

MULTIPLE REGRESSION ANALYSIS OF THE OCCUPATIONAL STATUS OF TWINS: A COMPARISON OF ECONOMIC AND BEHAVIOURAL GENETICS MODELS

Paul Miller, Charles Mulvey and Nick Martin

I INTRODUCTION

Research into the determinants of economic well-being using data on twins has addressed two major issues. The early research focused on attributing the variance in measures of earnings to heritability and shared environmental influences (see, for example, Behrman and Taubman, 1976). More recently, the focus has shifted to assessing the extent of bias in conventional rates of return to schooling associated with the failure to control adequately for genetics and common environment (see, for example, Ashenfelter and Krueger, 1994). Within this second line of research, two models have been used. The first is a fixed effects model wherein the difference in the earnings of members of a set of twins is related, in a regression framework, to differences in characteristics of the same individuals. Estimation of separate equations for monozygotic (MZ) and dizygotic (DZ) twins is equivalent to holding constant, in the first instance, genetic endowments and common environment, and, in the second instance, common environment influences only (see Behrman and Taubman, 1976). This amounts to an implicit control for these factors. Alternatively, a structural model that explicitly accounts for family effects (genetic endowments and common environment) through the inclusion in the estimating equation of information on a respondent's co-twin has also been applied (see Ashenfelter and Krueger, 1994). In Ashenfelter and Krueger (1994), information on the respondent's co-twin's educational attainment is used to capture family effects.

The selection effects model outlined in Ashenfelter and Krueger and familiar to economists is similar in design to the regression model proposed by DeFries and Fulker (1985) in the genetics literature. The latter model was devised to provide tests of the relative importance of genetic and shared environmental factors within a multiple regression framework, and has a more scientific foundation than the selection effects model. It can be readily augmented with various independent variables to

assess the effect of education while simultaneously providing estimates of heritability and of the influence of shared environmental factors.

The purpose of this note is to compare the performance of the model of DeFries and Fulker (hereafter 'the DFF model') with the conventional fixed effects models in the study of economic well-being. Economic well-being is measured here by the occupational rank of the individual's job, and the ranking in that regard is obtained using average occupational earnings. This study is motivated by two factors. First, there is a methodological advantage in estimating the impact of schooling on economic well-being within the more general model specified by DeFries and Fulker. Second, the utilization of an additional model permits us to assess the robustness of the estimates of the impact of schooling to choice of estimating equation. In this regard the results presented in Ashenfelter and Krueger for the fixed effects and selection effects models in their study of a sample of US twins differ appreciably, but the use of these models in a study of a sample of Australian twins by Miller, Mulvey and Martin (1995) yielded results that are quite robust to choice of model.

The structure of this paper is as follows. In Section II a brief outline is provided of the regression model developed by DeFries and Fulker and of related models. Estimates of the model parameters are reported in Section III using a large sample of Australian MZ and DZ twins. The same data are used to examine the determination of occupational status using the fixed effects models. Section IV concludes.

II. MULTIPLE REGRESSION MODELS OF TWINS' OCCUPATIONAL STATUS

To begin with, we specify the estimating equation in the following general form:

$$Y_j = \alpha_0 + \alpha_1 A_j + \alpha_2 E_j + \alpha_3 S_j + \alpha_4 G_j + \alpha_5 AG_j + \alpha_6 M_j + v_j \quad (1)$$

where Y_j denotes an index of the occupational status of individual j , A_j denotes the ability of the j th individual, E_j denotes the home environmental factors that influence the occupational rank of individual j , S_j , G_j , AG_j and M_j are, respectively, the educational attainment, gender, age and marital status of individual j , and v_j is a stochastic error term.

Both ability and shared environmental influences are generally unobserved and hence the estimated impact of education, α_3 , may be biased. The challenge for empirical researchers is to find ways of controlling for the influence on occupational status of these omitted variables. The approach adopted here is to use information from a sample of twins. This amounts to the estimation of the ability and shared environment biases in the impact of schooling using indirect means. As such, the estimates derived may be sensitive to the model used. It is this sensitivity that we analyse here.

