What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings
Author(s): Paul Miller, Charles Mulvey, Nick Martin
Source: *The American Economic Review*, Vol. 85, No. 3 (Jun., 1995), pp. 586-599
Published by: American Economic Association
Stable URL: http://www.jstor.org/stable/2118189
Accessed: 20/08/2010 03:39

# What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings

*By* PAUL MILLER, CHARLES MULVEY, AND NICK MARTIN*

Conventional estimates of rates of return to education are constrained because they are unable to isolate the returns to schooling from the contribution of individual ability and the influence of family background. Potentially, studies using a sample of twins may overcome these problems. Monozygotic twins are genetically identical and, if they have been reared together, share the same family background. Differences in income between identical twins can therefore be associated with differences in the amount of education they have undertaken, in order to estimate the independent influence of education. Comparisons of such estimates with estimates made from a sample of dizygotic twins, who are genetically similar but not identical, and who are also reared together, may permit evaluation of the biases associated with ability and family background in conventional estimates of the return to education.

In an important paper, Jere Behrman et al. (1977) analyzed data from a sample of male World War II veteran twins and found that, of the overall rate of return to schooling of 8 percent, 2.7 percentage points, could be attributed to schooling per se, 3.2 percentage points could be attributed to genetic factors, and 2.1 percentage points could be attributed to shared family environment. In a recent paper, however, Orley Ashenfelter and Alan Krueger (1994) have estimated the economic returns to education using data on a new sample of twins which permitted them to adjust their estimate for omitted ability variables and measurement error. On this basis, they find that family and genetic effects make virtually no contribution to the returns to schooling. Measurement error, however, biases the estimated return to schooling downward.

The findings of Ashenfelter and Krueger (1994) are very different from previous work in this field, and there is now some confusion as to what analysis of twins data really reveals about the economic returns to schooling. Ashenfelter and Krueger's results are based on analysis of data for a sample of 298 individuals (149 pairs of identical twins) who attended a twins festival in 1991. They remark at the end of their paper "Only additional data collection is likely to lead to better estimates of the returns to schooling" (p. 1171).

## I. The Australian Twin Register Data

In this analysis we utilize data from the Australian Twin Register which was gathered in two surveys, in 1980–1982 and 1988–1989, and which contains an exceptionally large sample of twins: 3,808 twin pairs in all, for which a complete set of data is available for 1,170 pairs.[1] This is a much larger and more representative sample than has been used in this type of analysis be-

*Miller: Department of Economics, University of Western Australia, Nedlands, Western Australia 6009; Mulvey: Department of Organisational and Labour Studies, University of Western Australia, Nedlands, Western Australia 6009; Martin: Queensland Institute of Medical Research, Brisbane, Queensland, Australia 4029. The analysis has been supported by a grant from the Australian Research Council, and the data collection was supported by a grant from the National Health and Medical Research Council.

[1]A description of the Australian Twin Register data is given in Appendix A. This Appendix also explains the relationship between the 1,170 pairs used in the analysis and the initial 3,808 pairs.

fore. In contrast to the data set utilized by Ashenfelter and Krueger (1994), this data set contains a large subsample of fraternal twins, which permits us, subject to some qualifications discussed below, to identify the separate contributions of ability and shared family environment to the bias due to family effects in conventional estimates of the relationship between education and income. Moreover, as in the data set employed by Ashenfelter and Krueger (1994), the presence in the Australian Twin Survey of multiple measures of schooling (reported by the twin and the co-twin) for each respondent permits the use of instrumental-variable techniques that minimize the adverse consequences of measurement error. While these features mean that the data set is exceptionally rich for analysis of the economic returns to education, there is one limitation which needs to be noted. The data on income gathered in the Australian Twin Register survey are categorical data which were unsuitable for a study of this kind. In place of these data we were able to generate an alternative income series following Zvi Griliches (1977) and Behrman et al. (1994) by attributing to each individual the mean income of the two-digit occupation in which that individual was employed. As noted by Griliches (1977), this approach ignores the returns to schooling within occupations. As Erica Groshen (1991) has shown that education must operate through job classification or establishment in order to affect wages, we view this as a minor limitation and, on Griliches's (1977) evidence, unlikely to affect the interpretation of our findings.

## II. The Analytical Framework

The income of an individual ($y_i$) may be defined as depending on the level of education undertaken by that individual ($S_i$) together with individual ability ($A_i$) and the influence of family background ($F_i$):

$$(1)\quad y_i = \beta_0 + \beta_1 S_i + \beta_2 A_i + \beta_3 F_i \ldots + \varepsilon_i.$$

Since serviceable data on $A$ and $F$ are generally not available, estimates of an equation such as this are seldom made. Accordingly, we look to data drawn from samples of twins in order to proceed. There are two general approaches that may be taken: the "within-twins" or fixed-effects estimator typically used and the selection-effect model outlined by Ashenfelter and Krueger (1994). A brief outline of each approach is provided below.[2]

Equation (1) may be rewritten as follows, where the subscript $i$ refers to the twin set and the subscript $j$ refers to the member of the twin set ($j = 1, 2$):

$$(2)\quad y_{ij} = \beta_0 + \beta_1 S_{ij} + \beta_2 A_{ij} + \beta_3 F_{ij} \ldots + \varepsilon_{ij}.$$

Consider identical (monozygotic) twins reared together. The model to explain the difference in income between the members of the twin pair ($y_{i1} - y_{i2}$) may be written as

$$(3)\quad (y_{i1} - y_{i2}) = \beta_1 (S_{i1} - S_{i2}) \ldots + (\varepsilon_{i1} - \varepsilon_{i2}).$$

Since identical twins reared together have, by definition, the same innate ability and family background the terms in $A$ and $F$ disappear from the fixed-effects version of the equation. In other words, in this model, relating the difference in the incomes of the twins to the difference in their educational attainments provides an estimate of the impact of education on income ($\beta_1$) which is not biased by the omission of the ability and family-background variables. The degree of bias generated by omission of these variables may be observed by comparing estimates of $\beta_1$ in equation (3) with estimates of $\beta_1$ in equation (1).

