Returns to Education: Evidence from U.K. Twins

By Dorothe Bonjour, Lynn F. Cherkas, Jonathan E. Haskel, Denise D. Hawkes, and Tim D. Spector*

This paper attempts to estimate returns to education using a new data set of identical U.K. twins.[1] We administered an initial questionnaire to around 6,600 individuals (3,300 same-sex twin pairs) in June 1999, all of whom are on the St. Thomas' U.K. Adult Twin Registry, based at the Twins Research and Genetic Epidemiology Unit, St. Thomas' Hospital, London, England. As well as the detailed medical information on the questionnaire, which covers age, birthweight, smoking, etc., we asked the twins additional socioeconomic questions on: earnings, occupation, and schooling; test scores; and the schooling of the other twin. This paper reports results on 1,364 identical twins, of whom 428 comprise 214 identical twin pairs with complete wage and schooling information. We also report results from a follow-up survey on test scores and additional schooling information for 67 pairs.

We believe our study is of interest for five main reasons. First, given the interest in genetics and economic success (see, e.g., Richard J. Herrnstein and Charles Murray, 1994), data on genetically identical individuals are of particular value.[2] Second, while there are many earnings/education studies, there are comparatively few based on identical twins.[3] Thus we add to this literature. Third, our study is the first for the United Kingdom to present within-twin-pair[4] estimates using identical twins. David G. Blanchflower and Peter Elias (1999) used a sample of 23 twin pairs from the U.K. National Child Development Study, but there was insufficient variation of education within each twin pair to perform any within-pair regressions. Fourth, we have followed Ashenfelter and Krueger's (1994) innovation of asking one twin to report on the schooling of the other, in order to examine possible measurement error. Fifth, our study has more data on twins than other studies including ability test scores, reading scores, smoking behavior, and schooling details.

The major criticism of within-twin-pair estimates is set out by John Bound and Gary Solon

* Bonjour: Policy Studies Institute, 100 Park Village East, London, England NW1 3SR (e-mail: d.bonjour@psi.org.uk); Cherkas: Twin Research & Genetic Epidemiology Unit, Guys and St. Thomas' Hospital Trust, St. Thomas' Hospital, Lambeth Palace Road, London, England SE1 7EH (e-mail: lynn.cherkas@gsst.sthames.nhs.uk); Haskel: Department of Economics, Queen Mary, University of London, London E1 4NS, England (e-mail: j.e.haskel@qmul.ac.uk; www.qmul.ac.uk/~ugte153); Hawkes: Centre for Longitudinal Studies (CLS), Bedford Group for Lifecourse and Statistical Studies, Institute of Education, University of London, 20 Bedford Way, London, England WC1H 0AL (e-mail: dh@cls.ioe.ac.uk); Spector: Twin Research & Genetic Epidemiology Unit, Guys and St. Thomas' Hospital Trust, St. Thomas' Hospital, Lambeth Palace Road, London, England SE1 7EH (e-mail: t.spector@kcl.ac.uk). We thank the DfEE and the ESRC (Grant No. R00429924188) for financial support, Anna Vignoles for help with data, a co-editor, two anonymous referees, Ian Walker, and seminar participants at Warwick and Bergen for comments, and Orley Ashenfelter and Alan Krueger for providing their questionnaires and encouragement. Any errors are our own. This is a shorter version of a working paper of the same title which is available on the Web site above. The data and programs to replicate the results in this paper are available on the AER Web site (http://www.aeaweb.org/aer/contents/).

[1] Other returns to education methods that attempt to control for ability and background use, for example, IQ tests and detailed family background data sets (e.g., Lorraine Dearden, 1999), or find an instrument, e.g., such as the raising of the school leaving age, proximity to college, or birth quarter, that is correlated with schooling but uncorrelated with earnings (see, e.g., Joshua Angrist and Alan B. Krueger, 1991; Colm Harmon and Ian Walker, 1995; David Card, 1997). See Card (1999) and Richard Blundell et al. (2001) for surveys.

[2] See, e.g., Orley Ashenfelter and David J. Zimmerman (1997) for a study based on brothers and father-son pairs.

[3] We are aware of seven: for the United States: the Twinsburg sample (Ashenfelter and Krueger, 1994; Ashenfelter and Cecilia Rouse, 1998; Rouse, 1999), the NAS study (Paul Taubman, 1976), and the Minnesota studies (Jere R. Behrman and Mark Rosenzweig, 1999), for Sweden (Gunnar Isacsson, 1999), and for Australia (Paul Miller et al., 1995).

