INTERNATIONAL ECONOMIC REVIEW
Vol. 17, No. 1, February, 1976

# THE EFFECT OF UNIONS ON EARNINGS AND EARNINGS ON UNIONS: A MIXED LOGIT APPROACH*

BY PETER SCHMIDT AND ROBERT P. STRAUSS[1]

## 1. INTRODUCTION

A LARGE NUMBER OF STUDIES have attempted to quantify the impact of unions on the wages or earnings of workers. See, for example, Lewis [9] or Ashenfelter and Johnson [1] for critical summaries of many of these studies. Several different research methodologies have been employed to ascertain the extent to which unions have raised relative wages. Empirically, time series and cross-sectional data at the firm, industry and economy-wide level have been examined. Theoretically, most of the analyses have been partial equilibrium in nature, although the recent papers by Johnson and Mieszkowski [7] and Diewert [4] investigate the impact of unionism in a general equilibrium setting. Virtually all of these studies have found a positive, significant effect of unions on wages, although there is considerable variation in the estimated size of the effect.

Common to all of these studies is the assumption that unionism exerts a unilateral and exogenous effect on wages. Unfortunately, the interesting issue of the determinants of union membership has been relatively unresearched, and by and large the matter of membership remains unrelated to the effects of unionism on wages. This is at least potentially a serious matter since it seems clear that there may be an effect of wages on unionism as well as an effect of unionism on wages. This may occur because relative wages affect the attractiveness of various industries to a potential union organizer, or because they may affect the probability of a worker voting for a union in a representation election.

That relative wages may affect the probability or extent of unionization has been previously noted—see, for example, Reder [12] and Wachter [20]. From a statistical point of view, this would imply that union membership, or extent of unionization, would more properly be viewed as jointly or simultaneously determined with wages, rather than being treated as exogenous. Nevertheless, most studies of the effects of unions on wages or earnings have treated unionization as exogenous.

A notable exception in this regard is the recent work of Ashenfelter and Johnson [1], which examined at the industry level the effect of unionization on wages *and* the effect of wages on unionization. In this paper we perform a similar analysis, except using individual observations. The two endogenous variables

* Manuscript received February 27, 1975; revised May 7, 1975.

[1] The authors wish to thank the Data and Program Library Service, University of Wisconsin, for an efficient version of the 1967 Survey of Economic Opportunity. Financial support from the U.S. Department of Labor is gratefully acknowledged. The constructive comments of two referees have led to substantial improvement in the paper, and are gratefully acknowledged.

are the individual's earnings and his membership or non-membership in a union. Since one of the endogenous variables is continuous while the other is binary, the formulation and estimation of such a model requires certain innovations in econometric technique, which are discussed in Section 2. We note that the technique of estimating models with jointly dependent qualitative and continuous dependent variables may be of interest beyond the issue of unions and earnings. Section 3 gives our empirical results, and Section 4 concludes.

## 2. THE MIXED LOGIT MODEL

Let $X_t$ be the $t$-th observation on the binary dependent variable, which takes on the values 0 and 1; let $Y_t$ be the $t$-th observation on the continuous dependent variable. (The observation index $t$ runs from 1 to $T$.) Then the specification is as follows:

$$\text{(1)} \qquad \log_e \left[ \frac{P(X_t = 1 \mid Y_t)}{P(X_t = 0 \mid Y_t)} \right] = Q_t\gamma + \alpha Y_t$$
$$Y_t \mid X_t \sim N(Z_t\beta + \delta X_t, \sigma^2)\,.$$

Here $Q_t$ and $Z_t$ are the $t$-th observations on row vectors of exogenous explanatory variables, $\gamma$ and $\beta$ are vectors of parameters, and $\alpha$, $\delta$, and $\sigma^2$ are scalar parameters.

The first equation is a standard logit specification, conditional on $Y$. It is precisely analogous to the specifications in Nerlove and Press **[11]** and Schmidt and Strauss **[16]**, the only difference being that in the models considered in those papers, $Y$ was also a qualitative variable. The second equation specifies the distribution of $Y$ given $X$ as normal with mean linear in the explanatory variables.

