THE DETERMINATION OF THE UNION STATUS OF WORKERS

Henry S. Farber

Massachusetts Institute of Technology

NUMBER 227

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

THE DETERMINATION OF THE UNION STATUS OF WORKERS

Henry S. Farber

Massachusetts Institute of Technology

NUMBER 227

Abstract

A model of the determination of the union status of workers is developed that incorporates the separate decisions of workers and potential union employers in a framework which recognizes the possibility of an excess supply of workers for existing union jobs. This theoretical framework results in an empirical problem of partial observability because information on union status is not sufficient to determine whether nonunion workers are nonunion because they do not desire union representation or because they were not hired by union employers despite a preference for union representation. The problem is solved by using data from the Quality of Employment Survey that have a unique piece of information on worker preferences which allows identification and estimation of the model.

The empirical results yield some interesting insights into the process of union status determination that cannot be gained from a simple logit or probit analysis of unionization. Chief among these relate to the unionization of nonwhites and southerners. The well-known fact that nonwhites are more likely to be unionized than otherwise equivalent whites is found largely to be due to a greater demand for union representation on the part of nonwhite workers. The equally well-known lower propensity to be unionized among southern workers is found to be due to a combination of a lower demand for union presentation on the part of southern workers and a supply of union jobs which is more constrained relative to demand than in the North.

This letter was drafted at the University of Warwick Summer Workshop in Labour Economics, supported by the SSRC.

This paper is circulated for discussion purposes only and its contents should be considered preliminary.

September, 1982

1. INTRODUCTION

A source of much confusion in the analysis of labor unions regards the process by which the union status of workers is determined. In most cases the union status of individual workers has been modeled as being the result solely of utility maximizing decisions by workers. (See, for example, Ashenfelter and Johnson [2], Lee [12] and Schmidt and Strauss [18]). On the other hand, it has been argued that any real effect of unions on compensation or other aspects of employment could be partially or even completely offset by union employers' ability to hire better workers. This argument, that union workers might be "better" than observationally equivalent nonunion workers, has led to the recent outpouring of research attempting to measure the "true" effect of unions in the United States.[2] It is clear that union employers must have some control over whom they hire in order for the true effect of unions to be offset by this mechanism, and such employer control is not consistent with the worker choice model of union status. Indeed, it is a major weakness of this literature that either a worker choice model or no explicit model is offered while the implicit reasoning suggests that employers are making relevant decisions. Given the centrality to these analyses of the process by which union status is determined, one must question any conclusions which are drawn in this context.

In this study it is argued that the union status of workers is determined as the result of separate decisions by workers and potential union employers. Workers decide whether they would prefer union or nonunion jobs based on the utilities that these jobs yield to them. At the same time, union employers are deciding which of the workers who want union jobs to hire given that workers differ in their productive characteristics and that these characteristics are compensated differently in the union and nonunion sectors. Essentially union employers are assumed to hire the workers who enable them to produce at minimum cost.

The presumption that union employers have some discretion in hiring results from the likelihood of queues for vacancies in existing union jobs.[3] These queues result from the facts that it is unlikely that dues and initiation fees completely offset the advantages of unionization for all workers and that it is expensive to create new union jobs by organizing nonunion jobs.[4] More fundamentally, the queues result from a distinction, arising from the process of unioniztion, which must be drawn between the union status of workers and the union status of jobs. Nonunion jobs become unionized through organization of the workers who hold them. This is a costly and uncertain process which can involve the holding of an election supervised by the National Labor Relations Board (NLRB).[5] These elections are often preceded by intense and closely monitored campaigns, and they may involve appeals by either or both sides to the NLRB regarding such issues as illegal campaign tactics and determination of the appropriate bargaining unit. However, once the jobs are successfully unionized, their union status is preserved even if the workers who made the investment in organization leave.[6] In addition, new jobs created through expansion of unionized establishments are unionized by definition. Union employers can hire whomever they wish to fill any vacancies, but all new hirees will be unionized.[7] Thus, unless dues or initiation fees are sufficiently large, there will be workers who desire vacancies in existing union jobs but who are not willing to undertake investment in new unionization. For these workers the benefits of unionization are larger than the costs of union membership but smaller than the costs of organizing nonunion jobs. The results are queues for union jobs.

In general, empirical analysis of a model of the determination of the union status of workers of the sort proposed here is hampered by the fact that only the outcome (union status) is observed so that it is impossible to discern whether nonunion workers did not desire union representation or desired union representation but were not selected from the queue by a union employer. Abowd and Farber [1] carry out with some success an empirical analysis of union status determination which is consistent with a queuing model, but they are hampered by just this partial observability problem. Poirier [14] presents an econometric approach to identification and estimation of such models. Unfortunately, his technique is heavily dependent on functional form for identification and to date has not proven very useful in applications. More successful are studies which use data from such sources as the Quality of Employment Survey (QES) and surveys of workers participating in NLRB-supervised representation elections to focus on worker preferences for union representation as distinct from actual union status. These include studies by Farber and Saks [8] and Farber [6,7]. The drawback of these studies is that they can shed no light on employer selection criteria, and as a result they cannot address the full question of the determination of the union status of workers.

The approach to estimation taken in this study is to utilize data from the QES on both the union status of workers and on the explicit preferences of nonunion workers for union representation. The crucial bit of information is the response elicited from nonunion workers as to whether or not they would vote for union representation on their current job were a secret ballot election to be held. While these data present some problems of their own, it is argued below that they provide enough information to allow identification of the queue and estimation of the full model of union status determination including both worker and employer decision criteria.

In the next section an explicit model of the determination of the union status of workers conditional on the locus of union jobs, incorporating both the worker and potential union employers as decision makers, is developed. Econometrically, the model is bivariate in nature which reflects the fact that there are two decision makers.

In Section III the data from the QES and the econometric framework are discussed. Particular attention is paid to the interpretation of the crucial question regarding nonunion worker preferences for union representation in the context of the problem of interest here. The data are censored with regard to this variable on the basis of the process of union status determination modeled in the previous section. It is argued that the censored QES information reflects current preferences for union representation while the model suggests that union status is a reflection of preferences for union representation at the time the worker began his current job. It is further argued that the structure of the workers' preference function for union representation does not change over time and that actual preferences will differ over time only to the extent that the measured and unmeasured characteristics of workers or their jobs change. In other words, age or seniority will vary over time and affect worker preferences, but the effect of a given level of age or seniority on preferences will not vary over time. In addition, unmeasured factors such as on-the-job relationships with co-workers or supervisors and unobserved factors which affect compensation can vary over time resulting in changes in preferences. An econometric framework which exploits this fixity of structure while accounting for the censored nature of the data is developed. Section IV contains the empirical analysis of the resulting trivariate discrete data model.