[4] We follow the medical literature and use the term "within-twin pair," or "within pair" to describe estimates using differences between twins of the same pair. These are variously referred to in the economics literature as between-twins estimates, within-family estimates, first-difference estimates, or within-twins estimates.

(1999) and David Neumark (1999), building on earlier work by Zvi Griliches (1979). They argue that while within-pair differencing removes genetic variation, differences might still reflect ability bias to the extent that ability is affected by more than just genes. To examine this, we follow and extend Ashenfelter and Rouse (1998). We calculate the correlation of average family education with those average family characteristics that might plausibly be correlated with ability or discount rates (e.g., birthweight, partner's characteristics, and smoking). This indicates expected ability bias in a pooled regression. We then calculate the correlation of within-twin-pair differences in education with within-twin-pair differences in characteristics. This indicates expected ability bias in a within-twin-pair regression. Using a range of variables, we find significant correlations in the pooled case, but no significant correlation in the within-pair case. This suggests that ability bias in pooled regressions is likely to be higher than that using within-pair regressions.

There are three other new contributions of this paper. First, we have data on twins' exam and reading scores. Like the other characteristics, we find no correlation between differences in these scores within twin pairs and differences in their education. Second, we also have data on smoking at ages 16 and 18 and we investigate whether smoking is a valid instrument for education. We find that smoking seems to be correlated with family background rather than reflecting individual discount rates and therefore is unlikely to be a valid education instrument. Third, we have information on differences in schools and school classes attended. Not only did the vast majority of twins in our data attend the same school but they also were in the same class.

The plan of the rest of this paper is as follows. In the next section we set out some simple theory. In Section II we describe the data and in Section III the results. Section IV contains concluding remarks.

I. Method

Suppose the wage of twin i in family f is determined by $\log w_{if} = \beta S_{if} + A_{if} + \varepsilon_{if}$ where S_{if} $(i = 1, 2)$ is schooling, A_{if} is "ability," broadly defined as all the other effects on wages outside those of schooling (intelligence, motivation, access to educational funds, etc.), and ε_{if} is an independently and identically distributed (i.i.d.) error.[5] A within-twin-pair estimator of β for identical twins, β_{WTP}, is based on

$$(1) \qquad \log w_{1f} - \log w_{2f} = \beta_{WTP}(S_{1f} - S_{2f}) + (a_{1f} - a_{2f}) + (\varepsilon_{1f} - \varepsilon_{2f})$$

where a_{if} is ability net of family and genetic effects.

There are two issues that arise with this method. First, Rouse (1999) estimates that 10 percent of variation in schooling is due to measurement error. Since measurement error in schooling will be exacerbated by the differencing, β_{WTP} will be downward biased due to the attention bias arising from measurement error (Griliches, 1979; Neumark, 1999). We therefore follow Ashenfelter and Krueger (1994) in instrumenting the reported schooling differences with differences based on reports from the other twin.[6]

The second question is what causes the differences in schooling between identical twins? Ashenfelter and Rouse (1998), Bound and Solon (1999), and Neumark (1999), following earlier arguments due to Griliches (1979), debate this at length in recent papers. Conventional ordinary least-squares (OLS) ability bias to β depends on the fraction of variance in schooling that is accounted for by variance in unobserved abilities that might also affect wages. Similarly, ability bias to β_{WTP} depends on the fraction of within-pair variance in schooling that is accounted for by within-pair variance in unobserved abilities that also affect wages. If the endogenous variation within families is smaller than the endogenous variation between families, then β_{WTP} is less biased than β. Hence even if there is ability bias in within-twin-pair regressions, β_{WTP} might still be regarded as an upper bound on the returns to education (if schooling and ability are positively correlated). However,

[5] i takes the numbers 1 and 2. We have one set of triplets on our data, which we dropped.

[6] Ashenfelter and Rouse (1998) and Rouse (1999) experiment with a number of different instrumentation methods using combinations of own and other twins reporting. Here we instrument using the report of one twin on the education of another. Other instrument configurations gave similar magnitudes to those reported below.

Bound and Solon (1999) argue there is no a priori reason to believe that β_{WTP} is less biased than β.

Ultimately the matter is of course an empirical one. Its investigation is subject to the central problem that ability is not observed. Ashenfelter and Rouse (1998) therefore look at the correlation between schooling and potential correlates of ability (e.g., employment status, tenure, and spouse's education). To investigate the covariance of schooling and ability between families they examine the correlation between the average level of schooling and the average level of characteristics across different families. To investigate the covariance of schooling and ability within families they examine the correlation between the difference in schooling and the differences in characteristics within twin pairs in the same family. They find the former is bigger than the latter and hence argue that most of the variation in ability is between families and not between twins within a family. We present similar investigations below and find similar results to Ashenfelter and Rouse (1998). We also extend their results by looking at twins exam performance and literacy test scores.