The model can be estimated by maximum likelihood if we can obtain the joint density of $X_t$ and $Y_t$. To do so, note that

$$\text{(2)} \qquad \frac{f(X_t = 1, Y_t)}{f(X_t = 0, Y_t)} = \frac{P(X_t = 1 \mid Y_t)}{P(X_t = 0 \mid Y_t)} = \exp\,(Q_t\gamma + \alpha Y_t)\,.$$

We also note that

$$\text{(3)} \qquad P(X_t = i) = \frac{f(X_t = i, Y_t)}{f(Y_t \mid X_t = i)} \qquad i = 0, 1$$

Since $P(X_t = 0) + P(X_t = 1) = 1$, we have

$$\text{(4)} \qquad \frac{f(X_t = 0, Y_t)}{f(Y_t \mid X_t = 0)} + \frac{f(X_t = 1, Y_t)}{f(Y_t \mid X_t = 1)} = 1\,.$$

Using (2),

$$\text{(5)} \qquad f(X_t = 0, Y_t)\left[ \frac{1}{f(Y_t \mid X_t = 0)} + \frac{\exp\,(Q_t\gamma + \alpha Y_t)}{f(Y_t \mid X_t = 1)} \right] = 1\,,$$

or

$$f(X_t = 0, Y_t) = \frac{f(Y_t \mid X_t = 0) f(Y_t \mid X_t = 1)}{f(Y_t \mid X_t = 1) + \exp(Q_t \gamma + \alpha Y_t) f(Y_t \mid X_t = 0)} \tag{6}$$

$$f(X_t = 1, Y_t) = \exp(Q_t \gamma + \alpha Y_t) f(X_t = 0, Y_t) .$$

Finally, we can substitute into (6) the facts that (from (1)):

$$f(Y_t \mid X_t = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{1}{2\sigma^2} (Y_t - Z_t \beta)^2 \right] \tag{7}$$

$$f(Y_t \mid X_t = 1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{1}{2\sigma^2} (Y_t - Z_t \beta - \delta)^2 \right] .$$

This gives the final expression for the joint distribution of $X$ and $Y$. We will not actually write it out, since it is more conveniently expressed as in (6) and (7).

Given that all observations are mutually independent, the likelihood function is

$$\mathrm{L} = \prod_{t \in \theta_0} f(X_t = 0, Y_t) \prod_{t \in \theta_1} f(X_t = 1, Y_t) , \tag{8}$$

where $\theta_0 = \{t \mid X_t = 0\}$, $\theta_1 = \{t \mid X_t = 1\}$. This can be maximized numerically with respect to $\gamma$, $\beta$, $\alpha$, $\delta$, and $\sigma^2$ to obtain the maximum likelihood estimates. Asymptotic variances of the estimates can be obtained from the inverse of the information matrix, which in this case is most conveniently found by numerical differentiation. (The analytic formulae for the second derivatives are unfortunately complicated.)

It is possible to extend the mixed logit model to cases where there are additional endogenous variables (and equations), and/or where the qualitative dependent variables take on more than two values. In the interest of brevity the details are omitted here; see Schmidt and Strauss [15].

## 3. EMPIRICAL RESULTS

Statistical analysis of the effect of unionization has been hampered at the individual level by a paucity of data on union membership. The questions posed in the decennial Census and available in the Public Use Samples or the recurrent Current Population Survey have not provided union membership information. Fortunately, the 1967 Survey of Economic Opportunity[2] did inquire about union membership in the private sector and accordingly is our point of departure. The model described in the previous section was applied to a random sample of 912 observations from the representative portion of the Survey, after it had been modified to include only those who were full-time workers over 14 years of age but less than 66 and who had non-zero annual earnings.

The binary dependent variable took on the value of 0 if the worker was not

[2] The characteristics of the data, especially for the union membership variable, are discussed in some detail in U.S. Bureau of Census [19].

TABLE 1

MEANS AND STANDARD DEVIATIONS OF VARIABLES

| Variable | Mean | Standard Deviation |
|---|---|---|
| Earnings | 6322.106 | 4601.21 |
| Union | 0.2456 | 0.4307 |
| Education | 11.384 | 3.206 |
| Experience | 23.978 | 13.200 |
| Race | 0.8958 | 0.3056 |
| Sex | 0.6798 | 0.4668 |
| N. E. | 0.2599 | 0.4388 |
| N. C. | 0.2664 | 0.4420 |
| West | 0.1941 | 0.3954 |

in a union, and the value of 1 if the worker was in a union. The continuous dependent variable was annual earnings in dollars. The explanatory variables chosen are typical regressors used in earnings functions:[3] education, measured as years of schooling; labor market experience, measured as calendar age minus education minus five; race, equal to zero for blacks and equal to one for whites; sex, equal to zero for females and one for males; and census region, represented by three binary dummy variables equal to one if the individual lived in the Northeast, North Central, or Western regions of the country, respectively, and zero otherwise. (The fourth region of the country is the South, which is chosen as the omitted category since it is typically thought to differ from the rest of the country in its attitudes towards unions.) The means and standard deviations of the variables are presented in Table 1.