When the data used in the analysis are supplied by twins, equation (1) needs to be written in the following form:

$$Y_{ji} = \alpha_0 + \alpha_1 A_{ji} + \alpha_2 E_{ji} + \alpha_3 S_{ji} + \alpha_4 G_{ji} + \alpha_5 AG_{ji} + \alpha_6 M_{ji} + v_{ji} \quad (2)$$

where subscript j now refers to the family and subscript i refers to the twin member ($i = 1, 2$).

Consider identical twins (reared together). The fixed effects model to explain the difference in average occupational earnings between the members of the twin pair ($Y_{ji} - Y_{j-i}$) can be written as:

$$(Y_{ji} - Y_{j-i}) = \alpha_3(S_{ji} - S_{j-i}) + \alpha_6(M_{ji} - M_{j-i}) + (v_{ji} - v_{j-i}) \quad (3)$$

where the subscript $-i$ refers to the co-twin of respondent i . Note that as identical twins reared together have, by definition, the same ability, family environment, gender and age, these variables disappear from the fixed effects version of the model. In other words, in this version of the model, relating the difference in the average occupational earnings of the twins to the difference in their educational attainments and marital status provides an estimate of the impact of education on earnings (α_3) that is not biased by the omission of the ability and family background variables. Comparison with estimates of α_3 derived from estimation of equation (1) provides an indication of the extent of ability and family background bias in the estimates typically generated.

Estimation of α_3 using equation (3) with data on the earnings of twins has provided a range of values, including 2.7 percent by Behrman, Taubman and Wales (1977) and 9.2 percent by Ashenfelter and Krueger (1994). The first estimate is suggestive of family effects being quite important when estimating the return to schooling while the latter estimate is consistent with a minor influence of the family in this respect.

In summary, the major feature of the fixed effects model is that genetic resemblance and common environment influences are held constant implicitly. This method of estimation will also net out of the estimated impact of schooling the compounding effects of any other fixed effects that affect earnings (e.g., race, possibly some affective characteristics such as motivation). This contrasts with the model of DeFries and Fulker (1985) where explicit controls are provided for genetic and shared environmental factors.

DeFries and Fulker (1985) focus their model on the situation where one twin, termed the proband, has a deviant score on a continuous variable. This could be tests of literacy, or, in the current application, average occupational income. It follows that if the condition that led to the lower occupational status for one twin (the proband) is inherited, the co-twins of DZ probands are expected to have higher occupational standing than co-twins of MZ probands. To test for heritability and shared environmentality, the DFF model suggests that the occupational rank of the high-status twin be regressed on the rank of the low-status

twin, a coefficient of genetic relationship, defined to equal 1.0 for MZ twins and 0.5 for DZ twins, and an interaction term between these variables. Thus, the regression model they propose is:

$$Y_{jt} = \beta_0 + \beta_1 Y_{j-t} + \beta_2 R_{jt} + \beta_3 Y_{j-t} R_{jt} + v_{jt} \quad (4)$$

where Y_{jt} is the occupational rank of the high-income member of twin set j , Y_{j-t} is the occupational rank of the low-income member of twin set j , and R_{jt} is the coefficient of genetic relationship for twin set j .

By construction, β_3 is twice the difference between the MZ and DZ regression coefficients, and, given the standard assumptions in the variance decomposition models, provides a direct estimate of heritability, conventionally labelled h^2 . (Behrman and Taubman (1976) list the main assumptions as: the model is additive, mating is random, non-common environment of a DZ twin is not correlated with his/her co-twins genes.) β_1 records twin resemblance that is independent of genetic resemblance as indexed by the interaction term. Hence, it provides a measure of common environmental influences, conventionally labelled c^2 .

