The Economic and Social Review, Vol. 21, No. 3, April, 1990, pp. 311-328

Occupational Endogeneity and Gender Wage Differentials for Young Workers: An Empirical Analysis using Irish Data

BARRY REILLY*
University of St Andrews, Scotland

Abstract: This paper presents estimates of the unexplained gender wage gap for young workers controlling for occupational endogeneity. Two contrasting econometric techniques are employed to control for occupational endogeneity. One is the Heckman two step procedure while the other is an IV estimator based on the work of Duncan and Leigh (1985). Statistical tests for the endogeneity hypothesis are provided for both estimators and mildly conflicting results are obtained.

I INTRODUCTION

In the past estimates of the wage effects of gender discrimination have focused on wage effects treating the occupational levels of males and females as exogenous. The literature is replete with examples of such studies where the unexplained differential between two reduced form wage equations is assumed to approximate a gender discrimination effect.[1] One of the major limitations of such studies is that the occupational effects are mediated through exogenous shifts in the wage equation. The estimation of separate occupational wage equations represents a clear advance, particularly if there is a suspicion that the mean discrimination effect conceals the presence of a larger intra-occupational effect.

*The author would like to thank Ben Knight, Wiji Narandranathan, Mark Stewart and Brian Main for comments on an earlier draft. The comments of two referees are gratefully acknowledged. The assistance of the ESRI in providing the data used in this analysis is also acknowledged. However, the usual disclaimer applies.

1. Gunderson (1989) provides a good survey of recent work in this area.

E

However, a major problem posed by the estimation of occupational wage equations relates to the possible existence of some selection process that determines the observed occupational sample. If the disturbance terms in the occupational wage equations are correlated with the disturbance terms in the occupational selection equation, then conventional estimation techniques, like OLS, provide biased and inconsistent parameter estimates. This has clear implications for the estimated discrimination effect. Methods designed to correct for such selectivity bias have been suggested in the literature and applied to the area of labour supply (Heckman (1976), migration (Robinson and Tomes (1982), and union endogeneity (Lee 1978) and Duncan and Leigh (1980)). Few studies have analysed the effects of selectivity bias on discrimination estimates placing particular emphasis on the effects of occupational selection.[2]

One of the main objectives of this paper is to explore gender and occupational wage differentials within a dichotomous non-manual/manual framework and to establish the effects, if any, of occupational sample selectivity on gender discrimination estimates. A second objective is to statistically test the proposition of occupational exogeneity or occupational sample selection bias.

Two contrasting econometric methods are employed to control and test for the potential endogeneity of occupational status. One is an Instrumental Variable (IV) estimator proposed by Dubin and McFadden (1984) and refined for use in the context of union endogeneity by Duncan and Leigh (1985). The other is the widely used Heckman (1976 and 1979) two-step estimator based on the Mill's ratio. Testing for occupational exogeneity in the former case is effected through the calculation of the Hausman (1978) test and in the latter through Melino's (1982) Lagrange Multiplier (LM) test for selection bias.

The econometric issues raised by the analysis should not hide important economic policy issues. Foremost among these is the question of whether the magnitude of the unexplained gender wage differential varies markedly across occupational sectors. A second question relates to the age of the sample under consideration in the analysis. Some theoretical models highlight the role played by female labour force interruption and subsequent skill depreciation in providing an explanation for female wage disadvantage.[3] In the context of young workers one may be surprised to detect evidence of wage based discrimination in any occupational sector. The detection of such an effect has clear implications for the transition of young female workers into the adult labour market.

2. Dolton, Makepeace and Van der Klaauw (1989) have examined the effects of sample selection on occupational wages in a polychotomous framework but without explicit reference to gender wage effects.

3. See Mincer and Polachek (1974) and Polachek (1981).

The layout of this paper is as follows: Section II outlines the econometric methodology employed and Section III describes the data set. Section IV discusses the empirical results and Section V concludes.

II ECONOMETRIC METHODOLOGY

The model describing the determination of non-manual and manual occupational attachment and wages is given by the following set of well known equations:

$$Y_i = K_i \gamma + \epsilon_i \tag{1}$$

$$w_{ni} = X_{ni}\beta_n + \eta_{ni} \tag{2}$$

$$w_{mi} = X_{mi}\beta_m + \eta_{mi} \tag{3}$$

where $i = 1, \ldots, T$,
T = number of individuals,
n and m subscripts refer to non-manual and manual occupational categories respectively,
Y_i = the latent dependent variable for the i^{th} individual capturing the determinants of occupational attachment,
w_i = the natural log of the i^{th} individual's hourly wage,
K_i = a vector of characteristics that determines the i^{th} individual's occupational attachment,
X_i = a vector of characteristics that determines the i^{th} individual's wage,
ϵ_i, η_{ni} and η_{mi} = error terms.