The resulting maximum likelihood estimates are given below:

$$\begin{aligned}
(9)\qquad \log_e \frac{P(\text{Union}\,|\,\text{Earnings})}{P(\text{Non-union}\,|\,\text{Earnings})} &= \underset{(-0.14)}{-0.0160} - \underset{(-6.11)}{0.1939}\ \text{Education} \\
&\quad \underset{(-0.31)}{-0.00215}\ \text{Experience} - \underset{(-1.95)}{0.5035}\ \text{Race} + \underset{(2.96)}{0.6186}\ \text{Sex} \\
&\quad \underset{(3.73)}{+0.8906}\ \text{N. E.} + \underset{(5.21)}{1.232}\ \text{N. C.} + \underset{(1.79)}{0.5062}\ \text{West} \\
&\quad \underset{(2.89)}{+0.000064}\ \text{Earnings}
\end{aligned}$$

$$(10)\qquad \text{Earnings} = \underset{(-2.98)}{-7240.} + \underset{(4.26)}{624.1}\ \text{Education} + \underset{(2.63)}{90.18}\ \text{Experience}$$

[3] With the exception of the absence of an industrial concentration measure, our model closely follows the Ashenfelter-Johnson model. Here, labor quality is controlled for through educational attainment and labor market experience. Whereas they employ proportion female in their model, we have at the individual level both race and sex information. Finally, we are able to control for region more extensively and compare the Northeast, West, and North Central regions to the South. We benefit from the much larger sample size available with individual data as well as the potentially richer effects which microdata may reveal.

$$+ \underset{(0.79)}{1058.}\ \text{Race} + \underset{(4.30)}{3757.}\ \text{Sex} + \underset{(0.80)}{894.3}\ \text{N. E.}$$

$$+ \underset{(0.54)}{614.9}\ \text{N. C.} + \underset{(1.17)}{1434.}\ \text{West} + \underset{(0.48)}{470.9}\ \text{Union}\,.$$

The numbers in parentheses under the estimates are the "asymptotic $t$ ratios," which are the ratio of the estimated coefficient to the estimated asymptotic standard error. They are asymptotically distributed as $N(0, 1)$ under the null hypothesis that the associated coefficient is zero. With 912 observations we can be relatively confident in using the $N(0, 1)$ critical points for such tests.

Of particular interest is the fact that the coefficient of earnings on union is positive and significant, while the coefficient of union on earnings is positive but insignificant. In other words, the "usual" statement that being in a union leads to higher earnings may have the causation reversed; it appears from these results that the correct statement is that higher earnings make one more likely to be unionized. This result supports the empirical results of Ashenfelter and Johnson [1]. In fact, this result is very similar to their main result, that the effect of union on earnings (actually, wages, in their case) is positive but insignificant when the reverse influence is allowed for.[4]

There are some other interesting results above as well. More education makes one less likely to be in a union, but raises earnings. More experience has no significant effect on the odds of being in a union, but raises earnings. To be white makes one less likely to be in a union, and has a positive but insignificant effect on earnings. To be male increases the odds of being in a union, and also increases earnings. Finally, to be in any region except the South increases one's odds of being in a union; the effects on earnings vary from region to region, but are insignificant.

## 4. SOME ADDITIONAL RESULTS

As a further aid to interpreting the empirical results given above, we report in Table 2 the partial derivatives of $P$ (union | earnings) with respect to the explanatory variables in the union equation, with all variables at their means.[5] Some of these are quite large—for example, to be male rather than female increases the probability of being in a union by .107; to live in the Northeast rather than the South increases this probability by .154. These are rather striking effects considering that the proportion of unionized workers in the sample is only .246. Similarly, in Table 3 we report the elasticities of $P$ (union | earnings) and of Earn-

[4] One possible explanation for the insignificant coefficient of union on earnings is the presence of the regional dummies in the earnings equation. If unionism is highly correlated with region (as it appears, given the results of the union equation) then we have a multicollinearity problem. One check on this is to rerun the model without the regional dummies. However, while this did increase the significance level of some other variables in the earnings equation, it did not have much of an effect on the union variable—its coefficient fell to 291.0, and its "$t$ ratio" increased only to 0.86.

[5] The partial derivatives of the unconditional probability $P$ (union) are not very different.