While the DFF model was initially developed for the situation where one twin had a deviant score on the variable of interest, it has subsequently been extended to random samples and a number of methods are available to accommodate this extension. The Cherny *et al.* (1992) double entry method is used in this study. This method involves entering each twin's score twice, once as proband and once as co-twin. All 't' statistics are adjusted for the correct degrees of freedom (Cherny *et al.*, 1992)

DeFries and Fulker note (p. 472) that their regression model can be extended to include other independent variables, such as gender, age, ethnicity and environmental indices. The additional variables considered for inclusion in this analysis are the age, gender, marital status and educational attainment variables included in the previous model. Hence, the specification of the DFF model employed in this study will be:

$$Y_{jt} = \beta_0 + \beta_1 Y_{j-t} + \beta_2 R_{jt} + \beta_3 Y_{j-t} R_{jt} + \beta_4 S_{jt} + \beta_5 G_{jt} + \beta_6 AGE_{jt} + \beta_7 M_{jt} + v_{jt} \quad (5)$$

This augmented specification of the estimating equation provides an estimate of the impact of schooling (β_4) that is excised of the influences of genetics and common environmental factors. It is important to note that the estimate of the impact of schooling on occupational attainment obtained through the DFF model may differ from that obtained with the fixed effects model as the common environmental influences (E) control is mediated through the co-twin's indicator variable in their model.

Comparison of the value for β_4 computed on the basis of equation (5) with the results from equation (3) permits discussion of the relative impacts of explicit (DeFries and Fulker) and implicit (fixed effects) controls for family effects.

III. DATA AND ESTIMATES

The Data

The data used in this study are drawn from an exceptionally large and representative sample of twins. The data begin with a mail survey undertaken in 1980–82 of all 5,967 twins aged over 18 years enrolled in the Australian National Health and Medical Research Council Twin Registry at that time. Joining this registry and responding to the survey were both voluntary. Replies were received from 3,808 complete pairs. In 1988–89 this sample was followed up and 2,934 twin pairs responded. Most of the data used in this study are derived from the follow-up survey.

Only sets of twins where each member responded to the questions used in the study are included in the sample. Moreover, to be included in the sample the twins needed to be 20–64 years of age, and employed on a part-time or full-time basis. 1,170 pairs met these requirements.¹

Data on income are an important element in the study of economic well-being and it is necessary to note that the income data were collected in categorical form, with only broad categories being used. This restricts the usefulness of that particular data for analysis of variations in income. However, as noted in Section I, the average earnings of the occupation of employment can be used in place of the categorical information. Similar measures have been used by Griliches (1977), Nickell (1982) and Behrman, Rosenzweig and Taubman (1994). Hence the study provides estimates of the impact of schooling and of other variables on the mean occupational earnings. It ignores the returns to schooling within occupations and therefore will not provide an unbiased estimate of the return to schooling.² Comparison of the estimates obtained with those derived using data on individual earnings (e.g. McNabb and Richardson, 1989) suggests that this bias is probably very small, a finding that is consistent with the research by Groshen (1991).

Descriptive statistics are presented in Miller, Mulvey and Martin (1995). These show that the average educational attainment of MZ twins is 12.51 years and that of DZ twins 12.70 years. These reports are a little more than one year higher than the national average recorded in the 1986 Australian Census of Population and Housing. The sample is one year younger than the 20–64 year old population (average age of 36 years compared to a national average of 37 years) and is more likely to be married than the general population (74.1 percent compared to 67.4 percent). In part the different characteristics of the sample are due to attrition bias that is related to age, education and marital status (see

¹Following Ashenfelter and Krueger (1994), both males and females are included in the sample selected for analysis

²Let the actual earnings of individual i be y_i^* and the mean occupational earnings assigned to individual i be y_i , where $y_i = y_i^* + \eta_i$. Then the true model of individual earnings is $y_i^* = \alpha + \beta x_i^* + v_i$. But here we estimate $y_i - \eta_i = \alpha + \beta x_i^* + v_i$ or $y_i = \alpha + \beta x_i^* + \epsilon_i$, where $\epsilon_i = v_i + \eta_i$. Unless $E(x_i \epsilon_i) = 0$, the estimates of β will be biased. An IV estimator is required in this case

Baker *et al.*, 1995). Baker *et al.*'s (1995) analysis of educational attainments shows that the attrition bias does not lead to increased twin similarity, which is important to a study of the type undertaken here.