Consider now fraternal (dizygotic) twins reared together. Fraternal twins are not identical in their genetic inheritance but, if they are reared together, share the same family background.[3] The model to explain

[2] A related model is that advanced by J. C. De Fries and D. W. Fulker (1985). Estimates from this model are presented in Miller et al. (1994).

[3] Fraternal twins are no more alike genetically than nontwin siblings. Moreover, they may differ in respect to other relevant characteristics, such as gender.

differences in their incomes can be written as

$$(4)\quad (y_{i1}-y_{i2})=\beta_1(S_{i1}-S_{i2}) + \beta_2(A_{i1}-A_{i2})\dots + (\varepsilon_{i1}-\varepsilon_{i2}).$$

The term in $F$ disappears from the fixed-effects equation, but since ability may differ between fraternal twins, the term $\beta_2(A_{i1} - A_{i2})$ remains. Data on ability were not collected in the Australian Twin Register survey. Accordingly, we cannot include the term $\beta_2(A_{i1} - A_{i2})$ in the fixed-effects version of the model. This means that relating the difference in the incomes of the twins to differences in their educational levels provides an estimate of the effect on income of education ($\beta_1$) which will be biased by the omission of individual ability but not biased by the omission of family background.[4]

We are therefore in a position to make three estimates of the effect of education on income:

(i) an estimate that is biased by the omission of measures of individual ability and family background [$\beta_1$ in equation (1)];
(ii) an estimate that is biased by the omission of a measure of ability [$\beta_1$ in equation (4)];
(iii) an estimate that is not biased by the omission of either a measure of ability or family background [$\beta_1$ in equation (3)].

Estimates of $\beta_1$ from equations (1) and (3) can be compared directly with estimates made by Ashenfelter and Krueger (1994), but they did not make an estimate of $\beta_1$ in equation (4) (presumably due to the very small subsample of nonidentical twins in their overall sample). In this regard, only comparisons with the earlier study by Behrman et al. (1977) can be made.

An alternative to the fixed-effects model is the selection-effects model which provides explicit consideration of the family effects in the earnings equation. Hence, in this model, the earnings of twin $i$ who is a member of family $j$ ($y_{ij}$) depend on variables that vary across families but not between twins (in this instance age), on individual-specific variables (education), and on unmeasured family effects ($\mu_j$). The unmeasured family effects are modeled as depending on the educational attainments of each twin member and on the age of the twins. Hence, the model is given as

$$(5)\quad y_{ij}=\alpha\,\mathrm{AGE}_j+\beta\,\mathrm{EDUC}_{ij}+\mu_j+\varepsilon_{ij} \qquad i=1,2 \quad j=1,n$$

$$(6)\quad \mu_j=\gamma\,\mathrm{EDUC}_{1j}+\gamma\,\mathrm{EDUC}_{2j} + \delta\,\mathrm{AGE}_j+\omega_j \qquad j=1,n.$$

Substitution for the $\mu_j$ term in the earnings equation results in the reduced form:

$$(7)\quad y_{ij}=(\alpha+\delta)\mathrm{AGE}_j+(\beta+\gamma)\mathrm{EDUC}_{ij} + \gamma\,\mathrm{EDUC}_{-ij}+\varepsilon_{ij} \qquad i=1,2 \quad j=1,n$$

where $-i$ indicates the co-twin of respondent $i$. In this equation the coefficient on the co-twin's educational attainment ($\gamma$) provides an estimate of the impact of family effects, which can be subtracted from the coefficient on the own-education variable ($\beta+\gamma$) to derive an estimate of the pure return to schooling.

## III. The Data and Empirical Results

The analysis of the data from the Australian Twin Survey proceeds as follows. First, we present basic descriptive statistics from the survey. Second, an analysis of data pooled across male and female respondents is undertaken. This analysis will permit comparison with Ashenfelter and Krueger (1994).

[4]For a discussion of the validity of this approach, see Griliches (1979) and Paul Taubman (1976).

TABLE 1—MEANS AND STANDARD DEVIATIONS OF SELECTED VARIABLES: AUSTRALIAN TWIN SURVEY

| Variable | Total sample | Identical twins | Nonidentical twins |
|---|---|---|---|
| Own education (years) | 12.602 | 12.514 | 12.695 |
| | (2.46) | (2.50) | (2.42) |
| Report of co-twin's education (years) | 12.421 | 12.372 | 12.472 |
| | (2.42) | (2.48) | (2.36) |
| Twins report same own level of education (proportion) | 0.471 | 0.558 | 0.379 |
| | (0.50) | (0.50) | (0.49) |
| Report by co-twin same as self-report (proportion) | 0.714 | 0.748 | 0.678 |
| | (0.45) | (0.43) | (0.47) |
| Male (proportion) | 0.486 | 0.468 | 0.505 |
| | (0.50) | (0.50) | (0.50) |
| Age (years) | 36.106 | 36.831 | 35.338 |
| | (8.19) | (8.28) | (8.03) |
| Married (proportion) | 0.741 | 0.759 | 0.722 |
| | (0.44) | (0.43) | (0.45) |
| Log of annual income | 10.001 | 9.996 | 10.007 |
| | (0.30) | (0.30) | (0.30) |
| Sample size | 2,340 | 1,204 | 1,136 |

*Note:* Figures in parentheses are standard deviations.

