Whose Wages do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men

Author(s): Francis Vella and Marno Verbeek

Source: *Journal of Applied Econometrics*, Mar. - Apr., 1998, Vol. 13, No. 2 (Mar. - Apr., 1998), pp. 163-183

Published by: Wiley

Stable URL: http://www.jstor.com/stable/223257

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **13**, 163–183 (1998)

WHOSE WAGES DO UNIONS RAISE? A DYNAMIC MODEL OF UNIONISM AND WAGE RATE DETERMINATION FOR YOUNG MEN

FRANCIS VELLA[a] AND MARNO VERBEEK[b,*]

[a] *Rutgers University and IFS, Department of Economics, New Jersey Hall, New Brunswick, NJ 00903-5055, USA*
[b] *KU Leuven and Tilburg University, Center for Economic Studies, Naamsestraat 69, B-3000 Leuven, Belgium*

SUMMARY

We estimate the union premium for young men over a period of declining unionization (1980–87) through a procedure which identifies the alternative sources of the endogeneity of union status. While we estimate the average increase in wages resulting from union employment to be in excess of 20% we find that the return to unobserved heterogeneity operating through union status is substantial and that the union premium is highly variable. We also find that the premium is sensitive to the form of sorting allowed in estimation. Moreover, the data are consistent with comparative advantage sorting. Our results suggest that the unobserved heterogeneity which positively contributes to the likelihood of union membership is associated with higher wages. We are unable, however, to determine whether this is due to the ability of these workers to extract monopoly rents or whether it reflects the more demanding hiring standards of employers faced by union wages. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

Empirical studies of the union impact on wages typically attempt to estimate how observationally equivalent workers' wages differ in union and non-union employment. This is known as the *union effect*. However, as the unobserved factors that influence the sorting into union and non-union employment may also affect wages it is necessary to incorporate how the unobserved heterogeneity responsible for the union/non-union decision is rewarded in the two sectors. While the average union effect is positive the incremental individual effects, which account for the endogeneity of union employment, may make the total union effect, or *union premium*, highly variable across individuals.

Panel data studies of the union effect generally control for this endogeneity through fixed effects or alternative instrumental variables estimators (for a survey see Robinson, 1989). These procedures are inflexible in their treatment of worker heterogeneity as they generally assume the endogeneity is individual-specific and fixed. A preferable approach would decompose the endogeneity underlying union status into an individual-specific component and an individual/time-specific effect. In this paper we adopt such an approach and focus on the following issues. First, what is the impact of unions on wages and how does it vary by worker characteristics? Second, which are the primary forms of worker heterogeneity generating the endogeneity of union status?

* Correspondence to: Marno Verbeek, KU Leuven and Tilburg University, Center for Economic Studies, Naamsestraat 69, B-3000 Leuven, Belgium. E-mail: Marno.Verbeek@econ.kuleuven.ac.be

Contract grant sponsor: NWO

CCC 0883–7252/98/020163–21$17·50

Received 7 February 1995
Revised 24 January 1997

Finally, with what form of economic sorting behaviour, in terms of union and non-union employment, are the data consistent?

We examine these issues through data for young males taken from the National Longitudinal Survey (Youth Sample) for the period 1980–87. We conclude that the selection bias in union models is not solely driven by fixed individual-specific effects. Moreover, focusing only on the individual-specific effects results in underestimates of the union effect. Furthermore, the union effect is accompanied by substantial individual-specific returns. Thus, while it is positive and large the overall impact, or premium, for many individuals is small. We also find that the estimate of the union effect is sensitive to the form of sorting imposed.

We limit our focus to male youth for the following reasons. First, to illustrate the sizable variation in the union premia we examine a relatively homogeneous group. If the estimated premium is small for some male youth, a group known to enjoy a large union effect, it is likely to be highly variable for members of other groups. Second, our estimation procedure requires we model the union membership decision. We are more confident that we can do so for a relatively homogenous group.

The following section introduces a model of wage determination and union status. The estimation procedure and results are presented in Sections 3 and 4. Section 5 explores how the union premium varies with worker characteristics. Section 6 presents conclusions.

2. A MODEL OF WAGE DETERMINATION AND UNION STATUS

We attempt to explain wage variation and union membership in terms of the individuals' characteristics. This approach has the primary shortcoming of ignoring the role of the employer in determining union status, although the individuals' characteristics will enter into the employer's decision-making process. While some employer features are captured through the industry and occupational variables our failure to include employer characteristics limits our ability to assign particular effects purely to unobserved individual heterogeneity.

We assume individuals locate in union or non-union employment on the basis of wages. These are determined by observed and unobserved characteristics and their associated prices. More explicitly,

$$w_{j,it} = \beta_{j,t}' X_{it} + \alpha_{j,i} + \varepsilon_{j,it} \qquad t = 1, \ldots, T; \quad i = 1, \ldots, N \tag{1}$$

where $w_{j,it}$ represents the (potential) wage of individual i in sector j ($j = 0, 1$) in time period t, where $j = 1$ corresponds to the union sector; β is an unknown parameter vector; and X_{it} is a vector of characteristics, including time dummies. The α and ε represent the unobserved random components of the individual's wage.

Union membership is also likely to be influenced by non-pecuniary benefits and individual preferences. Furthermore, even when wages are determined on a period-by-period basis the potential gains to union membership for young workers, through future non-pecuniary benefits and potential seniority, may possess some dynamic component. Individuals located in the union sector may have additional incentives to remain in that sector independent of wage movements. This introduces state dependence.[1] Accordingly, we write a reduced-form model for the choice of sector as

$$U_{it}^* = \gamma_1' Z_{it} + \gamma_2 U_{i,t-1} + \theta_i + \eta_{it} \qquad t = 1, \ldots, T; \quad i = 1, \ldots, N \tag{2}$$

[1] A disadvantage of introducing a lagged dependent variable in the union membership equation is that it introduces the 'initial conditions' problem. We address this issue below.

$$U_{it} = I(U_{it}^* > 0) \qquad t = 1, \ldots, T; \quad i = 1, \ldots, N \tag{3}$$

$$w_{it} = w_{j,it} \quad \text{if } U_{it} = j \qquad t = 1, \ldots, T; \quad i = 1, \ldots, N \tag{4}$$

where U_{it}^* is a latent variable capturing the benefits of union employment; w_{it} is the log of the actual hourly real wage rate; U_{it} denotes the sector chosen and is a dummy variable indicating that individual i's wage in period t was determined through collective bargaining; the γ's denote unknown parameters; and Z_{it} is a vector of exogenous variables. The composite error term captures the unobserved heterogeneity driving union membership.

While the above model assumes that individuals locate in the sector in which they prefer, union membership is also determined by the employer's willingness to hire the worker. Accordingly we interpret equation (2) as the reduced-form representation of the employer's and employee's decisions.

The random components in equations (1) and (2), respectively, correspond to an individual-specific effect and an individual/time-specific effect. We assume these are i.i.d. drawings from a multivariate normal distribution, where the effects from different equations are potentially correlated. In particular, we allow the four covariances $\sigma_{j,\alpha\theta}$ and $\sigma_{j,\varepsilon\eta}$ to be non-zero. The co-variances between the effects in the two wage equations are left unspecified, while all other covariances are set equal to zero. These covariances indicate that the random components in the wage equation are potentially correlated with those in the membership equation. This generates the potential endogeneity of union status in the wage equation.