The Fixed Effects Model of Earnings

The fixed effects model estimated on the sample of identical twins will, in principle, remove the influence of genetic and shared environmental influences. Table 1 lists relevant results. For comparison purposes estimates are also presented for the sample of non-identical twins.

The OLS estimates for when the sample is treated as one of individuals listed in columns (i) and (iii) of Table 1 indicate that the impact of schooling on occupational rank is around 6.5 percent in Australia. There is a positive and significant relationship between age and average occupational earnings. This relationship is much weaker than the conventional estimate of age on individual earnings, though this is to be expected given the grouped nature of the dependent variable. There is mixed evidence on the relative well-being of the married and non-married, with there being a small positive premium for the married among identical twins and no difference between the occupational status of the two groups in the

TABLE 1
*OLS Estimates of Log Average Occupational Earnings Equations
(Australian Twins Sample)*

	<i>Identical</i>		<i>Non-identical</i>	
	<i>Individuals (i)</i>	<i>Twins (ii)</i>	<i>Individuals (iii)</i>	<i>Twins (iv)</i>
Constant	8.986 (193.49)	0.011 (1.11)	8.993 (192.29)	0.032 (2.67)
Own education	0.064 (26.64)	0.025 (4.92)	0.066 (25.64)	0.045 (9.31)
Age	0.002 (2.54)	(a)	0.002 (2.43)	(a)
Married	0.035 (2.64)	0.037 (1.86)	0.003 (0.22)	-0.016 (0.79)
Male	0.231 (18.47)	(a)	0.214 (15.88)	0.226 (11.35)
Sample size	1204	602	1136	568
R ²	0.51	0.05	0.47	0.34

Note

(a) = variable not relevant

Heteroscedasticity-consistent *t* statistics in parentheses.

case of fraternal twins. Australian studies typically report a positive wage premium for being married among males and no difference between the status of the two groups in the case of females (Miller, 1995). Finally, the results indicate that males have average occupational earnings over 20 percent higher than females, *ceteris paribus*. This result is typical of the Australian studies that use measures of annual income (see, for example, Jones, 1983).

The estimates in columns (ii) and (iv) that use the twins nature of the data show that the bias in the impact of schooling associated with failure to control adequately for family effects may be substantial. In the column (ii) results where both genetics and shared environment are implicitly held constant, the impact of schooling drops to 2.5 percent. In the column (iv) results where only the shared environment influences are held constant the impact of schooling is 4.5 percent. The gender effect is of the same order of magnitude as for when the sample is treated as one of individuals.

We conclude from these estimates that failure to control for family effects when estimating the impact of schooling will result in estimates that are severely biased. The true impact of schooling appears to be 2.5 percent rather than the range 6–7 percent conventionally estimated.

DeFries and Fulker's Model of Earnings

Table 2 contains relevant estimates of the DFF model. The first column lists OLS estimates of equation (4). The interaction term between the proband's average occupational earnings and the genetic relationship variable is statistically significant and the coefficient is 0.72. This provides an estimate of the fraction of the variance in mean occupational earnings due to heritable influences. This estimate is slightly higher than those reported in the literature for individual earnings (e.g. Taubman, 1976). The coefficient on the proband's mean occupational earnings is statistically insignificant. Recall that this provides an estimate of the fraction of the variance in occupational rank attributable to common environment. Hence our estimate is slightly lower than the 8–15 percent reported by Taubman (1976).

Column (ii) augments the model with a number of individual characteristics. The impact of schooling is estimated to be 5.3 percent, some 1 percentage point lower than the conventional estimate. Once education is held constant in the analysis, the direct impact of heritability on earnings drops substantially, from 72 percent to 26 percent. In other words, much of the impact of heritability on earnings works via its effect on educational outcomes. The estimate of shared family environment remains insignificantly different from zero in the column (ii) results.

Does Measurement Error Matter?