Table 1 lists descriptive statistics for the main variables used in the analysis. Details on the construction of some of the key variables are to be found in Appendix B. These and all other statistics in this study are computed for the subset of the sample aged 20–64 who report data on each of the variables used in the analysis. The major feature of the table is that there are only minor differences between the samples of identical and nonidentical twins. For example, the mean level of education of identical twins is 12.51 years, and that of nonidentical twins is 12.70 years.[5] In each instance the report of the respondent's level of education by his or her co-twin is slightly lower than the self-reported measure of educational attainment. This is also a feature of the sample used by Ashenfelter and Krueger (1994). There is a slightly greater representation of females among the identical-twins sample than among the fraternal-twins sample, and identical twins are, on average 1.5 years older. Both of these characteristics are shared by the sample analyzed by Ashenfelter and Krueger (1994). The annual earnings of the two types of twins are quite similar.

Table 2 presents correlation coefficients among the main variables of interest. There are two items to look for in this table, the first being the actual correlation between variables and the second being the system-

[5]The average educational attainment of the twins recorded in the survey is a little more than one year higher than the national average recorded in the 1986 Australian Census of Population and Housing. On average the twins are one year younger than the population of 20–64 year olds and are more likely to be married than the total population (74.1 percent compared to 67.4 percent). In part, the different characteristics of the sample are due to attrition bias that is age-, education-, and marital-status-related (see Laura A. Baker et al., 1995). Baker et al's analysis of educational attainments shows that the attrition bias does not lead to increased similarity, which is important to the study we undertake here. The samples used by Ashenfelter and Krueger (1994) and Behrman et al. (1977) also have mean educational attainments in excess of the national average.

TABLE 2—CORRELATION COEFFICIENTS BETWEEN SELECTED VARIABLES: AUSTRALIAN TWIN SURVEY

A. *Identical Twins (Sample Size = 602 Twin Pairs):*

| | Variable | | | | | |
|---|---|---|---|---|---|---|
| Variable | EDUC$_1^1$ | EDUC$_2^1$ | EDUC$_2^2$ | EDUC$_1^2$ | Income 1 | Income 2 |
| EDUC$_1^1$ | 1.000 | | | | | |
| EDUC$_2^1$ | 0.776 | 1.000 | | | | |
| EDUC$_2^2$ | 0.703 | 0.884 | 1.000 | | | |
| EDUC$_1^2$ | 0.870 | 0.733 | 0.776 | 1.000 | | |
| Income 1 | 0.593 | 0.518 | 0.557 | 0.597 | 1.000 | |
| Income 2 | 0.508 | 0.582 | 0.601 | 0.522 | 0.680 | 1.000 |

B. *Nonidentical Twins (Sample Size = 568 Twin Pairs):*

| | Variable | | | | | |
|---|---|---|---|---|---|---|
| Variable | EDUC$_1^1$ | EDUC$_2^1$ | EDUC$_2^2$ | EDUC$_1^2$ | Income 1 | Income 2 |
| EDUC$_1^1$ | 1.000 | | | | | |
| EDUC$_2^1$ | 0.527 | 1.000 | | | | |
| EDUC$_2^2$ | 0.408 | 0.819 | 1.000 | | | |
| EDUC$_1^2$ | 0.827 | 0.510 | 0.477 | 1.000 | | |
| Income 1 | 0.603 | 0.363 | 0.317 | 0.571 | 1.000 | |
| Income 2 | 0.331 | 0.541 | 0.571 | 0.322 | 0.321 | 1.000 |

*Notes:* EDUC$_1^1$ = twin 1's report on own educational attainment; EDUC$_2^1$ = twin 1's report on twin 2's educational attainment; EDUC$_2^2$ = twin 2's report on own educational attainment; EDUC$_1^2$ = twin 2's report on twin 1's educational attainment; Income 1 = twin 1's income; Income 2 = twin 2's income.

atic difference between identical and nonidentical twins. The first twin encountered in the sample in any twin pair has been arbitrarily labeled as the first member of the twin set in the construction of this table.[6]

The correlation between the educational attainments of twin pairs is 0.70 in the case of identical twins, but only 0.41 in the case of nonidentical twins. That is, identical twins are more alike in terms of their educational outcomes than are nonidentical twins, a finding which would be expected on the basis of their greater genetic similarity. A more detailed examination of the data behind this correlation reveals that 56 percent of identical twins and 38 percent of nonidentical twins report the same level of education. The figure for identical twins is slightly higher than the 49 percent reported by Ashenfelter and Krueger (1994), and the 38-percent figure for nonidentical twins is lower than the 43 percent reported by Ashenfelter and Krueger on the basis of their small sample (92 individuals) of fraternal twins. The correlation between the self-reported measure of educational attainment and the report by the co-twin[7] is around 0.87 for identical twins and around 0.82 for nonidentical twins.[8] These correlation co-

[6] These are Pearson correlation coefficients. Polychoric correlations, which are often used with categorical data, are higher (see Baker et al., 1995).

[7] The simple correlation coefficient between the self-reported and co-twin-reported measures of educational attainment is the fraction of the variance in the reported measures of schooling that is due to true variation in educational attainment (see Ashenfelter and Krueger, 1994).