The covariances provide insight into the form of sorting (see Roy, 1951). For simplicity, consider where the endogeneity operates purely through the individual effects $\alpha_{j,i}$ and θ_i. First note, however, that the θ_i's are constructed below such that their average value for the union sector is positive while their average value for the non-union sector is negative. If either covariance between α and θ is non-zero then the unobserved factors that influence union membership also affect wages. If both covariances are positive the individuals with high values of θ are, on average, the 'better' workers, in terms of their endowment of unobserved productivity, irrespective of which sector they are located. This is referred to as *hierarchical* sorting. An extreme case of this is $\alpha_{0,i} = \alpha_{1,i}$, i.e. when unobserved factors are identical across sectors. Another outcome is where workers perform differently in the two sectors and locate accordingly. That is, there is a negative correlation between relative productivity in the union sector and non-union sector. This is known as *comparative advantage* or *positive* sorting. As this requires that the contribution of the unobserved heterogeneity increases wages in both sectors it is necessary that $\sigma_{1,\alpha\theta}$ is positive and $\sigma_{0,\alpha\theta}$ is negative.

By restricting the returns to the observed characteristics to be invariant across time and sector we can write the wage equation as

$$w_{it} = \beta' X_{it} + \delta U_{it} + e_{it} \tag{5}$$

where $e_{it} = U_{it}(\alpha_{1,i} + \varepsilon_{1,it}) + (1 - U_{it})(\alpha_{0,i} + \varepsilon_{0,it})$, and δ captures the union effect.

3. ESTIMATION

If the heterogeneity generating the endogeneity of union status operates only through the individual 'fixed' effects $\alpha_{j,i}$'s, which are further restricted to be identical for the two sectors, the union effect can be consistently estimated via the fixed effects estimator (see Jakubson, 1991 for a recent example). Alternative instrumental variables procedures (see, for example, Hausman and Taylor, 1981, and Amemiya and MaCurdy, 1986) can also be employed. However, these procedures, in their standard form, do not allow for endogeneity operating through the other error components. They also restrict the unobserved heterogeneity to be identical across the two sectors. Robinson (1989) and Vella and Verbeek (1993) show this imposes that the ordering of the workers' productivity within each sector is invariant to sector.[2]

An appealing alternative is described in Heckman (1979) and Lee (1978).[3] We employ a similar approach while exploiting the panel nature of the data. We derive estimates of the unobserved heterogeneity underlying union status to include in the wage equation to account for the endogeneity of union status. Following Vella and Verbeek (1996), we rewrite equation (5) conditional on the vector U_i, of length T, containing the union status of individual i in each period:[4]

$$E[w_{it} \mid U_i] = \beta' E[X_{it} \mid U_i] + \delta E[U_{it} \mid U_i] + E[\alpha_{j,i} \mid U_i] + E[\varepsilon_{j,it} \mid U_i] \tag{6}$$

To obtain estimates of the unobserved heterogeneity requires estimation of the sectorial choice equation, which is a dynamic random effects probit model with likelihood function

$$\prod_i \int \prod_t \Phi\left(\frac{\gamma' W_{it} + \theta}{\sigma_\eta}\right)^{u_{it}} \Phi\left(-\frac{\gamma' W_{it} + \theta}{\sigma_\eta}\right)^{1-U_{it}} \frac{1}{\sigma_\theta}\phi(\theta/\sigma_\theta)\, \mathrm{d}\theta \tag{7}$$

where $\gamma = (\gamma_1', \gamma_2)'$, $W_{it} = [Z_{it}, U_{i,t-1}]$ and Φ and ϕ denote the cumulative probability function and the density function of the standard normal distribution, respectively.

Given the presence of the individual-specific effects θ_i one cannot validly assume that union status in the first period is truly exogenous. We employ an approximate solution, suggested by Heckman (1981), in which the reduced-form marginal probability of the initial state is approximated by a probit function using all pre-sample information on the exogenous variables. We do not impose any restrictions on the relationship between the structural parameters and those from the approximate reduced form for the initial state.[5] Assuming that the model is exact as $N \to \infty$, the maximum likelihood estimator from equation (7) augmented with a reduced form for the initial state is consistent for γ, σ_θ^2 and σ_η^2, provided some normalization is employed.

For the conditional expectations in equation (6) it can be shown that[6]

$$E[\alpha_{j,i} \mid U_i] = \sigma_{j,\alpha\theta}\left[\frac{T}{\sigma_\eta^2 + T\sigma_\theta^2} E[\bar{v}_{i.} \mid U_i]\right] = \sigma_{j,\alpha\theta} C_i \tag{8}$$

$$E[\varepsilon_{j,it} \mid U_i] = \sigma_{j,\varepsilon\eta}\left[\sigma_\eta^{-2} E[v_{it} \mid U_i] - \frac{T\sigma_\theta^2}{\sigma_\eta^2(\sigma_\eta^2 + T\sigma_\theta^2)} E[\bar{v}_{i.} \mid U_i]\right] = \sigma_{j,\varepsilon\eta} C_{it} \tag{9}$$

where the C_i and C_{it} are defined above; $v_{it} = \theta_i + \eta_{it}$ and $\bar{v}_{i.} = (1/T)\sum_{t=1}^{T} v_{it}$.

[2] While Robinson (1989) shows that instrumental variables rules out positive (or negative) selection in both sectors. Vella and Verbeek (1993) show that it imposes the degenerate hierarchical sorting described above. It is straightforward to verify that this is imposed in fixed-effects estimation.

[3] For applications of this approach in the cross-sectional case see, for example, Willis and Rosen (1979) and Heckman and Sedlacek (1985).

[4] All conditional expectations below are also conditional upon the exogenous variables in Z_{it}, which is assumed to include X_{it}.

[5] Monte Carlo evidence in Heckman (1981) suggests that this solution is relatively successful.

[6] The derivation of these terms is provided in the Appendix.

From our joint normality assumption the conditional expectations are linear in the covariances. The remaining expressions in equations (8) and (9) are known functions of the parameters in the union model equation (2).[7] With estimates of the conditional expectations we obtain a form of equation (5) that can be estimated by least squares. The normality assumption allows us to express the latent effects in the wage equation as linear functions of the random latent effects in the union equation. However, following Lee (1984), Gallant and Nychka (1987) and Pagan and Vella (1989) we capture departures from normality in equation (5) by expressing the latent effects in the wage equations as higher-order functions of the latent effects from the union equation.

The assumptions regarding the errors identify all parameters in equation (5) since the correction terms given in equations (8) and (9) are non-linear functions of the exogenous variables and observed values from other periods. However, the imposition of exclusion restrictions is desirable. Since these are usually disputable and frequently internally inconsistent with the economics of union membership (see Vella and Verbeek, 1993), the only variable excluded from the wage equation is lagged union status. This identifies equation (5) provided γ_1 is different from zero.

Consider how the lagged value of union status influences the union decision while not affecting the current wage. First, lagged union status may capture movement costs. This is not specific to union employment but reflects that individuals are only likely to change union status if they change jobs. Accordingly, lagged union status will influence current status while having no impact on wages. Second, union employment often produces long-term advantages, thereby generating a commitment to union employment. As the impact of these long-term benefits on wages, inasmuch as they represent compensating differentials, are likely to be small it would seem that lagged union status will have a minor impact on current wages. Finally, if the union premium is large there may be queuing for union employment. A predictor of whether one is able to acquire union employment in the current period is whether one was able to do so in the previous period.

4. EMPIRICAL RESULTS

We now present the empirical results. The data, taken from the National Longitudinal Survey (Youth Sample), comprise a sample of full-time working males who have completed their schooling by 1980 and then followed over the period 1980 to 1987. We exclude individuals who fail to provide sufficient information to be included in each year leaving a sample of 545 observations. The summary statistics for the total period are reported in Table I. Union membership is based on the question reflecting whether or not the individual had his wage set in a collective bargaining agreement.[8] This measure displays variation over the period indicating movement in and out of union membership.[9] The unconditional union premium is around 15%.