One of the econometric problems that arises in the estimation of the effect of schooling is the effect of mismeasurement of the schooling variable. Griliches (1977) showed that taking account of measurement error using an instrumental variables (IV) estimator can result in marked changes to the estimates.³ Where measurement error poses a problem for the estimation of the impact of schooling, it will be more serious in the case of the fixed effects model where the measures of schooling of the respondent and his/her co-twin will both be prone to error. Ashenfelter and Krueger (1994) use respondent's reports on the schooling of their co-twins as instruments in response to this problem of estimation. However,

TABLE 2
OLS and IV Estimates of DeFries and Fulker's Model of Average Occupational Earnings Determination

	<i>OLS</i>		<i>IV</i>	
	(i)	(ii)	(iii)	(iv)
Constant	10 403 (11.94)	8.895 (12.99)	8.813 (12.88)	8.971 (12.81)
Co-twin's (proband's) earnings (Y_{-i})	-0.039 (0.45)	0.026 (0.39)	0.019 (0.28)	0.003 (0.05)
Genetic Relationship (R)	-7 197 (6.98)	-2.617 (3.04)	-2.223 (2.56)	-2.452 (2.78)
$R \times Y_{-i}$	0.719 (6.97)	0.263 (3.06)	0.223 (2.58)	0.246 (2.80)
Own education	(a)	0.053 (19.46)	0.064 (18.54)	0.064 (18.06)
Age	(a)	0.002 (3.15)	0.002 (2.80)	0.002 (2.50)
Married	(a)	0.018 (1.27)	0.020 (1.41)	0.023 (1.59)
Male	(a)	0.186 (13.97)	0.184 (13.79)	0.187 (13.58)
Sample size	1170	1170	1170	1113
R^2	0.29	0.53		

Note

(i) = DFF model

(ii) = DFF augmented model.

(iii) = IV estimates of DFF augmented model, full sample

(iv) = IV estimates of DFF augmented model, restricted sample.

³The IV method of estimation produces consistent estimates. This consistency property, however, is obtained at the cost of a larger variance than might be obtained from alternative econometric techniques.

instrumenting the difference between the twins' schooling by the difference between their cross-reported measures of schooling will result in consistent estimates only if the own-report and the co-twin report do not have a common measurement error. Ashenfelter and Krueger (1994) outline an IV estimator that is consistent in the presence of correlated measurement error. Thus, they express the earnings difference between twins as a function of the difference between the respondent's own level of education and his/her report of their co-twin's level of education, and instrument this using the difference between the co-twin's report of the first twin's level of education and the co-twin's report on his/her own level of education. Both Ashenfelter and Krueger (1994) and Miller, Mulvey and Martin (1995) report findings indicative of correlated measurement errors being important. Accordingly, Ashenfelter and Krueger's estimator which is consistent in the presence of measurement error is employed below in addition to the conventional IV estimator.

In contrast to the fixed effects model, the DFF model includes only one schooling variable that may be subject to error. Consequently, the use of the co-twin's report on the respondent's level of schooling as an instrument in this instance will give consistent estimates. Moreover, in the case of the model of DeFries and Fulker, family background variables (father's educational attainment, mother's educational attainment, father's occupational status) can be used as alternative/additional instruments.

Relevant IV results are presented in Table 3 for the fixed effects model and in columns (iii) and (iv) of Table 2 for the DFF model. The first two columns of Table 3 list results for when the samples of MZ and DZ twins are treated as samples of individuals and the earnings equation estimated using IV with the co-twin's report on the respondent's schooling being used as an instrument for the schooling variable. According to these results, the impact of education in Australia is between 7.3 and 7.9 percent.

The third column of Table 3 presents results from estimation of the fixed effects model on the sample of MZ twins using an IV estimator that is consistent in the presence of correlation between the errors of measurement in the own-reports and cross-reports of schooling. The coefficient on the education variable (at 0.045) is two percentage points higher in this set of results than in the Table 1 results. Column (iv) lists results for the fixed effects model using the conventional IV estimator. In this instance the estimated impact of education is 0.083.⁴ Similarly, the coefficients on the education variable in columns (v) and (vi) for the sample of DZ twins are also several percentage points higher than that reported in Table 1. Thus, the main conclusion to be drawn here is that the true impact of education in Australia is between 5 and 8 percent, which is very close to the conventional OLS estimate. Thus, the results