In Section V the substantive results are discussed in the context of the theoretical framework derived in Section II. Important insights into well known relationships between union status and such characteristics as race, region, occupation, and age are gained from the results through the decomposition of these relationships separately into components due to workers and to employers. For example, it is found that the low probability of working on union jobs for southern workers is the result of a combination of a somewhat lower worker demand for union representation combined with a supply of union jobs which is more constrained relative to demand than in the North. On the other hand, the relatively high probability for nonwhite workers of working on union jobs, even after standardizing for education and occupation, is found largely to be due to a substantially higher demand for union representation among nonwhite workers.

The final section contains a summary of the results along with a discussion of their implications both with regard to the process of unionization and with regard to analysis of the "true" effects of labor unions.

II. A MODEL OF UNION STATUS DETERMINATION

The determination of the union status of workers is the result of decisions made separately by workers and union employers. Essentially, a worker will be unionized only if he both wants a union job and is hired by a union employer. It is assumed that the workers make their decisions regarding preference for union representation based on the relative utilities derived from union and nonunion employment. In addition, it is assumed that employers decide which workers to hire based on a comparison of the unit costs of effective (productivity adjusted) labor input yielded by different workers.

The decision of an individual worker to desire union representation is based on a comparison of the worker's utilities in the union and nonunion sectors. The worker will desire employment in the sector which yields the highest level of satisfaction. More formally, if M represents the difference between the worker's utility on a union job and his utility on a nonunion job then the criterion for the worker to desire union representation is that $M>0$. Given that workers are heterogeneous in their preference for union representation to the extent that workers of different characteristics derive different amounts of pecuniary and nonpecuniary benefit from unionization, M will vary across workers. A convenient parameterization for the worker preference criterion as a function of individual characteristics is

(1) $$M = XG_1 + u_1$$

where X is a vector of observable individual characteristics, G_1 is a parameter vector, and u_1 represents unobservable individual characteristics which affect worker preference for union representation.[8]

The union employer decision criterion regarding which workers to hire is the result of a comparison by the employer of the relative cost of "producing" effective labor using workers of differing characteristics and hence differing productivities. The cost of producing effective labor in the union sector will vary with worker characteristics as long as compensation differentials in the union sector do not accurately reflect productivity differentials, and since compensation in the union sector are set through the collective bargaining process there is no reason to expect compensation and productivity to be so precisely related.[9] Given that union employers are cost minimizing producers of output, they will wish to hire those workers who enable them to produce effective labor, and hence output, most cheaply. The structure of compensation in the union sector relative to productivity combined with the distribution of workers who desire union representation relative to the supply of unionized jobs defines a threshold level of effective labor cost which represents the maximum that union employers will be willing to pay for effective labor. In this context an individual worker will be hired by a union employer only if his effective labor cost in the union sector is less than this threshold.

In more formal terms, the criterion for a union employer in a given geographic or occupational labor market to hire a particular worker is that the union effective labor cost of that worker (C) be smaller than the treshold (K) in that labor market. Let H=C-K represent the difference between union effective labor cost and the threshold so that the union employer criterion for hiring a particular worker is that H<0. A convenient parameterization for this employer criterion as a function of individual characteristics (X) is

$$H = XG_2 + u_2 \qquad (2)$$

where G_2 represents a vector of parameters and u_2 represents unobservable individual characteristics which affect the employer decision process. The factors which affect H reflect variation in the supply of union jobs across different labor goegraphic and occupational labor markets as well as variation in effective labor cost of different workers.

The unobserved components of the model (u_1 and u_2) can be assumed to be random variables which may be correlated for any particular individual but are distributed independently across different individuals. These random variables have zero mean and covariance matrix[10]

$$V = \begin{bmatrix} v_1^2 & \\ v_{12} & v_2^2 \end{bmatrix}. \tag{3}$$

In order to understand how the model can be implemented, it is useful to express formally what can be inferred from data on union status alone. If a worker reports that he is working on a union job then it can be inferred that at the time he took the job he both desired a union job and was hired by a union employer. Alternatively, if a worker reports that he is working on a nonunion job then it can be inferred that at the time he started the job he either desired a union job but was not hired by a union employer or he did not desire a union job. However, for neither union nor nonunion workers can this information be used to make inferences about current preferences for union representation or current ability to be hired by a union employer. Consider the following examples. First regarding the preferences of union workers, it is possible that a union worker may no longer desire union representation but not be willing to quit his union job and sacrifice the nonportable benefits of seniority in order to take a nonunion job. A similar argument can be made concerning the preferences of nonunion workers. Next regarding the ability of nonunion workers to be hired by a union employer, a nonunion worker who desired a union job but was not hired by a union employer at the time he started his current job may now be able to be hired by a union employer but not be willing to sacrifice his nonunion seniority to take a union job. These examples suggest that both worker and employer decisions can change over time and that inferences based on the union status of workers must be restricted to preferences of workers and employers at the time of hire.

In the context of the model developed here, the probability that a worker is observed in a union job is the joint probability that he desired a union job at the time of hire ($M_0 > 0$) and he was hired by a union employer ($H_0 < 0$). The "$_0$" subscript denotes that the relevant quantities are measured at the time of hire. On this basis, the probability of observing a worker on a union job is written in terms of the random variables as

$$(4)\quad Pr(U=1) = Pr(u_1 > -X_0G_1,\ u_2 < -X_0G_2).$$

Similarly, the probability of observing a worker in a nonunion job is $1 - Pr(U=1)$, which can be expressed as

$$(5)\quad Pr(U=0) = Pr(u_1 > -X_0G_1,\ u_2 > -X_0G_2) + Pr(u_1 < -X_0G_1)$$

where the first term represents the probability that the worker desired a union job at the time he took his current job but was not hired by a union employer while the second term represents the probability that the worker did not desire a union job at the time he took his current job. The exogenous variables are time-subscripted to reflect conditions at the start of the job, and the random components (u_1 and u_2), while not subscripted, are considered to be specific to the time of hire. The crucial point to note is that the structural parameters (G_1 and G_2) are not time-subscripted and are assumed to be stable over time.