Before proceeding, consider our treatment of industry and occupational choice and labour supply. Industry of employment is treated as exogenous. While we believe that workers match with employers on the basis of industry-specific skills and technology we argue that this sorting

[7] The forms of these functions are given in the Appendix and involve one-dimensional numerical integration.
[8] We will refer to those who responded yes to this question as being union members.
[9] About 50% of the individuals in our sample changed union status at least once between 1980 and 1987. Overall, the sample shows a decline in the unionization rate in the private sector from 25% in 1980 to 22% in 1985, and 21% in 1986. In 1987 an increase to 26% was observed.

Table I. Descriptive statistics, 1980–87

Variable	Definition	Mean	Standard deviation
School	Years of schooling	11·76	1·75
Exper	Age-6-School	6·51	2·83
Exper2	Experience Squared	50·42	40·78
LogExper	Log(1 + Experience)	1·94	0·42
Union	Wage set by collective bargaining	0·24	0·43
Mar	Married	0·44	0·50
Black	Black	0·12	0·32
Hisp	Hispanic	0·16	0·36
Health	Has health disability	0·02	0·13
Rural	Lives in rural area	0·20	0·40
NE	Lives in North East	0·19	0·39
NC	Lives in Northern Central	0·26	0·44
S	Lives in south	0·35	0·48
Wage	Log of hourly wage	1·65	0·53
EWage	Hourly Wage ($)	5·91	3·20
Wdif	Union differential	0·87	
Industry dummies			
AG	Agricultural	0·03	
MIN	Mining	0·02	
CON	Construction	0·08	
TRAD	Trade	0·27	
TRA	Transportation	0·06	
FIN	Finance	0·04	
BUS	Business & Repair Service	0·08	
PER	Personal Service	0·02	
ENT	Entertainment	0·02	
MAN	Manufacturing	0·28	
PRO	Professional & Related Service	0·08	
PUB	Public Administration	0·04	
Occupational dummies			
OCC1	Professional, Technical and kindred	0·10	
OCC2	Managers, Officials and Proprietors	0·09	
OCC3	Sales Workers	0·05	
OCC4	Clerical and kindred	0·11	
OCC5	Craftsmen, Foremen and kindred	0·21	
OCC6	Operatives and kindred	0·20	
OCC7	Laborers and farmers	0·09	
OCC8	Farm Laborers and Foreman	0·01	
OCC9	Service Workers	0·12	

mechanism is more relevant for relatively experienced workers.[10] It does not seem plausible, however, to make a similar argument for occupational status and this raises some difficulties. Ideally, we would account for this endogeneity and include occupation in both the union membership and wage equations. However, this would complicate our analysis considerably.

[10] Murphy and Topel (1987), Jovanovic and Moffitt (1990) and Topel and Ward (1992) find that workers experience multiple short-lived jobs early in their careers before sorting into longer-term employment arrangements after 10 to 15 years.

Accordingly, we can either exclude occupational status or include it and treat it as exogenous. We adopt the first strategy as we prefer to allow the occupational effects to be captured through their correlation with the included exogenous variables. However, as occupational status is a measure of ability, and could contaminate our conclusions regarding the role of unobserved heterogeneity, we ensure that our final results are robust to the inclusion of occupational status. We also do not address the issue of attrition bias. This is not an oversight as our aim is to illustrate the variation of the union effect for a homogeneous group. However, the conclusions that follow are conditional on continuous employment throughout the sample period.

4.1. The Model for Union Membership

The estimates for the dynamic random effects probit model, with and without occupational status, are reported in Table II.[11] Focus first on those which exclude occupation. Several of the explanatory variables have a statistically significant impact on the probability of union membership. The negative coefficients on the industry dummies reflect the sizable unionization rate in the omitted group which is the public sector. The time effects display an increasingly negative pattern, consistent with the aggregate data which indicate sizable decreases in unionization over this period. The individuals in our data, however, display a weaker tendency to leave unions than is revealed by aggregate data. The coefficients on the time dummies indicate that the time effect on union membership is negative in spite of the fact that union membership did not show a substantial decline for these data over this period. This indicates that the experience effects for our sample partly offset the overall economy's tendency to leave unions. Although we can separately identify the aggregate time effects from an experience effect by including a logarithmic transformation of experience rather than linear and quadratic terms, this is not very satisfactory. We employ this specification, however, to allow the two effects to operate separately.

The highly significant estimate of 0·611 for the coefficient on lagged union status indicates a substantial degree of positive state dependence. The estimate for σ_θ^2 of 0·57 indicates that 57% of the total variance is due to across individual variation.[12] The coefficients on the dummy variables denoting that the individual is black or hispanic are both positive and statistically significant and are large in magnitude. This is consistent with earlier studies. These groups may choose to bargain through union membership rather than on an individual basis. They may also experience a higher degree of labour market discrimination and may reduce its impact via union membership. Few of the other individual-related variables appear to have a statistically significant impact on union membership. There does appear, however, to be some role for marital status and regional differences.

Table II also contains the corresponding results while including eight dummy variables to capture occupational status. They indicate that occupational status does appear to influence the probability of union membership. Individuals in the blue-collar industries display a higher propensity to acquire union membership.

For the model excluding the occupational variables the conditional moment tests of normality of θ_i and η_{it} resulted in values of 0·90 and 4·30, respectively. Under the null hypothesis the test statistics are Chi-squared distributed with 2 degrees of freedom and, consequently, we do not take

[11] We do not report estimates for the parameters of the reduced-form model for the initial state as they have no direct interpretation.
[12] The normalization used is $\sigma_\theta^2 + \sigma_\eta^2 = 1$.

Table II. Random effects probit estimates of union membership

Variable	Estimate	(st. error)	Estimate	(st. error)
Constant	−0·801	(0·484)	−1·132*	(0·513)
Lagged union status	0·611*	(0·073)	0·632*	(0·077)
LogExper	0·222	(0·164)	0·292	(0·173)
School	−0·001	(0·028)	0·033	(0·030)
Mar	0·114*	(0·056)	0·122*	(0·060)
Black	0·477*	(0·133)	0·423*	(0·133)
Hisp	0·291*	(0·133)	0·249	(0·132)
Rural	0·002	(0·080)	−0·017	(0·085)
Health	−0·213	(0·149)	−0·261	(0·167)
NE	0·232*	(0·118)	0·192	(0·121)
S	−0·009	(0·113)	−0·003	(0·116)
NC	0·161	(0·100)	0·093	(0·104)
Industry dummies				
AG	−0·517*	(0·150)	−0·510*	(0·207)
MIN	−0·104	(0·184)	−0·065	(0·199)
CON	−0·435*	(0·131)	−0·432*	(0·148)
MAN	−0·148	(0·101)	−0·095	(0·120)
TRA	−0·046	(0·115)	−0·020	(0·133)
TRAD	−0·538*	(0·104)	−0·476*	(0·118)
FIN	−1·143*	(0·243)	−1·004*	(0·249)
BUS	−0·762*	(0·139)	−0·724*	(0·159)
PER	−0·566*	(0·211)	−0·562*	(0·237)
ENT	−0·291	(0·180)	−0·256	(0·192)
PRO	−0·151	(0·114)	−0·061	(0·125)
Occupational dummies				
OCC1	=		−0·493*	(0·101)
OCC2	=		−0·589*	(0·128)
OCC3	=		−0·644*	(0·143)
OCC4	=		−0·226*	(0·101)
OCC5	=		−0·162	(0·096)
OCC6	=		−0·155	(0·092)
OCC7	=		−0·013	(0·107)
OCC8	=		−0·119	(0·243)
Time dummies				
1981	−0·328*	(0·087)	−0·338*	(0·091)
1982	−0·346*	(0·104)	−0·353*	(0·109)
1983	−0·464*	(0·132)	−0·487*	(0·139)
1984	−0·485*	(0·148)	−0·496*	(0·154)
1985	−0·628*	(0·169)	−0·646*	(0·175)
1986	−0·709*	(0·186)	−0·736*	(0·190)
1987	−0·482*	(0·188)	−0·491*	(0·196)
σ_θ^2	0·567*	(0·039)	0·541*	(0·040)
Log-likelihood value	−1537·47		−1512·99	

our results as evidence against the null. The joint test on normality of both components, which corresponds to a χ^2 with 4 degrees of freedom, yields the insignificant value of 5·65. The model containing the occupational variables showed slightly more evidence of non-normality with test statistics of 1·10 and 6·74 for the univariate tests, and 8·49 for the two-variate normality test.