[8] In 75 percent of cases for identical twins, the respondent's reported level of education corresponds with the report of the co-twin on the respondent's educational attainment. Among nonidentical twins the comparable figure is 68 percent.

efficients provide a measure of the reliability ratio of the measure of educational attainment. While these coefficients appear high, they are slightly lower than the values reported by Ashenfelter and Krueger (1994). They are, however, more consistent than those reported by Ashenfelter and Krueger, where the reliability ratio differs between identical and fraternal twins depending on which comparison is made (twin 1 is the reference twin for the reported measure of schooling versus twin 2 being the reference twin for the reported measure of schooling).

The correlation between education and income is about the same and is similar for the different types of twins. It is noted that the correlation between level of education and income is considerably higher than that reported in earlier studies. This, however, is due to our use of an average measure of income: individual variations in earnings will reduce the correlation coefficient. As with previous studies, the correlation between the incomes of identical twins is higher than that for nonidentical twins. Thus, in Table 2, the correlation between the incomes is 0.68 for identical twins and only 0.32 for nonidentical twins.

### A. *A Traditional Analysis of the Australian Twins Data*

In this subsection we briefly present a traditional analysis of the twins data. As outlined in the Introduction, by conducting comparable analyses for identical and nonidentical twins, insights into the role of ability and shared family environment in earnings determination may be gained. This will provide a basis for our update of the earlier work of Behrman et al. (1977) using the methodology proposed by Ashenfelter and Krueger (1994).

As mentioned earlier, the Australian twins data are slightly deficient in terms of the information collected on income, and so an alternative income variable was constructed using the average earnings of the occupation in which the individual was employed. This procedure calls for some comment. First, the average income of 60 minor group occupations was obtained from the 1986 Census. Only individuals working full time were used in the construction of the income measure. Accordingly, when the Australian twins data are used, all employed individuals, whether they are employed on a full-time or part-time basis, are included in the sample. Thus, we effect a control for hours of leisure along the lines of Richard S. Eckaus (1973). Second, because the dependent variable is average income, the distribution of income across educational attainments, age, and other characteristics will be compressed compared to a situation where individual incomes were used. Accordingly, the independent effect of these variables will be diluted unless there is a high degree of occupational segregation on the basis of the characteristic (such as appears to be the case with gender). Indeed, the estimated partial effect of age is quite small in most specifications. A linear age variable is, however, included in all models estimated in levels rather than differenced form.

Table 3 lists estimates for identical twins (first two columns) and for fraternal twins (second two columns). Column (i) presents estimates of the selection-effects model considered by Ashenfelter and Krueger (1994). This model is estimated by generalized least squares. The column-(ii) results are for the fixed-effects version of the model.[9] In this instance, for each model, the returns to schooling are estimated to be 0.025, an estimate which provides a measure of the returns to schooling net of ability and shared family environment.[10] It is considerably lower than the 6.5 percent obtained when the sample is treated as one of individuals (see Table 4). This is an almost identical relative effect to that estimated by Behrman

[9]A constant term has been included in the fixed-effects model. Including this has a negligible impact on the results.

[10]The structural estimate of the return to schooling that controls for omitted-variables bias in the column-(i) estimates is the coefficient on the own-education variable ($\beta + \gamma = 4.8$ percent) less that on the co-twin's educational attainment ($\gamma = 2.3$ percent). It thus equals 2.5 percent.

TABLE 3—ESTIMATES OF TWINS MODELS OF LOG ANNUAL EARNINGS: AUSTRALIAN TWIN SURVEY

| | Identical twins | | Fraternal twins | |
|---|---|---|---|---|
| Variable | (i) Selection effects | (ii) Fixed effects | (iii) Selection effects | (iv) Fixed effects |
| Constant | 8.878 | 0.011 | 8.860 | 0.032 |
| | (161.57) | (1.11) | (150.96) | (2.67) |
| Own education | 0.048 | 0.025 | 0.060 | 0.045 |
| | (16.65) | (4.92) | (21.44) | (9.31) |
| Co-twin's education | 0.023 | —[a] | 0.014 | —[a] |
| | (7.90) | | (5.17) | |
| Age | 0.002 | —[a] | 0.002 | —[a] |
| | (2.83) | | (2.80) | |
| Married | 0.040 | 0.037 | 0.005 | −0.016 |
| | (2.83) | (1.86) | (0.37) | (0.79) |
| Male | 0.223 | —[a] | 0.218 | 0.226 |
| | (15.85) | | (17.06) | (11.35) |
| $R^2$: | 0.54 | 0.05 | 0.49 | 0.03 |
| Sample size: | 1,204 | 602 | 1,176 | 568 |

*Note:* Numbers in parentheses are *t* statistics.
[a]Variable not relevant.

