The Economic and Social Review, Vol. 18, No. 4, July 1987, pp. 271-305

Wages, Sex Discrimination and the Irish Labour Market for Young Workers

BARRY REILLY*
University of Warwick

Abstract: Human capital wage equations are estimated using individual level data on single males and single females in the Irish labour market for young workers. The results are broadly consonant with theoretical predictions. Returns to on-the-job training and educational qualifications are reported and, in the latter case, are on average higher for the single males. Estimates of sex discrimination based on the unexplained differential from a reduced form equation are also reported. The findings suggest that the observed differential in wages that exists between young males and young females in the sample is, in large part, explained by differing characteristics. A small unexplained residual, interpreted under certain assumptions as an estimate of sex discrimination, is detected. However, the discrimination element is not found to be statistically significant.

I INTRODUCTION

The purpose of this study is to analyse some of the determinants of wages at the individual level in the Irish youth labour market. To the author's knowledge this has been the focus of no attention to date. Even for the adult labour market in Ireland little research effort has been expended on analysing the determinants of adult earnings.[1] An exception in this regard is Walsh and Whelan (1976) who used a sample of redundant adult workers to estimate human capital earnings functions in an attempt at analysing male-female earnings differentials. The emphasis in the present study is on young workers using a data set derived from a relatively recent national survey. It is more

*This paper has benefited from the anonymous comments of two referees which are gratefully acknowledged. Comments and suggestions made by Ben Knight, Steve Machin, Wiji Narandranathan, Mark Stewart and Brendan Walsh on earlier drafts are gratefully acknowledged. The assistance provided by Jerry Sexton, Brendan Whelan and James Williams in making the data set used in this analysis available is also gratefully acknowledged. However, in all cases, the usual disclaimer applies.

1. This is in large part explained by either the unavailability or lack of access to suitable data sets.

general than that of Walsh and Whelan (1976) and could not be interpreted as containing the biases associated with the Walsh and Whelan (1976) data. However, the data set used in this study cannot be interpreted as being free of bias.[2]

The estimation of human capital wage equations provides an important insight into some relevant policy issues. Such estimation allows inferences to be made regarding returns to work experience and educational qualifications and how they vary across males and females. A major focus of attention in terms of this study is the magnitude of the male-female wage differential. An important question in this regard is to what extent it is explained by differing characteristics and/or by differing returns given the same characteristics. Under certain assumptions this latter case may be interpreted as sex discrimination. Becker (1971) formalises the notion of sex discrimination by proposing the concept of a discrimination coefficient. Oaxaca (1973) provides an empirical methodology known as the "index number" approach to estimate the discrimination coefficient.

The following section outlines the methodology employed. Section III describes the data set used and Section IV, the estimation. Results based on estimating the wage equations appear in Section V and Section VI provides some statistical tests of the underlying specifications. Section VII contains the estimates of sex discrimination and Section VIII offers some conclusions.

II METHODOLOGY

Human capital theory provides one framework inside which the effects of sex discrimination can be examined. The reduced form wage (or earnings) equations to be estimated are assumed to be derived from some underlying structural model of labour supply and demand. The widely applied reduced form Mincerian earnings function (see Mincer (1974)) may be expressed as a function of schooling, post-schooling (or on-the-job) investment, and a vector of personal characteristics. Algebraically, it may be given by (1) as:

$$\ln(w_i) = \alpha_i + \beta_1 S_i + \beta_2 X_i + \sum_{j=3}^{n} \beta_j Z_{ij} \qquad (1)$$

where

w_i is the i^{th} individual's wage,

S_i is years of schooling,

X_i is labour force experience and

Z_{ij} is a vector of personal characteristics (e.g., marital status, region of residence, occupation and industry etc.).

2. For example, the sample excludes young persons still in the education system.

The theory predicts that in a long-run equilibrium with perfect labour mobility and perfect competition in the labour market, wages reflect an individual's characteristics in terms of schooling, post-schooling investments and compensating differentials commensurate with differing job characteristics. In the context of firm specific training, human capital theory predicts short-run departures from the above equilibrium condition. Individuals who obtain such training are awarded a remuneration initially above their marginal product during training and below it after its completion in order that firms may recoup their outlayed training costs. Because of this one might not expect young workers' wages to be exclusively based on productivity variables particularly where firm specific training is involved.

Discriminatory practices can also be interpreted as inducing departures from equilibrium. As long as rigidities exist in such a way that the different sexes are paid different market prices for comparable productivity characteristics, then, in the long-run, females will be paid less than their marginal product. These discriminatory practices may take the form of employer discrimination, employee discrimination (through monopolistic organisations, e.g. trade unions) or consumer discrimination. This last form of discrimination is most likely to impinge upon the wage differentials of the self-employed.

Oaxaca (1973) following Becker (1971) defines the discrimination coefficient, (D), as:

$$D = \frac{\left(\frac{w_m}{w_f}\right) - \left(\frac{w_m}{w_f}\right)^{o}}{\left(\frac{w_m}{w_f}\right)^{o}} \qquad (2)$$

where

$\left(\frac{w_m}{w_f}\right)$ = observed male-female wage,

$\left(\frac{w_m}{w_f}\right)^{o}$ = male-female wage ratio in the absence of discrimination (which is unobserved).

Re-writing (2) as follows:

$$\ln(1 + D) = \ln\left(\frac{w_m}{w_f}\right) - \ln\left(\frac{w_m}{w_f}\right)^{o} \qquad (3)$$

The problem with this expression is that the last term in (3) is unobserved.

An observable analogue for this particular expression is required. Define the observed mean wage differential as:

$$G = \frac{\bar{w}_m - \bar{w}_f}{\bar{w}_f} \tag{4}$$

This may be re-expressed as:

$$\ln(1 + G) = \ln(\bar{w}_m) - \ln(\bar{w}_f) \tag{5}$$

Assume a semi-logarithmic functional form equation that in general may be given by:

$$\ln(w_i) = Z_i\beta + u_i \tag{6}$$

where
w_i = the net hourly wage of the i^{th} worker,
Z_i = a vector of individual and job characteristics etc.,
β = a vector of unknown parameters,
u_i = a disturbance term.

Then, from the properties of OLS the wage equations may be expressed as:

$$\ln(w_m) = \bar{Z}_m\hat{\beta}_m \tag{7}$$

$$\ln(w_f) = \bar{Z}_f\hat{\beta}_f \tag{8}$$

where the subscripts m and f refer to males and females respectively, the bars over the variables denote means and the circumflexes estimates. Clearly, with constant terms in Equations (7) and (8) the mean of the actual $\ln(w_i)$ equals the mean of the predicted $\ln(w_i)$. Substituting (7) and (8) into (5) yields:

$$\ln(1 + G) = \bar{Z}_m\hat{\beta}_m - \bar{Z}_f\hat{\beta}_f \tag{9}$$

Now define

$$\Delta\bar{Z} = \bar{Z}_m - \bar{Z}_f$$
$$\Delta\hat{\beta} = \hat{\beta}_f - \hat{\beta}_m.$$

Now assume that in the absence of sex discrimination the male wage structure obtains. Substitute $\hat{\beta}_f = \Delta\hat{\beta} + \hat{\beta}_m$ and obtain:

$$\ln(1 + G) = \overline{Z}_m \hat{\beta}_m - \overline{Z}_f(\Delta\hat{\beta} + \hat{\beta}_m). \tag{10}$$

This reduces to:

$$\ln(1 + G) = \Delta\overline{Z}\hat{\beta}_m - \overline{Z}_f\Delta\hat{\beta} \tag{11}$$

This expresses one plus the observed mean differential in terms of differences in characteristics and in coefficients assuming a male wage structure prevails in the absence of sex discrimination. The latter term is interpreted as the expression for ln(1 + D) since it gives the difference between what females earn and what they would earn if confronted by a male wage structure. It is this component that is of interest here in terms of the calculations presented for sex discrimination in Section VII. Thus, G and D may be tied up as follows:

$$\ln(1 + G) = \Delta\overline{Z}\hat{\beta}_m - \ln(1 + D) \tag{12}$$

The first component of expression (12) refers to that portion of the observed mean wage differential which is explained by differing characteristics or endownments. The second component refers to that portion of the differential which is explained by the payment of different prices for the same characteristics or endowments, ln(1 + D), and is interpreted as an estimate of the natural log of one plus the discrimination coefficient.

Any genuine attempt at calculating the effects of sex discrimination should focus on those males and females who could best be described as possessing a similarity in the accumulation of human capital. Greenhalgh (1980) attempts to estimate sex discrimination by reference to the unexplained differential between single men and single women under the assumption that both groups have a similar attitude to human capital investment. Though this may be interpreted as a reasonably plausible assumption in the context of the adult labour market, it need not necessarily be the case for workers under the age of 25 who feature in this particular study. Nevertheless, a similar strong assumption to that of Greenhalgh (1980) is assumed here.

A number of problems are associated with the use of the "index number" approach in calculating the effects of sex discrimination. Chief among these is the "index number" problem. This stems from the fact that the results based on the above approach are contingent on which wage structure is initially assumed in the absence of sex discrimination. The sex discrimination estimates reported here are based on an assumed male wage structure. This may be regarded as a relatively innocuous assumption since it could be plausibly argued that males are paid a wage that is discrimination-free.