Table III. Wage regressions with union effects

Variable	[1] OLS	[2] OLS	[3] FE	[4] FE	[5] OLS	[6] OLS	[7] OLS	[8] OLS
Constant	0·224	0·388*			0·273	0·360*	0·282	0·363*
	(0·128)	(0·158)			(0·156)	(0·158)	(0·156)	(0·159)
Union	0·146*	0·177*	0·079*	0·080*	0·392*	0·389*	0·285*	0·311*
	(0·026)	(0·026)	(0·018)	(0·018)	(0·087)	(0·084)	(0·088)	(0·085)
School	0·090*	0·073*			0·083*	0·070*	0·082*	0·070*
	(0·008)	(0·010)			(0·010)	(0·010)	(0·010)	(0·010)
Exper	0·076*	0·057*	0·112*	0·111*	0·051*	0·049*	0·053*	0·050*
	(0·011)	(0·018)	(0·008)	(0·008)	(0·018)	(0·018)	(0·018)	(0·018)
Exper2	−0·0022*	−0·0018	−0·0041*	−0·0041*	−0·0016	−0·0015	−0·0017	−0·0016
	(0·0008)	(0·0009)	(0·0005)	(0·0006)	(0·0009)	(0·0009)	(0·0009)	(0·0009)
Hisp	−0·059	−0·047			−0·079	−0·061	−0·063	−0·049
	(0·042)	(0·042)			(0·042)	(0·042)	(0·042)	(0·042)
Black	−0·155*	−0·126*			−0·189*	−0·154*	−0·171*	−0·141*
	(0·044)	(0·044)			(0·046)	(0·045)	(0·046)	(0·046)
Rural	−0·131*	−0·114*	0·050	0·048	−0·131*	−0·113*	−0·131*	−0·116*
	(0·031)	(0·032)	(0·032)	(0·027)	(0·032)	(0·032)	(0·032)	(0·032)
Mar	0·110*	0·102*	0·040*	0·038*	0·102*	0·094*	0·107*	0·097*
	(0·024)	(0·024)	(0·017)	(0·017)	(0·024)	(0·024)	(0·024)	(0·024)
Health	−0·058	−0·032	−0·017	−0·010	−0·036	−0·011	−0·037	−0·011
	(0·062)	(0·062)	(0·044)	(0·044)	(0·062)	(0·042)	(0·062)	(0·042)
C_i	=	=			−0·050*	−0·030	−0·051*	−0·033
					(0·024)	(0·023)	(0·025)	(0·023)
C_{it}	=	=	=	=	−0·109*	−0·113*	−0·072*	−0·084*
					(0·031)	(0·031)	(0·032)	(0·032)
C_i^2	=	=			=	=	0·034*	0·027*
							(0·013)	(0·012)
C_{it}^2	=	=	=	=	=	=	−0·0005	−0·0008
							(0·0058)	(0·0061)
Adj R^2	0·260	0·274	0·186	0·187	0·264	0·279	0·268	0·281

Note: Standard errors in parentheses. All regressions include industry and region dummy variables; all, except FE, include time dummies; in addition, columns [2], [4], [6] and [8] include occupational dummies (in both steps where applicable).

4.2. The Wage Equation Under Hierarchical Sorting

Columns [1] and [2] of Table III report the estimates from equation (5) including and excluding the eight occupational variables. The estimated union effects are 15% and 18%, respectively, and, given the evidence in Robinson (1989), appear low. These estimates, however, are contaminated by the endogeneity.

Columns [3] and [4] present the fixed effects estimates of equation (5).[13] These are consistent with previous results, (see, for example, Angrist and Newey, 1991; Jakubson, 1991), and show the estimated union effect falls markedly to 7·9% and 8·0%, respectively. Recall the shortcomings of this approach. First, the fixed-effect approach only eliminates the endogeneity operating through the individual-specific effects $\alpha_{j,i}$'s. Thus, any time-varying endogeneity continues to contaminate our estimates. Second, even if the relevant unobserved heterogeneity is individual-specific and

[13] These within regressions do not include time dummies. When included, all time effects were insignificant, both individually and jointly.

time invariant the fixed-effects estimator requires that the price of the heterogeneity is invariant to sector. This imposes that the covariance $\sigma_{\alpha\theta}$ is constant across sector.

Columns [5] and [6] report the estimates when the correction terms are included, based on the estimated union model without and with the occupational dummies, respectively. The coefficients on the union dummy increase dramatically to 0·39, reflecting a union effect of 48%. While these estimates are in the range of previous estimates they seem somewhat high.[14] The corrections terms are statistically significant and negative.[15] This indicates that both the individual and individual/time effects influence the union effect. This questions the appropriateness of the fixed-effects procedure. Moreover, it appears that the large union effects are due to the individual-specific time effects.[16] We next include powered-up correction terms to capture the possibility of non-normality. The order of the non-linear terms was chosen by cross-validation (CV). The CV criterion is the sum of squares of prediction errors from predicting each observation using coefficient estimates based on all other observations.[17] Table IV gives the CV criteria for values of k, the power of the highest order, between 0 and 7. The k which minimizes the CV criterion is between 2 and 5. Since the criterion value for $k = 2$ is close to the minimal value, and the differences in point estimates for larger values of k are small, we choose $k = 2$. This preserves degrees of freedom and reduces collinearity.

Table IV. Cross-validation for order of correction terms

Highest order k	Occupation dummies excluded		Occupation dummies included	
	Estimated union effect	CV value	Estimated union effect	CV value
0	0·148	921·5	0·177	901·1
1	0·392	917·4	0·389	897·9
2	0·286	912·7	0·311	896·9
3	0·285	912·6	0·349	896·8
4	0·323	912·0	0·346	896·2
5	0·278	912·6	0·336	896·9
6	0·266	913·8	0·309	897·4
7	0·274	915·3	0·272	897·5

The estimates with linear and squared correction terms included are shown in columns [7] and [8] of Table III. The union dummy coefficients are now 0·28 and 0·31. The coefficients on the included correction terms are jointly statistically significant, revealing selection bias, while the higher-order terms indicate non-normality.

The statistically significant and negative coefficients on the selection terms indicate that the workers who receive lower wages, after conditioning on their characteristics and in the absence of unions, are those most likely to be in the union. This is consistent with the findings of Heywood (1990) that minorities displayed a greater tendency to queue for union jobs than whites. If these

[14] Robinson (1989) reports an estimate of 43% and Linnemann and Wachter (1986) present estimates in excess of 50%.

[15] Standard errors in Tables III and V are computed taking into account the covariance structure of the error terms and, for the two-step results, using the appropriate formulae in Newey (1984). Ignoring the first-stage estimation produces standard errors that underestimate the correct standard errors by between 1% and 11%.

[16] We report only the results from models where we treat the time effects as fixed. An alternative approach is to treat them as random. When we did so there were no substantive differences in our results. However, under the random time effects assumption, the covariance between the random time components in the two equations is identified. Without exception, this covariance was estimated to be negative.