⁴See Miller, Mulvey and Martin (1995) for detailed discussion of these results

TABLE 3
IV Estimates of Models of Average Occupational Earnings Determination

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Constant	8.850 (171.79)	8.809 (162.85)	0.004 (0.39)	0.008 (0.79)	0.008 (0.64)	0.026 (2.08)
Own education	0.073 (25.84)	0.079 (24.31)	0.045 (4.87)	0.083 (4.18)	0.074 (9.36)	0.078 (8.93)
Age	0.003 (3.17)	0.003 (3.21)	(a)	(a)	(a)	(a)
Married	0.038 (2.85)	0.005 (0.29)	0.034 (1.73)	0.024 (1.04)	-0.022 (1.06)	-0.024 (1.13)
Male	0.222 (17.42)	0.204 (14.91)	(a)	(a)	0.229 (11.69)	0.203 (9.82)
Sample size	1204	1136	602	602	568	568

Note

- (i) MZ twins treated as individuals
- (ii) DZ twins treated as individuals
- (iii) MZ twins: fixed effects model, correlated measurement errors
- (iv) MZ twins: fixed effects model, conventional IV method
- (v) DZ twins: fixed effects model, correlated measurement errors
- (vi) DZ twins: fixed effects model, conventional IV method
- (a) Variable not relevant.

t statistics in parentheses are heteroscedasticity consistent

suggest only a small role for ability and family effects. The finding in respect to family background is consistent with findings derived when direct measures of family background (father's educational attainment, mother's educational attainment and father's occupational status) are included in the estimating equation (see Behrman, Taubman and Wales, 1977, Table 4).

Column (iii) of Table 2 lists estimates of the DFF model obtained using the co-twins' report on the respondent's level of education as an instrument for the own education variable. Here the impact of education is indicated as being 6.4 percent. The remaining estimates are not affected by the change in method of estimation. In column (iv) of Table 2 we use a linear combination of the co-twin's report on the respondent's educational attainment, mother's level of schooling, father's level of schooling, father's occupational status and number of siblings as an instrument.⁵ Again, the impact of education is indicated to be 6.4 percent. This result is not materially different from that derived using conventional measures and supports the conclusion of Miller, Mulvey and

⁵The sample is smaller than in other estimations owing to the deletion of individuals who failed to report information on their parents

Martin (1995) and Ashenfelter and Krueger (1994) that the usual OLS estimates are not biased upwards by the omission of family effects.

IV. CONCLUSION

Conventional estimates of the impact of schooling in Australia centre on the 7 percent mark. These estimates may be biased, however, owing to the omission of controls for genetic factors and environmental influences. Estimation of a fixed effects model, in which the genetic and environmental influences are held constant implicitly, gives an estimate of the impact of schooling of 2.5 percent. This implies a family effects bias in the conventional estimate of four to five percentage points. In comparison, the impact of education obtained through application of DeFries and Fulker's model is 5.3 percent. Whereas this estimate is one to two percentage points lower than conventional estimates of the effect of education, it is almost three percentage points higher than the estimates obtained through application of the fixed effects model. This pattern could be explained through measurement error being an important source of bias in the estimates of the effect of education: in this situation, the downward bias in the fixed effects estimator, where two schooling variables are measured with error, will be greater than in the model of DeFries and Fulker, where there is only one schooling variable that is prone to measurement error.

Re-estimation of the fixed effects model and the DFF model using IV estimators results in sets of findings where the impact of education is several percentage points higher. Thus, within this framework, the fixed effects model yields an impact of education of between 5 and 8 percent, and the DFF model an impact of 6.4 percent whereas the return computed in more conventional methods of analysis is in the range of 7–8 percent. Shared-environmental influences apparently have a minor role. This evidence from the study of twins is in agreement with the findings from studies that include measures of family background in the earnings equation.

The weight of the evidence from estimation of the two models considered in this paper is that the impact of education in Australia, holding constant genetic and shared environmental factors, is of the same order of magnitude as that estimated in studies that do not take account of these factors. The minor role established for shared environmental influences on average occupational earnings has important implications for the impact that many social reforms might be expected to have on economic inequality. The similarity of the results computed for the different models employed is reassuring, and suggests that reliable controls for the omitted genetic and shared environment variables are obtained through these indirect methods.