Using the same framework we also investigate the suggestion that smoking be used as an instrument for education, since it might proxy discount rates (Victor R. Fuchs, 1986). Daniel Hamermesh (2000) suggested however that youth smoking is a measure of family background and thus not a valid instrument for education. We believe that our twins data allows us to shed some light on the smoking debate. Again, this is based on a comparison of correlations between and within families. A high correlation between family smoking behavior and educational attainment is consistent with both views. However, significant within-twin-pair correlation is only consistent with the hypothesis that smoking reflects an individual's discount rate. Finding no within-twin-pair correlation provides indirect evidence for the family background view.

II. Data

A. *Data Set*

The Twins Research Unit, St. Thomas' Hospital, London, has built up a list of (mainly female) identical and nonidentical twins. The data we have used in this paper are derived from a mailing list of about 6,600 individuals. They are mailed questionnaires on mostly medical information (including birthweight, birth order, gestation period) plus socioeconomic questions on sex, age, presence of children, age of mother, etc. We added more detailed socioeconomic questions to the most recent questionnaire which went out in June 1999. We asked the twins to report their qualifications, their twin's qualifications, the age they finished full-time education, their occupation, their spouse's occupation, their employment status, earnings, and household income (see Data Appendix at http://www.aeaweb.org/aer/contents/, for more details). We should note that response rates are very high (above 80 percent) on these questionnaires, kept up by remailing and telephoning nonrespondents.

Full details of our various measures are set out in the Data Appendix. To calculate wages we asked twins to report normal earnings before taxes and deductions and then asked whether this was hourly, daily, weekly, monthly, or yearly. We also asked how many hours were usually worked (excluding meals and paid overtime). From these questions we converted the wage data into an hourly rate. To measure schooling, we asked each twin to report their qualification and their twin's qualification. Qualifications were split into 12 groups (e.g., University, A levels, 5+ O levels, 1–4 O levels, etc.; see Data Appendix). We then assigned years of education to each qualification.[7]

B. *Descriptive Statistics and Comparisons with Other Work*

We have completed questionnaires from 1,364 individuals who are one of an identical twin pair, aged 21 to 59, all of whom are women. Due to use of postal questionnaires, we do not necessarily have replies from both members of a twin pair. Of the identicals, therefore, we have 621 complete pairs, i.e., 1,242 individuals. For 214 of these pairs (428 individuals) we have complete wage information on both twins in the pair. Thus our sample

[7] See Data Appendix. We refer to this education measure as "estimated" years of schooling. In our regressions we use estimated years. We tried different imputations for estimated years and found similar results. Regressions with reported years, based on the age they left full-time education, gave similar coefficients but were less precisely determined (likely due to recall error).

size is between the Ashenfelter and Krueger (1994; 298 individuals) and Ashenfelter and Rouse (1998; 680 individuals) and Rouse (1999; 906 individuals) studies. Our study is somewhat special as we only have data on female twins. Most of the other studies have both male and female twin pairs, although they do not attempt to estimate wage equations separately for men and women. Our sample size is less than Taubman (1976; 2,038 individuals), Miller et al. (1995; 1,204 individuals), Behrman and Rosenzweig (1999; 1,440 individuals), and Isacsson (1999; 4,984 individuals). However Taubman (1976) had no measurement error correction, Miller et al. (1995) impute earnings from two-digit occupations, and Behrman and Rosenzweig (1999) impute earnings for nonworking women.

How do our data compare with Blanchflower and Elias (1999) (the only other U.K. twins study we are aware of)? They identify 267 (individual) twins from the National Child Development Study (the NCDS, a panel study of all U.K. births between March 3–9, 1958). This is a potentially very rich data set since it contains detailed information about, for example, test scores. There are however two difficulties with the study. First, due to high twin infant mortality and subsequent panel attrition, only 59 pairs have complete wage and education information and, of these, 23 pairs are classified as identical twins (see their figures 1 and 2). They therefore have too little variance among their 23 identical pairs to estimate within-pair equations. Second, the twins were identified as identical at birth, but "... from the documentation we have available to us we are unclear how such designations were made in practice" (their footnote 6). The usual method at that time was to see if there were one or two placenta present and identify identicals as coming from one placenta. Unfortunately recent research indicates that as much as one-third of identicals can come from double placentas (Elizabeth Bryan, 1992). Thus it seems likely that their sample of identicals is identified with substantial error.[8]

An important innovation of the Ashenfelter and Krueger (1994) study is to ask each twin his/her own and their co-twin's education. If self-reported education is measured with error this provides a potential instrument since the report of the other twin should be correlated with the self-reported education level but uncorrelated with the equation regressand. This strategy was adopted in the subsequent Twinsburg, Miller et al. (1995), Behrman and Rosenzweig (1999) studies, and we use it too. Isacsson (1999) uses the comparison of reported education and registry information to control for measurement error.