Another problem lies in the fact that the estimated equation is a reduced form. The estimate of sex discrimination is taken as the unexplained differential that emerges after controlling for the standard set of productivity and other relevant variables. The interpretation of the unexplained residual differential as the pure discrimination effect is only valid if supply side factors, (e.g. female preferences for low paying occupations) etc., do not impinge upon the unexplained differential. In the context of a reduced form equation it proves difficult to disentangle the demand side (discrimination) influences from the supply side influences associated with preferences, etc.

Since the focus of this paper is wage discrimination, defined as differences in pay given occupations, the potential discriminatory effect that arises through occupational segregation is ignored. This may be viewed as an unsatisfactory approach assuming, as it does, that there exists no discriminatory restrictions on access to certain occupations. It might be argued that if most sex discrimination is occupation based, then, the above framework is clearly lacking. Brown *et al.* (1980) argue that if the same characteristics that determine wages also determine occupation, then, the "index number" approach may be viewed as an appropriate methodology. In general, they argue that this is not likely to be the case and propose controlling for occupations by incorporating an occupational attainment model into the analysis of wage differentials as a method of overcoming this problem.

For the purposes of this study two "second best" alternatives are presented. First, occupations and industries are controlled for in the standard manner through the use of dummy variables. This approach implicitly treats the occupational distribution of males and females as exogenous and ignores the potential endogeneity of occupational attainment. Regardless of whether the potential endogeneity is a consequence of sample selection on the part of employers or self-selection on the part of workers, a potential bias is likely in the estimates. Furthermore, the inclusion of occupation dummies will also lead to an understatement of the returns to formal educational qualifications since access to certain occupations is contingent on possessing certain formal qualifications. Though this simplistic treatment of occupations and industries is accepted as unsatisfactory, a failure to control for occupations or industries (in some way or another) leads to a confusion between the effects of sex discrimination and those of legitimate occupation or industry compensating differentials.

The second alternative is to estimate the wage equations excluding the occupation and industry dummies. The potential endogeneity problem associated with occupations and industries is thus by-passed in a rather *ad hoc* manner. The consequent estimates of sex discrimination are overstated but the returns to the formal educational qualifications are not under-stated as is the case when occupations are included.[3] Therefore, the results of Section V contain estimates obtained from models that both include and exclude occupation and industry controls.

III DATA

The data used in this study are obtained from an EEC commissioned survey carried out by The Economic and Social Research Institute in 1982 and titled "Youth Employment and the Transition from Education to Working Life". The target group in the survey were males and females between the ages of 15 and 24 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 those single individuals in the survey who defined their main economic activity as either working for payment or profit in non-agricultural activities. The total number of such cases for which no missing values are present is 1,022 (449 males and 573 females). Full details of the variables used in the analysis are contained in Appendix 1.

IV ESTIMATION

Linear equations are estimated separately for both the single males and single females in the sample. In the estimation of standard wage (or earnings) functions quadratic terms in labour force experience are introduced allowing for the effects of such experience to vary over its own range. In the context of the youth labour market it may be unreasonable to impose a priori such a restrictive functional form on the data. In view of this, linear splines (following Stewart and Wallis (1981), pp. 201-204) are constructed from the labour force experience variable. This allows for returns to experience to vary over different ranges of an individual's labour force experience. The choice of the appropriate nodes for the linear splines was determined by using standard F-tests. These tests suggest for the male equations the inclusion of two linear splines, one for less than or equal to five years of experience, the other for more than five years of experience.[4]

3. However, if educational qualifications are chosen endogenously there is latitude for selectivity bias in this regard and the returns to educatonal qualifications (with or without occupation and industry dummies) are potentially biased.

4. The null hypothesis in this case is the male equation strictly linear in experience. This null is rejected in favour of a piece-wise linear alternative having a five year split by an F-statistic of 5.256 distributed as $F(1,417)$. The associated critical value is 3.840 at the 5 per cent level of significance. For the female equation the null hypothesis of linearity in experience cannot be rejected in favour of piece-wise linearity with a five year split. The computed $F(1,541)$ is 3.170. The validity of choosing a similar cut-off point for both males and females may, therefore, be justifiably questioned. However, in terms of this analysis, it is necessary for obvious reasons to estimate comparable male and female equations.

In general, the estimating equation may be expressed as:[5]

$$\ln(w_i) = \alpha_i + \beta_1 \text{EXP1} + \beta_2 \text{EXP2} + \sum_{j=3}^{n} \beta_j \text{DUMMIES} + u_i \tag{13}$$

where

$\ln(w_i)$ is the natural logarithm of the net hourly wage. The net wage is constructed from the respondents' answers to the questions on the usual number of weekly hours worked in their current job and the usual net weekly earnings from this job (excluding overtime and other abnormal payments). EXP1 and EXP2 are the linear splines in experience. The labour force experience variable is constructed from the respondents' answers to the questions concerning the start dates and terminal dates of all jobs held by the respondents since leaving full-time education. It is thus a precisely calculated variable in contrast to the proxy constructs used by Oaxaca (1973) and Greenhalgh (1980) which could be interpreted as possessing a greater measurement error potential. DUMMIES represent the controlling educational qualifications, plant size, promotion, occupational, industrial and region of residence dummies and u_i is an error term for which assumptions are made below.

Alternatively, experience may be expressed in terms of (0,1) dummies. Eight such dummies are also constructed from the experience variable for each of the eight potential years of labour force experience. Though the results of this analysis are not recorded extensively here, this alternative specification is used for purposes of comparison in terms of the diagnostics of Section VI.

Returning to Equation (13) it should be noted that the standard assumptions are made concerning the error terms:

(i) $E(u_i) = 0$ (an assumed mean of zero).

(ii) $E(u_i^2) = \sigma^2$ (constant variance).

(iii) $E(u_i X_i) = 0$ (the explanatory variables are assumed to be exogenous and omitted variables are assumed orthogonal to the included ones).

(iv) The parameters are assumed constant.

(v) u_i is assumed normally distributed.

The violation of the first assumption may occur through equation misspecification thus inducing bias in the coefficient estimates. Since these coefficient estimates are used in quantifying the magnitude of sex discrimination the correct specification of the equation is of crucial importance.

5. Since estimation is based on those individuals who are currently working another potential selectivity bias problem should be noted.

Biases may lead to inappropriate conclusions in regard to the nature and magnitude of any sex discrimination element.

A priori one would expect violation of the second assumption. In the presence of heteroscedasticity OLS provides unbiased coefficient estimates but a biased and inconsistent covariance matrix estimator. Since consistent standard errors are required for hypothesis testing White's (1980) heteroscedasticity-consistent covariance matrix estimator is employed. The White estimator possesses the appropriate asymptotic properties for hypothesis testing.

Departure from the assumption of normality must also be viewed with some concern. Such a departure may be indicative of some form of mis-specification and may also vitiate the use of standard statistical tests (e.g., the t-statistics and the F-statistics, etc.) which are based on the normal distribution.

In view of the foregoing and the need for some confidence in the results, Section VI contains a number of diagnostic tests that statistically test for model mis-specification, heteroscedasticity and the underlying assumption of normality.

V RESULTS

Tables 1 to 4 contain the male and female regression estimates for the single individuals based on Equation (13). The first two tables contain estimates based on the inclusion of occupation and industry dummies. The latter two tables contain estimates based on the exclusion of these controls. For purposes of comparison the OLS and White-adjusted standard errors are reported in these tables. Statistical inference is based on the White-adjusted standard errors.

Since the dependent variable is in logarithmic form and the explanatory variables are either in levels (the splines) or expressed as (0,1) dummies, some caution must be exercised in their respective interpretations. The coefficients of the experience variables (the splines) are defined as the proportional returns to having an additional year of labour force experience. The dummy coefficients possess a slightly different interpretation. The estimates, themselves, give the differential effect of being in the included category relative to the excluded category (i.e., the reference category). Since the dependent variable is expressed in logarithmic terms, the β_j dummy coefficient is interpreted as e^{β_j}.