The dichotomous realisation of the unobserved latent dependent variable, Y_i, is provided by a dummy indicator variable, I_i, which equals 1 if the observed individual is non-manual and 0 otherwise. It would be difficult to portray the model based on Equations (1) to (3) as an occupational choice model. Clearly, many manual workers, given their educational qualifications, may not have a choice in terms of their occupational attachment. However, if one wishes to investigate how the gender wage gap varies across manual and non-manual workers the issue of selection bias must be addressed. The endogeneity of an occupation may not be determined by some choice process but through the fact that a sub-sample of either manual or non-manual workers may not represent a random drawing from the population of workers as a whole. This raises an important econometric issue to which we now turn.

The wage equations of (2) and (3) cannot be validly estimated separately by OLS since estimation would be on the basis of a truncated sample. The truncation follows from the fact that the non-manual wage is unobserved for the manual worker and vice-versa.[4] However, Heckman (1976 and 1979) provides a method for estimation in the presence of such truncation. The regression Equations (2) and (3) may be expressed as:

$$E(w_{ni} \mid X_{ni}, Y_i \geq 0) = X_{ni}\beta_n + E(\eta_{ni} \mid Y_i \geq 0) \qquad (2^*)$$

$$E(w_{mi} \mid X_{mi}, Y_i < 0) = X_{mi}\beta_m + E(\eta_{mi} \mid Y_i < 0) \qquad (3^*)$$

where all the elements are as defined above with E( ) denoting the expectations' operator. Heckman (1979) points out that the straight application of OLS to the Equations (2) and (3) suffers from two sources of misspecification; one due to omitted variables, the other to heteroscedasticity. He proposes the use of proxy constructs designed to take into consideration the truncated nature of the error terms in (2*) and (3*). The regression equations may be re-written as follows:

$$E(w_{ni} \mid X_{ni}, Y_i \geq 0) = X_{ni}\beta_n + \theta_n\lambda_{ni} \qquad (4)$$

$$E(w_{mi} \mid X_{mi}, Y_i < 0) = X_{mi}\beta_m + \theta_m\lambda_{mi} \qquad (5)$$

where

$$\lambda_{ni} = \frac{\phi(K_i\gamma)}{\Phi(K_i\gamma)} \qquad (6)$$

and

$$\lambda_{mi} = -\frac{\phi(K_i\gamma)}{1 - \Phi(K_i\gamma)} \qquad (7)$$

$\phi(\ )$ and $\Phi(\ )$ are the density and cumulative distribution functions of a standard normal variable.

The standard approach to estimating the above model is to apply probit analysis to the reduced form criterion function of (1) yielding estimates for γ. These estimates are then inserted into (6) and (7) to obtain the proxy constructs designed to control for the truncated nature of the error terms in the wage equations. OLS is then applied to the heteroscedastic regression equations of (4) and (5).[5]

4. As Duncan (1983) points out if both sectoral wages are observed simultaneously for each individual drawn at random from the population, then the application of OLS with the standard set of caveats is valid. Such circumstances rarely, if ever, occur.

5. The appropriate variance-covariance matrix for the wage equations is outlined in Greene (1981).

Olsen (1982) highlights the necessity of imposing some form of structure on the problem of correcting for selectivity bias with particular importance placed on the identification of the selectivity effect. In the context of the empirical union endogeneity literature identification creates a clear problem. All the variables that influence the wage also influence union attachment and identification of the selectivity effect relies on the functional form. Since the Mill's ratio is a non-linear function of the exogenous variables in the reduced form probit the same set of regressors can be used in (1) as in (2) and (3). However, a condition required for identification of the selectivity effect in the two-step framework outlined is the availability of some variable that shifts the probability of observing the dependent variable without shifting the mean of the dependent variable. For the purposes of this study a set of parental background dummy variables are included designed to shift the probability of occupational attachment but do not enter the wage equations.[6]

The statistical test for occupational exogeneity in this framework is provided by Melino (1982) who shows that the Heckman test (i.e., the t-statistic on the selection term) is equivalent to a Lagrange multiplier test of the null hypothesis of no sample selection bias. This Lagrange multiplier test is calculated on the basis of the square of the t-statistic on the selection variable (using the uncorrected OLS variance). This test is shown to be distributed as a χ^2 variate with one degree of freedom and possesses superior asymptotic properties to those exhibited by the Heckman test.

This two-step procedure has not been free of criticism. In particular Lee (1982) suggests that the imposed normality assumption on the error term of the criterion function may have serious implications for the detection of selectivity bias. A failure to detect such bias may be related to a distributional misspecification in the error term. Lee (1983) outlines a correction method based on more general distributional assumptions. However, the problem of having to make some distributional assumption is not avoided. It is this particular problem that has forced attention to turn towards distribution free estimators. The IV estimator provides one such alternative.