Table 1 sets out some descriptive statistics for our data along with comparative data from the U.K. Labour Force Survey (LFS) as a check on the representativeness of our sample. Column (1) shows data from the 1999 LFS for all women and all women who report a wage. These women average 12.1 years of schooling, are aged 39, and 59.5 percent are married. Column (3) sets out data for all identical twins. They have 12.6 years of schooling, are aged 44.3, and 65.1 percent are married. So our twins are slightly more educated and slightly older, but our data do not seem to be too far from the average for women. Column (4) shows the data for our working twins, who earn, on average, £10.17 per hour, have worked in the present job for 11.7 years and 58.2 percent are part time. Comparing this to column (2), which shows the LFS data for working women, wages and tenure are slightly lower. These lower LFS figures presumably reflect the somewhat more educated and older twins sample. The figures are very similar if we only consider twin pairs [columns (5) and (6)].

III. Results

A. *Returns to Education*

Table 2 sets out our estimates. Column (1) shows an OLS regression using all working women from the LFS, entering schooling, age, and age squared. The return to education is quite precisely estimated at 7.8 percent. The rest of the columns are estimates for twins. Column (2) is an OLS pooled regression using all identicals for whom we have

[8] Note in passing they also find the sample of identicals have no significant within-twin-pair differences for math and reading scores; see their Table 8.

TABLE 1—DESCRIPTIVE STATISTICS

	LFS 1999		Identical twins		Identical twin pairs	
	All (1)	Working (2)	All (3)	Working (4)	All (5)	Both work (6)
Reported years of schooling[a]	12.1	12.3	12.6	13.0	12.6	13.2
	(2.37)	(2.39)	(2.89)	(2.92)	(2.89)	(3.04)
Estimated years of schooling[b]	12.5	12.9	13.5	13.9	13.5	14.1
	(2.32)	(2.35)	(2.52)	(2.48)	(2.54)	(2.50)
Age	38.9	38.6	44.3	42.7	44.8	42.5
	(11.08)	(10.72)	(10.40)	(10.15)	(10.3)	(10.0)
Married (percent)	59.5	60.3	65.1	61.4	65.3	61.3
Nonparticipating (percent)	29.0	0	18.2	0	18.6	0
Hourly wage rate		7.09		10.17		10.03
		(4.37)		(10.36)		(9.12)
Tenure		6.9		11.7		11.9
		(6.84)		(9.64)		(9.15)
Full time (percent)		58.5		58.2		60.8
Self-employed (percent)		4.8		5.1		4.9
Sample size (individuals)	7,729	4,226	1,364	748	1,242	428

Notes: Standard deviations are in parentheses.
[a] Based on age when finished full-time education minus five.
[b] Based on highest qualification (see Data Appendix).

complete wage information, 428 individuals, and schooling, age, and age squared as regressors. This gives a return to education of 7.7 percent, similar to the figure in column (1). Column (3) maintains a pooled specification, but instruments education with reported level of the other twin. This should control for measurement error in reported education, which would bias down the returns estimate in column (2). As column (3) shows, returns rise to 8.5 percent when this is done.

Column (4) estimates the within-pair equation (1). Figure 1 illustrates data in this case. The cluster around zero is due to the fact that 55 percent of the twin pairs have the same education years. Since the pooled estimates do not control for ability bias we would expect the within-pair returns estimates to be less.[9] As column (4) shows, the return is indeed less, at 3.9 percent, but is poorly determined. This figure might however also reflect downward bias due to exacerbated measurement error in the differenced equation. To check this column (5) instruments reported schooling. As expected the point estimate rises to 7.7 percent, with a standard error of 0.033. Comparison of the pooled IV and the within-pair IV estimates therefore provide an estimate of the magnitude of ability bias as both control for measurement error; comparing columns (3) and (5) suggests ability bias is positive.

The right-hand panel of the table repeats the exercise controlling for marriage, current job tenure, part-time status, and region.[10] The pattern of point estimates on the regressors is similar. As before, measurement error biases returns down (OLS returns are less than IV returns) but here the within-pair IV estimates are slightly higher than the pooled IV estimates suggesting negative ability bias.