In order to implement the model a functional form must be selected for the random variables. Therefore, it is assumed that v_1 and v_2 are distributed as bivariate normal with zero mean and covariance matrix as defined in equation (3). Not all of the parameters of the covariance matrix errors (v) are estimable. Due to its discrete choice nature, the model is identified only up to the ratio of the parameter vectors to the standard deviations of their respective errors. For this reason the variances of u_1 and u_2 are normalized to one. Thus, the only element of the covariance matrix which is estimable is the correlation between the reduced form errors (p_{12}). In addition, the probabilities in equations (4) and (5) become standardized normal probabilities.

The model is theoretically identified and can be estimated using data on union status alone where the probability of a worker being unionized is defined as Pr(U=1) in equation (4). However, the two distinct elements in Pr(U=0) in equation (5) highlight the fundamental partial observability problem which stems from not knowing whether nonunion workers are nonunion because they desired a union job but were not hired by a union employer or because they did not desire a union job. Poirier [14] discusses estimation of partial observability bivariate probit models of this sort and argues that the model is identified and estimable. However, identification relies heavily on nonlinearities in the functional form of the probability distribution, and this is not terribly satisfactory. In addition, some experience with estimation of partial observability models in this context suggests that there are convergence problems and that where convergence is reached the parameters are not estimated with useful precision.[11] In view of these factors, the empirical analysis proceeds using a different approach: additional information on worker preferences, available from the Quality of Employment Survey, is used to aid in the identification and estimation of the model. The discussion turns now to a description of the data and the development of the appropriate econometric framework for estimation of the model utilizing the auxiliary information on worker preferences.

III. THE DATA AND ECONOMETRIC FRAMEWORK

The data used are from the 1977 cross-section of the Quality of Employment Survey (QES) developed by the Survey Research Center of the University of Michigan. The QES contains data for approximately 1500 randomly selected workers (both union and nonunion) on their personal characteristics and job attributes.[12] The particular sample for use in this study was derived from the QES by selecting those workers for whom the survey contained valid information on the variables listed in Table I. Self-employed workers, managers, sales workers, and construction workers were deleted from the sample due to the fact that the union status of these workers is determined by a different process than that outlined in the previous section. For example, self-employed workers will not be unionized by definition, while union employment in the construction industry is characterized by hiring halls where the union effectively makes the hiring decisions for employers. The remaining sample contains 915 workers. Table I contains descriptions of the variables used in the study as well as their means and standard deviations for the entire sample and the union and nonunion subsamples. The base group for the dichotomous variables consists of white, nonsouthern, unmarried, male, blue collar workers with twelve years of education. On average, the 37 percent of the sample who are unionized are slightly older and are more likely to be male, married, nonwhite, nonsouthern, and in a blue collar occupation. Unionization is defined as working on a job which is covered by a collective bargaining agreement. This is appropriate in light of the fact that it is collective bargaining as opposed to union membership which alters the employment relationship.

The crucial bits of information for this study are data on the union status of the jobs held by the individuals and the response to the question asked only of nonunion workers, "If an election were held with secret ballots, would you vote for or against having a union or employee association represent you?". This latter variable, called VFU, is the piece of information which is unique to this data set, and it will serve as the basis for identification of the queue for union jobs. It is interpreted here as the current preference of a worker for union representation on his current job. Thus, it holds all job characteristics fixed, including seniority, except those which the worker expects the union to affect. Fully 37 percent of the nonunion sample answered this question in the affirmative so that there is substantial variation in the response.

It was noted in the previous section that the partial observability problem is the cause of difficulty in identifying and estimating the model strictly from data on union status. The information on VFU can be used to solve this problem in a rather straightforward fashion. It is argued that the probability that a worker currently desires union representation on his job ($Pr(VFU=1)$) is a result of the same decision calculus derived in the previous section. This probability is $Pr(M_c > 0)$ where the subscript "c" refers to the current time. In terms of the underlying random variables,

(6) $M_c = X_c G_1 + u_3$,

and the probability that a worker currently desires union representation is

(7) $Pr(VFU=1) = Pr(u_3 > -X_c G_1)$

where X_c represents the exogenous variables measured at the current time and u_3 represents the random component in the worker preference function measured at the current time.[13]

If the data on VFU were available for all workers it would be straightforward to estimate G_1 from a simple probit likelihood function derived from equation (7) under the assumption that u_3 was normally distributed. However, data on VFU are available only for nonunion workers so that the data are censored on the basis of a variable which is obviously related. The standard approach to estimating a censored data model is to specify the censoring process along with the joint stochastic structure of the censored and censoring processes. The model can then be estimated jointly using maximum likelihood techniques. In the case at hand, the censoring process is the model of union status determination derived in section II and expressed probabilistically in equations (4) and (5). Assuming that u_3 is distributed as standard normal and using the earlier assumption regarding the joint normality of u_1 and u_2, the implication is that u_1, u_2, and u_3 have a trivariate normal distribution with zero mean and covariance matrix

(8) $$\begin{bmatrix} 1 & p_{12} & p_{13} \\ p_{12} & 1 & p_{23} \\ p_{13} & p_{23} & 1 \end{bmatrix}$$

where the variances are normalized to one as required for identification of this class of discrete data models and where p_{ij} represents the correlation between u_i and u_j.

Three distinct events are possible in this framework. The first is that the worker is unionized, in which case there is no information regarding current preferences for union representation. The probability of this event is the probability that at the time the worker started his union job he desired a union job ($M_0 > 0$) and he was hired by a union employer ($H_0(S) < 0$). From equation (4) this is

$$(9) \quad Pr(U=1) = Pr(u_1 > -X_0G_1,\ u_2 < -X_0G_2) .$$

The second event is that the worker is nonunion and currently desires union representation. The probability of this event is derived from equations (5) and (7) as

$$(10) \quad Pr(U=0,\ VFU=1) = Pr(u_1 > -X_0G_1,\ u_2 > -X_0G_2,\ u_3 > -X_cG_1) + Pr(u_1 < -X_0G_1,\ u_3 > -X_cG_1) .$$

The first term represents the joint probability that the worker is nonunion because he desired a union job but was not hired and that the worker currently desires a union job. The second term represents the joint probability that the worker is nonunion because he did not desire a union job at the time he started his job and that he currently desires a union job.