[17] A discussion on the optimality of several cross-validation criteria is given in Andrews (1991).

groups are lower paid for discriminatory reasons they may seek union employment. It is also consistent with Robinson (1989), who concluded that there was no support for the popular argument that better workers are chosen from a queue by the union. It is possible, however, that the less productive workers queue to join the union and the union then chooses the better workers.

While it is possible to construct an argument supporting the existence of negative coefficients on the basis of individual behaviour they are somewhat troublesome if a role for the employer is incorporated. While the less productive employees may pursue union employment, in order to obtain some share of any accrued monopoly rents, it seems unlikely that the employers will hire them. Accordingly, we explore the effect of relaxing this restricted form of sorting. These results are reported in Table V and discussed in the following sub-section.

4.3. The Wage Equation under Unrestricted Sorting

As there is no evidence that the higher-order terms of the latent effects are statistically significant, we report, in columns [2] and [4] of Table V, the results with the quadratic terms included, and we do not employ cross-validation to determine the length of the polynomial. We focus on columns [1] and [3] under the assumption of normality. The non-normality detected in Table III appears to be an artifact of the restriction that the unobserved heterogeneity was equally valued in each sector.

The primary feature of Table V is related to our 'preferred' specifications in columns [1] and [3]. The restriction that the random components are equally rewarded in each sector is rejected. First, focus on the individual-specific effects denoted by C_i. Table VI, which reports the descriptive statistics for the random components, indicates that due to the signs on the coefficients in Table V the average contribution of the individual-specific components to each sector is positive. Thus, sorting into union and non-union employment appears to be done on the basis of comparative advantage. That is, individuals are located in the sector where the price associated with their C_i increases their wage. As the C_i's clearly display differently signed coefficients for the two-sectors estimation of this model by instrumental variables or fixed (individual) effects is inappropriate.

An examination of Table V, however, reveals that the sign of the coefficients for the C_{it}'s remains negative for both sectors. Moreover, the descriptive statistics in Table VI reveal that the average contribution to the union wage operating through this effect is negative although it is positive for the non-union sector. This is consistent with the discussion in the previous section which supported the conjecture that the union increases the wages of those who would be relatively lower paid in their absence. However, given the conflicting impact of these two random components we evaluate the total impact of union membership on each individual's wage in the following section.

Finally, the relaxation of the restrictions on the coefficients for the random components has a substantial impact on the estimated union effect. For the model without occupational dummies the estimated union coefficient is 0·21 while that for the model including the occupational variables is 0·23.

Why does the hierarchical sorting pattern increase the estimated union effect so drastically? Table V indicates that the appropriate return to the fixed individual effect is approximately 0·04, noting that the majority of union workers have positive values for this random component. The corresponding return for the non-union workers is $-0{\cdot}06$, noting that the majority in this sector have negative values. Accordingly, when we examine two observationally equivalent workers,

Table V. Wage regressions with union effects

Variable	Occupation excluded [1]	Occupation excluded [2]	Occupation included [3]	Occupation included [4]
Union	0·214*	0·312*	0·232*	0·341*
	(0·102)	(0·108)	(0·097)	(0·106)
School	0·082*	0·082*	0·070*	0·070*
	(0·010)	(0·010)	(0·010)	(0·010)
Exper	0·053*	0·053*	0·051*	0·050*
	(0·018)	(0·018)	(0·018)	(0·018)
Exper2	−0·0017*	−0·0017	−0·0016	−0·0015
	(0·0009)	(0·0009)	(0·0009)	(0·0009)
Hisp	−0·065	−0·059	−0·048	−0·044
	(0·042)	(0·043)	(0·042)	(0·043)
Black	−0·167*	−0·169*	−0·136*	−0·138*
	(0·046)	(0·047)	(0·046)	(0·046)
Rural	−0·131*	−0·130*	−0·115*	−0·114*
	(0·032)	(0·032)	(0·032)	(0·032)
Mar	0·106*	0·107*	0·098*	0·099*
	(0·024)	(0·024)	(0·024)	(0·024)
Health	−0·037	−0·035	−0·012	−0·009
	(0·062)	(0·062)	(0·062)	(0·063)
C_i	−0·060*	−0·050	−0·039	−0·037
	(0·026)	(0·028)	(0·025)	(0·027)
C_{it}	−0·090*	−0·060	−0·093*	−0·080
	(0·040)	(0·100)	(0·040)	(0·102)
$C_i * Union$	0·103*	−0·104	0·093*	−0·075
	(0·045)	(0·079)	(0·042)	(0·074)
$C_{it} * Union$	0·029	0·026	0·022	−0·031
	(0·043)	(0·153)	(0·042)	(0·147)
C_i^2	=	0·021	=	0·006
		(0·022)		(0·021)
C_{it}^2	=	0·007	=	0·005
		(0·040)		(0·042)
$C_i^2 * Union$	=	0·050	=	0·052
		(0·041)		(0·038)
$C_{it}^2 * Union$	=	−0·025	=	0·001
		(0·046)		(0·047)
Adj.R^2	0·266	0·268	0·281	0·282

Note: Standard errors in parentheses. All regressions include a constant, industry, region and time dummies; in addition columns [3] and [4] include occupational dummies (in both steps).

one union and one non-union, with values for C_i of zero, the predicted log wage difference is approximately 0·21 and this represents the union effect if the other random effects have no contribution. Now consider the implication of imposing the same trade-off for both sectors. As the majority of workers are non-union the least squares criteria generates a return of −0·05. While this has no substantial impact on the non-union sector it completely distorts the returns in the union sector. In fact, union members with relatively low fixed individual effects will have predicted wages far above their actual wages. Accordingly, a union member with a fixed effect of zero will have a predicted wage above his actual wage. As the union effect is evaluated at this point, it is consequently overestimated.

Table VI. Descriptive statistics of latent effects by union membership

	Union members			Non-union members		
Model with occupation excluded:						
Latent effect	Mean	Minimum	Maximum	Mean	Minimum	Maximum
C_i	1·332	−0·875	2·994	−0·429	−1·740	2·708
C_{it}	0·955	0·095	3·693	−0·308	−2·889	−0·003
C_i^2	2·367	0·000	8·968	0·750	0·000	7·335
C_{it}^2	1·515	0·009	13·64	0·267	0·000	8·346
Model with occupation included:						
Latent effect	Mean	Minimum	Maximum	Mean	Minimum	Maximum
C_i	1·332	−1·003	3·100	−0·430	−1·726	2·853
C_{it}	0·922	0·098	3·852	−0·298	−2·754	−0·002
C_i^2	2·411	0·000	9·622	0·764	0·000	8·139
C_{it}^2	1·420	0·010	14·84	0·252	0·000	7·584

Note: The table contains the sample averages, minima and maxima of the (squared) correction terms included in the wage equation.

While the restriction that the fixed individual specific random components are equally valued across sectors increases the union effect the estimate in column [1] of Table V, which relaxes this restriction, is greater than that in column [2] of Table III which imposes it. This is due to the substantial impact on the union effect operating through the individual-specific time effects. This further highlights the inadequacy of the fixed effects estimator. It also partially explains the conclusion by others (see Robinson, 1989, for example), that the longitudinal estimates of the union impact are smaller than those from cross-sectional studies.

The comparative advantage sorting result, operating through the C_i's, suggests that individuals most likely to join unions benefit most from doing so. This does not necessarily imply they are more productive in union employment. It simply indicates that they do relatively better in the union sector. One possibility is that the union premium indicates that these workers prefer to rely more on any monopoly power that results from unionization rather than their individual skills to gain higher wages.