TABLE 2

PARTIAL DERIVATIVES OF $P$ (UNION/EARNINGS), EVALUATED AT MEANS

| Variable | Partial Derivative |
|---|---|
| Education | −0.03354 |
| Experience | −0.00037 |
| Race | −0.08711 |
| Sex | 0.1070 |
| N. E. | 0.1541 |
| N. C. | 0.2131 |
| West | 0.08757 |
| Earnings | 0.000011 |

TABLE 3

ELASTICITIES, EVALUATED AT MEANS

| Variable | Union Equation | Earnings Equation |
|---|---|---|
| Education | −1.555 | 1.124 |
| Experience | −0.03612 | 0.3420 |
| Race | −0.3177 | 0.1499 |
| Sex | 0.2962 | 0.4039 |
| N. E. | 0.1630 | 0.03676 |
| N. C. | 0.2312 | 0.02591 |
| West | 0.06921 | 0.04403 |
| Union | | 0.01829 |
| Earnings | 0.2832 | — |

ings, with respect to their respective explanatory variables, with all variables at their means. Naturally these reflect the same effects as the original coefficients, but they have the advantage of eliminating the effect of the scale of the variables. The most striking thing in looking at the elasticities is the extent to which the elasticities with respect to education dwarf all others. It is also clear that the predominant effect of experience is on earnings, not unionism, while the opposite is true of region.

The reader may also be interested in seeing how the results from the mixed logit technique compare with those gotten by methods which ignore the simultaneous nature of the model. Ignorning the simultaneous nature of the model would lead one to estimate the union equation by the simple logit technique, and the earnings equation by ordinary least squares. The logit and OLS results for these two equations are given in Table 4. For ease of comparison the mixed logit results are reproduced there as well. A glance at Table 4 shows that there is very little difference between the simple logit results and the mixed logit results for the union equation. Furthermore, for the earnings equation, there is little difference in the coefficients yielded by OLS and the mixed logit techniques.

TABLE 4

RESULTS BY ALTERNATIVE ESTIMATION TECHNIQUES

Union Equation

| | Logit | | Mixed Logit | |
|---|---|---|---|---|
| Variable | Coefficient | "$t$ ratio" | Coefficient | "$t$ ratio" |
| Constant | .0000 | 0.00 | .0160 | −0.14 |
| Education | −.2019 | −5.82 | −.1939 | −6.11 |
| Experience | −.00201 | −0.27 | −.00215 | −0.31 |
| Race | .4940 | 1.88 | .5035 | −1.95 |
| Sex | .6331 | 2.99 | .6186 | 2.96 |
| N. E. | .9121 | 3.78 | .8906 | 3.73 |
| N. C. | 1.255 | 5.26 | 1.232 | 5.21 |
| West | .5282 | 1.85 | .5062 | 1.79 |
| Earnings | .000059 | 2.79 | .000064 | 2.89 |

Earnings Equation

| | OLS | | Mixed Logit | |
|---|---|---|---|---|
| Variable | Coefficient | "$t$ ratio" | Coefficient | "$t$ ratio" |
| Constant | −7231. | — | −7240. | −2.98 |
| Education | 617.3 | 13.53 | 624.1 | 4.26 |
| Experience | 88.53 | 8.36 | 90.18 | 2.63 |
| Race | 1117. | 2.70 | 1058. | 0.79 |
| Sex | 3779. | 13.99 | 3757. | 4.30 |
| N. E. | 894.9 | 2.61 | 894.3 | 0.80 |
| N. C. | 603.1 | 1.74 | 614.9 | 0.54 |
| West | 1450. | 3.83 | 1434. | 1.17 |
| Union | 648.3 | 2.12 | 470.9 | 0.48 |

However, the OLS $t$ ratios are invariably three to four times bigger than the mixed logit "$t$ ratios." For example, the effect of union on earnings would be significant if we looked at the OLS results. This shows the importance of considering the simultaneous nature of the model.

A final thing to note is that our results imply rates of return of education and experience on earnings. The rate of return of education on earnings is typically defined as

$$\frac{d \log (\text{Earnings})}{d\,\text{Education}} = \frac{d\,\text{Earnings}}{d\,\text{Education}} \cdot \frac{1}{\text{Earnings}}.$$

This depends in our specification on the level of Earnings. Evaluated at the sample mean of 6332.1, we find a rate of return of education of 9.87%. Similarly, the rate of return of experience is 1.43%. Since (for a person of a given age) each extra year of education decreases experience by a year, we have a net return to education of 8.44%.