The final event is that the worker is nonunion and currently does not desire union representation. The probability of this event is derived from equations (5) and (7) as

$$(11) \quad Pr(U=0,\ VFU=0) = Pr(u_1 > -X_0G_1,\ u_2 > -X_0G_2,\ u_3 < -X_cG_1) + Pr(u_1 < -X_0G_1,\ u_3 < -X_cG_1) .$$

The first term represents the joint probability that the worker is nonunion because he desired a union job but was not hired by a union employer and that he currently does not desire union representation. The second term represents the joint probability that the worker is nonunion because he did not desire union representation at the time he started his job and that he currently does not desire union representation.

The three probabilities defined in equations (9) through (11) appropriately account for the union status of a particular worker along with his current preference for union representation where it is observed. Identification is clearly aided by the assumption that the parameters of the model which determines worker preferences at the start of the job are the same as the parameters of the model which determines current preferences (G_1). This is a prior theoretical restriction which provides "real" identification of the model and does not rely unduly on the functional form of the probability distribution. It is interesting to note that censored data models are generally estimated in order to obtain consistent estimates of the parameters of the censored process, while in this case the censored data are used to help identify and estimate the parameters of the censoring process.

Although the parameters of the model are fixed over time, the framework allows considerable flexibility in preferences over time. This comes from two sources. The first is that the unobserved components in worker preferences at the start of the job (u_1) and currently (u_3) can and likely do differ while the real possibility of correlation is allowed for. The second source of flexibility comes from the fact that the exogenous variables can change over time. In the empirical work which follows, the major time-varying variables are age and seniority.[14] Overall, the framework allows fluctuations over time in both the measured and unmeasured characteristics of workers and their jobs to have effects on worker preferences for union representation. These effects are consistent with the theoretical framework while at the same time preserving the fundamental identification of the model.

IV. ESTIMATION

The log-likelihood function for the trivariate censored data model is defined using equations (9) through (11) as

$$
\begin{aligned}
(12)\quad L = \sum_{i=1}^{n} \{ & U_i \ln \Pr(u_1 > -X_{0i}G_1,\ u_2 < -X_{0i}G_2) \\
& + (1-U_i)VFU_i \ln[\Pr(u_1 > -X_{0i}G_1,\ u_2 > -X_{0i}G_2,\ u_3 > -X_{ci}G_1) \\
& \qquad + \Pr(u_1 < -X_{0i}G_1,\ u_3 > -X_{ci}G_1)] \\
& + (1-U_i)(1-VFU_i)\ln[\Pr(u_1 > -X_{0i}G_1,\ u_2 > -X_{0i}G_2,\ u_3 < -X_{ci}G_1) \\
& \qquad + \Pr(u_1 < -X_{0i}G_1,\ u_3 < -X_{ci}G_1)]\} ,
\end{aligned}
$$

where i indexes observations. The dichotomous variable U_i equals one for union workers and is zero otherwise, and the dichotomous variable VFU_i equals one if the worker responded to the VFU question affirmatively and is zero otherwise. The likelihood function and its derivatives are composed of univariate, bivariate, and trivariate normal cumulative distribution functions which, while they cannot be evaluated in closed form, can be approximated numerically to the required accuracy. The likelihood function was maximized numerically with respect to G_1, G_2, and the three correlations between u_1, u_2, and u_3 using the algorithm described by Berndt, Hall, Hall, and Hausman [3]. This was a process which consumed large amounts of computational resources but was not marked by any particular difficulty in convergence. Various starting values were used to ensure convergence to a consistent set of parameters.

The maximum likelihood estimates of the parameters are contained in Table II. The value of the log-likelihood function at the maximum is -897.2. This is compared to a log-likelihood value for a constrained model with two parameters which represent constant probabilities of observing a worker in each of the three possible states of -983.3. This model embodies twenty-eight constraints on the structural model and can be rejected using a likelihood ratio test at any reasonable level of significance. This suggests that the model explains a significant portion of the variation in the data.

Table II also contains estimates of a simple univariate probit model of the union status of workers using the same variables as the queuing model. The time dependent variables are measured at the start of the workers' current jobs. These estimates are included simply as an illustration of the conventional approach to estimating models of union status determination, and they are best interpreted as indicative of the partial correlations between the exogenous variables and union status.

It is clear from the estimates in Table II that two of the three estimated correlations are estimated very imprecisely. These are the correlation (p_{12}) between the errors in the start-of-job worker preference equation and in the employer selection equation and the correlation (p_{23}) between the errors in the current worker preference equation and in the employer selection equation. This suggests that the likelihood function is very flat in these dimensions, which implies that there is little information in the data regarding whether workers who are more likely on the basis of their unobservable attributes to desire union representation are more or less likely to be hired by union employers. Further evidence for this is that when two versions of the model which constrain these correlations were estimated, the results did not change substantially. The first special case was to impose the constraint that $p_{12} = p_{23}$ so that the correlations between the unobservables affecting worker and employer preferences are time invariant. The maximum log-likelihood value of this model was -897.3 which implies using a likelihood ratio test that it is not possible to reject the constraint at any reasonable level of significance. The second special case was to impose the double constraint that $p_{12} = p_{23} = 0$ so that the unobservables affecting worker and employer preferences are uncorrelated. The maximum log-likelihood value for this model was -897.3 which again implies using a likelihood ratio test that the constraint cannot be rejected at any reasonable level of significance. The estimates of the other parameters of the model are virtually unchanged, although the precision with which they are estimated is improved somewhat by the imposition of the constraints. Nonetheless, to be conservative, the discussion of the results will focus on the estimates obtained for unconstrained model and contained in Table II.

The remaining correlation (p_{13}) between the unobservable factors affecting worker preferences at different points in time is asymptotically significantly greater than zero at conventional levels. This is consistent with the expectation that there are unmeasured attributes of jobs and workers which affect preferences for union representation and which persist over time.