The IV procedure employed here follows closely that proposed by Duncan and Leigh (1985). Retaining the notation used above the full sample wage equation may be written as:

6. An alternative solution usually adopted is to use non-linearities in the exogenous variables (e.g., squared or interactive terms) in order to identify the relationship. Since an investigator rarely possesses any intuition regarding the appropriate functional form Olsen (1980) dismisses this as relatively unappealing.

$$w_i = I_i w_{ni} + (1 - I_i) w_{mi} \tag{8}$$

Substituting in for the non-manual and manual wages using (2) and (3) yields:

$$w_i \ (I_i X_{ni})\beta_n + ((1 - I_i) X_{mi})\beta_m + \nu_i \tag{9}$$

or more compactly

$$w_i = Z_{ni}\beta_n + Z_{mi}\beta_m + \nu_i \tag{10}$$

where $\nu_i = I\eta_{ni} + (1 - I_i)\eta_{mi}$ and the error terms are assumed to possess the properties $E(\nu_i) = 0$ and $\text{var}(\nu_i) = \sigma_\nu^2$.

The fully interactive model described by (10) allows returns to the variables to vary across occupational sectors. However, the use of OLS is invalidated by the fact that $E((Z_{ni}:Z_{mi})\nu_i)$ is not zero. As Duncan and Leigh (1985) show in order to estimate (10) using the IV procedure the stringent condition that the joint density functions $g(\epsilon_i, \eta_{ni})$ and $g(\epsilon_i, \eta_{mi})$ are equal is imposed. This implies that the error generating process that characterises the wage equations in both sectors is approximately the same for the first two moments of the distribution. This is a necessary condition (as the authors show in an appendix) to ensure that the error term of (10) is mean zero and constant variance.

A necessary criterion for admissable instruments is high correlation with the regressor in question i.e., occupation. Duncan and Leigh (1985) suggest that natural instruments to use in this case are the expected values of the explanatory variables, $E(Z_{ni}) = P_i X_{ni}$ and $E(Z_{mi}) = (1 - P_i) X_{mi}$ where $P_i = \text{prob}(I_i = 1)$ with P_i calculated from the reduced form of (1) using probit. Instruments are then formed by interacting the predicted probabilities $\hat{P}_i$ with the actual X_{ni} and X_{mi} variables. Define the instruments calculated in this manner by the matrix W and the natural log of the hourly wage by the vector y. Redefine the matrix $(Z_{ni}:Z_{mi})$ more simply by Z; then the well known IV coefficient estimator is given by $\hat{\beta}_{iv} = (W^T Z)^{-1} W^T y$. Following White (1982) the estimator for the variance covariance can be modified to take into consideration the potential presence of heteroscedasticity. The corrected variance may be given by $\text{var}(\hat{\beta}_{iv}) = (W^T Z)^{-1} W^T \Omega W (Z^T W)^{-1}$ where $\Omega = \text{diag}((y - Z^T \hat{\beta}_{iv})^T (y - Z^T \hat{\beta}_{iv}))$.

The advantage possessed by the IV approach over the Heckman procedure is the fact that no distributional assumptions enter the second stage of estimation. Though in this study a normality assumption is necessary to obtain the predicted probabilities used in the IV case, this assumption does not enter the wage equation estimation as with the Heckman procedure. The statistical test for the occupational endogeneity or occupational sample selection is provided by Hausman (1978).[7]

III DATA

The data used in this study are obtained from an EC commissioned survey carried out by the ESRI in 1982. The target group in the survey were young people in the 15 to 24 age-group who had left full-time education and were either actively engaged in employment or actively searching for work. The sub-sample employed in this analysis is composed of individuals of single marital status who defined their main economic activity as either working for payment or profit in non-agricultural activities. Only those who classified themselves as full-time workers are included. It could be stated that the use of single individuals raises another selectivity problem, that of participation selectivity bias. However, excluding married individuals (both males and females) does not atrophy the sample to any great degree. Furthermore, there is no econometric evidence that the sample used in estimating the wage equations represents a non-random sample. This was confirmed by the estimation of participation equations with wage equations corrected by the Heckman procedure.

The sub-sample was allocated between the broad non-manual/manual occupational category on the basis of the Census of Population Classification of Occupations (1981). The total number of observations for which no missing values were recorded was 2,827. Of these 1,566 were non-manual (of which 568 were male) and 1,261 manual (of which 937 were male) workers.

The variables used in estimation of the wage and occupational equations were:

Wage: Net hourly wage expressed in logs.

Experience: Total labour force experience expressed in years. In the wage equations this variable is transformed into two linear splines (see Reilly (1987)).

Previous Experience: Experience prior to the current job expressed in years.