While the above explanation highlights the ability of the workers to capture some monopoly rents through unionization there is another explanation consistent with the results. When unions are able to impose higher wages the employers are likely to impose more demanding hiring standards on their workers. Accordingly, those individuals who have more productive capabilities, after controlling for their observed characteristics, are more likely to gain union employment. This is also consistent with the queues one observes for union employment.

While the comparative sorting story for the union sector reflects either the workers' ability to accrue rents or the more demanding hiring standards of the employers faced by union wages the sorting result for the non-union sector probably reflects that individuals who do not have the characteristics conducive to union employment do better in the non-union sector. This result may reflect our inability to capture other measures of productivity. It is thus useful to consider column [3] of Table V, which incorporates a role for occupational status.

The primary feature of the estimates in column [3] is their similarity to those in column [1]. The notable differences are in the coefficients capturing the return to education and those reflecting the race of the worker. This is not unexpected as these variables are likely to be correlated with occupational status. However, while these coefficients display some changes the coefficients related to the impact of union membership on wages and the role of the estimates of the unobserved heterogeneity remain unchanged. The inclusion of the occupational variables reduces the magnitude of the coefficients on the individual effects. This indicates that some component of the unobserved ability determining union status is correlated with occupational status.

4.4. Specification Issues

The probability of union employment is treated as a function of the single index appearing in the union equation. An alternative approach, discussed in Abowd and Farber (1982), suggests that the sorting follows a multiple indices rule. For example, perhaps the relatively productive individuals do not join the union while the less productive pursue union jobs. However, among those seeking union employment there are those who are unsuccessful as the employers choose the better workers. This suggests that the use of a single index, and the manner it is generated, are inappropriate. However, provided the correct exogenous variables are included in the index an incorrect selection model will only impose incorrect weights for the exogenous variables in the construction of the index. Accordingly, a higher-order polynomial of the single index will partly capture the true random effects.

To explore this issue we interacted the correction term, and its higher-order values, with the level of schooling and entered these values in equation (5). We chose schooling as the variable of interaction as union effects often differ by education level. While the interaction effects were statistically significant the estimated union effects were similar to those in Table V. Accordingly, we are confident that our single-index approach is close to the true model. It is likely that the issue of multiple indices is more troublesome when the group is more heterogeneous. For example, Card (1996) shows the union effect varies significantly by location in the income distribution. This indicates a multiple indices model should be employed when analysing the impact of union membership over a more heterogeneous sample.

We also estimated the model as a switching regression, similar to that of Lee (1978), while accounting for the panel nature of the data. The relaxation of the restriction that the coefficients on the full set of regressors are identical across sectors had no major impact on the results. Accordingly, we continue to focus on the more parsimonious representation in order to directly evaluate the union effect.

To examine how well our model fits the data we followed the two-step idea from our estimation approach. To see whether our model tracks the dynamics of union membership well, we computed the implied conditional probabilities of being in the union sector, given each individual's history, for various classes of individuals.[18] Table VII presents the average probabilities as well as the actual relative numbers of union members in each of the cells, for the model which excludes occupation. The results for the model with the occupational variables are only marginally different and therefore not reported. Overall, the probability of being in the union sector, conditional upon being a union-worker in the previous period, is 72%, while for

[18] To compute the conditional probabilities of being in the union sector we use

$$P\{U_{it} = j \mid U_{i,t-1} = j_{t-1}, \ldots, U_{i1} = j_1\} = \frac{P\{U_{it} = j, U_{i,t-1} = j_{t-1}, \ldots, U_{i1} = j_1\}}{P\{U_{i,t-1} = j_{t-1}, \ldots, U_{i1} = j_1\}}, \tag{10}$$

where the probabilities in the latter term can be computed along the lines of equation (7).

Table VII. Average estimated probabilities and actual relative cell frequencies of union membership 1981–87. Model excluding occupational dummies

Characteristic	Lagged members			Lagged non-members		
	Prob.	Freq.	*N*	Prob.	Freq.	*N*
(a) Years of schooling						
< = 9 years	0·709	0·750	72	0·087	0·074	243
10 years	0·730	0·754	69	0·125	0·119	260
11 years	0·726	0·717	138	0·096	0·113	506
12 years	0·717	0·743	335	0·119	0·119	1282
13 years	0·730	0·705	78	0·143	0·127	300
14 years	0·735	0·649	57	0·135	0·122	230
15 years	0·693	0·702	47	0·092	0·088	170
16 years	0·779	0·875	8	0·082	0·050	20
(b) Race						
White	0·723	0·718	560	0·123	0·123	2219
Black	0·703	0·756	86	0·100	0·093	355
Hispanic	0·723	0·753	158	0·083	0·076	437
(c) Region						
North east	0·728	0·688	144	0·142	0·137	577
Northern central	0·715	0·714	189	0·145	0·129	782
South	0·715	0·730	296	0·089	0·095	1050
West	0·729	0·777	175	0·094	0·100	602
(d) Industry						
Agricultural	0·672	0·571	21	0·116	0·074	95
Mining	0·752	0·929	14	0·074	0·104	48
Construction	0·712	0·846	65	0·111	0·115	217
Trade	0·722	0·751	205	0·116	0·103	774
Transportation	0·693	0·588	51	0·101	0·137	204
Finance	0·718	0·771	35	0·120	0·096	114
Business and Repair Serv.	0·726	0·648	54	0·120	0·121	239
Personal Services	0·755	0·643	14	0·125	0·125	48
Entertainment	0·738	0·727	11	0·148	0·083	48
Manufacturing	0·722	0·741	224	0·118	0·132	879
Professional Services	0·737	0·676	74	0·121	0·102	215
Public	0·714	0·750	36	0·091	0·054	130

non-union-workers this is only 11%, which illustrates the strong persistence in union membership. In most cells, the average probability of having a union-job, as implied by our model, is a good approximation for the actual proportion of union members. For the second step, we report, in Table VIII, the averages of predicted and actual (log) wages for union and non-union workers for different subsets of the explanatory variables. These results are based on our preferred specification in column [1] of Table V. For most reasonably populated cells the predicted average wage is quite close to the actual average.[19] Union wages appear more difficult to predict although this may reflect the smaller cell sizes. Overall, Tables VII and VIII do not suggest any obvious misspecification.

[19] To gauge the accuracy of the numbers in Table VIII, the conditional standard deviation of an individual's log wage, as estimated by the root mean squared error of the second-step regression, is about 0·45. For a cell size of 500, this implies a standard deviation of the average actual log wage of 0·02.

Table VIII. Predicted versus actual average log wages per cell

Characteristic	Union members			Non-union members		
	Predicted	Actual	N	Predicted	Actual	N
(a) Years of schooling						
< = 9 years	1·46	1·46	47	1·36	1·41	313
10 years	1·50	1·47	89	1·42	1·36	287
11 years	1·69	1·74	181	1·52	1·47	555
12 years	1·85	1·85	601	1·62	1·61	1247
13 years	1·81	1·88	59	1·71	1·75	373
14 years	1·90	1·94	51	1·79	1·83	277
15 years	2·00	1·85	36	1·88	1·96	212
16 years	—	—	0	1·91	1·81	32
(b) Race						
White	1·83	1·82	691	1·63	1·63	2485
Black	1·64	1·70	187	1·45	1·42	317
Hispanic	1·77	1·74	186	1·57	1·58	494
(c) Region						
North east	1·82	1·79	218	1·69	1·69	611
Northern central	1·80	1·84	313	1·56	1·54	811
South	1·74	1·72	350	1·56	1·56	1179
West	1·82	1·81	183	1·67	1·67	695
(d) Industry						
Agricultural	1·56	1·45	22	1·26	1·28	118
Mining	1·94	2·17	23	1·91	1·79	45
Construction	1·83	1·77	65	1·57	1·59	262
Trade	1·63	1·73	203	1·47	1·45	966
Transportation	1·97	1·99	128	1·83	1·81	158
Finance	1·96	2·11	16	1·87	1·85	145
Business and Repair Services	1·75	1·66	27	1·65	1·66	304
Personal Services	1·48	1·43	9	1·56	1·57	64
Entertainment	1·05	0·87	10	1·21	1·24	56
Manufacturing	1·84	1·82	389	1·75	1·76	842
Professional Services	1·60	1·57	90	1·51	1·52	243
Public	1·88	1·80	82	1·69	1·77	93