Thus we can conclude the following. First, ability bias appears to bias the pooled estimates upwards [with the exception of column (9)]. Second, measurement error appears to bias all estimates downwards especially in the case of the within-pair estimate. Third, female

[9] If the variation within twin pairs is uncorrelated with ability, or if there is more between-family ability bias than within-family bias.

[10] Region is only identified where twins live in different regions. With finely defined regions this rapidly exhausts degrees of freedom and we therefore opted for a London and South East region dummy (the area where U.K. wages are significantly higher). Note we cannot control for ethnicity in the differences since identical twins have the same ethnicity.

TABLE 2—OLS, IV, AND WITHIN-TWIN-PAIR ESTIMATES OF THE RETURN TO EDUCATION FOR IDENTICAL U.K. TWINS

	LFS	Twins							
		Without other covariates				Controlling for other covariates			
	Pooled	Pooled		Within pair		Pooled		Within pair	
	OLS (1)	OLS (2)	IV (3)	OLS (4)	IV (5)	OLS (6)	IV (7)	OLS (8)	IV (9)
Education	0.078	0.077	0.085	0.039	0.077	0.073	0.077	0.039	0.082
	(0.002)**	(0.011)**	(0.012)**	(0.023)	(0.033)*	(0.011)**	(0.018)**	(0.024)	(0.036)*
Age	0.058	0.078	0.077			0.059	0.059		
	(0.004)**	(0.021)**	(0.021)**			(0.024)*	(0.024)*		
Age^2 (÷100)	−0.062	−0.097	−0.095			−0.001	−0.001		
	(0.049)**	(0.027)**	(0.027)**			(0.000)**	(0.000)**		
London and South East						0.065	0.066	0.110	0.124
						(0.054)	(0.055)	(0.122)	(0.123)
Married						−0.004	−0.004	−0.053	−0.048
						(0.059)	(0.059)	(0.091)	(0.092)
Tenure (years)						0.012	0.012	−0.001	0.001
						(0.003)**	(0.003)**	(0.006)	(0.006)
Part time						−0.091	−0.089	−0.099	−0.102
						(0.065)	(0.065)	(0.097)	(0.098)
Observations	4,398	428	428	214	214	374	374	187	187
R^2	0.31	0.15	0.15	0.01	0.0009	0.21	0.21	0.03	0.009

Notes: Pooled twins regressions are based on log $w_{if} = \alpha + \beta S_{if} + \gamma X_{if} + \varepsilon_{if}$ where w_{if} is the wage of twin i in family f, S_{if} is years of schooling, and X_{if} other covariates where indicated. Within-pair regressions are based on log w_{if} − log w_{2f} = $\beta(S_{1f} - S_{2f}) + \gamma(X_{1f} - X_{2f})$. Pooled estimates in column (1) are for all working women reporting a wage in the 1999 U.K. Labour Force Survey. Standard errors are in parentheses. Columns (1), (2), (3) and (6), (7) include a constant (not reported), the other columns exclude a constant. For the pooled IV estimates, twin 1's education is instrumented by twin 2's report of twin 1's education and vice versa. For the within-twins IV estimates, the difference in self-reported education is instrumented by the difference in the co-twin's report of the other's education.

* Significant at the 5-percent level.

** Significant at the 1-percent level.

returns to education appear to be about 7.7 percent, very similar to OLS estimates, suggesting the measurement error and ability bias

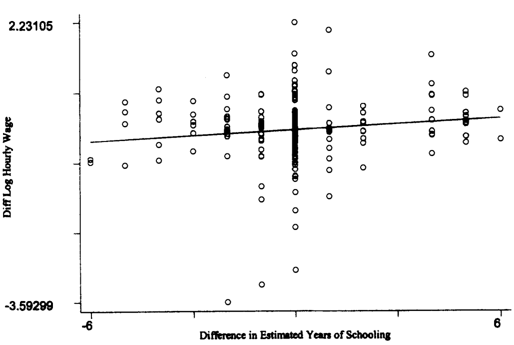

FIGURE 1. DIFFERENCES IN LOG HOURLY EARNINGS AGAINST DIFFERENCES IN SCHOOLING (Schooling Based on Highest Qualifications)

roughly cancel out. Fourth, Dearden (1998) obtains returns of 8.3 percent for women using covariates from the NCDS to control for ability and family background (see her Table 4.3, column 4). Thus our results are similar to hers.

Fifth, these above comments on the twins' results refer to our point estimates. It is worth noting that our within-pair results are insufficiently precise to state that such differences are statistically significant.[11] Nonetheless the pattern of point estimates is that suggested by theory and very similar to the pattern in other twin studies; see the conclusion for a summary.

[11] Using Hausman tests for differences between the IV and OLS coefficients in columns (2) and (3), (4) and (5), (6) and (7), (8) and (9), we could not reject the null of no difference between them.