University of Western Australia
Queensland Institute of Medical Research

Date of Receipt of Final Manuscript April 1995

REFERENCES

- Ashenfelter, O. and Krueger, A. (1994). 'Estimates of the Economic Returns to Schooling from a New Sample of Twins', *American Economic Review*, Vol 84, pp 1157-73.
- Baker, L. A., Treloar, S. A., Reynolds, C., Heath, A. and Martin, N. G. (1995) 'Genetics of Educational Attainment in Australian Twins: Sex Differences and Secular Changes', *Behavior Genetics* (forthcoming)
- Behrman, J. and Taubman, P. (1976). 'Intergenerational Transmission of Income and Wealth', *American Economic Review*, Vol 66, pp 436-40.
- Behrman, J., Taubman, P. and Wales, T. (1977) 'Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success', in Taubman, P. (ed), *Kinometrics: The Determinants of Socioeconomic Success Within and Between Families*, North Holland, Amsterdam
- Behrman, J., Rosenzweig, R. and Taubman, P. (1994) 'Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment', *Journal of Political Economy*, Vol 102, pp 1131-74
- Cherny, S. S., Cardon, L. R., Fulker, D. W. and DeFries, J. C. (1992) 'Differential Heritability Across Levels of Cognitive Ability', *Behavior Genetics*, Vol. 22, pp 153-62.
- DeFries, J. C. and Fulker, D. W. (1985) 'Multiple Regression Analysis of Twin Data', *Behavior Genetics*, Vol. 15, pp. 467-73
- Griliches, Z. (1977) 'Estimating the Returns to Schooling: Some Econometric Problems', *Econometrica*, Vol 45, pp 1-22
- Groshen, E. L. (1991) 'Sources of Intra-Industry Wage Dispersion. How Much Do Employers Matter?', *The Quarterly Journal of Economics*, Vol 106, pp. 869-84
- Jones, F. L. (1983). 'Sources of Gender Inequality in Income: What the Australian Census Says', *Social Forces*, Vol 62, pp 134-53
- McNabb, R. and Richardson, S. (1989) 'Earnings, Education and Experience: Is Australia Different?', *Australian Economic Papers*, Vol 28, pp. 57-75
- Miller, P. W., Mulvey, C. and Martin, N. (1995). 'What Do Twins Studies Tell Us About the Economic Returns to Education?: A Comparison of Australian and US Findings', *American Economic Review*, Vol. 85, pp. 586-99
- Miller, P. W. (1994). 'Effects on Earnings of the Removal of Direct Discrimination in Minimum Wage Rates: A Validation of the Blinder Decomposition', *Labour Economics*, Vol 1, pp. 347-63.
- Nickell, S. (1982) 'The Determinants of Occupational Success in Britain', *Review of Economic Studies*, Vol. 49, pp. 43-53
- Taubman, P. (1976). 'The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins', *American Economic Review*, Vol. 66, pp. 858-70.

DATA APPENDIX

Details on the construction of a number of key variables are set out below.

Educational Attainment: All the education variables were coded in the survey to a seven-point scale: <7 years of schooling; 8–10 years of schooling; 11–12 years of schooling; Apprenticeship, Diploma, Certificate; Technical or Teachers' College; University — first degree; University — postgraduate degree. These categories have been recoded as 5, 9, 11.5, 11.5, 13, 15 and 17 years of education respectively. The general thrust of the paper's conclusions is not sensitive to reasonable variations in the assumed mean levels of education for each category.

Marital Status: Individuals reporting that they are married, are living in a *de-facto* relationship (common law marriage) or have remarried are classified as married.

Average Occupational Earnings: The average income of full-time workers in each two-digit occupational group was computed from the 1986 Census of Population and Housing. A distinction was made between males and females in deriving these averages. This approach follows Griliches (1977), Nickell (1982) and Behrman, Rosenzweig and Taubman (1994).