These rates of return agree roughly with those of other investigators. The rate of return of education has been estimated by Becker [2], Becker and Chiswick [3], Eckaus [5], Johnson [8], and Mincer [10], among others; the results range from 5%—30%, with most between 10% and 20%. Mincer's results include some that are easily compared to our own. For example, he reports the following two equations, based on 1959 data on white nonfarm workers:

$$\text{(11)} \quad \log \text{Earnings} = 6.20 + .107 \text{ Education} + .081 \text{ Experience} - .0012 \text{ Experience}^2$$

$$\text{(12)} \quad \log \text{Earnings} = 7.43 + .110 \text{ Education} - 1.651\, e^{(-.15 \text{ Experience})} .$$

His rates of return to education are therefore 10.7% and 11.0% for the two specifications, and are fairly close to our results. His rate of return to experience depends on the level of experience. At the sample mean of 23.978, his rates of return from the two specifications are 2.35% and 0.68% respectively; our result lies between these two.

## 5. CONCLUSIONS

In this paper we have developed a method of analysis for models with jointly dependent qualitative and continuous variables. This mixed logit model may be of general use, as such models should be widespread. We have applied it here to the joint determination of union membership and earnings.

We find union membership to have a positive but insignificant effect on earnings, and earnings to have a positive and significant effect on the probability of union membership. This result is similar to that of Ashenfelter and Johnson [1], who also allowed for the effect of earnings on union membership. The finding of an insignificant effect of union membership on earnings contrasts with most earlier studies, which treated union membership as exogenous. This may suggest that the common statment that unions raise wages may suffer from an incomplete notion of causation.

*The University of North Carolina, U. S. A.*

## REFERENCES

[1] ASHENFELTER, O. AND G. E. JOHNSON, "Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries," *International Economic Review*, XIII (October, 1972), 488-508.

[2] BECKER, G. S., *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (New York: Columbia University Press, 1964).

[3] ——— AND B. R. CHISWICK, "Education and the Distribution of Earnings," *American Economic Review*, LVI (May, 1966), 358-369.

[4] DIEWERT, W. E., "The Effects of Unionization on Wages and Employment: A General Equilibrium Analysis," *Economic Inquiry*, XII (September, 1974), 319-339.

[5] ECKAUS, R. S., "Estimation of the Returns to Education with Hourly Standardized In-

comes," *Quarterly Journal of Economics*, LXXXVII (February, 1973), 121–131.

[6] Fuchs, V. R., *The Service Economy* (New York: National Bureau of Economic Research, 1968).

[7] Johnson, H. G. and P. Mieskowski, "The Effects of Unionization on the Distribution of Income: A General Equilibrium Approach," *Quarterly Journal of Economics*, LXXXIV (November, 1970), 539–561.

[8] Johnson, T., "Returns From Investment in Human Capital," *American Economic Review*, LX (September, 1970), 546–560.

[9] Lewis, H. G., *Unionism and Relative Wages in the United States* (Chicago: University of Chicago Press, 1963).

[10] Mincer, J. B., *Schooling, Experience and Earnings* (New York: Columbia University Press, 1974).

[11] Nerlove, M. and S. J. Press, *Univariate and Multivariate Log-linear and Logistic Models* (Santa Monica: Rand R1306-NIH/EDA, 1973).

[12] Reder, M. W., "Unions and Wages: The Problems of Measurement," *Journal of Political Economy*, LXXIII (April, 1965), 188–196.

[13] Rosen, S., "On the Interindustry Wage and Hours Structure," *Journal of Political Economy*, LXXVII (March/April, 1969), 249–273.

[14] ———, "Unionism and the Occupational Wage Structure in the United States, "*International Economic Review*, XI (June, 1970), 269–286.

[15] Schmidt, P. and R. P. Strauss, "Estimation of Models with Jointly Dependent Qualitative Variables," presented at 1974 Econometric Society meetings.

[16] ——— and ———, "Estimation of Models with Jonitly Dependent Qualitative Variables: A Simultaneous Logit Approach," *Econometrica*, XLIII (July, 1975), 745–755.

[17] Stafford, F., "Concentration and Labor Earnings: Comment," *American Economic Review*, LVIII (March, 1968), 174–180.

[18] Throop, A. W., "The Union-Nonunion Wage Differential and Cost-Push Inflation," *American Economic Review*, LVIII (March, 1968), 79–99.

[19] U.S. Bureau of the Census, "Labor Union Membership in 1966," *Current Population Reports*, Series P-20, Number 216, (Washington, D. C.: Government Printing Office, 1971).

[20] Wachter, M. L., "Cyclical Variation in the Interindustry Wage Structure," *American Economic Review*, LX (March, 1970), 75–84.

[21] Weiss, L., "Concentration and Labor Earnings," *American Economic Review*, LVI (March, 1966), 96–117.