TABLE 3—BETWEEN-FAMILY AND WITHIN-FAMILY TWIN-PAIR CORRELATIONS OF EDUCATION AND OTHER VARIABLES

Correlation of average family education with average family characteristics		Correlation of within-twin differences in education with within-twin difference in other characteristics	
	Education		Δ Education
Birthweight	0.2153***	Δ Birthweight	−0.0765
Married	−0.1279***	Δ Married	−0.031
Self-employed	−0.0876*	Δ Self-employed	−0.03
Part time	−0.2067***	Δ Part time	0.0379
Partner's tenure	−0.2124***	Δ Partner's tenure	−0.0093
Partner's occupation	0.4908***	Δ Partner's occupation	0.0305
For reduced sample:			
Passing 11+	0.1095	Δ Passing 11+	−0.0556
Adult reading score[a]	0.4933***	Δ Adult reading score	0.2111

[a] The reading score used is the National Adult Reading Test (NART).
* Significant at the 10-percent level.
*** Significant at the 1-percent level.

B. *Ability Differences Within Twin Pairs*

To investigate ability biases within and between families Table 3 shows the results of the correlation analyses described in the introduction. Consider the first column, first row. This shows that the correlation between average family education and average family birthweight is 0.22 and is highly significant. It suggests that families with low average birthweight have low average schooling, consistent with ability and family background affecting schooling choice. The second column shows an insignificant correlation between differences in education within twin pairs and differences in birthweight within twin pairs. To the extent that birthweight measures ability therefore, between-family differences in education are more affected by ability bias than the within-pair education differences.

The rest of the first column shows other family correlations. This shows strong correlations between average family education and average family marriage status, self-employment, part-time status, partner's tenure, and partner's occupation. The second column shows the correlations between within-pair differences in education and within-pair differences in characteristics. None of them is significant. In sum, within-pair education differences are uncorrelated with any other within-twin difference in observables. Of course, these characteristics are incomplete measures of ability, but the evidence is suggestive, especially as it mirrors that found by Ashenfelter and Rouse (1998).

For a subsample of twins we managed to collect more detailed data on characteristics that are also likely to be highly correlated with ability. For these twins we have their reading score on the National Adult Reading Test (NART) and whether the twins passed the 11+ exam (an exam taken at age 11). Before the introduction of comprehensive schools, the 11+ was universally applied across Britain as a means of selecting which secondary school to attend. If the pupil passed the 11+ (around 25 percent of the population) this meant that they were selected to attend a grammar school where education was largely academically based. If the pupil did not pass they were selected to attend a secondary modern school where education was more vocationally based. As a result this 11+ test result can be regarded as an early ability test.[12] However, we only have data on 43 pairs (86 individuals) who reported the answer to this question in a short follow-up questionnaire we conducted. Of these 43 pairs only 3 pairs actually received a different result in the 11+ test.[13] As shown in the lower panel of

[12] The 11+ consists of four multiple choice tests of English, math, verbal, and nonverbal reasoning. The aim of the test is to elicit academic versus vocational ability.

[13] Fifteen pairs both passed and 25 pairs both failed: in only 3 pairs did one pass and one fail.

Table 3, correlations between families and within twins show a pattern similar to the upper panel.

In addition to this early ability measure an adult ability measure is also available for a subset of twins. The measure is the National Adult Reading Test (NART) which is based on the ability to read and correctly pronounce each word from a list of 25 words. Of course, being an adult reading test, the result may be affected by the schooling the respondent has received. However, John Crawford et al. (2001) compared NART results of 77-year-olds with IQ tests taken when the same individuals were 11 and obtained a correlation of 0.69. We have NART test results for 108 identical twin pairs (the twins' scores had a correlation of 0.71). As shown in Table 3, the NART results confirm the same pattern as the other ability correlates: there is a high and significant correlation between average family NART and average family education but the corresponding correlation of within-twin differences is insignificant. This is additional evidence that educational differences within twin pairs are likely to be less correlated with ability difference than across families.[14]

C. *Smoking as an Instrument?*

A strength of our data is that we have information on the smoking behavior of the twins at the age of 16 and 18. Smoking has been suggested as an instrument for education, since it might proxy discount rates (Fuchs, 1986) and subsequently been used by William Evans and Edward Montgomery (1994) for the United States and Arnaud Chevalier and Ian Walker (1999) for the United Kingdom. This was criticized by Hamermesh (2000) who suggests that a youth's smoking behavior is a measure of family background and thus not a valid instrument for education.