5. WHOSE WAGES DO UNIONS RAISE?

While the union effect of 23% represents a sizable increase in wages an examination of Tables V and VI reveals this may not be an accurate estimate of the union premium for most individuals in our sample. Furthermore, while the coefficients on the individual-specific effects supported the comparative advantage story the coefficients on the individual-specific time effects appeared to support the possibility of another form of sorting. Accordingly, we examine which effect is dominant in the data. We employed the estimates from column [1] of Table V to evaluate the potential return to union membership for each observation in the sample. The descriptive statistics are reported in Table IX and highlight the variation in the union premium. The first column presents the increase in log wages, due to union membership, for all workers while columns 2 and 3 present the increase by union membership. While we do not report the corresponding findings from replicating this exercise using the estimates from column [3] of Table V an examination of these results indicated that they were almost identical.

Table IX. Descriptive statistics of predicted increase in log wages from union membership

	Whole sample	Union members	Non-union members
Mean	0·213	0·378	0·161
Minimum	0·023	0·195	0·023
5% quantile	0·089	0·276	0·086
10% quantile	0·100	0·295	0·095
25% quantile	0·120	0·330	0·112
Median	0·168	0·381	0·132
75% quantile	0·300	0·424	0·211
90% quantile	0·392	0·462	0·269
95% quantile	0·434	0·492	0·297
Maximum	0·541	0·541	0·432

A number of features are worth noting from the first column of Table IX. First, given that the latent effects have zero mean the average increase is equal to the estimated effect of 0·213. However, the increases range from 0·023 to 0·541. Furthermore, the descriptive statistics indicate that the majority of the higher values are distributed over a relatively small share of the workers. While the average increase is 0·213 the median increase is 0·168 and the lowest quartile is 0·120. In contrast, the highest potential decile increases are in excess of 0·433.

Columns 2 and 3 of Table IX report the premium by sector. Column 2 indicates that the average potential increase in log wages to non-union members was 0·160 with a median of 0·132 and a range from 0·022 to 0·432. For union members the average increase was 0·378 with a median of 0·381 and a range of 0·195 to 0·541. This indicates that many of those who have the most to gain from union membership have obtained union employment. It also reveals substantial variation within each sector. While the maximum value for the non-union members indicates a substantial incentive for union membership, note that the potential increase in log wages from union membership for non-union members at the 90th percentile of 0·268 is less than that at the 5th percentile of the union members (0·275). Thus while Table IX strongly indicates that 'appropriate' workers have sought union membership there is still a substantial number of non-union workers who would gain from union membership.

Table X reports the mean and median of the increase in log wages from union membership by various characterizations of the data. As the apparent sorting pattern is consistent with comparative advantage it is likely that the variables with positive coefficients in the union employment equation will be the variables associated with the biggest union premia. Note, however, the estimates in Table X reflect premia computed for the entire workforce. They also do not reflect marginal differences as we compute the increase for all those with a specified worker characteristic. We also include in Table X, for the sake of comparison, the implied union effects associated with the various characteristics obtained by least squares estimation of a model in which all the regressors, except time, are interacted with the union dummy. Note that these estimates, in the third column of Table X, are not adjusted for the endogeneity of union status. This specification also excludes the occupational dummies to avoid difficulties arising from categories with very small numbers of observations.

We first report the premium by years of schooling noting that we have constructed the lowest education group into those with nine or less years of schooling to avoid small group sizes. Table X(a) indicates that the mean union premium shows no strong relationship with years of schooling except for a substantial decrease at 16 years. The median premium, however, indicates

Table X. Average and median union effect on log wages by worker characteristics

Characteristic	OLS with corrections Mean	OLS with corrections Median	OLS with Interactions
(a) Years of schooling			
< = 9 years	0·17	0·14	0·07
10 years	0·22	0·20	0·05
11 years	0·22	0·19	0·04
12 years	0·24	0·21	0·02
13 years	0·18	0·13	0·01
14 years	0·19	0·14	−0·01
15 years	0·19	0·12	−0·02
16 years	0·12	0·11	−0·03
(b) Race			
White	0·21	0·15	0·19
Black	0·21	0·21	0·28
Hispanic	0·21	0·20	0·13
(c) Region			
North east	0·21	0·17	0·22
Northern central	0·21	0·13	0·27
South	0·22	0·18	0·11
West	0·22	0·18	0·19
(d) Industry			
Agricultural	0·21	0·20	0·21
Mining	0·22	0·12	0·73
Construction	0·22	0·14	0·26
Trade	0·21	0·15	0·48
Transportation	0·25	0·26	0·43
Finance	0·22	0·18	0·53
Business and Repair Services	0·20	0·15	0·19
Personal Services	0·21	0·17	0·21
Entertainment	0·17	0·13	0·00
Manufacturing	0·21	0·19	0·26
Professional Services	0·20	0·13	0·26
Public	0·25	0·27	0·19

Note: The first entry in column 3 is evaluated at 9 years of schooling.

that while those with 12, or less, years of schooling benefit substantially from union membership the premium decreases drastically for those with 13 or more years. This indicates that the union premia are higher for those with relatively less education. The results in the third column are somewhat puzzling as they indicate the union premium is low even for those with low schooling levels and turns negative at 14 years of school. One suspects this simple approach adopted may be generating unreliable estimates.

Table X(b) indicates that the mean union premia for whites, blacks and hispanics are all 21%. However, the median premia are 15%, 20% and 21%, respectively. This indicates that the union premia are relatively robust for the minority workers and the mean white premium is inflated by relatively few workers whose wages are increased substantially via unionism. Column 3 confirms that blacks benefit most from union employment although the effect for hispanics appears low.

The union equation revealed that workers located in the north east displayed a greater tendency to pursue union employment. Table X(c) reveals, however, that the union premium varies very little by region. With the exception of those workers in the northern central enjoying a smaller median value the remaining medians and means are quite similar. In contrast, the simple approach again appears to generate quite large differences by regions although it does produce quite large effects for the north.

Finally, we consider union premia by industry. These are reported in Table X(d) noting that the estimates are frequently based on a small number of observations. This table reveals relatively little variation in industry means with only an 8% difference between the highest (public sector transportation, 0·25) and the lowest (entertainment services, 0·17). The median estimates, however, show substantially more variation with a range of 0·15. This suggests that while there are substantial union premia to be enjoyed in the public sector and transport industries several of the other industries offer no real gains. Furthermore, the fact that the medians are almost always below the mean values suggests that the substantial gains are distributed among relatively fewer workers. The estimates in Table V and the figures in Table IX suggest these workers are those who are union members. The results in column 3 produce quite drastic differences by industry. While this is partially due to the small numbers of observations in some cells it reveals, along with some of the earlier numbers in this column, the danger in adopting this simplistic approach.

6. CONCLUSIONS

The primary focus of this paper is the estimation of the union premium for young males during a period of declining unionization. We employ a methodology which controls for the individual- and time-specific effects operating through the union membership decision. We also test for sources of endogeneity, thus gaining greater insight into the mechanisms driving union membership.