V. ANALYSIS OF RESULTS

The estimates of G_1 contained in Table II reflect variation in worker preferences for unionization. In particular, the probability that a worker desires union representation is $Pr(u_1 > -XG_1)$ so that a positive coefficient on a variable in XG_1 implies that workers with higher values of that variable are more likely to desire union representation. Similarly, the estimates of G_2 reflect variation in the propensity of union employers to hire particular workers. The probability that a given worker will be hired by a union employer is $Pr(u_2 < -XG_2)$ so that a positive coefficient on a variable in XG_2 implies that workers with higher values of that variable are less likely to find union employment.

The estimates of the simple probit model of union status determination contained in Table II highlights a number of interesting empirical relationships. Chief among these are that nonwhites are more likely while southern workers less likely to be union workers. In addition, older workers are less likely to be unionized while blue collar workers are significantly more likely to be unionized than any of the other three occupational groupings. These results, while typical, are not easily interpreted with regard to the behavior of workers or employers. For example, the fact that southern workers are less likely to be unionized does not provide any information regarding the extent to which this is a result of less preference for union representation on the part of workers as opposed to a relative lack of supply of union jobs.

The estimates of the queuing model of union status determination can be used to resolve these behavioral issues. The important quantities are the probability that a worker desires union representation ($Pr(DES=1)$), the probability that a worker who desires union representation will be hired by a union employer ($Pr(HIRE=1|DES=1)$), and the probability that a worker is unionized ($Pr(U=1)$). These probabilities are easily constructed from the parameter estimates as

$$
\begin{aligned}
&\Pr(\text{DES}=1) &&= \Pr(u_1 > -XG_1); \\
&\Pr(\text{U}=1) &&= \Pr(\text{DES} = 1, \text{HIRE} = 1), \\
(13)\quad & &&= \Pr(u_1 > -XG_1,\ u_2 < -XG_2); \text{ and} \\
&\Pr(\text{HIRE}=1 \mid \text{DES}=1) &&= \frac{\Pr(\text{U}=1)}{\Pr(\text{DES}=1)},
\end{aligned}
$$

where the last relationship follows from application of Bayes' Law and where $\Pr(\text{HIRE}=1) = \Pr(u_2 < -XG_2)$. Note that by itself the probability that a worker will be hired by a union employer ($\Pr(\text{HIRE}=1)$) does not have a clear interpretation because it does not account for whether or not the particular worker is even interested in a union job. The relevant decision from the union employer's standpoint is which workers to hire from the pool or workers who desire union representation. In this context the quantity $\Pr(\text{HIRE}=1|\text{DES}=1)$ measures the ability of a worker to be hired by a union employer, and it reflects (inversely) the extent to which there are queues for vacancies in existing union jobs.

The parameter estimates will be discussed considering the effect of one variable at a time for a thirty year old worker in the base group consisting of white single male blue collar nonsouthern workers with twelve years of education and zero seniority. The first row of Table III contains the probabilities defined in equation (13) computed for a worker in the base group using the parameter estimates contained in Table II for the queuing model. The predicted probability of unionization based on the simple probit model is also presented for the purpose of comparison. The asymptotic standard errors contained in this and succeeding tables are approximations based on a first order expansion of the relevant function around the estimated parameter values and, as such, they are constructed using the entire covariance structure of the parameters.

Table III also contains the predicted probabilities for otherwise observationally equivalent nonwhite and southern workers. The second half of the table contains the differences between the predicted probabilities for nonwhites and southerners and those for workers in the base group along with the asymptotic standard errors of these differences.

It is clear from the estimated probabilities in Table III that nonwhite workers are significantly more likely to be working on a union job. This result is found both with the queuing model and with the simple probit model. The results using the queuing model suggest that differential between nonwhites and whites in their probability of unionization is due almost entirely to the significantly higher probability of nonwhites of desiring union representation. Quantitatively, nonwhites have a probability of desiring union representation which is approximately 45 percent higher (25.6 percentage points) than that for observationally equivalent whites. At the same time the conditional probability of a nonwhite being hired by a union employer given that he desires union representation is not significantly different at conventional levels from that for whites. Thus, the effective "length" of the queue for union jobs does not seem to differ significantly by race.

The results contained in Table III highlight sharp distinctions which emerge on the basis of region. Using the estimates of both the queuing model and the simple probit model, southern workers are significantly less likely to be working on union jobs than are observationally equivalent nonsouthern workers. The results using the queuing model suggest that this difference is due to two factors. First, southern workers are significantly less likely to desire union representation. The second factor is that the conditional probability of a southern worker being hired by a union employer given that he desires a union job is significantly and substantially (26 per cent) lower than that for nonsouthern workers. In other words, despite the fact that southern workers demand somewhat less unionization, the length of the queue for union jobs relative to demand is much longer in the south than outside that region. This no doubt reflects supply constraints on union jobs which may be due to a social and legal climate (typified by Right-to-Work laws common in the South) which makes union organizing and administration in the South more difficult and expensive than outside that region.

Table IV contains the predicted probabilities defined in equation (13) for base group workers in the various occupational groups. The differences in these probabilities for each occupational group relative to blue collar workers are also presented. It is clear that workers in each of the three occupational groups including clerical, service, and professional and technical workers are significantly and substantially less likely than blue collar workers to be working on union jobs. While no distinction can be drawn among the first three groups based on the simple probit results, some interesting distinctions can be drawn using the queuing model. These are discussed in turn.

Clerical workers are significantly less likely than blue collar workers to desire union representation. At the same time clerical workers who desire union representation are significantly less likely to be hired by a union employer than are blue collar workers who desire union representation. In other words the queue for union jobs is relatively longer for clerical workers than for blue collar workers. This may reflect higher costs of organizing among clerical workers as a result of market conditions or employer resistance. The conclusion to be drawn is that clerical workers are less likely to be unionized than blue collar workers as a result of both a lower desire for union representation and a relative inability to translate demand for union representation into a union job.

Service workers show a somewhat different pattern. Service workers do not differ significantly from blue collar workers in their desire for union representation. The relatively low extent of unionization among service workers is largely due to a significantly and substantially (29 per cent) lower probability of being hired by a union employer conditional on desiring a union job. Again, this relatively long queue, which reflects supply constraints on the number of union jobs, may be the result of higher costs of creating new union jobs as a result of market conditions or employer resistance. Simply put, service workers are less unionized than blue collar workers largely as a result of an inability to be hired by a union employer in spite of an equivalent demand for union jobs.