Evans and Montgomery (1994) show that smoking is highly correlated with educational outcomes and use it as an instrument in estimating returns to education. Their IV estimate of the returns to education lies about 10 percent above the OLS estimate.[15] This would indicate negative ability bias, unlike twins studies where ability bias is small or positive. Evans and Montgomery present indirect evidence that the correlation of smoking and educational attainment is due to differences in time preferences. However, they acknowledge that there is no possibility to test this directly against the alternative hypothesis that the observed correlation is due to unobserved "ability" in a very broad sense including genes, family, and social background as well as peers.

While not able to perform a direct test, our twin data allow us to advance indirect evidence which relies on the correlation method used in Table 3. A significant negative correlation between average family smoking and average family education is consistent with either smoking reflecting discount rates or family background. However, if smoking reflects individual's discount rates, differences in smoking within families should be correlated with within-family differences in education. But the within-pair correlation should be insignificant if the cross-sectional correlation between smoking and education is due to family background.

Table 4 shows the correlation results for smoking. There is a strong significant negative correlation between average family smoking and average family education. However, there is no significant correlation between within-twin-pair smoking and within-twin-pair education. This suggests smoking is more likely to reflect family background than discount rates. Furthermore, if the family background view is true and if ability bias is positive—as is the case for our data—then using smoking as an instrument is

[14] We note however that measurement error likely biases downwards the within-twins correlation coefficients in column (2) of Table 3. To get some idea of this bias, note that the measured correlation coefficient can be written as $r^M = [\text{cov}(\Delta S^M, \Delta X^M)]/[\sqrt{\text{var}(\Delta S^M)}\sqrt{\text{var}(\Delta X^M)}]$ where ΔS^M and ΔX^M are the measured differences in schooling and differences in characteristics respectively. We wish to calculate the correlation between the true quantities $r^T = [\text{cov}(\Delta S^T, \Delta X^T)]/[\sqrt{\text{var}(\Delta S^T)}\sqrt{\text{var}(\Delta X^T)}]$ where the T superscript denotes the true value. Assuming classical measurement error so that $\Delta S^M = \Delta S^T + \Delta v_1$ and $\Delta X^M = \Delta X^T + \Delta v_2$ we can show that $r^M = r^T\sqrt{(rr_{\Delta S})}\sqrt{(rr_{\Delta X})}$ where $(rr_{\Delta S})$ and $(rr_{\Delta X})$ are the reliability ratios of differences in schooling and X respectively (i.e., $rr_{\Delta S} = \text{var}(\Delta S^T)/\text{var}(\Delta S^M)$ and $rr_{\Delta X} = \text{var}(\Delta X^T)/\text{var}(\Delta X^M)$). The ratio of the OLS to IV estimates in Table 2 columns (4) and (5) gives $rr_{\Delta S} = 0.039/0.077 = 0.506$ in this case. Therefore, assuming that $rr_{\Delta S} = rr_{\Delta X}$, the r^M data in Table 2, column (3) are biased down by about ½ .

[15] The difference is higher in their estimates for females only.

TABLE 4—BETWEEN-FAMILY AND WITHIN-FAMILY TWIN-PAIR CORRELATION OF EDUCATION AND SMOKING

Correlation of average family education with average family characteristics		Correlation of within-twin differences in education with within-twin difference in other characteristics	
	Education		Δ Education
Smoking at 16	−0.2680***	Δ Smoking at 16	−0.0241
Smoking at 18	−0.2699***	Δ Smoking at 18	−0.0541

*** Significant at the 1-percent level.

likely to exacerbate ability bias. Table 5 investigates this. Column (1) upper panel shows, for comparison, the pooled OLS results from column (2) of Table 2. The return of 0.077 compares closely with the two smoking studies in the literature set out in the lower panel, Evans and Montgomery (1994; 0.079) and Chevalier and Walker (1999; 0.099). The second and third columns show returns to education when using smoking at ages 16 and 18 as an instrument. The returns rise just as in the Chevalier/Walker and Evans/Montgomery studies (see lower panel), consistent with an exacerbation of positive ability bias. In sum, evidence seems to suggest that smoking reflects family background rather than discount rates. Thus the higher estimated returns in studies using smoking as an instrument are more likely caused by an augmentation of (positive) ability bias than the existence of negative ability bias.[16]

D. *Selection Bias*

How are the returns to education estimates affected by possible selection bias? There are several different selection stages in order to appear in the regression sample: a pair have to volunteer to be on the database, both have to respond to the questionnaire, both have to be working, and both have to report wages.[17]

Our main concern is with the within-twin-pair estimates and we shall argue that selection is not a problem as long as returns are linear. Consider first the effects on the *pooled* estimates. The profile of our twins in Table 1 suggests better-educated twins seem more likely to volunteer to be on the database and return the questionnaire. However, if returns to education are linear in schooling, then having a sample of highly schooled individuals should not matter for pooled estimates. If there are diminishing marginal returns[18] then, since we have a slightly above average education group, our pooled estimates would understate the "average" marginal returns. As in all studies that are concerned with wages there is the potential of selection bias due to the participation decision. As our sample consists of female twins, selection issues of this kind do potentially affect our pooled estimates. We therefore experimented on the pooled regressions with traditional Heckman-correction models (using children and husband's occupation in the participation equation) but found no evidence that selection affected our estimates significantly.