Education: Years in post-compulsory education. This variable was preferred to an educational qualifications' variable since it proved difficult to construct

7. A logit model could have been used to obtain the predicted probabilities. This would have involved estimation of the reduced form of (1) by assuming a logistic distribution for the error term in (1) as opposed to a normal distribution. Since the main difference between these two distributions lies in their tails (the logistic has thicker tails than the probit transformation) one would expect little difference in the predicted probabilities. This is confirmed for this study where a logit model was used to estimate the criterion function of (1). This information was then used to correct the wage equations for selectivity bias using the Lee (1983) procedure. The predicted probabilities obtained from the logit model were also used to construct the instrument set used in the IV procedure. Overall the results reported here are not sensitive to the use of alternative distributional assumptions.

the latter on a comparable basis over the gender and occupational categories employed in the analysis. The exogenous treatment of education is not free of criticism (see Willis and Rosen (1979)). It could be argued that occupation and education are jointly and endogenously determined. The econometric model necessary to address this particular issue is of a far more complicated nature than the one outlined in this study and such a treatment is avoided here.

Occupation: Coded 1 if the individual holds a non-manual job, 0 otherwise. This variable serves as the dichotomous realisation of the latent dependent variable of Equation (1).

Region of Schooling: A dummy variable adopting a value of 1 if the individual's region of schooling is in Dublin City or county, 0 otherwise. A set of dummies for type of school (i.e., vocational, secondary, etc.) was also employed but to little effect in the sectoral wage equations and are not reported here.

Establishment Size: A set of three mutually exclusive dummy variables capturing the size of the establishment an individual works in. The two dummies included in estimation are for workers in establishments with between 50 and 400 and greater than 400. The omitted dummy is for less than 50.

Unemployment: The number of months of unemployment since leaving full-time education.

Move Residence: A dummy adopting a value of 1 if the individual changed residence to take their current job.

Father's Occupation: A set of three dummies capturing the occupational status of an individual's father. The two categories included in estimation are manual and non-manual with agriculture the omitted category.

Appendix II contains means, etc., for the variables used in estimation.

IV EMPIRICAL RESULTS

Probit estimates for the reduced form occupational attachment equations are contained in Appendix I and are not commented on here. Tables 1 and 2 contain OLS, IV and Heckman estimates for the male and female wage equation estimates respectively. Across all three estimators the returns to labour force experience are relatively robust and suggest greater returns to labour force experience in the early years of such experience. The private rates of return to education appear more sensitive to the estimator used and especially so for the female manual workers. Rates of 1.5 per cent and 8.5 per cent for the Heckman and IV techniques respectively are recorded for the female manual category as compared to 3.6 per cent for the OLS estimate. In both the IV and the Heckman cases neither estimate is statistically significant at a satisfactory level which is in contrast to the manual male estimates.

The non-manual estimates for the private rates of return vary less across the estimator used and are broadly compatible across gender.

Table 1: *Male Wage Coefficient Estimates*

Variable	*OLS*	*Std.error*	*IV*	*Std.error*	*Heckman*	*Std.error*
Manual						
Constant	−0.2853***	0.0356	−0.3810***	0.0813	−0.3268***	0.0387
Exp. ≤ 4 yrs.	0.1510***	0.0108	0.1750***	0.0195	0.1505***	0.0095
Exp. > 4 yrs.	0.0568***	0.0081	0.0294	0.0203	0.0537***	0.0104
Education	0.0634***	0.0086	0.0438**	0.0209	0.0418***	0.0142
50 ≤ Estab. < 400	0.1246***	0.0261	0.1171**	0.0568	0.1311***	0.0255
Estab. ≥ 400	0.1421***	0.0258	0.2287***	0.0605	0.1330***	0.0266
Schooling in Dublin	0.0537**	0.0225	−0.0379***	0.0666	−0.0021***	0.0381
Unemployment (months)	0.0056***	0.0437	0.0079***	0.0020	0.0067***	0.0015
Move Residence	0.0391	0.0437	0.1407*	0.0729	0.0247	0.0510
Selectivity Bias	–	–	–	–	0.1621*	0.0854
Non-Manual	–	–	0.1407	0.2020	–	–
Constant	−0.1474***	0.0437	–	–	−0.1119	0.1265
Exp. ≤ 4 yrs.	0.1003***	0.0120	0.0565*	0.0294	0.0998***	0.0122
Exp. > 4 yrs.	0.0484***	0.0140	0.0870**	0.359	0.0480***	0.0141
Education	0.0683***	0.0104	0.0492***	0.0248	0.0646***	0.0150
50 ≤ Estab. < 400	0.1388***	0.0337	0.1605	0.1081	0.1391***	0.0343
Estab. ≥ 400	0.1987***	0.0309	0.0479	0.1102	0.1954***	0.0328
Schooling in Dublin	0.1242***	0.0267	0.1693**	0.0762	0.1158***	0.0391
Unemployment (months)	0.0032	0.0024	−0.0023	0.0064	0.0034	0.0034
Move Residence	0.0723*	0.0427	0.0099	0.0820	0.0697	0.0498
Selectivity Bias	–	–	–	–	0.0242	0.0808
Observations	1,505		1,505		1,505	

*** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

Table 2: *Female Wage Coefficient Estimates*

Variable	*OLS*	*Std.error*	*IV*	*Std.error*	*Heckman*	*Std.error*
Manual	–	–				
Constant	-0.0316	0.0633	-0.5117**	0.1987	-0.0981	0.1098
Exp. ⩽4 yrs.	0.0616***	0.0151	0.1067***	0.0251	0.0601***	0.0127
Exp. >4 yrs.	0.0151	0.0106	-0.0016	0.0259	0.0157	0.0118
Education	0.0362***	0.0126	0.0849*	0.0503	0.0157	0.0326
50⩽Estab.<400	0.1196***	0.0403	0.4865***	0.1695	0.1717**	0.0847
Estab.⩾400	0.2820***	0.0418	0.6542***	0.1689	0.3072***	0.0550
Schooling in Dublin	0.0435*	0.0262	0.0270	0.0626	0.0229	0.0435
Unemployment (months)	0.0062***	0.0019	0.0054	0.0044	0.0019***	0.0007
Move Residence	0.1188**	0.0553	-0.0714	0.3396	0.1006	0.0863
Selectivity Bias	–	–	–	–	0.0666	0.0999
Non-Manual	–	–	-0.1209	0.0704*	–	–
Constant	-0.2298***	0.0361	–	–	-0.2357***	0.0534
Exp. ⩽4 yrs.	0.0883***	0.0092	0.0734***	0.0122	0.0885***	0.0084
Exp. >4 yrs.	0.0435***	0.0117	0.0588**	0.0277	0.0433***	0.0117
Education	0.0879***	0.0079	0.0686***	0.0140	0.0884***	0.0127
50⩽Estab.<400	0.1933***	0.0244	0.1036*	0.0571	0.1888***	0.0424
Estab.⩾400	0.2714***	0.0223	0.2151***	0.0294	0.2698***	0.0255
Schooling in Dublin	0.1391***	0.0190	0.1575***	0.0281	0.1407***	0.0231
Unemployment (months)	-0.0055**	0.0023	-0.0018	0.0036	-0.0014**	0.0006
Move Residence	0.0115	0.0378	0.0456	0.0474	0.0124	0.0338
Selectivity Bias	–	–	–	–	-0.0103	0.0779
Observations	1,322		1,322		1,322	

*** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

The coefficients on establishment size are, like the experience coefficients, relatively robust to the estimator used. In general, the reported returns are characterised by a monotonic increase with establishment size. The coefficient on the schooling in Dublin dummy follows a similar pattern regardless of which estimator is employed. Interpretation of this coefficient is difficult since it proxies not only individual attributes but also job attributes. However, there is a clear contrast between manual and non-manual workers in regard to this coefficient. For non-manual workers the reported effects are well over 10 per cent in most cases in comparison to a negligible return for the manual workers. This effect could be just proxying the large number of relatively well paid non-manual jobs in the Dublin area. Since the reduced form probit estimates suggest that young workers educated in Dublin have a higher probability of non-manual occupational attachment this result should not come as a surprise. Most of the other coefficient estimates are not well determined and are not deemed worthy of additional comment.

Attention now turns to an examination of the exogeneity hypothesis of the broad occupational categories employed in the analysis. The results based on the Melino (1982) test outlined in Section II suggest little evidence of occupational endogeneity for the females. The comparable LM test for the non-manual male category also suggests little or no evidence of selectivity bias. However, the LM test for the male manual sector suggests a rejection of the null of exogeneity.[8] The latter result needs to be interpreted in terms of the direction of the selection bias. Due to the construction of the proxy variables, outlined in Section II, a positive coefficient on the selection terms implies negative selectivity (or negative truncation) in terms of manual jobs. In fact, the data suggest that young workers selecting the manual category earn on average 8.5 per cent lower wages than an individual drawn at random from the population with identical observable characteristics would be expected to earn in that sector. A suggestive interpretation for this result may lie in the fact that young workers in the manual sector sacrifice wages for training in return for greater life-cycle earnings. Seventy per cent of workers in the manual category, in this study, belong to the skilled category and this interpretation could be viewed as reasonably plausible.

The Hausman test results based on the IV estimates of Tables 1 and 2 are in slight contrast to those obtained using the Melino LM test.[9] The χ^2 statistics are 7.82 and 8.88 for the male and female equations respectively. Taking the LM and the Hausman tests together there appears little evidence of occupational endogeneity with the exception of the male manual category.