At the other extreme, professional and technical worker are significantly less likely to desire union representation than are blue collar workers. However, there is at best a weak difference between the probabilities of being hired by a union employer conditional on desiring a union job for professional and technical workers and for blue collar workers. In other words, the queues for union jobs are of relatively the same length for professional and technical workers and for blue collar workers. The conclusion to be drawn is that the lower probability of unionization of professional and technical workers is largely due to a lower desire for union representation.

Table V contains the predicted probabilities defined in equation (13) for workers in the base group of various ages. The differences in these probabilities for workers of various ages are also presented. It is clear on the basis of both the queuing model results and the simple probit results that older workers are significantly less likely to be unionized. Examination of the results of the queuing model yields the conclusion that this is due to a significantly lower probability of desiring union representation on the part of older workers. A contributing factor may be that older workers have a lower probability of being hired by a union employer conditional on desiring a union job. However, this latter conclusion must be interpreted with caution due to the fact that the hypothesis that there is no difference in this conditional probability by age can be rejected at best at the ten percent level using an asymptotic t-test. On its face the result that older workers are less likely to desire union representation seems to contradict the notion that union employers provide more fringe benefits, such as pensions, which ought to be valued more by older workers than do nonunion employers.[15] However, this result is consistent with evidence presented by Farber and Saks [8], based on an entirely different data set, which shows a similar inverse relationship between age and worker preferences for union representation.

Nonunion seniority can affect only the desire for union representation in this model. Workers with more nonunion seniority are significantly less likely to desire union representation than are workers with less nonunion seniority. To illustrate this, the probability that a worker in the base group with no nonunion seniority at age 40 desires union representation is .531, while the same probability for an otherwise equivalent worker with 10 years seniority is .429. The difference between these probabilities is .102 with an asymptotic standard error of .068. Note that the result refers to the effect of seniority on the desire for union representation on the <u>current</u>

job so that it is not caused by a reluctance of high seniority nonunion workers to quit their jobs in order to take union jobs.

The remaining set of variables relates to the educational attainment, sex, and marital status of workers. No systematic patterns emerge from the estimates regarding the relationship between these variables and the process by which the union status of workers is determined.

VI. SUMMARY AND CONCLUSIONS

In this study a model of the determination of the union status of workers was developed which differs substantially from the standard worker choice model. The decisions of both workers and potential union employers were incorporated in the model, recognizing the possibility of an excess supply of workers for existing union jobs. In this context, workers make explicit decisions regarding their desire for union representation which do not necessarily result in employment on a union job. Only if the worker is hired by a union employer out of the queue of workers who desire union representation will the worker's preference actually result in unionization. This theoretical framework results in an empirical problem of partial observability because data on union status are not sufficient to determine whether nonunion workers are nonunion because they do not desire union representation or because they were not hired by a union employer despite their preference for such a job.

In order to solve this problem without relying unduly on distributional assumptions for identification, a rather unique data set from the Quality of Employment Survey (QES) was used. These data contain information that, for nonunion workers, provides information on their current preferences for union representation. Using these data, a trivariate econometric model which accounts for the censored nature of these data as well as for the union status of workers was derived explicitly from the theoretical framework. This empirical specification embodies the separate decisions of workers and potential union employers regarding the determination of the union status of workers.

The empirical results yield some interesting insights into the process of union status determination which cannot be learned from a simple probit or logit analysis of unionization. Chief among these relate to unionization of nonwhites and southerners. The well-known fact that nonwhites are more likely to be unionized compared to otherwise equivalent whites was found largely to be the result of a greater preference for union representation. The equally well-known lower propensity to be unionized among southern workers was found to be due to a combination of a somewhat lower demand for union representation on the part of workers and a supply of unionized jobs which is substantially more constrained than outside the South relative to demand. The longer queues in the South for vacancies in existing union jobs implied by the latter result are attributed to higher costs of organization and administration of labor unions in the South. Other dimensions along which the results interpreted in the context of the model yielded behavioral insights include occupational status and age.

The model and estimates presented here have important implications for measuring the true effect of unions (as opposed to the union-nonunion differential) on such quantities as wages, turnover, and productivity. The wealth of studies (surveyed and critiqued by Freeman and Medoff [10]) that attempt to estimate this true effect rely on econometric techniques which posit that union status is determined through a single equation/single decision-maker process. To the extent that this process is inadequately modeled, the estimates of the true effects of unions which rely on them will be misleading.

To be more explicit, consider the example of the widely used Mills' ratio technique presented by Heckman [11] to correct for sample selection bias. This technique proceeds on the assumption that the log of wages, for example, is distributed normally and that union status can be modeled as determined by a simple probit. Under the assumption of joint normality of the errors, estimates can be derived for the mean of the error(s) in the wage equation(s) conditional on union status as a function of the reduced form probit estimates on union status. These estimated conditional means are the basis of the correction of the union-nonunion differential to yield estimates of the true effect of unions. This correction is crucially dependent on a range of assumptions, not the least of which is that union status can be modeled correctly as a simple univariate probit. If this particular assumption fails, then the conditional means of the wage functions will have a different form from that derived from a simple probit so that the correction will be unreliable.

It should be clear from the results of this study that the determination of union status cannot be modeled adequately as a simple probit and that an approach to estimating the true effects of unions consistent with the model developed here would be preferable. Unfortunately, the data problems outlined above make implementation of this model for such purposes difficult. As far as can be determined, only the QES has the data required to estimate the model, and previous experience with estimating union and nonunion wage equations using these data is not typical of similar experience with more widely used data sources such as the Current Population Survey or the Panel Study of Income Dynamics.[16] A topic for future research is the development of techniques for estimating models of the sort presented here which use data solely on union status and which do not rely to an undue extent on the functional form of the error distribution for identification.

Massachusetts Institute of Technology September 1982

REFERENCES

[1] Abowd, John M. and Henry S. Farber: "Job Queues and the Union Status of Workers," Industrial and Labor Relations Review 35, No. 3 (April 1982).

[2] Ashenfelter, Orley and George Johnson: "Unionism, Relative Wages, and Labor Quality in U.S. Manufacturing Industries," International Economic Review (October 1972): 488-508.

