In The Role of Health Insurance in the Health Services Sector, edited by Richard N. Rosett. New York: The National Bureau of Economic Research, 1976.
- Hirsch, Barry. "The Interindustry Structure of Unionism, Earnings and Earnings Dispersion." Industrial and Labor Relations Review 36(1) (October 1982): 22-39.
- Hyclak, Thomas. "Unionization and Urban Differentials in Income Inequality." Journal of Economics 3 (1977): 205-207.
- _____. "The Effect of Union on Earnings Inequality in Local Labor Markets." Industrial and Labor Relations Review (October 1979): 77-84.
- Leigh, Duane E. "The Effect of Unionism on Workers' Valuation of Future Pension Benefits" Industrial and Labor Relations Review Vol 34, (July 1981): 510-521.
- Lewis, H. Gregg. "Fixed Effects or Measurement Errors in Panel Data?" Durham, N.C.: Duke University Graduate School of Business, June 1983.
- Marquis, K.H.; Duan, N.; Marquis, M.S.; Polich, J.M.; with Meslkoff, J.E.; Shwartzbach, D.S.; and Stasz, C.M. "Response Errors in Sensitive Topic Surveys: Estimates, Effects, and Correction Options." Santa Monica, CA: Rand Corporation, April 1981. Rand/R-2710/2-HHS.
- Mellow, Wesley. "Unionism and Wages: A Longitudinal Analysis." Review of Economics and Statistics 63(1) (February 1981): 43-52.
- Mincer, Jacob. "Union Effects: Wages, Turnover, and Job Training." National Bureau of Economic Research Working Paper 808, November 1981.
- Plotnick, Robert D. "Trends in Male Earnings Inequality." Southern Economic Journal Vol 48 (January 1982): 724-732.
- Solnick, Loren M. "Unionism and Employer Fringe Benefit Expenditures." Industrial Relations Vol 17 (February 1978): 102-107.
- Viscusi, W. Kip. "Unions, Labor Market Structure and the Welfare Implications of the Quality of Work." Journal of Work Research Vol 1 (Spring 1980): 175-92.