TABLE 4—ESTIMATES OF IDENTICAL-TWINS MODEL OF LOG ANNUAL EARNINGS: RESULTS FROM ASHENFELTER AND KRUEGER (1994) AND BEHRMAN ET AL. (1977) AND ESTIMATES FROM AUSTRALIAN TWIN SURVEY

| | Ordinary least squares | | | Fixed effects | | | Selection effects[b] | |
|---|---|---|---|---|---|---|---|---|
| Variable | AK | BTW[a] | MMM | AK | BTW | MMM | AK | MMM |
| Own education | 0.084 | 0.080 | 0.064 | 0.092 | 0.027 | 0.025 | 0.088 | 0.048 |
| | (6.00) | (32.4) | (26.64) | (3.83) | (3.6) | (4.92) | (5.87) | (16.65) |
| Co-twin's education | | | | | | | −0.007 | 0.023 |
| | | | | | | | (0.47) | (7.90) |
| Married | | | 0.035 | | | 0.037 | | 0.040 |
| | | | (2.64) | | | (1.86) | | (2.83) |
| Age | 0.088 | | 0.002 | | | | 0.090 | 0.002 |
| | (4.63) | | (2.54) | | | | (3.91) | (2.83) |
| $(\text{Age})^2$ | −0.087 | | | | | | −0.090 | |
| | (3.78) | | | | | | (3.10) | |
| Male | 0.204 | | 0.231 | | | | 0.206 | 0.223 |
| | (3.24) | | (18.47) | | | | (2.67) | (15.85) |
| White | −4.10 | | | | | | −0.424 | |
| | (3.23) | | | | | | (2.94) | |
| $R^2$: | 0.260 | 0.20 | 0.510 | 0.092 | 0.1 | 0.05 | 0.219 | 0.535 |
| Sample size: | 298 | 3,852 | 1,204 | 149 | 1,019 | 602 | 298 | 1,204 |

*Notes:* AK = Ashenfelter and Krueger (1994); BTW = Behrman et al. (1977); MMM = present study.
[a]Estimates from pooled sample of identical and nonidentical twins.
[b]Estimates not available for Behrman et al. (1977).

et al. (1977): the return to education estimated from the fixed-effects model for identical twins is about 40 percent of the conventional estimate.

Columns (iii) and (iv) of Table 3 list results for nonidentical twins. Again, both models yield similar results with the income returns to schooling estimated at 4.5 percent. The 2.0-percentage-point difference between this estimate and the conventional estimate of 6.5 percent provides an indirect measure of the bias in conventional measures of the return to schooling associated with failure to control for family background. Comparison of the 4.5-percent return with the 2.5 percent estimated in the fixed-effects model for identical twins provides a measure of the impact of omitted-variables bias associated with innate ability.

Hence, the conclusion from the traditional twins model is that, while the return to schooling typically estimated in Australia is around 6–7 percent, around one-third of this is due to shared family environment, one-third to ability, and only one-third to schooling per se.[11]

Table 4 lists estimates derived from the identical-twins sample of the Australian Twin Survey alongside the closest equivalent estimates from Ashenfelter and Krueger (1994) and Behrman et al. (1977). The first two columns list the simple ordinary least-squares (OLS) estimates of the whole sample of identical twins (from table 3 in Ashenfelter and Krueger [1994] and table 4 in Behrman et al. [1977]); the next three columns list the estimates of the fixed-effects model for identical twins (from table 3 of Ashenfelter and Krueger [1994], table 4a of Behrman et al. [1977], and table 3 of this paper); and the final two columns list the estimates of the selection-effects model for identical twins (from table 3 of Ashenfelter and Krueger [1994] and table 3 of this paper).

Inspection of the results listed in Table 4 shows that the estimated return to schooling computed from the whole sample of identical twins treated as individuals, before adjustments are made for selection effects or measurement error, is almost one-third higher in the U.S. samples, at 8.4 percent (Ashenfelter and Krueger, 1994) and 8.0 percent (Behrman et al., 1977), than in the Australian sample, at 6.4 percent. This difference is presumably associated with the more centralized system of wage determination in Australia and the more egalitarian distribution of income which results from it (see also Robert McNabb and Sue Richardson, 1989). Exploitation of the twins nature of the data shows that as the influences of natural ability and shared family environment wash out of the Australian estimate of the return to schooling it falls from 6.4 percent to only 2.5 percent. A similar pattern is evident in the study by Behrman et al. (1977). In the Ashenfelter and Krueger (1994) study, however, as ability and family effects are washed out, the estimated rate of return increases from 8.4 percent to 9.2 percent. Estimates of the selection-effects model also reveal important and unexplained differences between the results from the U.S. sample of Ashenfelter and Krueger and those from the Australian one. While the estimated rate of return made from the Australian sample again falls, as expected, to 2.5 percent as the influence of family effects (omitted-variables bias) is controlled for, in the U.S. sample of Ashenfelter and Krueger (1994) the estimated rate of return remains almost the same at 9.5 percent.

The results from Table 4 made from the Australian sample are broadly in line with the findings of Behrman et al. (1977). However, the 2.5-percent return and the equivalent estimate from Behrman et al. (1977) may not provide a sound basis for policy, as no allowance has been made for the potential effects of mismeasurement of the schooling variable. In his presidental address to the Econometrics Society, Griliches (1977) alerted labor economists to the importance of this issue. His use of an instrumental-variables estimator demon-

[11]Note that substantial assortative mating may attenuate the estimated impact on genetic factors and increase the apparent importance of common environmental factors (see Lindon Eaves et al., 1978).