[3] Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman: "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement 3/4 (1974): 653-665.

[4] Dunlop, John T.: Wage Determination Under Trade Unions. New York: Augustus M. Kelley, Inc., 1950.

[5] Farber, Henry S.: "Individual Preferences and Union Wage Determination: The Case of the United Mine Workers," Journal of Political Economy (October 1978): 923-942.

[6] _______.: "Worker Preferences for Union Representation," Research in Labor Economics (1982), forthcoming.

[7] _______.: "The Demand for Union Representation," Working Paper No. 295, Department of Economics, M.I.T. (February 1982).

[8] _______, and Daniel H. Saks.: "Why Workers Want Unions: The Role of Relative Wages and Job Characteristics," Journal of Political Economy (April 1980): 349-369.

[9] Freeman, Richard B.: "The Effect of Trade Unionism on Fringe Benefits," Industrial and Labor Relations Review 34, No. 4 (July 1981): 489-509.

[10] ______, and James L. Medoff.: "The Impact of Collective Bargaining: Illusion or Reality?". In Stieber, McKersie, and Mills, eds., U.S. Industrial Relations 1950-1980: A Critical Assessment. Madison, Wisc.: Industrial Relations Research Association, 1981.

[11] Heckman, James J.: "Sample Selection Bias as Specification Error," Econometrica (January 1979): 153-162.

[12] Lee, Lung-Fei.: "Unionism and Wage Rates: Simultaneous Equations Model with Qualitative and Limited Dependent Variables," International Economic Review (1978): 415-433.

[13] National Labor Relations Board. Annual Report, 1979. Washington: Government Printing Office, 1979.

[14] Poirier, Dale J.: "Partial Observability in Bivariate Probit Models," Journal of Econometrics (1980): 209-217.

[15] Quinn, Robert P. and Graham L. Staines.: The 1977 Quality of Employment Survey: Descriptive Statistics, with Comparison Data from the 1969-70 and the 1972-73 Surveys. Ann Arbor, Mich.: Institute for Social Research, 1979.

[16] Raisian, John.: "Union Dues and Wage Premiums," mimeo, August 1981.

[17] Ross, Arthur M.: Trade Union Wage Policy. Berkeley: University of California Press, 1948.

[18] Schmidt, Peter and Robert P. Strauss.: "The Effects of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach," International Economic Review 17 (February 1976): 204-212.

[19] Webb, Sidney and Beatrice Webb.: Industrial Democracy. 1920. Reprint edition. New York: Augustus M. Kelley, Inc., 1965.

FOOTNOTES

1. This research was supported by grants No. SES-7924880 and No. SES-8207703 from the National Science Foundation. The author also received support from the Alfred P. Sloan Foundation as a Sloan Research Fellow. Helpful comemnts on an earlier version were received from Katharine Abraham, Rebecca Blank, Jerry Hausman, Daniel McFadden, Peter Schmidt, Robert Solow, and participants in workshops at Michigan State University, the National Bureau of Economic Research, and the University of Warwick.

2. See Freeman and Medoff [10] for an interesting summary of this literature as well as a critique from a unique perspective.

3. This analysis is not applicable to industries, such as construction, where hiring is controlled by the union through a hiring hall. Workers in such industries are excluded from both the theoretical and empirical analyses throughout.

4. Raisian [16] investigates the issue of the magnitude of union dues relative to the union-nonunion wage differential.

5. The particular set of institutions described here refer to private sector nonagricultural and nonmanagerial workers in the United States who are covered by the National Labor Relations Act (NLRA). Organization of workers not covered by the NLRA proceeds along different, but equally costly and uncertain, lines.

6. It is possible for union jobs to revert to nonunion status through an NLRB-supervised decertification election. However, these are relatively rare and can safely be ignored in this analysis. For example, according to the NLRB [13], during fiscal 1979 7266 certification elections involving 538,404 workers were officially decided while only 777

decertification elections involving 39,538 workers were officially decided.

7. In states with Right-to-Work laws, new hirees cannot be forced to join the union or pay dues, but they do share in any benefits of unionization. This issue will be raised again in interpreting the empirical results.

8. The foregoing analysis is considerably complicated by recognition that certain individual characteristics which affect skill level are determined at least in part through investment decisions made by the individual. However, explicit consideration of this factor is beyond the scope of this study, and the current assumptions that individual characteristics are determined exogenously to union status is sufficient for the problem at hand.

9. In the union sector compensation is determined through the collective bargaining process where market and other factors serve as constraints. It is beyond the scope of this study to model the determination of the compensation schedule in the union sector, though a major factor along with labor market forces is likely to be the internal political processes of the union. See the Webbs [19], Ross [17], and Dunlop [4] for early discussions of market and political forces in the determination of union bargaining goals. Farber [5] develops and estimates a simple voting model of union wage determination.

10. The assumption of a zero mean is neutral due to the presence of constant terms in the parameter vectors which capture the mean unobserved effect.

11. These models have been estimated in this context using samples from the Panel Study of Income Dynamics in excess of 1500 observations and from the Current Population Survey in excess of 19,000 observations.

12. See Quinn and Staines [15] for a detailed description of the survey design.
13. A more cumbersome notation would define u_3 as u_{1c} and u_1 and u_2 in equations [1] and [2] as u_{10} and u_{20} respectively.
14. Other variables, such as marital status, which can change over time are assumed not to vary due to lack of information on such variation.
15. See Freeman [9] for an empirical analysis of the relationship between unionization and fringe benefits.
16. See Farber [6].