Table 1: *Wage Equation Estimates for Single Males*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	0.0219	0.0977	0.1138
Experience (5 or less yrs)	0.0748	0.0132	0.0158***
Experience (more than 5 yrs)	0.0125	0.0186	0.0131
Educational qualifications			
Intermediate Certificate	-0.0406	0.0438	0.0406
Leaving Certificate	0.0424	0.0508	0.0475
University degree	0.2899	0.1160	0.1425**
Other qualifications			
Group Certificate	0.0517	0.0489	0.0554
Apprenticeship	0.0631	0.0400	0.0370*
Basic training qualification	0.0688	0.0364	0.0353*
Job characteristics			
$15 \leqslant$ Firm < 50	0.1664	0.0479	0.0534***
$50 \leqslant$ Firm < 100	0.0968	0.0587	0.0528*
Firm $\geqslant 100$	0.2260	0.0422	0.0474***
Promotion on job	0.0743	0.0323	0.0320***
Occupations			
Professional	0.0531	0.1081	0.1094
Self-employed	-0.0697	0.1110	0.1987
Salaried employees	0.1157	0.1386	0.1363
Intermediate non-manual	-0.1867	0.0847	0.1179
Other non-manual	-0.2174	0.0961	0.1112*
Skilled manual	-0.1509	0.0778	0.1035
Semi-skilled manual	-0.1188	0.0950	0.1089
Industries			
Building & engineering	0.0753	0.0538	0.0491
Transport & communication	0.0927	0.0705	0.0615
Banking & insurance	-0.0561	0.0805	0.0555
Public admin. etc.	0.1055	0.0648	0.0803
Metal manufacturing	0.0492	0.0568	0.0575
Other manufacturing	-0.0221	0.0540	0.0563
Extractive & chemicals	0.0886	0.0726	0.0648
Regions			
Dublin county	0.1270	0.0573	0.0490***
Southern counties	0.1112	0.0532	0.0521**
Midlands counties	0.0319	0.0571	0.0624
Leinster (excl. Dublin)	0.1334	0.0681	0.0686*
Urban	0.0968	0.0394	0.0362***
$\bar{R}^2$	0.312		
Standard error	0.319		
Number of cases	449		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

Table 2. *Wage Equation Estimates for Single Females*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	-0.0611	0.1884	0.0738
Experience (5 or less yrs)	0.0631	0.0078	0.0077***
Experience (more than 5 yrs)	0.0247	0.0176	0.0165
Educational qualifications			
Intermediate Certificate	-0.0653	0.0419	0.0422
Leaving Certificate	-0.0194	0.0417	0.0393
University degree	0.2218	0.1194	0.2261
Other qualifications			
Group Certificate	-0.0524	0.0837	0.1143
Apprenticeship	-0.0111	0.0521	0.0576
Basic training qualification	0.0017	0.0244	0.0244
Job characteristics			
$15 \leqslant$ Firm < 50	0.1385	0.0348	0.0367***
$50 \leqslant$ Firm < 100	0.1491	0.0428	0.0449***
Firm $\geqslant 100$	0.2222	0.0293	0.0302***
Promotion on job	0.0675	0.0220	0.0244***
Occupations			
Professional	0.0640	0.1896	0.0765
Self-employed	0.1227	0.2018	0.1495
Salaried employees	0.3567	0.2351	0.2027*
Intermediate non-manual	0.0534	0.1836	0.0594
Other non-manual	-0.1122	0.1863	0.0815
Skilled manual	-0.0590	0.1880	0.0664
Semi-skilled manual	-0.0688	0.1830	0.0531
Industries			
Building & engineering	-0.0283	0.0914	0.0888
Transport & communication	0.0682	0.0586	0.0487
Banking & insurance	0.1594	0.0349	0.0350***
Public admin. etc.	0.1090	0.0338	0.0343***
Metal manufacturing	0.1213	0.0501	0.0465***
Other manufacturing	0.0289	0.0431	0.0400
Extractive & chemicals	0.1382	0.0648	0.0436***
Regions			
Dublin county	0.0992	0.0397	0.0354***
Southern counties	0.0647	0.0380	0.0350*
Midlands counties	0.0498	0.0459	0.0392
Leinster (excl. Dublin)	0.0051	0.0551	0.0547
Urban	0.0600	0.0282	0.0270**
$\bar{R}^2$	0.391		
Standard error	0.247		
Number of cases	573		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

As Tables 1 and 2 indicate, the returns to labour force experience do indeed vary over an individual's work experience. This is confirmed by examining the coefficients on the splines in both equations and testing to ascertain whether they are statistically different from each other. For both equations this is found to be the case.[6] The actual annual returns to labour force experience are found to be higher for the males in the first five years, 7.8 per cent as compared to 6.5 per cent for the females. For the subsequent years the fall is more dramatic for the males declining to around 1 per cent on average for the years after the fifth year of experience. The decline for the females is less dramatic. They record returns of 2.5 per cent on average for each year subsequent to the fifth. In both cases, however, the higher returns to on the job training in the early years (leading to a steeper earnings profile in the initial years of experience) is a finding compatible with the predictions of human capital theory.

The educational dummy coefficients allow for some conclusions to be drawn concerning returns to educational qualifications. It should be borne in mind that the inclusion of the occupational dummies leads to an under-estimate of the returns to the formal education qualifications as stated above. It is clear from the results of Tables 1 and 2 that possessing an Intermediate Certificate as one's highest qualification has an effect that is not significantly different from zero for either males or females. The returns to a Leaving Certificate are on average higher for males than for females, however, in both equations the coefficients are neither well determined nor statistically significant from zero. On the other hand, returns to a university degree or its equivalent are quite large and in the case of the males statistically significant. In this case the returns are of the order of 33.6 per cent as compared to the females' 24.8 per cent. In general, therefore, males appear to gain more from educational qualifications than do females. This is in direct contrast to the Greenhalgh (1980) findings in regard to returns to educational qualifications. In a sub-sample of single males and single females under the age of 30 derived from the 1975 UK General Household Survey (GHS) Greenhalgh found that the returns to educational qualifications were, in general, higher on average for the single females. The reverse findings obtained here may be indicative of traditional discriminatory practices operating within the Irish educational system which affect the subject uptake of females at secondary and hence tertiary level.

At this point it would be of interest to contrast these results with returns based on the specification excluding occupation and industry dummies. The exclusion of this set of control variables provides a more realistic interpretation

6. The t-statistic associated with the male equation is 3.97*** and that for the female equation 2.35** (where *** denotes 1 per cent and ** denotes 5 per cent significance).

of the returns since the possession of formal educational qualifications is a pre-requisite for admission to certain occupations. As Greenhalgh (1980) points out the inclusion of occupation dummies leads to an understatement of the returns to these qualifications.

For both males and females (Tables 3 and 4, respectively) the returns to a Leaving Certificate and a university degree are more well determined than in the larger specification of Tables 1 and 2. The annual returns to a male holder of either a Leaving Certificate or a university degree is 8.5 per cent and 42.3 per cent respectively. The comparable female figures are 1.9 per cent and 24.8 per cent. In both cases these are greater in magnitude than those from the larger specification with again the female returns being on average lower than the males.

The use of educational qualifications as opposed to years in education (as some studies have employed) is based on a desire to control for the qualitative

Table 3. *Wage Equation Estimates for Single Males*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	-0.1059	0.0650	0.6900
Experience (5 or less yrs)	0.0808	0.0132	0.0157***
Experience (more than 5 yrs)	0.0080	0.0188	0.0132
Educational qualifications			
Intermediate Certificate	-0.0287	0.0439	0.0426
Leaving Certificate	0.0815	0.0475	0.0486*
University degree	0.3528	0.1160	0.1335***
Other qualifications			
Group Certificate	0.0518	0.0491	0.0562
Apprenticeship	0.0465	0.0389	0.0349
Basic training qualification	0.0709	0.0355	0.0327**
Job characteristics			
$15 \leqslant$ Firm < 50	0.1485	0.0474	0.0520***
$50 \leqslant$ Firm < 100	0.1063	0.0568	0.0523**
Firm $\geqslant 100$	0.2289	0.0377	0.0395***
Promotion on job	0.0619	0.0319	0.0298**
Regions			
Dublin county	0.1239	0.0566	0.0515**
Southern counties	0.1236	0.0526	0.0542**
Midlands counties	0.0505	0.0574	0.0638
Leinster (excl. Dublin Co.)	0.1287	0.0687	0.0692*
Urban	0.0954	0.0396	0.0357***
$\bar{R}^2$	0.290		
Standard error	0.324		
Number of cases	449		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

aspects of years in education. However, specifications using years in post-compulsory education as an explanatory variable in the wage equation have also been estimated. Though the full set of results is not reported here, the findings are broadly consistent with those obtained from the reported specifications. The returns to an additional year in post-compulsory education for males and females, when industries and occupations are included, are 4.3 per cent and 1.4 per cent respectively. The comparable returns when excluding these controls are 5.2 per cent and 2.3 per cent.

The above discussion concerning the returns to education is contingent on the assumption that any variables omitted from the specification are orthogonal to the included ones. In terms of education and unobserved ability this may not necessarily be the case. If ability is positively correlated with education and determines wages then OLS yields biased estimates of the education coefficients. Taubman (1976) using US survey data on twins (in

Table 4. *Wage Equation Estimates for Single Females*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	-0.0544	0.0575	0.0605
Experience (5 or less yrs)	0.0643	0.0079	0.0076***
Experience (more than 5 yrs)	0.0298	0.0178	0.0173*
Educational qualifications			
Intermediate Certificate	-0.0236	0.0420	0.0429
Leaving Certificate	0.0697	0.0390	0.0396*
University degree	0.2784	0.1212	0.2369
Other qualifications			
Group Certificate	-0.0279	0.0861	0.1380
Apprenticeship	-0.0544	0.0513	0.0607
Basic training qualification	0.0293	0.0242	0.2406
Job characteristics			
$15 \leq$ Firm < 50	0.1243	0.0351	0.0372***
$50 \leq$ Firm < 100	0.1364	0.0430	0.0447***
Firm ≥ 100	0.2300	0.0281	0.0284***
Promotion on job	0.0736	0.0226	0.0214***
Regions			
Dublin county	0.1287	0.0405	0.0363***
Southern counties	0.0677	0.0392	0.0381*
Midlands counties	0.0477	0.0474	0.0428
Leinster (excl. Dublin)	0.0114	0.0567	0.0575
Urban	0.0599	0.0290	0.0277**
$\bar{R}^2$	0.340		
Standard error	0.256		
Number of cases	573		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

an attempt to control for environmental and genetic ability) found that failing to control for ability causes a large upward bias on the educational coefficients. Hausman and Taylor (1981) exploit the use of US panel data to econometrically control for individual-specific unobservable effects which are assumed correlated with explanatory variables in their wage equation. In marked contrast to Taubman (1976) their empirical results suggest that econometric methods which control for correlation with the latent individual effects increase the schooling coefficient. On the other hand, Chowdhury and Nickell (1985) also use US panel data and extend the Hausman and Taylor (1981) econometric methodology but are unable to obtain any precise estimates for the schooling coefficients. The empirical evidence is clearly divided as to whether controlling for ability bias leads to an increase or a decrease in the returns to education. Since no attempt is made in this study to control for unobservables the cross-sectional estimates on returns to education reported here should be interpreted in a cautionary context.