However, a number of caveats need to be inserted regarding the reliability of both sets of exogeneity tests. For example, the LM test is contingent on a normally distributed error term in the reduced form probit of (1).[10] Departures from normality may lead to erroneous conclusions regarding selectivity bias (see Lee (1982)). On the other hand, the Hausman test possesses its own limitations. If the instruments used are orthogonal to the regressors being instrumented the power of the test is zero and not rejecting the null when untrue is certain. Therefore, both sets of results should be couched in terms of the above provisos and any conclusion regarding occupational exogeneity/endogeneity must remain relatively neutral.

8. The χ^2 statistics for the females are 0.017 and 0.429 for the manual and non-manual sectors respectively. The comparable estimates for the males are 0.088 and 3.482. The latter result is marginally outside the 5 per cent level of significance but well inside the 10 per cent level.

9. The quadratic form of the Hausman (1978) test was employed here rather than the auxiliary regression approach. The resultant test statistic has a χ^2 distribution with the number of restrictions under the test given by the number of parameters estimated.

10. Chesher and Irish (1987) provide easily computable diagnostics to test, among other things, the assumption of normality. The χ^2 statistics based on their suggested normality tests are 656.6 and 846.5 for the male and female reduced form versions of (1). The decisive rejection of normality brings the validity of the LM test into question but the rejection may also be attributable to the poor finite properties of the tests used.

The observed manual differential is relatively small suggesting that on average manual male workers get 1.7 per cent more than female manual workers. In contrast the non-manual observed differential is considerably larger with males, on average, earning 9.5 per cent more than their female counterparts. Tables 3 to 5 report OLS, IV and Heckman based explained and unexplained differentials in an attempt to establish how much of the gross differential is explained by characteristics and how much by differing coefficients.[11] In terms of all three estimators little evidence of an unexplained gender wage differential exists in the manual sector.

Table 3: *OLS Wage Differentials by Occupational Sector*

Sector	$\Delta\bar{W}$	$\Delta\bar{X}\hat{\beta}^m$	$\bar{X}^f\Delta\hat{\beta}$
Non-Manual	0.0904	0.0327***	0.0576***
		(0.0036)	(0.0170)
Manual	0.0167	-0.0121	0.0288
		(0.0083)	(0.0189)

Asymptotic standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

Table 4: *IV Wage Differentials by Occupational Sector*

Sector	$\Delta\bar{W}$	$\Delta\bar{X}\hat{\beta}^m$	$\bar{X}^f\Delta\hat{\beta}$
Non-Manual	0.0904	0.0269**	0.1521
		(0.0106)	(0.0890)
Manual	0.0167	-0.0040	0.0044
		(0.0182)	(0.1002)

Asymptotic standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

11. The explained and unexplained differentials are calculated as in Reilly (1987). The Heckman based estimates of Table 5 are unconditional and are calculated by setting the selection effects to zero. Reimers (1983) describes these wage differentials as the wage offer differentials. The standard errors are also calculated as in Reilly (1987).

Table 5: *Heckman Wage Offer Differentials by Occupational Sector*

Sector	$\Delta \bar{W}^c$	$\Delta \bar{X} \hat{\beta}^m$	$\bar{X}^f \Delta \hat{\beta}$
Non-Manual	0.1143	0.0323***	0.0819*
		(0.0022)	(0.0471)
Manual	-0.0076	-0.0175	0.0099
		(0.0112)	(0.1059)

$\Delta \bar{W}^c$ is the observed wage differential corrected for selectivity bias, i.e., the wage offer differential. Asymptotic standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

The overall picture is slightly different when the unexplained differentials are calculated for the non-manual sector. The OLS estimates of Table 3 imply that males in the non-manual sector receive, on average, 5.9 per cent more than females with comparable characteristics. Moreover, this estimate is statistically significant. The comparable IV based estimate is considerably larger suggesting a male mark-up of over 16 per cent. Given its statistical insignificance a cautious interpretation is required. The gulf in estimates is surprising given the non-rejection of the null of exogeneity by the Hausman test. An argument that the consistent IV estimates are purchased at the price of efficiency could be made. The large variances recorded for the IV based estimates may reflect the use of a relatively poor set of instruments. This may again bring the Hausman test results back into question and raise the question of whether there exists endogeneity but that the instrument set used is not adequate for the task of detecting it. A belief in this argument could explain the vast disparity in estimates between OLS and IV.

Table 5 reports estimates based on the Heckman procedure. The non-manual sector again exhibits evidence of wage discrimination of the order of 8.5 per cent (significant at the 10 per cent level). This estimate is an unconditional estimate (like the IV based estimate) and its interpretation differs somewhat from that attached to the OLS interpretation. The latter represents a differential conditional on an individual's sectoral attachment. The former represents a differential in wages for an individual drawn at random from the population unconditional on occupational attachment. Thus both the IV and the Heckman estimates are unconditional estimates with the Heckman estimates derived by setting the selectivity effects to zero (see Gyourko and Tracy (1988)). Despite this different interpretation the Heckman estimates and the OLS estimates are relatively close. This is not surprising given the numerically small and statistically insignificant coefficients reported for the female selectivity coefficients in particular.