TABLE 5—IV ESTIMATES OF MODELS OF LOG ANNUAL EARNINGS, IDENTICAL TWINS: AUSTRALIAN TWINS SAMPLE

| | | Twins | | |
|---|---|---|---|---|
| | (i) | (ii) | (iii) | (iv) |
| Variable | Individuals | Selection effects | Fixed effects | Fixed effects |
| Constant | 8.850<br>(171.79) | 8.859<br>(169.13) | 0.008<br>(0.79) | 0.004<br>(0.39) |
| Own education | 0.073<br>(25.84) | 0.078<br>(8.41) | 0.083<br>(4.18) | 0.045<br>(4.87) |
| Co-twin's education | —[a] | −0.005<br>(0.53) | —[a] | —[a] |
| Age | 0.003<br>(3.17) | 0.003<br>(3.11) | —[a] | —[a] |
| Married | 0.038<br>(2.85) | 0.037<br>(2.71) | 0.024<br>(1.04) | 0.034<br>(1.73) |
| Male | 0.222<br>(17.42) | 0.222<br>(17.35) | —[a] | —[a] |
| Sample size: | 1,204 | 1,204 | 602 | 602 |

*Notes:* The IV estimator in column (iv) is robust to correlated measurement errors. Numbers in parentheses are $t$ statistics.
[a]Variable not relevant.

strated that estimates of the return to schooling derived using ordinary least squares may have "seriously under-estimated rather than overestimated it" (Griliches, 1977 p. 16). The availability of suitable instruments in their sample of twins permitted Ashenfelter and Krueger (1994) to adopt alternative estimators to minimize the impact of measurement error. Their instrumental-variables approach is pursued in the following subsection.[12]

### B. *Instrumental-Variables Analysis of the Australian Twins Data*

Two clear lessons are contained in the research of Griliches (1977) and that of Ashenfelter and Krueger (1994). First, taking account of measurement errors using an instrumental-variables (IV) estimator can result in marked changes in the estimates. For example, Griliches's (1977) use of an IV estimator in place of ordinary least squares resulted in a 50-percent increase in the returns to schooling. The instruments used by Griliches were family background factors (e.g., mother's education, father's occupation).[13] In Ashenfelter and Krueger (1994), two measures of each individual's schooling are available, allowing one measure to be used as an instrument for the other. The use of the IV estimator was associated with large increases in the estimated returns to schooling, although it was shown that the results were sensitive to the type of IV model.

Table 5 lists results from the application of IV estimators to data from the sample of

[12]An alternative approach is to use the average of the multiple measures of schooling. Using this method, the estimated coefficient on the education variable for the total sample (of individuals) was 7 percent, and the pure effect of education from the fixed-effects model estimated on the sample of identical twins was 4.2 percent. The estimated impact of education in the fixed-effects model estimated from the sample of fraternal twins was 6 percent.

[13]Following this procedure here, and using mother's educational attainment, father's educational attainment, father's occupational status, and number of siblings as instruments for the respondent's level of education results in the return to schooling rising by about one-third, to 8.1 percent.

identical twins. Column (i) contains the IV results for the conventional model that relates income to education, gender, marital status, and age, treating the sample as one of individuals. In this model the report on twin 1's schooling by twin 2 is used as an instrument for twin 1's self-reported level of educational attainment. The estimated coefficient on the schooling variable increases from 0.064 to 0.073. The coefficients on the age, marital-status, and gender variables change slightly, but the changes recorded are not of any material consequence.

Column (ii) lists results for the selection-effects model, where each twin's self-reported schooling level is instrumented by the cross-reported measure. These results are quite interesting: the coefficient on the self-reported measure of schooling is 0.078 and is statistically significant ($t = 8.41$), while the coefficient on the co-twin's level of schooling is statistically insignificant ($t = 0.53$). The suggestion derived from this equation, therefore, is that family selection effects or equivalently omitted-variables bias comprises only a negligible component of the conventional estimate of the returns to schooling.

Column (iii) lists estimates of the fixed-effects twins model. Here the difference in actual (i.e., self-reported) schooling levels is instrumented by the difference in cross-reported schooling levels. The estimated return to schooling is 0.083, which is of the same order of magnitude as that derived from the standard specification in column (i). Again, the conclusion that would be reached from this equation is that ability and family background contribute relatively minor amounts to the gross effect of education estimated in the standard study.

The final column of Table 5 lists estimates for a model that takes into account the possibility that the errors of measurement in the respondent's self-reported measure are correlated with the report of the co-twin's level of schooling (see Ashenfelter and Krueger, 1994).[14] Application of this type of IV estimator results in a reduction in the estimated impact of schooling on earnings to a 4.5-percent increase in earnings per year of schooling. As with Ashenfelter and Krueger (1994), this alternative IV estimator provides estimates of the returns to education that are smaller than the conventional IV estimates.

Table 6 sets out a comparison of the U.S. and Australian estimates of two IV models. The first two columns present estimates for the selection-effects model. Correcting for measurement error in the self-reported schooling level in the selection-effects model for the Australian sample increases the estimated rate of return from 4.8 percent to 7.8 percent, a 3-percentage-point increase. The estimates in Table 6 indicate a negligible role for family selection effects in that the coefficient on the co-twin's educational attainment is insignificant. The U.S. estimates of the selection-effects model also suggest that correction for measurement error is important, since the rate of return also increases by about 3 percentage points from 8.8 percent to 11.6 percent. The final four columns of Table 6 illustrate the impact of correction for measurement error in the fixed-effects model. The Australian findings indicate a negligible role for ability and family effects, so that the estimated return to education (5–8 percent) is very similar to the conventional OLS estimate. The U.S. estimates reveal that the rate of return rises to 12–17 percent, depending on the form of the IV estimator. Comparison of these estimates with the estimates in Table 4 suggests that measurement error related to self-reported schooling levels imparts considerable downward bias to the estimates in traditional twins studies. Thus a reinterpretation of the role of family effects is required. Prior to offering this, however, we complete

[14]Ashenfelter and Krueger (1994) show that where the two reports by a twin contain a common measurement error, the usual IV estimator of the fixed-effects model will be biased. A consistent estimator is obtained by expressing the earnings difference as a function of the difference between the respondent's own level of education and his or her report on the co-twin's level of education, and instrumenting this using the difference between the co-twin's report on the first twin's level of education and the co-twin's report on his or her own level of education.