Returning now to the remaining results of Tables 1 and 2. The effects of the vocational training qualifications are also recorded in Tables 1 and 2. The three examined are the Group Certificate, apprenticeship qualifications and basic training qualifications.[7] In all three cases the female coefficients are badly determined and statistically insignificant. This is in no small part due to the small number of females in these particular categories. In terms of the males the returns to the above are 5.3 per cent, 6.5 per cent and 7.1 per cent respectively.

Variables controlling for firm size were also allowed to enter the analysis. It might be argued that this variable acts as a proxy for unionisation. The larger the plant size the more likely is the possibility of union influence in regard to, for example, pay. The reference group in terms of this set of dummies is firms with less than fifteen individuals. Thus, the coefficients of these dummies should be interpreted in relation to this reference group. It is evident from the results that being employed in a firm with more than one hundred employees has a large wage effect relative to a firm with less than fifteen. The effect is of the order of 25.0 per cent for males and 24.8 per cent for females. One surprising feature of this set of results is the fact that the wage effects for males working in the medium-sized category of firm is less than the effects associated with working in the smaller-sized category. This finding is not, however, repeated for females. The effects, in this case, are found to increase directly with firm size.

Promotion on-the-job is also deemed as having a positive and statistically

7. The basic training is the initial training which provides an individual with the means to exercise a particular trade or profession. This might include, for example, basic training programmes which enable an individual to become an electrician, carpenter, computer programmer, etc. Thus, the variable used here records any qualifications received for this purpose.

significant effect on wages. In terms of their magnitude the effects are not much different across the sexes recording returns of 7.7 per cent and 7.0 per cent for the males and females respectively.

The inclusion of occupation, industry and regional dummies facilitates a ranking of these categories in terms of their effects on wages. The top occupational category for males is salaried employees. The differential for this category relative to the unskilled reference group is of the order of 12.3 per cent. The salaried employees category also ranks top for the females. The estimated differential between this group and the unskilled reference group is of the order of 42.8 per cent. Skilled manual workers are anchored near the bottom of both the male and female occupational rankings. This may be viewed in terms of the firm specific human capital theory alluded to above. It might be argued that skilled workers in the youth labour market are in receipt of a relatively low wage in order that firms can recoup their cost outlays associated with training and this may explain the poor ranking for both sexes.

The industry rankings indicate that males in public administration, etc., are better off than in any other industry. The differential here relative to the distributive trades reference group is of the order of 11.1 per cent. The top-ranking industry for females is the banking and insurance category. Industries with a large concentration of new foreign firms, e.g., those in chemical type industries, feature prominently in the top half of the industry rankings supporting the notion that jobs in these industries are relatively well paid for both males and females.

Reference to the coefficients on the urban and regional dummies suggest some dramatic regional differences in wages. Residing in an urban area with 1,000 or more individuals is recorded as having a statistically significant effect for both males and females, 10.2 per cent and 6.2 per cent respectively. Furthermore, being resident in Dublin county relative to the North-West is worth a differential of nearly 13.5 per cent for the males and 10.4 per cent for the females with the most disadvantaged region in wage terms being the reference group itself, the North-West. This finding reflects the more favourable labour market conditions obtaining in Dublin relative to the rest of the country and in no small part could be attributed to the large number of public administration employees in this area. This finding is resonant of one of the Walsh and Whelan (1976) findings.

In general, the remaining results that feature in Tables 3 and 4 are broadly in agreement with those obtained from the larger specification and require no further comment.[8]

8. The industry and occupation dummies are statistically tested on the basis of F-tests for both the male and female equations. The F-test associated with the male equation is 1.989** and that of the female equation 4.338*** (where *** denotes 1 per cent and ** denotes 5 per cent significance).

VI DIAGNOSTIC TESTS

The existence of specification error in any of the above equations has serious consequences for the coefficient estimates and hence the discrimination coefficients as estimated below. This would be particularly so if the misspecification were due to the omission of relevant variables. The prime example in the literature of such an omitted variable is the motivation for work or ability which does not explicity enter the estimating equations.

In order to have confidence in the discrimination estimates it is important to test the underlying equations. The purpose of this section is to present an array of diagnostics that explicitly test the underlying assumptions of the estimated models. These assumptions are the standard ones of the classical linear regression model as referred to in Section IV. The error terms are assumed to have a zero mean, a constant variance and are normally distributed. The statistical tests employed here provide tests of these assumptions.

Tables 5 and 6 report the diagnostic results for the male and female equations respectively. For each sex four equations are estimated and compared in terms of these diagnostics. In these tables (A) and (C) refer to Equation (13) with and without the occupation and industry dummies, respectively. The results of these estimated models have been the subject of extensive discussion in Section V. (B) and (D) refer to an alternative version of Equation (13) that uses dummy variables for the different years of labour force experience instead of splines. The former includes occupation and industry dummies and the latter excludes them.[9]

The first diagnostic to be examined is the J-test developed by Davidson and MacKinnon (1981) and suggested by the authors as a specification test. It is a non-nested test procedure and is interpreted in the context of augmenting the conditional mean (see Pagan (1984)). The proxy variable used in this case is based on predictions from the alternative or competing model. The validity of the null hypothesis (H_0) is evaluated in terms of the t-statistic associated with the proxy variable. If H_0 is true, the coefficient on the proxy variable will not be significantly different from zero. This can be tested by a simple t-statistic. As Davidson and MacKinnon(1981) point out the t-statistic which is valid for testing the truth of H_0 is not valid for testing the truth of H_1. Therefore, the roles of H_0 and H_1 are reversed and the test is carried out again. Thus, it is possible that both hypotheses may be rejected or both not rejected or one not rejected over the other. The J-tests recorded

9. Though the coefficient estimates of these alternative models are not reported in the text they are available on request to the author.

Table 5. *Diagnostics for Male Equations*

Test	*(A)*	*(B)*	*(C)*	*(D)*
J-test	1.354	2.355*	1.535	2.827**
RESET	2.765*	3.377**	1.396	1.203
Breusch-Pagan	171.053**	187.555**	95.101	113.132
Skewness	1.721	2.093	5.412*	5.695*
Kurtosis	788.880**	686.165**	862.732**	783.078**

(A) and (C) refer to the estimated equations, using splines in experience, including and excluding occupation and industry dummies respectively. (B) and (D) refer to the estimated equations, using dummies in experience, including and excluding occupation and industry dummies respectively. The non-nested J-test value is interpreted as a t-statistic. The values for the RESET mis-specification test is interpreted as an F-test. The Breusch-Pagan test for heteroscedasticity is interpreted as a χ^2 variate with degrees of freedom equal to the number of parameters estimated in the original equations less one. The normality tests of skewness and kurtosis are based on the Keifer-Salmon test statistic and are interpreted as χ^2 variates with one degree of freedom each. ** and * denotes significance at the 1 per cent and 5 per cent level of significance respectively.

Table 6. *Diagnostics for Female Equations*

Test	*(A)*	*(B)*	*(C)*	*(D)*
J-test	1.429	1.483	2.304*	1.041
RESET	8.698**	8.084**	5.582**	4.760**
Breusch-Pagan	149.649*	156.552*	112.673	112.528
Skewness	0.004	0.001	0.553	0.853
Kurtosis	88.426**	78.011**	169.106**	142.407**

See Table 5 for a full description of these tests. ** and * denotes significance at the 1 per cent and 5 per cent level of significance respectively.

in Tables 5 and 6 are the t-statistics associated with the proxy variables generated from the competing model. The null hypothesis alternates between the treatment of the labour force experience variable in terms of splines or in terms of dummy variables. The results provide unambiguous support in favour of the specification using splines for the male equations. The female results are, however, more ambiguous. In general, however, one would expect the continuous spline variables to dominate the discrete dummy variables and the J-test results simply confirm this.

The second specification diagnostic is the RESET test due to Ramsey (1969). As Pagan (1984) points out, it again can be interpreted in terms of augmenting the conditional mean with a proxy variable (or proxy variables) assumed to be closely correlated with the omitted part of the conditional mean. The proxy variables, in this case, are based on the predicted values of the dependent variable from the original specification raised to a number of arbitrary powers. Simulation studies suggest that the optimal number of powers is four. The RESET test then reduces to an F-test of the significance from zero of the proxy variables. The success of the test is strongly dependent on the closeness of the correlation between these proxy constructs and the omitted part of the conditional mean.[10]

In terms of the male equations the null hypothesis that there is no omitted part of the conditional mean cannot be rejected for two of the four estimated versions at the 1 per cent level of significance. For the female equation the same null hypothesis is decisively rejected for all the estimated versions confirming some form of potential mis-specification in the female equation.