Finally, Table 6 reports the non-manual/manual mark-up within gender groups. No standard errors are reported for this set of differentials since covariances between the non-manual and manual wage equations within each gender group cannot be identified (see Lee (1979)). In general, the mark-ups are considerably higher for males ranging from 3.3 per cent to over 25 per cent depending on the estimator used. The range of female estimates is from close to 0 per cent to 10 per cent. These results are not surprising given the estimates of Tables 3 to 5.

Table 6: *Non-Manual/Manual Wage Differentials by Gender*

Gender	*OLS*	*IV*	*Heckman*
Male	0.0324	0.2242	0.1550
Female	0.0074	0.1027	0.0994

V CONCLUSIONS

Two contrasting econometric methods were employed to control for the potential endogeneity or selection bias associated with occupational attachment. The Heckman procedure provided some evidence of sample or self selectivity bias in terms of the allocation of young workers to manual jobs. In contrast the IV procedure provided little of such evidence. The caution expressed in terms of interpreting these results was prompted by possible departures from assumed normality in the Heckman procedure and by the orthogonality of instruments and regressors in the IV case.

Three salient conclusions emerge from this study. First, no evidence of wage based discrimination exists in the manual sector. In the context of this sample of young workers this need represent no surprise. A large proportion of males (70%) are in the skilled sub-group within the manual category and in receipt of relatively low wages. This could be rationalised in terms of the sacrifice of wages for training in the return for a greater life-cycle earnings. If female access to training is limited then one would expect greater unexplained differentials to emerge with the passage of these young workers into the adult labour market.

Secondly, there is strong evidence of wage discrimination in the non-manual sector ranging from around 6 per cent to over 16 per cent depending on the estimator used. This fact should be viewed with some concern given the youth of the workers and the fact that 75 per cent of female workers, in this study, are in the non-manual category.

Thirdly, the low observed aggregate gender wage differential of 2.8 per cent provides a misleading figure and disguises vaster gender differences in wages by occupational sectors. Invoking the non-manual/manual framework allowed a more worthwhile insight into the detection of unexplained gender wage differentials. This in itself may be interpreted as a vindication of the exercise undertaken.

REFERENCES

CHESHER, A., and M. IRISH, 1987. "Residual Analysis in the Grouped and Censored Normal Linear Model", *Journal of Econometrics*, Vol. 34, pp. 33-61.

DOLTON, P., G. MAKEPEACE, and W. VAN DER KLAAUW, 1989. "Occupational Choice and Earnings' Determination: The Role of Sample Selection and Non-Pecuniary Factors", *Oxford Economic Papers*, Vol. 41, pp. 573-594.

DUBIN, J., and D. McFADDEN, 1984. "An Econometric Analysis of Residential Electrical Appliance Holdings and Consumption", *Econometrica*, Vol. 52, pp. 345-362.

DUNCAN, G., 1983. "Sample Selectivity as a Proxy Variable Problem: On the Use and Misuse of Gaussian Selectivity Corrections", in J. Reid (ed.), *New Approaches to Labor Unions*, Research in Labor Economics, Suppl. 2, JAI Press.

DUNCAN, G., and D. LEIGH, 1980. "Wage Determination in the Union and Nonunion Sectors; A Sample Selectivity Approach", *Industrial and Labor Relations Review*, Vol. 34, pp. 24-34.

DUNCAN, G., and D. LEIGH, 1985. "The Endogeneity of Union Status: An Empirical Test", *Journal of Labor Economics*, Vol. 3, pp. 385-402.

GREENE, W., 1981. "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 49, pp. 795-798.

GUNDERSON, M., 1989. "Male-Female Wage Differentials and Policy Responses", *Journal of Economic Literature*, Vol. 27, pp. 46-72.

GYOURKO, J., and J. TRACY, 1988. "An Analysis of Public- and Private-sector Wages Allowing for Endogenous Choices of Both Government and Union Status", *Journal of Labor Economics*, Vol. 6, pp. 229-253.

HAUSMAN, J., 1978. "Specification Tests in Econometrics", *Econometrica*, Vol. 46, pp. 1251-1271.

HECKMAN, J., 1976. "The Common Structure of Statistical Models of Truncation, Self-Selection and Limited Dependent Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement*, Vol. 5, pp. 475-492.

HECKMAN, J., 1979. "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, pp. 153-161.

LEE, L-F., 1978. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables", *International Economic Review*, Vol. 19, pp. 415-433.

LEE, L-F., 1979. "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables", *Econometrica*, Vol. 47, pp. 977-996.