TABLE I
Means (Standard Deviations) of Data
Quality of Employment Survey, 1977

Variable	Description (Dichotomous variables = 0 otherwise)	Combined Sample (n=915)	Union Sub-Sample (n=337)	Non-Union Sub-Sample (n=578)
U	= 1 if works on union job	.368	--	--
VFU	= 1 if desires union represent.	--	--	.370
Age_c	age in years	36.8 (13.1)	38.2 (12.6)	35.9 (13.3)
Sen_c	firm seniority in years	6.90 (7.49)	9.48 (8.18)	5.40 (6.60)
Age_o	$Age_c - Sen_c$	29.9 (10.8)	28.7 (9.28)	30.5 (11.5)
Fe	= 1 if female	.419	.329	.471
Marr	= 1 if married w/spouse present	.640	.709	.600
Marr*Fe	= 1 if Fe = 1 and Marr = 1	.198	.181	.208
NW	= 1 if nonwhite	.137	.160	.123
South	= 1 if worker resides in South	.353	.237	.420
Fd < 12	= 1 if <12 years education	.223	.258	.202
Ed=12	= i if = 12 years education	.364	.374	.358
12<Ed<16	= 1 if >12 years & <16 years educ.	.212	.166	.239
Ed ≥ 16	= 1 if ≥16 years education	.201	.202	.201
Blue	= 1 if occupation is blue collar	.415	.564	.317
Cler	= 1 if occupation is clerical	.205	.116	.258
Serv	= 1 if occupation is service	.156	.119	.178
Prof&Tech	= 1 if occupation is professional or technical	.234	.211	.247

TABLE II
Estimates of Union Status Model

	Queue Model G_1	Queue Model G_2	Simple Probit
Constant	.526 (.275)	-1..31 (2.65)	.364 (.181)
NW	.771 (.220)	.148 (1.70)	.316 (.134)
Fe	.252 (.164)	.345 (.780)	-.0269 (.159)
Marr	.118 (.135)	-.290 (.270)	.272 (.136)
Marr*Fe	-.264 (.195)	-.0713 (.702)	-.0571 (.197)
South	-.224 (.105))	.735 (.271)	-.542 (.0965)
Cler	-.444 (.150)	.742 (.702)	-.689 (.140)
Serv	-.148 (.152)	.782 (.290)	-.509 (.138)
Prof & Tech	-.420 (.166)	.506 (.748)	-.506 (.168)
Ed < 12	.0441 (.125)	-.179 (.234)	.0922 (.126)
12 < Ed < 16	-.138 (.119)	.149 (.323)	-.156 (.125)
Ed > 16	.174 (.161)	-.0900 (.444)	.145 (.172)
Age	-.0112 (.00434)	.0146 (.0209)	-.0141 (.00472)
Sen	-.0257 (.0174)	---	---
ρ_{12}		-.220 (3.88)	
ρ_{13}		.765 (.287)	
ρ_{23}		.241 (2.48)	
n	915		915
ln L	-897.2		-546.3

The numbers in parentheses are asymptotic standard errors. The base group consists of single white nonsouthern male blue collar workers with twelve years of education.

TABLE III

Predicted Probabilities by Race and Region

	Queue Model			Simple Probit
	Pr(DES=1)	Pr(HIRE=1|DES=1)	Pr(U=1)	Pr(U=1)
Base group	.575 (.0937)	.851 (.121)	.489 (.0524)	.477 (.0564)
Nonwhite	.831 (.0855)	.789 (.122)	.656 (.0795)	.602 (.0700)
South	.486 (.0925)	.628 (.155)	.305 (.0584)	.274 (.0531)

Predicted Differences in Probabilities by Race and Region

	Queue Model			Simple Probit
	ΔPr(DES=1)	ΔPr(HIRE=1|DES=1)	ΔPr(U=1)	ΔPr(U=1)
Nonwhite-Base group	.256 (.0619)	-.0622 (.0623)	.167 (.0663)	.125 (.0521)
South-Base group	-.0889 (.0416)	-.223 (.0669)	-.184 (.0408)	-.203 (.0347)

The numbers in parentheses are approximate asymptotic standard errors derived from a first order expansion of the relevant function around the estimated parameter values contained in Table II. The Base group consists of thirty year old, white, single, male, blue collar workers with twelve years education who live outside the south and who have no seniority.

TABLE IV
Predicted Probabilities by Occupation

	Queue Model			Simple Probit
	Pr(DES=1)	Pr(HIRE=1|DES=1)	Pr(U=1)	Pr(U=1)
Blue Collar	.575 (.0937)	.851 (.121)	.489 (.0524)	.477 (.0564)
Clerical	.399 (.0952)	.638 (.178)	.255 (.0558)	.227 (.0530)
Service	.516 (.102)	.606 (.154)	.313 (.0631)	.285 (.0586)
Professional and Technical	.408 (.107)	.722 (.190)	.295 (.0679)	.286 (.0713)

Predicted Differences in Probabilities by Occupation

	Queue Model			Simple Probit
	ΔPr(DES=1)	ΔPr(HIRE=1|DES=1)	ΔPr(U=1)	ΔPr(U=1)
Clerical- Blue Collar	-.178 (.0584)	-.213 (.107)	-.235 (.048)	-.249 (.0473)
Service- Blue Collar	-.0585 (.0603)	-.246 (.087)	-.177 (.053)	-.192 (.0493)
Professional and Technical- Blue Collar	-.167 (.0647)	-.129 (.122)	-.194 (.056)	-.190 (.0584)

The numbers in parentheses are approximate asymptotic standard errors derived from a first order expansion of the relevant function around the estimated parameter values contained in Table II. All workers are thirty year old white, single, and male with twelve years of education who live outside the south and have zero seniority.

TABLE V

Predicted Probabilities by Age

	Queue Model			Simple Probit
	Pr(DES=1)	Pr(HIRE=1|DES=1)	Pr(U=1)	Pr(U=1)
20 years	.618 (.0931)	.880 (.108)	.544 (.0519)	.533 (.0560)
30 years	.575 (.0937)	.851 (.121)	.489 (.0524)	.477 (.0564)
50 years	.486 (.100)	.782 (.156)	.380 (.064)	.367 (.0693)

Predicted Differences in Probabilities by Age

	Queue Model			Simple Probit
	ΔPr(DES=1)	ΔPr(HIRE=1|DES=1)	ΔPr(U=1)	ΔPr(U=1)
50 years- 20 years	-.133 (.0512)	-.0984 (.0824)	-.164 (.0487)	-.166 (.0539)
50 years- 30 years	-.0891 (.0346)	-.0697 (.0585)	-.110 (.0324)	-.110 (.0351)
30 years- 20 years	-.0434 (.0167)	-.0287 (.0224)	-.0547 (.0166)	-.0562 (.0188)

The numbers in parentheses are approximate asymptotic standard errors derived from a first order expansion of the relevant function around the estimated parameter values contained in Table II. All workers are white single male blue collar workers with twelve years of education who live outside the south and have zero seniority.