TABLE 6—IV ESTIMATES OF IDENTICAL-TWINS MODELS OF LOG ANNUAL EARNINGS: RESULTS FROM ASHENFELTER AND KRUEGER (1994) AND ESTIMATES FROM THE AUSTRALIAN TWIN SURVEY

| | Selection effects | | Fixed effects | | Fixed-effects correlated measurement error | |
|---|---|---|---|---|---|---|
| Variable | AK | MMM | AK | MMM | AK | MMM |
| Own education | 0.116<br>(3.87) | 0.078<br>(8.41) | 0.167<br>(3.88) | 0.083<br>(4.18) | 0.129<br>(4.30) | 0.048<br>(4.87) |
| Co-twin's education | −0.037<br>(1.27) | −0.005<br>(0.53) | | | | |
| Age | 0.088<br>(4.63) | 0.003<br>(3.11) | | | | |
| $(\text{Age})^2/100$ | −0.087<br>(3.62) | | | | | |
| Male | 0.206<br>(3.22) | 0.222<br>(17.35) | | | | |
| White | −0.428<br>(3.34) | | | | | |
| Married | | 0.037<br>(2.71) | | 0.024<br>(1.04) | | 0.034<br>(1.73) |
| Sample size: | 298 | 1,204 | 149 | 602 | 149 | 602 |

*Note:* Numbers in parentheses are $t$ statistics.

TABLE 7—IV ESTIMATES OF MODELS OF LOG ANNUAL EARNINGS, FRATERNAL TWINS: AUSTRALIAN TWIN SURVEY

| | | Twins | | |
|---|---|---|---|---|
| Variable | (i) Individuals | (ii) Selection effects | (iii) Fixed effects | (iv) Fixed effects |
| Constant | 8.809<br>(162.83) | 8.812<br>(147.07) | 0.026<br>(2.08) | 0.008<br>(0.64) |
| Own education | 0.079<br>(24.31) | 0.079<br>(18.05) | 0.078<br>(8.93) | 0.074<br>(9.36) |
| Co-twin's education | —[a] | −0.001<br>(0.09) | —[a] | —[a] |
| Age | 0.003<br>(3.21) | 0.003<br>(3.16) | —[a] | —[a] |
| Married | 0.005<br>(2.29) | 0.004<br>(0.29) | −0.024<br>(1.13) | −0.022<br>(1.06) |
| Male | 0.204<br>(14.91) | 0.204<br>(14.83) | 0.203<br>(9.82) | 0.022<br>(11.69) |
| Sample size: | 1,136 | 1,136 | 568 | 568 |

*Notes:* The IV estimator in column (iv) is robust to correlated measurement errors. Numbers in parentheses are $t$ statistics.

[a]Variable not relevant.

our update of the Behrman et al. (1977) study by presenting an analysis of the sample of nonidentical twins. Table 7 presents relevant results.

Estimates of the standard specification of the earnings equation for nonidentical twins by instrumental variables yields a return on schooling of around 8 percent (cf. the OLS estimate of 6.6 percent). The results from estimation of the selection-effects model using the IV approach are consistent with those obtained from the sample of identical twins in that the coefficient on the self-reported schooling variable is statistically significant and the same as the estimate obtained in column (i), while the coefficient on the co-twin's schooling level is statistically insignificant. The conclusion from this type of model, therefore, is that omitted variables (in this instance family background) do not matter. Estimation of the fixed-effects model using instrumental variables yields an estimate of the return to schooling of 7.8 percent ($t = 8.93$). Consistent with the previous results, the comparison of this coefficient with that presented in column (i) suggests that the omission of family background from the standard earnings equation will not cause the estimated return to schooling to be badly biased, as is sometimes suggested. Finally, results from the IV estimator that is robust to correlated measurement errors are presented in column (iv). Here the estimated return to schooling is 7.4 percent.

Relating these results to the model derived above permits the following conclusions. First, the application of an IV estimator that does not take account of the correlation between the errors of measurement in the schooling variable and its instrument yields results which are consistent with those reported by Ashenfelter and Krueger (1994): they suggest that there is little role for family effects (i.e., genetic factors and family background) in the determination of the returns to schooling. Second, the application of an IV estimator that is robust to correlated measurement errors indicates that genetic factors (but not family background) may have a modest impact on the relationship between schooling and earnings in Australia. Third, regardless of the form of the IV estimator, the within-twins estimates presented in Ashenfelter and Krueger (1994) and in the present study reveal that there is little evidence of upward bias in the typical OLS estimate of the return to education.

## IV. Conclusions

In this paper we have presented an analysis of the economic returns to schooling using a sample of Australian twins and placed this in the context of previous twins studies. In particular, we have compared our findings with those of Ashenfelter and Krueger (1994) and those of Behrman et al. (1977).