LEE, L-F., 1982. "Some Approaches to the Correction of Selectivity Bias", *Review of Economic Studies*, Vol. 49, pp. 355-372.

LEE, L-F., 1983. "Generalised Models with Selectivity", *Econometrica*, Vol. 51, pp. 507-512.

MELINO, A., 1982. "Testing for Sample Selection Bias", *Review of Economic Studies*, Vol. 49, pp. 151-153.

MINCER, J., and S. POLACHEK, 1974. "Family Investments in Human Capital: Earnings of Women", *Journal of Political Economy*, Vol. 82, pp. s76-s108.

OLSEN, R., 1980. "A Least Squares Correction for Selectivity Bias", *Econometrica*, Vol. 48, pp. 1815-1820.

OLSEN, R., 1982. "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator", *International Economic Review*, Vol. 23, pp. 223-240.

POLACHEK, S., 1981. "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure", *Review of Economics and Statistics*, Vol. 63, pp. 60-69.

REILLY, B., 1987. "Wages, Sex Discrimination and the Irish Labour Market for Young Workers", *The Economic and Social Review*, Vol. 18, pp. 271-305.

REIMERS, C., 1983. "Labor Market Discrimination Against Hispanic and Black Men", *Review of Economics and Statistics*, Vol. 45, pp. 570-579.

ROBINSON, C., and N. TOMES, 1982. "Self-selection and Inter-Provincial Migration in Canada", *Canadian Journal of Economics*, Vol. 15, pp. 474-502.

WHITE, H., 1982. "Instrumental Variables Regression with Independent Observations", *Econometrica*, Vol. 50, pp. 483-499.

WILLIS, R., and S. ROSEN, 1979. "Education and Self-Selection", *Journal of Political Economy*, Vol. 87, pp. s7-s36.

APPENDIX I

Table I.1: *Reduced Form Probit Estimates for Male and Female Equations*

Variable	*Male*	*Female*
Constant	-1.2403***	0.3767**
	(9.334)	(2.260)
Exp. ⩽ 4 yrs.	0.0217	0.0396
	(0.657)	(0.969)
Exp. > 4 yrs.	0.0366	-0.0205
	(1.003)	(-0.467)
Education	0.2487***	0.4514***
	(9.600)	(11.780)
50 ⩽ Estab. < 400	-0.1135	-1.246***
	(1.254)	(11.879)
Estab. ⩾ 400	0.1066	-0.5452***
	(1.231)	(4.585)
Schooling in Dublin City & Co.	0.4744***	0.5166***
	(5.911)	(5.028)
Unemployment (months)	-0.0130**	-0.0373***
	(2.109)	(4.381)
Move Residence	0.2252	0.2915
	(1.456)	(1.406)
Father Non-Manual	0.5895***	0.1170
	(5.755)	(0.884)
Father Manual	0.1547	-0.2765**
	(1.589)	(2.307)
Observations	1,505	1,322

Values in parentheses are |t| values. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 10% level using two tailed tests.

APPENDIX II

Data Appendix

Table II.1: *Means for Males and Females by Manual and Non-Manual Category*

Variable	*Male*[a]	*Male*[b]	*Female*[a]	*Female*[b]
ln (wage)	0.4853	0.3685	0.3949	0.3517
	(0.3722)	(0.4012)	(0.3783)	(0.2773)
Previous Exp. (years)	0.9854	0.9869	0.7834	0.9796
	(1.584)	(1.598)	(1.936)	(1.616)
Experience (years)	3.322	3.517	3.105	3.482
	(2.019)	(2.127)	(1.936)	(2.258)
Education (years)	2.310	1.376	2.407	1.139
	(1.696)	(1.288)	(1.387)	(1.215)
Schooling in Dublin	0.4577	0.2327	0.3717	0.2685
	(–)	(–)	(–)	(–)
$50 \leqslant$ Estab. < 400	0.2025	0.2412	0.1974	0.5586
	(–)	(–)	(–)	(–)
Estab. $\geqslant 400$	0.3556	0.2220	0.3026	0.1944
	(–)	(–)	(–)	(–)
Unemployment (in months)	2.838	4.009	2.173	3.887
	(4.466)	(7.268)	(4.065)	(6.521)
Move Residence	0.0792	0.0427	0.0922	0.0278
	(–)	(–)	(–)	(–)
Father Non-Manual	0.3627	0.4739	0.3687	0.5772
	(–)	(–)	(–)	(–)
Father Manual	0.4842	0.2391	0.3697	0.2346
	(–)	(–)	(–)	(–)
λ	0.8560	-0.5213	0.2935	-0.9041
	(0.3797)	(0.2580)	(0.2774)	(0.5281)

a denotes the non-manual sector and *b* the manual sector. Standard deviations for the continuous variables are in parentheses.