The results of the RESET tests again establish the necessity for exercising extreme caution in the interpretation of sex discrimination estimates. It is obvious that regardless of which wage structure is assumed in the absence of sex discrimination the potential bias in the female coefficients will impinge upon the discrimination estimates. It should also be pointed out, in the light of the discussion in Section V, that the estimated returns to education and work experience for the female equations may also be rendered somewhat dubious.

The third diagnostic focuses on the assumption of constant variance.[11] The heteroscedasticity test employed here is the Lagrangean Multiplier (LM) statistic developed by Breusch and Pagan (1979). This again may be interpreted in the context of augmenting a conditional moment of the original specifica-

10. The test was originally described in terms of BLUS residuals by Ramsey (1969) but it was subsequently shown by Ramsey and Schmidt (1976) that this was equivalent to carrying out the above F-test in terms of the OLS residuals.

11. It might be argued that the appropriate test statistic to use in this context is one derived from White's (1982) information matrix test. This information matrix test provides a test of the information matrix identity. Since LR tests, LM tests and Wald tests (which are related to F-tests and t-tests) are derived under the assumption that this identity holds, White (1982) suggests that the information matrix test should be used as a preliminary test to inference. Hall (1984) provides a decomposition of the information matrix test into the sum of three independent χ^2 variates, the first of which is White's (1980) direct test for heteroscedasticity with the remaining two components independent test statistics for skewness and kurtosis. However, the computation of the White heteroscedasticity test involves, for the largest specification reported here (that of Table 10), the estimation of an auxiliary regression with potentially 630 explanatory variables. In view of this the more easily computable Breusch-Pagan LM test is instead used.

tion. In this case the conditional variance. The potential heteroscedasticity to be tested is of the form:

$$\sigma_i^2 = \bar{G}_i \alpha \qquad (14)$$

where

σ_i^2 = the variance and

$\bar{G}_i$ =a matrix of variables, the first column of which is ones. In the context of this study the $\bar{G}$ matrix is simply composed of all the variables from the original specification. The test statistic is calculated by first obtaining the OLS residuals from the original equation. The squared residuals from this equation are then deflated by the Maximum Likelihood estimate of the error variance from the original regression. This newly constructed variable is then regressed on the $\bar{G}$ variables and the actual test statistic is half the explained sum of squares from this regression. The resultant LM statistic is a χ^2 variate with the k - 1 degrees of freedom where k is the number of parameters from the original equation.

The results of the Breusch-Pagan tests indicate a general rejection of the null hypothesis of homoscedasticity for most of the estimated equations. Since the standard errors employed in the analysis are the White adjusted standard errors this need not present a problem in the context of inference. The decisive rejection may be related more to mis-specification. However, due to the limitations of the data set employed obtaining a handle on a more appropriately specified model has proved almost impossible. The models estimated represent the best description of the data given this set of limitations.

A further test also aimed at detecting mis-specification has been proposed by Kiefer and Salmon (1983). Like the RESET test its advantage lies in its ease of computation but unlike the RESET test it focuses on departures from the normality assumption. This specification test is based on an Edgeworth expansion and is written as:

$$S = \frac{N}{6}(\hat{\mu}_3 - 3\hat{\mu}_1)^2 + \frac{N}{24}(\hat{\mu}_4 - 6\hat{\mu}_2 + 3)^2 \qquad (15)$$

where N is the number of observations and the $\hat{\mu}_i$'s are the estimated sample moments based on the residuals. Following Davidson and MacKinnon (1985) the estimated sample moments are standardised by the Maximum Likelihood estimate of the standard error from the original regression. The test statistic is composed of the sum of two asymptotically independent χ^2 variates each with one degree of freedom. The test statistic has the advantage that attention can exclusively focus on either the third moment (the first part of (15)) or the fourth moment (the second part of (15)) or both. Thus one can examine the null hypothesis of skewness which has a χ_1^2 distribution independently of the null hypothesis of kurtosis (which also has a χ_1^2 distribution).

In terms of the Kiefer-Salmon normality test the female equations perform better than the male equations. Skewness is upheld for all the estimated female equations but rejected for two of the four male equations. The imposition of the restrictions on the occupation and industry dummies leads to an increase in the skewness and kurtosis test values. This is as one would expect if these restrictions are invalidly imposed. The joint test of normality is, however, clearly rejected for all equations.

The violation of the normality assumption has clear implications for the reliability of the classical statistical tests (t-tests and F-tests) which are based on the normal distribution. The coefficients, however, remain unbiased. Accepting this, however, the violation of the kurtosis assumption may be again reflecting some underlying mis-specification. In terms of the normal distribution the fourth moment is equal to three times the square of the second moment (the variance). The decisive rejection of kurtosis may thus be suggestive of some underlying heteroscedasticity again supporting the findings of the Breusch-Pagan test.

The performance of the estimated equations in terms of the diagnostics is clearly mixed. In view of this, caution should be exercised in the interpretation and use of the parameter estimates. This cautionary note should be borne particularly in mind in the context of the following section.

VII ESTIMATES OF SEX DISCRIMINATION

Section II outlined in detail the methodology to be adopted in estimating the wage effects of sex discrimination. A male wage structure is assumed to represent best the hypothetical set of conditions that obtain in the absence of sex discrimination. The observed differential stated in (11) is repeated here:

$$\ln(1 + G) = \Delta\overline{Z}\hat{\beta}_m - \overline{Z}_f\Delta\hat{\beta} \qquad (11'')$$

It is obvious that one could equally well have assumed a female wage structure as prevailing in the absence of sex discrimination. In that case the observed differential would be given by:

$$\ln(1 + G) = \Delta\overline{Z}\hat{\beta}_f - \overline{Z}_m\Delta\hat{\beta} \qquad (16)$$

In both (11″) and (16) the first term in the expression represents the portion of the observed differential explained by characteristics. The latter term represents that portion due to the existence of differing coefficients given the same set of characteristics. Due to the "index number" problem estimates of (11″) and (16) will be different and sensitive to the assumed wage structure (see Sloane (1985)). The results presented in this section are based on the assumption of a male wage structure. This seems a more appropriate approach to adopt since it purports to ask what females would receive in terms of wages for a given set of characteristics if they were presented with a set of male opportunities.

Estimates of the explained and unexplained portions of the differential are calculated for the models including occupations and industries and for those excluding these controlling characteristics. Standard errors are also calculated for both parts of the observed differential (see Appendix 2).

Examining the estimates based on (11″) above reveals that the observed differential $\ln(1 + G) = 0.1037$ at the mean. The portion unexplained by differing characteristics (i.e., discrimination) $\ln(1 + D) = 0.0305$. This constitutes a little under 30 per cent of the observed differential. Thus, according to these results young single male workers are interpreted as being in receipt of wages that are on average approximately 11 per cent greater than their young female counterparts. The greater proportion of this differential is explained by the possession of differing characteristics with a relatively small amount due to the presence of unexplained factors which is taken, in terms of this analysis, to be sex discrimination. The asymptotic standard error associated with the $\ln(1 + D)$ of 0.0305 is 0.0389 suggesting that the unexplained differential is not statistically significant. That part of the observed differential explained by characteristics (i.e., the difference between $\ln(1 + G)$ and $\ln(1 + D)$) is equal to 0.0732 and its associated standard error is 0.0300. The explained differential is, on the basis of this normally distributed test statistic, statistically and significantly different from zero.

Examining the average may provide a misleading picture of how the differential behaves across different types or categories of individuals. Table 7 contains differential calculations based on a male wage structure for different stylised female workers. The first such stylised individual falls into the base group (i.e., has an amount of labour force experience equal to the female mean in the sample and scores zero on all the binary variables controlled for in the analysis). The differential due to discrimination in this case is estimated at 12.9 per cent but again is not recorded as being statistically significant. The remainder of the table reports deviations from this base set of characteristics. Each deviation is examined by itself alone with the objective of trying to establish whether there exists a large variation in the differential across differing characteristics. Asymptotic standard errors are also recorded to establish statistical significance. Though the variation is found to be large in regard to some characteristics all but one is found to be statistically insignificant. The differential is seen to decline with labour force experience. The more experience an individual possesses the smaller is the unexplained differential. Again, however, this effect is not found to be statistically significant.

The only significant (and the largest) unexplained differential is obtained for the Leinster region (excl. Dublin County). The differential is of the order of 28 per cent and its magnitude is explained by the vast difference between the male and female coefficients associated with this regional dummy. This large differential could be explained by the fact that males resident in the Leinster region are more able to commute to well paying jobs in Dublin county than are the Leinster resident females. This, however, can only be offered as a tentative explanation for what is a rather odd result.

Another interesting feature of Table 7 is that there exists some evidence of "reverse" discrimination in terms of three occupational categories; the self-employed, salaried employees and the intermediate non-manual categories. In terms of the self-employed discrimination usually takes the form of consumer motivated discrimination. An interpretation for the result recorded here is that consumers are more likely to discriminate against young self-employed male workers than their female counterparts. The magnitude of the effect in this case is a little under 7 per cent but is not statistically significant.