The analysis in this paper utilizes a large set of data on Australian twins which is more representative of the population than the data sets used in the studies by Behrman et al. (1977) or Ashenfelter and Krueger (1994). Our OLS estimates are similar to those of Behrman et al. (1977): one-third of the overall return to schooling is due to education per se, one-third to ability, and one-third to shared family environment. However, when we replicate the analysis of Ashenfelter and Krueger (1994) by estimating both fixed-effects and selection-effects models, and after correction of measurement error in self-reported schooling levels, we find a much more modest role for ability and family environment in the relationship between schooling and income. Moreover, as in Ashenfelter and Krueger (1994), after correcting for measurement error in the fixed-effects model for identical twins, our estimate of the return to schooling increases considerably, rising from 2.5 percent to between 5 percent and 8 percent. While these returns are lower than those reported by Ashenfelter and Krueger (1994), this is largely due to the fact that rates of return to schooling in Australia are lower than in the United States, because of differences in the dispersion of the distribution of income. The finding of the current study that the usual OLS estimates are not biased upward by the omission of family effects mirrors the major conclusion of Ashenfelter and Krueger (1994).

## Appendix A: The Australian Twin Register Data

The data used in this study begin with a mail survey undertaken in 1980–1982 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 1988–1989 this sample was followed up, and 2,934 twin pairs responded.

Only sets of twins where each member responded to the questions used in the study are used. Hence, of the 2,943 twin pairs, 170 were eliminated because they were older than 64 years, 351 were eliminated because information was not provided on the co-twin's educational attainment, 580 were eliminated because of missing data on occupational status, and 672 were eliminated because at least one member of the pair was not employed on either a full-time or part-time basis. This left 1,170 complete pairs.

Most of the data used in this analysis are from the second survey, and the discussion in the present paper is focused mainly on this. The survey gathered information on the extent of contact between twin pairs during childhood and during the eight years prior to the survey, extensive information on the respondent's family background (parents, siblings, marital status, and children), socioeconomic status (education, employment status, income, and occupation), personal details (body size, smoking and drinking habit, and general health), personality, and feelings and attitudes.

There are some aspects of the data which require explanation. First, some of the information that is central to any analysis of returns to education was collected in a less than ideal form. The education data were collected in seven categories, though these categories are broadest at the tail and hence capture most of the variation across the education levels that matter most in the schooling decisions in Australia. Each respondent was asked to provide information on the highest education level completed, and each respondent was also asked to provide similar information on the education level completed by his or her twin, mother, father, and spouse. The availability of this information permits estimates of the reliability of the education measures usually included in the human-capital earnings function.

Second, the income data were collected in categorical form, with eight categories being specified on the survey instrument. Unfortunately, in this case the intervals selected are quite broad in the middle of the income distribution (for example, \$15,000–\$25,000, \$25,000–\$35,000, \$35,000–\$50,000). A consequence of this is that over one-third of the employed are indicated as being in the same income category as their twin, a factor which appears to preclude the application of any fixed-effects estimator to the raw income data.

Third, the data on occupation of the usual/regular lifetime occupation are available at the four-digit level and were collected for the respondent, his or her twin, mother, father, and spouse. These data can also be linked to the average income for occupations from the 1986 Census of Population and Housing. Following Griliches (1977) and Behrman et al. (1994), this is done by assigning each individual the average or median income of the occupation in which he or she is employed.

## Appendix B

Details of 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 recorded as 5, 9, 11.5, 11.5, 13, 15, and 17 years of education, respectively. While a number of individual findings (e.g., reliability ratios) are obviously sensitive to the assumptions made here, 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 spousal relationship (common-law marriage), or have remarried are classified as married.

*Income:* 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 is made between males and females in deriving these averages.

## REFERENCES


**Ashenfelter, Orley and Krueger, Alan.** "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review*, December 1994, *84*(5), pp. 1157–73.

**Baker, Laura A.; Treloar, Susan A.; Reynolds, Chandra; Heath, Andrew and Martin, Nick G.** "Genetics of Educational Attainment in Australian Twins: Sex Differences and Secular Changes." *Behaviour Genetics*, 1995 (forthcoming).

**Behrman, Jere; Taubman, P. and Wales, Terence J.** "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success," in Paul Taubman, ed., *Kinometrics: The determinants of socioeconomic success within and between families*. Amsterdam: North-Holland, 1977, pp. 35–96.

**Behrman, Jere; Rosenzweig, Mark R. and Taubman, Paul.** "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy*, December 1994, *102*(6), pp. 1131–74.

**De Fries, J. C. and Fulker, D. W.** "Multiple Regression Analysis of Twins Data." *Behaviour Genetics*, September 1985, *15*(5), pp. 467–73.

**Eaves, Lindon; Last, Krystyna A.; Young, Phillip A. and Martin, Nick G.** "Model-Fitting Approaches to the Analysis of Human Behaviour." *Heredity*, June 1978, *41*(3), pp. 249–320.

**Eckaus, Richard S.** "Estimates of the Returns to Education with Hourly Standardised Incomes." *Quarterly Journal of Economics*, February 1973, *87*(1), pp. 121–31.

**Griliches, Zvi.** "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica*, January 1977, *45*(1), pp. 1–22.

______. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy*, October 1979, *87*(5), pp. S37–S64.

**Groshen, Erica L.** "Sources of Intra-industry Wage Dispersion: How Much Do Employers Matter?" *Quarterly Journal of Economics*, August 1991, *106*(3), pp. 869–84.

**McNabb, Robert and Richardson, Sue.** "Earnings, Education and Experience: Is Australia Different?" *Australian Economic Papers*, June 1989, *28*(52), pp. 57–75.

**Miller, P. W.; Mulvey, C. and Martin N.** "Estimates of the Return to Education Using De Fries and Fulker's Model for Twins Data." Mimeo, University of Western Australia, 1994.

**Taubman, Paul.** "The Determinants of Earnings: Genetics, Family and Other Environments; A Study of White Male Twins." *American Economic Review*, December 1976, *66*(5), pp. 858–70.