Our empirical work identifies several important results. First, for the data period examined, the union effect is approximately 21%. However, the random effects contribute significantly, making the total union premium highly variable across individuals. Second, the random effects are valued differently by sector and the pattern of sorting into union employment is consistent with that of comparative advantage. Moreover, individuals with characteristics typically associated with lower wages are the recipients of the larger premia. This is consistent with these workers extracting some monopoly rents through unionization. It is also consistent with those employers facing union wages imposing higher hiring standards on their workforce.

APPENDIX

In this appendix we follow Vella and Verbeek (1996) and sketch the estimation method and derive the appropriate correction terms. Represent the respective error terms as follows:

$$e_{j,it} = \alpha_{j,i} + \varepsilon_{j,it} \qquad v_{it} = \theta_i + \eta_{it} \tag{A1}$$

We need to compute the conditional expectation of the elements of $e_{j,it}$ given the T vector U_i (i.e. given the inequality constraints on all T elements of v_{it}) and given $\varphi = (\varphi_1, \ldots, \varphi_T)'$. Employing our assumption of joint normality the conditional expectation of $e_{j,it}$ given the vector

v_i can be derived from the standard formulae for the conditional expectation of normally distributed vectors. This results in

$$E[\alpha_{j,i} \mid v_i] = \sigma_{j,\alpha\theta}\left[\frac{T}{\sigma_\eta^2 + T\sigma_\theta^2}\bar{v}_{i.}\right] \tag{A2}$$

$$E[\varepsilon_{j,it} \mid v_i] = \sigma_{j,\varepsilon\eta}\left[\frac{1}{\sigma_\eta^2}v_{it} - \frac{T\sigma_0^2}{\sigma_\eta^2(\sigma_\eta^2 + T\sigma_\theta^2)}\bar{v}_{i.}\right] \tag{A3}$$

To obtain the conditional expectations, given the vector U_i, replace the v_{it}'s in equations (A2) and (A3) by their conditional expectations given U_i.

Next we use

$$E[\theta_i + \eta_{it} \mid U_i] = \int_{-\infty}^{\infty}[\theta_i + E[\eta_{it} \mid U_i, \theta_i]]f(\theta_i \mid U_i)\,\mathrm{d}\theta_i \tag{A4}$$

where $E[\eta_{it} \mid U_i, \theta_i] = E[\eta_{it} \mid U_{it}, \theta_i]$ is the usual generalized residual of the probit model given by

$$E[\eta_{it} \mid U_i, \theta_i] = (2U_{it} - 1)\sigma_\eta \frac{\phi(b_{it})}{\Phi(b_{it})} \tag{A5}$$

where $b_{it} = (2U_{it} - 1)(\gamma' W_{it} + \theta_i)/\sigma_\eta$. In equation (A4) we integrate over the conditional distribution of θ_i given U_i, which is given by

$$f(\theta_i \mid U_i) = \frac{\prod_{s=1}^{T}\Phi(b_{is})\sigma_\theta^{-1}\phi(\theta_i/\sigma_\theta)}{\int \prod_{s=1}^{T}\Phi(b_{is})\sigma_\theta^{-1}\phi(\theta_i/\sigma_\theta)\,\mathrm{d}\theta_i} \tag{A6}$$

Consequently, given the parameter estimates for the probit model (including the variance components) the generalized residual for the random effects probit model can be computed from equation (A4) using equations (A5) and (A6). This requires numerical integration over one dimension (in both equation (A6) and equation (A4)).

ACKNOWLEDGEMENTS

This paper was partially written while Vella was a visitor at the CentER for Economic Research at Tilburg University. We have benefited from useful discussions with David Card, Bertrand Melenberg, Whitney Newey, Peter Hartley, Theo Nijman, Adrian Pagan and Jörn-Steffen Pischke. We are also grateful to John Rust, the co-editor, and an anonymous referee for detailed comments and suggestions, and to Craig Strain for preparing the data. Vella acknowledges the financial assistance from the Netherlands Organization for Scientific Research (N.W.O.). We alone are responsible for any remaining errors.

REFERENCES

Abowd, J. and H. Farber (1982), 'Job queues and the union status of workers', *Industrial and Labor Relations Review*, **35**, 354–67.

Amemiya, T. and T. MaCurdy (1986), 'Instrumental variable estimation of an error-components model', *Econometrica*, **54**, 869–81.

Andrews, D. W. K. (1988), 'Chi-square diagnostic tests for econometric models: introduction and applications', *Journal of Econometrics*, **37**, 135–56.
Andrews, D. W. K. (1991), 'Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors', *Journal of Econometrics*, **47**, 359–77.
Angrist, J. D. and W. K. Newey (1991), 'Over-identification tests in earnings functions with fixed effects', *Journal of Business and Economic Statistics*, **9**, 317–23.
Card, D. (1996), 'The effects of unions on the structure of wages: a longitudinal analysis', *Econometrica*, **64**, 957–79.
Gallant, A. and D. Nychka (1987), 'Semi-nonparametric maximum likelihood', *Econometrica*, **55**, 363–90.
Hausman, J. and W. Taylor (1981), 'Panel data and unobservable individual effects', *Econometrica*, **49**, 1377–98.
Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica*, **47**, 153–61.
Heckman, J. J. (1981), 'The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process', in C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.
Heckman, J. J. and G. Sedlacek (1985), 'Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market', *Journal of Political Economy*, **93**, 1077–1125.
Heywood, J. S. (1990), 'Who queues for a union job?' *Industrial Relations*, **29**, 99–127.
Jakubson, G. (1991), 'Estimation and testing of the union wage effect using panel data', *The Review of Economic Studies*, **58**, 971–91.
Jovanovic, B. and R. Moffitt (1990), 'An estimate of a sectorial model of labor mobility', *Journal of Political Economy*, **98**, 827–52.
Lee, L. F. (1978), 'Unionism and wage rates: a simultaneous equations model with qualitative and limited dependent variables', *International Economic Review*, **19**, 415–33.
Lee, L. F. (1984), 'Tests for the bivariate normal distribution in econometric models with selectivity', *Econometrica*, **52**, 843–63.
Linneman, P. and M. Wachter (1986), 'Rising union premiums and the declining boundaries among noncompeting groups', *American Economic Review*, **76**, 103–8.
Murphy, K. M. and R. H. Topel (1987), 'The evolution of unemployment in the United States: 1968–1985', in S. Fischer (ed.), *NBER Macroeconomics Annual*, Vol. 2, MIT Press, Cambridge, MA.
Newey, W. K. (1984), 'A method of moments interpretation of sequential estimators', *Economics Letters*, **14**, 201–6.
Pagan, A. and F. Vella (1989), 'Diagnostic tests for models based on unit record data: a survey', *Journal of Applied Econometrics*, **4**, s29–60.
Robinson, C. (1989), 'The joint determination of union status and union wage effects: some tests of alternative models', *Journal of Political Economy*, **97**, 639–67.
Roy, A. (1951), 'Some thoughts on the distribution of earnings', *Oxford Economic Papers*, **3**, 135–46.
Topel, R. H. and M. P. Ward (1992), 'Job mobility and the careers of young men', *Quarterly Journal of Economics*, **107**, 439–79.
Vella, F. and M. Verbeek (1993), 'Estimating and interpreting models with endogenous treatment effects: the relationship between competing estimators of the union impact on wages', CentER Discussion Paper 9351, Tilburg University.
Vella, F. and M. Verbeek (1996), 'Two-step estimation of simultaneous equation panel data models with censored endogenous variables', CES Discussion Paper 96.02, KU Leuven.
Willis, R. J. and S. Rosen (1979), 'Education and self-selection', *Journal of Political Economy*, **87**, S7–36.