The negative effect in the intermediate non-manual category (an effect of over 11 per cent in favour of the females) helps explain the relatively low average value recorded in the first row of Table 7. Since the mean discrimination coefficient, in this case, is the difference in coefficients weighted by the female mean characteristics, the mean estimate will clearly be influenced by the proportions in certain categories. Since over 70 per cent of all the females in the sample are in the intermediate non-manual category this clearly has a dampening effect on the mean estimate. This clearly highlights one of the dangers associated with using this approach.[12]

Table 8 contains calculations of the sex differential based on the estimated models of Tables 3 and 4 (i.e., those excluding the occupation and industry dummies). As anticipated in Section II, the mean estimate of the unexplained differential increases dramatically and becomes statistically significant. The $\ln(1 + D)$ estimate is 0.0847 and with a standard error of 0.0264 is statistically significant and different from zero. The portion of the differential explained by characteristics falls and becomes statistically insignificant. This clearly illustrates the role played by occupation and industry compensating differentials and their effects on the observed differential. Failure to control for

12. A more sensitive treatment of occupations along the lines of estimating within occupation wage equations while correcting for selectivity bias is clearly required and is to be the subject of further investigation.

Table 7. *Estimates of Sex Differential for Different Types of Stylised Individuals Including Industries and Occupations*

Characteristic	*ln(1 + D)*	*1 + D*	*ASE of ln(1 + D)*
Mean	0.0305	1.0310	0.0389
Base	0.1212	1.1288	0.1252
Educational qualifications			
Intermediate Certificate	0.1459	1.1571	0.1267
Leaving Certificate	0.1829	1.2007	0.1317
University degree	0.1892	1.2083	0.3004
Other qualifications			
Group Certificate	0.2253	1.2527	0.1708
Apprenticeship	0.1954	1.2158	0.1477
Basic training qualification	0.1882	1.2071	0.1358
Job characteristics			
$15 \leqslant$ Firm < 50	0.1490	1.11607	0.1291
$50 \leqslant$ Firm < 100	0.0689	1.0713	0.1351
Firm $\geqslant 100$	0.1250	1.1331	0.1381
Promotion on job	0.1280	1.1365	0.1349
Occupations			
Professional	0.1103	1.1166	0.1136
Self-employed	-0.0713	0.9312	0.2158
Salaried employees	-0.1198	0.8871	0.2309
Intermediate non-manual	-0.1190	0.8878	0.1065
Other non-manual	0.0159	1.0160	0.1066
Skilled manual	0.0292	1.0296	0.1080
Semi-skilled manual	0.0711	1.0737	0.1098
Industries			
Building & engineering	0.2248	1.2521	0.1469
Transport & communication	0.1341	1.1567	0.1341
Banking & insurance	-0.0943	0.9100	0.1316
Public admin. etc.	0.1176	1.1248	0.1280
Metal manufacturing	0.0491	1.0503	0.1266
Other manufacturing	0.0702	1.0727	0.1292
Extractive & chemicals	0.0716	1.0742	0.1286
Regions			
Dublin county	0.1489	1.1606	0.1277
Southern counties	0.1677	1.1826	0.1204
Midlands counties	0.1033	1.1088	0.1270
Leinster (excl. Dublin)	0.2495	1.2834	0.1296*
Urban	0.1579	1.1710	0.1239
Work experience			
Five years	0.1418	1.1523	0.1300
Six years	0.1296	1.1384	0.1291
Seven years	0.1174	1.1246	0.1315
Eight years	0.1053	1.1110	0.1372

See Table 8 for a full explanation. *** denotes significance at 1 per cent level, ** significance at the 5 per cent level and * significance at the 10 per cent level.

these effects clearly distorts the wage discrimination estimates. In view of this, the more realistic estimates of discrimination are assumed obtained from the larger occupation/industry specification.

Table 8. *Estimates of Sex Differential for Different Types of Stylised Individuals Excluding Industries and Occupations*

Characteristic	*ln(1 + D)*	*1 + D*	*ASE of ln(1 + D)*
Mean	0.0847	1.0884	0.0264***
Base	0.0016	1.0016	0.0800
Educational qualifications			
Intermediate Certificate	-0.0035	0.9965	0.0781
Leaving Certificate	0.0133	1.0134	0.0806
University degree	0.0759	1.0788	0.2768
Other qualifications			
Group Certificate	0.0813	1.0847	0.1513
Apprenticeship	0.1025	1.1079	0.0916
Basic training qualification	0.0432	1.0441	0.0860
Job characteristics			
$15 \leqslant$ Firm < 50	0.0257	1.0260	0.0806
$50 \leqslant$ Firm < 100	-0.0285	0.9719	0.0843
Firm $\geqslant 100$	0.0004	1.0004	0.0781
Promotion on job	-0.0101	0.9899	0.0787
Regions			
Dublin county	-0.0033	0.9967	0.0854
Southern counties	0.0574	1.0591	0.0872
Midlands counties	0.0044	1.0044	0.0949
Leinster (excl. Dublin)	0.1188	1.1261	0.0995
Urban	0.0370	1.0377	0.0781
Work experience			
Five years	0.0301	1.0306	0.0889
Six years	0.0083	1.0083	0.0854
Seven years	-0.0136	0.9865	0.0872
Eight years	-0.0354	0.9652	0.0943

The first row records the main differential estimate. The second row records the differential estimate for an individual with a base set of characteristics. The third and subsequent rows record deviations from the base characteristics and are allowed to occur singly. The fourth column in the above table reports the asymptotic standard errors (ASE) of the estimated differentials. *** denotes significance at the 1 per cent level, ** signifiance at the 5 per cent level and * significance at the 10 per cent level.

As in Table 7 base calculations and deviations from the base are also reported. Again, despite the significance of the mean differential, none of the remaining set of calculations is statistically significant. The findings of Table 7 are, more or less, repeated here. However, in contrast to Table 7 it appears evident that the larger the firm size the less likely is there of a wage differential in favour of males. Furthermore, the Leinster differential becomes insignificant.

It might be asked how these estimates compare with the UK evidence. Since Greenhalgh (1980) employed a roughly similar methodology to the one adopted in this study the estimates contained therein will be used for comparison purposes. The data set used was the 1975 GHS and the closest comparable group Greenhalgh examined of interest in terms of this study is the under 30 single men/single women subset. In the context of this group an unexplained residual estimate of 10 per cent was obtained. However, no standard errors were calculated for this or any other of the estimates recorded there. Since this is a 1975 estimate and is based on a group with an older terminal age than the one employed here a cross-country comparison between the two must remain at least tentative and at most crude. It must remain a matter of conjecture as to how much of the 7 percentage points that separate the two estimates could be explained by (i) the passing of time and the influence of equal pay legislation, (ii) differences in the structure of the Irish and UK labour markets and (iii) the fact that the specification estimated here controls explicitly for certain job characteristics (e.g., firm size and on-the-job promotion) in a way that the Greenhalgh model did not. One may speculate but it would be a major surprise if more recent UK estimates did not more closely mirror the Irish estimates.

In general Tables 7 and 8 reveal little statistical difference between the coefficients of the male and female equations. In the light of this a more parsimonious model using the pooled sample of males and females with a sex dummy is estimated. The sex dummy adopts a value of 1, if male and 0 otherwise.[13] The full sample specification is estimated with occupation and industry controls and the results of this exercise are contained in Table 9. Most of the coefficient estimates are in line with those reported in Tables 1 and 2. The coefficient of most interest in Table 9 is the sex coefficient. This suggests that males on average earn 7.5 per cent more than females. This figure is in contrast to the 11 per cent wage differential obtained from the above analysis. Sex interactions are also estimated using the full sample. Of the full set of interactions attempted only two are statistically significant at conventional levels of significance. As Table 10 indicates the two significant interactive terms are sex and the Banking and Insurance industry category and sex and the intermediate non-manual occupational category. The co-

13. Chow tests are carried out to statistically test whether the two separate sex models fit the data better than the pooled sample model constrained by the inclusion of a sex dummy. Chow tests for the specifications including occupation and industry dummies and excluding these controls are calculated. The null hypotheses are the sex dummy constrained models. The resultant F-tests are 1.046 ($F(31,958)$) and 0.645 ($F(31,986)$). Neither null hypotheses can be rejected.

Table 9. *Wage Equation Estimates for Pooled Sample*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	0.0022	0.0762	0.1041
Sex	0.0722	0.0257	0.0282**
Experience (5 or less yrs)	0.0651	0.0069	0.0074***
Experience (more than 5 yrs)	0.0217	0.0123	0.0099**
Educational qualifications			
Intermediate Certificate	-0.0316	0.0293	0.0304
Leaving Certificate	0.0284	0.0307	0.0312
University degree	0.3049	0.0798	0.1166***
Other qualifications			
Group Certificate	0.0471	0.0372	0.0465
Apprenticeship	0.0560	0.0278	0.0292*
Basic training qualification	0.0424	0.0197	0.0196**
Job characteristics			
$15 \leqslant$ Firm < 50	0.1465	0.0284	0.0322***
$50 \leqslant$ Firm < 100	0.1255	0.0347	0.0344***
Firm $\geqslant 100$	0.2251	0.0242	0.0263***
Promotion on job	0.0682	0.0185	0.0183***
Occupations			
Professional	-0.0636	0.0745	0.1034
Self-employed	-0.0472	0.0823	0.1644
Salaried employees	0.0997	0.1044	0.1353
Intermediate non-manual	-0.1120	0.0654	0.1035
Other non-manual	-0.2507	0.0699	0.1034**
Skilled manual	-0.1420	0.0636	0.0953
Semi-skilled manual	-0.1520	0.0704	0.1004
Industries			
Building & engineering	0.0776	0.0407	0.0405*
Transport & communication	0.0792	0.0445	0.0378**
Banking & insurance	0.1224	0.0326	0.0312***
Public admin. etc.	0.0894	0.0304	0.0329***
Metal manufacturing	0.0683	0.0364	0.0365*
Other manufacturing	-0.0032	0.0334	0.0342
Extractive & chemicals	0.1199	0.0469	0.0397***
Regions			
Dublin county	0.1174	0.0329	0.0299***
Southern counties	0.0893	0.0311	0.0312***
Midlands counties	0.0463	0.0356	0.0371
Leinster (excl. Dublin)	0.0805	0.0427	0.0457*
Urban	0.0762	0.0231	0.0220***
$\bar{R}^2$	0.357		
Standard error	0.281		
Number of cases	1,022		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

Table 10. *Wage Equation Estimates for Pooled Sample with Sex Interactions*

Variable	*Coefficient*	*OLS se*	*White se*
Constant	-0.0646	0.0784	0.1018
Sex	0.1434	0.0330	0.0333***
Experience (5 or less yrs)	0.0660	0.0069	0.0074***
Experience (more than 5 yrs)	0.0215	0.0122	0.0099**
Educational qualifications			
Intermediate Certificate	-0.0375	0.0292	0.0299
Leaving Certificate	0.0220	0.0307	0.0303
University degree	0.2831	0.0796	0.1165**
Other qualifications			
Group Certificate	0.0411	0.0371	0.0465
Apprenticeship	0.0619	0.0279	0.0298**
Basic training qualification	0.0320	0.0198	0.0199
Job characteristics			
15 ≤ Firm < 50	0.1480	0.0283	0.0318***
50 ≤ Firm < 100	0.1244	0.0346	0.0339***
Firm ≥ 100	0.2264	0.0241	0.0262***
Promotion on job	0.0705	0.01838	0.0181***
Occupations			
Professional	-0.0023	0.0759	0.1003
Self-employed	-0.0181	0.0825	0.1599
Salaried employees	0.1559	0.1051	0.1292
Intermediate non-manual	-0.0368	0.0701	0.0988
Other non-manual	-0.2082	0.0706	0.1005**
Skilled manual	-0.1395	0.0633	0.0945
Semi-skilled manual	-0.1208	0.0707	0.0980
Industries			
Building & engineering	0.0704	0.0405	0.0405*
Transport & communication	0.0781	0.0442	0.0374**
Banking & insurance	0.1451	0.0351	0.0335***
Public admin. etc.	0.0945	0.0303	0.0329***
Metal manufacturing	0.0714	0.0362	0.0361***
Other manufacturing	-0.0017	0.0332	0.0340
Extractive & chemicals	0.1166	0.0467	0.0392***
Regions			
Dublin county	0.1207	0.0328	0.0295***
Southern counties	0.0939	0.0310	0.0307***
Midlands counties	0.0460	0.0354	0.0368
Leinster (excl. Dublin)	0.0806	0.0425	0.0450*
Urban	0.0783	0.0230	0.0220***
Interaction terms			
Sex x Banking & insurance	-0.1688	0.0573	0.0685**
Sex x Intermediate non-manual	-0.1361	0.0501	0.0525***
$\bar{R}^2$	0.364		
Standard error	0.279		
Number of cases	1,022		

Statistical inference is based on the White standard errors. Two-tailed test of significance are employed and *** denotes significance at the 1 per cent level, ** at the 5 per cent level and * at the 10 per cent level.

efficients of the interactive terms are interpreted as the *ceteris paribus* differences between the male and female coefficients for the given categories. For both the categories in question the female coefficients are statistically and significantly larger than the male coefficients. In terms of Table 7 above a similar finding is reported for these two categories with the effect, however, statistically insignificant. The difference in statistical significance is due to the fact that the results of Table 7 explicitly assume a male wage structure. For completeness, diagnostics for these two equations are reported in Table 11.

Table 11. *Diagnostics for Full Sample Equations Including Industries and Occupations*

Test	*(A)*	*(B)*	*(C)*	*(D)*
J-test	0.688	2.483**	0.315*	2.782**
RESET	8.740**	9.656**	8.022**	8.824**
Breusch-Pagan	301.398**	326.377**	319.699**	345.164**
Skewness	3.646	3.467	2.713	2.506
Kurtosis	1,403.277**	1,402.493**	1,316.547**	1,081.530**

(A) and (C) refer to the estimated equations using splines in experience with occupation and industry dummies, the results for which are reported in Tables 9 and 10 respectively. (B) and (D) refer to the estimated equations using dummies in experience with occupation and industry dummies, the results for which are not reported here but are available on request. The non-nested J-test value is interpreted as a t-statistic. The values for the RESET mis-specification test is interpreted as an F-test. The Breusch-Pagan test for heteroscedasticity is interpreted as a χ^2 variate with degrees of freedom equal to the number of parameters estimated in the original equations less one. The normality tests of skewness and kurtosis are based on the Keifer-Salmon test statistic and are interpreted as χ^2 variates with one degree of freedom each. ** and * denotes significance at the 1 per cent and 5 per cent level respectively.

VIII CONCLUSIONS

The estimation of human capital wage equations has provided an opportunity to examine the returns to both labour force experience and educational qualifications. By and large, the results are broadly compatible with the predictions of human capital theory. The returns to on-the-job training are greater for young males in the first five years of labour force experience than for young females. However, the returns are found to diminish for the males by a more dramatic amount in the subsequent years than is the case for the young females. Returns to educational qualifications are on average higher for males than females and this might be interpreted as reflecting some form of discrimination in terms of female access to certain subjects within the Irish educational system. The differences in returns to educational qualifications are not found to be statistically significant.

The general view that emerges from the above exercise is that the data do not provide any convincing evidence in support of wage discrimination in the context of the labour market for young workers. This should not be interpreted as suggesting that sex discrimination *per se* is an absent phenomenon from the Irish labour market for young workers. It can be stated that the results of the above analysis provide scant statistical evidence in support of a discrimination effect that originates through wage differences. However, the approach adopted would not be expected to detect forms of sex discrimination that occur through, for example, the existence of barriers to occupational entry or employer motivated discrimination in terms of on-the-job promotion offers. One may conjecture that the small magnitude recorded for the unexplained wage differential may be attributable to the success of equal pay legislation. Wages are the one obvious variable that can be easily regulated by anti-discrimination legislation. Introducing and implementing legislation to remove wage discrimination may be a far easier task than removing certain other forms of employer motivated discrimination that manifest themselves through, for example, promotional offers to females. Failure to detect sex discrimination in the form of wage effects does not necessarily imply the absence of sex discrimination in its other forms.

Therefore, the results obtained cannot claim to represent the definitive statement on sex discrimination in the labour market for young workers in Ireland. Since the focus of attention has been young single workers the discrimination effect measured here does not relate to any discrimination that may occur as a consequence of female labour force intermittency. Nor has the focus here been on other forms of discrimination that may arise as a consequence of either occupational segregation or unequal access to promotion. The absence of wage sex discrimination cannot be interpreted as *prima facie* evidence against the existence of any of these other types of discrimination. These forms of discrimination must be the subject of investigation along the lines suggested in Section VII before any firm conclusions on this particular issue are allowed to be drawn.

REFERENCES

BECKER, G.S., 1971. *The Economics of Discrimination*, University of Chicago Press, second edition.

BREUSCH, T.S., and A.R. PAGAN, 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica*, Vol. 47, pp. 1287-1294.

BROWN, R.S., M. MOON, and B.S. ZOLOTH, 1980. "Incorporating Occupational Attainment in Studies of Male/Female Earnings Differentials", *Journal of Human Resources*, Vol. 15, pp. 3-28.

CHOWDHURY, G., and S. NICKELL , 1985. "Hourly Earnings in the United States: Another Look at Unionisation, Schooling, Sickness, and Unemployment Using PSID Data", *Journal of Labor Economics*, Vol. 3, pp. 38-69.

DAVIDSON, R., and J.G. MacKINNON, 1981. "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, Vol. 49, pp. 781-793.

DAVIDSON, R., and J.G. MacKINNON, 1985. "The Interpretation of Test Statistics", *Canadian Journal of Economics*, Vol. 18, pp. 38-57.

GREENHALGH, C., 1980. "Male/Female Wage Differentials in Great Britain: Is Marriage an Equal Opportunity?", *The Economic Journal*, Vol. 90, pp. 751-775.

HALL, A.R., 1984. "The Information Matrix Test for the Linear Model", *Warwick Economic Research Papers*, No. 250.

HAUSMAN, J.A., and W.E. TAYLOR, 1981. "Panel Data and Unobservable Individual Effects", *Econometrica*, Vol. 49, pp. 1377-1397.

KIEFER, N.M., and M. SALMON, 1983. "Testing Normality in Econometric Models", *Economics Letters*, Vol. 11, pp. 123-127.

MINCER, J., 1974. "Schooling, Experience and Earnings", *National Bureau of Economic Research.*

OAXACA, R.L., 1973. "Male/Female Wage Differentials in Urban Labour Markets", *International Economic Review*, Vol. 14, pp. 693-709.

PAGAN, A., 1984. "Model Evaluation by Variable Addition", Chapter 3 in Hendry, D. and K.F. Wallis (eds), *Econometrics and Quantitative Economics*, Blackwell.

RAMSEY, J.B., 1969. "Tests for Specification in Classical Linear Least Squares Regression Analysis", *Journal of Royal Statistical Society*, Vol. 31, pp. 350-371.

RAMSEY, J.B., and P. SCHMIDT, 1976. "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error", *Journal of the American Statistical Association*, Vol. 67, pp. 389-390.

SLOANE, P.J., 1985. "Discrimination in the Labour Market", Chapter 2 in *Labour Economics*, D. Carline, *et al.*, Longman.

STEWART, M.B., 1987. "Collective Bargaining Arrangements, Closed Shops and Relative Pay", *The Economic Journal*, Vol. 97, pp. 140-156.

STEWART, M.B., and K.F. WALLIS, 1981. *Introductory Econometrics*, Blackwell, second edition.

TAUBMAN, P., 1976. "Earnings, Education, Genetics and Environment", *The Journal of Human Resources*, Vol. 11, pp. 447-461.

WALSH, B.M., and B.J. WHELAN, 1976. "A Micro-Economic Study of Earnings in Ireland", *The Economic and Social Review*, Vol. 7, pp. 199-217.

WHITE, H., 1980. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica*, Vol. 48, pp. 817-838.

WHITE, H., 1982. "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, Vol. 50, pp. 1-25.

APPENDIX 1

The dependent variable used in this analysis is the natural logarithm of the average net hourly wage ($\ln(w_i)$). This variable thus excludes abnormal hours worked and overtime payments.

The independent or control variables used are:

(i) Labour force experience: This was calculated as the aggregate duration of all jobs (including the current job) held by the i^{th} individual. The unit of measurement used for the experience variable is, in this case, years. Since this is a precisely calculated variable the usual errors-in-variables problem that characterises similar studies does not exist in regard to this particular variable.

(ii) Three (0,1) dummies for the highest educational qualifications obtained by the i^{th} individual. These qualifications are the Intermediate Certificate, the Leaving Certificate and a university degree or its equivalent. Individuals who have commercial course diplomas are allocated to either the Intermediate or the Leaving Certificate category depending on whether they possess one or other of these public examinations. The coefficients on these variables will allow some judgement to be made on the returns to educational qualifications. In terms of estimation the obvious reference group in this case are those individuals in the sample who have no such qualifications.

(iii) A (0,1) Group Certificate dummy variable adopting a value of 1 if the individual possesses such a certificate, 0 otherwise.

(iv) A (0,1) apprenticeship dummy variable adopting a value of 1 if the individual successfully completed an apprenticeship scheme, 0 otherwise.

(v) A (0,1) basic training qualification dummy variable adopting a value of 1 if the individual successfully completed such a qualification, 0 otherwise. These training qualifications refer to formal instruction outside the context of the educational system leading to formal qualifications.

(vi) A (0,1) promotion dummy adopting a value of 1 if the individual received promotion on the current job, 0 otherwise.

(vii) Four (0,1) dummy variables designed to capture the effects of variation in plant, firm or organisation size. The four categories are plants with less than fifteen individuals (including the proprietor), plants with fifteen and over but less than fifty individuals (including the proprietor), plants with fifty and over but less than a hundred individuals (including the proprietor), and plants with a hundred or over individuals (including the proprietor). The dummy variable adopts a value of 1 if the individual falls into any of these mutually exclusive categories, 0 otherwise. The reference group in terms of this set of dummy variables is the first category of less than fifteen individuals.

(viii) A (0,1) urban dummy variable adopting a value of 1 if the individual resides in a town of 1,000 or more, 0 otherwise.

(ix) Eight (0,1) industry dummies which are:

(a) Energy, water, extraction and processing of non-energy-producing minerals and derived products and chemicals.
(b) Metal manufacturing (mechanical, electrical and instrument engineering).
(c) Other manufacturing industries (food, drink, tobacco, leather, footwear and clothing, textiles, timber and wooden furniture, paper and paper products, printing and publishing, rubber and plastics, etc.).
(d) Building and civil engineering.
(e) Distributive trades, hotels, catering and repairs.
(f) Transport and communication.
(g) Banking and finance, insurance, business services and renting.
(h) Other services (public administration, education, research and development, medical and other health services, government recreational services and personal services).

In terms of estimation the reference group for this set of dummies is the distributive trades industry group.

(x) Eight (0,1) occupational dummies which are:
(a) Professional.
(b) Self-employed: employs others and managers.
(c) Salaried employees, i.e., insurance and financial agents, auctioneers and valuers, ships' officers and pilots, etc.
(d) Intermediate non-manual.
(e) Other non-manual.
(f) Skilled manual workers.
(g) Semi-skilled manual workers.
(h) Unskilled.

In terms of estimation the reference group for this set of dummies is the unskilled category.

(xi) Five regional dummies which are:
(a) Dublin County.
(b) North-West (Counties Sligo, Mayo, Roscommon, Donegal, Leitrim, Monaghan, Cavan and Galway).
(c) Southern (Counties Cork, Waterford, Limerick, Kerry, Tipperary and Clare).
(d) Midlands (Counties Carlow, Laois, Kilkenny, Kildare, Westmeath and Offaly).
(e) Rest of Leinster (Counties Wicklow, Wexford and Louth).

In terms of estimation the reference group for this set of dummies is the North-West region.

Table A.1 provides information on the means of the continuous variables and the proportions of individuals in the relevant binary variable categories, etc., for males and females.

Table A.1: *Mean Values of Variables for the Individuals in the Sample*

Variable	*Male*	*Female*
Sex	0.4393	0.5607
Net hourly wage (ln)	0.5866	0.5089
Experience in years	4.2820	3.2486
Educational qualifications		
Intermediate Certificate	0.3140	0.2112
Leaving Certificate	0.2683	0.6562
University degree	0.0223	0.0087
Other qualifications		
Group Certificate	0.1938	0.0192
Apprenticeship	0.3720	0.0558
Basic training qualification	0.4165	0.6771
Job characteristics		
Firm $<$ 15	0.3318	0.3192
15 $\leqslant$ Firm $<$ 50	0.1626	0.1571
50 $\leqslant$ Firm $<$ 100	0.0980	0.0890
Firm $\geqslant$ 100	0.4076	0.5096
Promotion on job	0.4632	0.4014
Occupations		
Professional	0.0512	0.0593
Self-employed	0.0312	0.0174
Salaried employees	0.0200	0.0052
Intermediate non-manual	0.1559	0.7103
Other non-manual	0.0757	0.0942
Skilled manual	0.5501	0.0349
Semi-skilled manual	0.0779	0.0750
Unskilled	0.0380	0.0037
Industries		
Building & engineering	0.1559	0.0140
Transport & communication	0.0735	0.0454
Banking & insurance	0.0557	0.2024
Public admin. etc.	0.0891	0.2967
Metal manufacturing	0.1336	0.0820
Other manufacturing	0.1693	0.1344
Extraction & chemicals	0.0668	0.0314
Distributive trades	0.2561	0.1937
Regions		
Dublin county	0.3541	0.4537
North-West counties	0.1462	0.1153
Southern counties	0.2695	0.2792
Midlands counties	0.1470	0.0977
Leinster (excl. Dublin)	0.0846	0.0541
Urban	0.6147	0.6911

In terms of dummy variables the values reported represent the sample proportions in the relevant categories.

APPENDIX 2

The calculation of the asymptotic standard errors associated with the explained and unexplained differentials follows Stewart (1987). If the mean unexplained wage differential is expressed as:

$$\hat{\bar{\lambda}} = \bar{X}^{f}(\hat{\beta}_m - \hat{\beta}_f),$$

then its variance may be expressed as:

$$\mathrm{var}(\hat{\bar{\lambda}}) = \bar{X}^{f'} V \bar{X}^{f}$$

where $V = \mathrm{var}(\hat{\beta}_m - \hat{\beta}_f)$,

$= \mathrm{var}(\hat{\beta}_m + \hat{\beta}_f)$.

Since $\hat{\beta}_m$ and $\hat{\beta}_f$ are estimated from separate non-overlapping samples no covariance exists between them. The asymptotic standard error is thus given by $\sqrt{\mathrm{var}(\hat{\bar{\lambda}})}$. The consequent test statistic is normally distributed.

If the mean explained wage differential is expressed as:

$$\hat{\bar{\Psi}} = (\bar{X}^{m} - \bar{X}^{f})\hat{\beta}_m,$$

its variance may be expressed as

$$\mathrm{var}(\hat{\bar{\Psi}}) = (\bar{X}^{m} - \bar{X}^{f})\bar{V}(\bar{X}^{m} - \bar{X}^{f})$$

where $\bar{V} = \mathrm{var}(\hat{\beta}_m)$.

The asymptotic standard error is $\sqrt{\mathrm{var}\hat{\bar{\Psi}}}$. The consequent test statistic is again normally distributed.