Turning to the within-twin-pair differences, in conventional wage regressions the selection problem is that the observed sample consists of individuals with a high wage plus low-wage individuals with a positive wage shock. Here, the analogous problem is that very low-wage individuals with adverse shocks are not likely to be observed, thus likely removing those with low levels of education. But this does not

[16] As before, using Hausman tests for differences between the IV and OLS coefficients in Table 5, we could not reject the null of no difference between them.

[17] The response to the wages section is optional (although in practice almost all of our working sample responded).

[18] The higher marginal returns in IV studies are often attributed to high marginal returns for a low educated group whose behavior is frequently the source of variation of the instrument (Card, 1999).

TABLE 5—SMOKING AS AN INSTRUMENT: OLS AND IV ESTIMATES OF THE RETURN TO EDUCATION FOR IDENTICAL TWINS
(Dependent variable log wages)

		Smoking at 16	Smoking at 18
Instrument:	OLS	IV	IV
Education	0.077	0.110	0.104
	(0.011)**	(0.044)*	(0.045)*
Age	0.078	0.074	0.074
	(0.021)**	(0.022)**	(0.022)**
Age^2 (÷100)	−0.097	−0.089	−0.091
	(0.027)**	(0.029)**	(0.029)**
Observations	428	428	428
R^2	0.15	0.13	0.14
Evans and Montgomery (1994)[a]			
Education	0.079		0.122
	(0.003)**		(0.030)**
Chevalier and Walker (1999)[b]			
GHS: Education	0.099	0.163	
	(0.003)**	(0.011)**	

Note: Regressions are for identical twins of log $w_{if} = \alpha + \beta S_{if} + \gamma X_{if} + \varepsilon_{if}$ where w_{if} is the wage of twin i in family f, S_{if} is years of schooling, and X_{if} covariates in age as indicated.

[a] Results from their Table 10.

[b] Results from their Table 23 (General Household Survey, GHS).

* Significant at the 5-percent level.

** Significant at the 1-percent level.

remove an individual in the differenced regressions, but rather a pair with one twin who has low education. If differences in education are random we are as likely to delete positive or negative differences as the twins are ordered randomly. In this case, then, if selection removes some differences it affects the precision of the estimates.

What if differences in education are nonrandom? The sample correlation between average family education levels and differences in education was −0.02 (and insignificant). This suggests there are (very weakly) smaller differences in education in more educated families. Thus we are somewhat more likely to remove pairs of twins with low average levels of education. However, there is no bias as long as returns to education are linear, since it makes no difference if we observe differences in schooling years between high-educated or low-educated pairs. If returns are diminishing, however, and since highly educated twins have somewhat smaller differences, then our estimates might underestimate the "average" marginal return.

While this argument suggests there is unlikely to be a selection bias problem for the within-pair estimates, we tried some more formal testing. Following Insan Tunali's (1986) work on double selection (in nondifferenced models) we estimated a bivariate probit model for participation of twin pairs, which returned predicted probabilities of each pair both working, neither working, and one or other working. We used the predicted probabilities of each twin working to form two Heckman selection terms in the twins differenced wage equation. The terms were insignificant in the within-twin-pair regressions and the return to education parameter was unaffected. Thus there is no evidence that selection bias is a problem for our estimates, although the modest sample size and absence of a compelling instrument for the probability of twins working gives us limited power.

E. *Other Aspects of Twins Schooling*

To further explore this we asked a subsample of the twins a number of other questions.[19] First, we asked about the twins' schooling experience. These are very similar. Reading down the panels in Table 6, of our sample of 67 twin pairs, 100 percent of them went to the same primary school, with 90 percent of them in the same class. Ninety-six percent went to the same secondary school, with 62 percent in the same class. We asked why the remaining 38 percent were separated. As the table shows, 17 percent were separated because it was school policy to keep them apart, and 16 percent were separated due to ability. Differences were somewhat more marked in higher education, although in 60 percent of cases neither twin went to higher education, and 16 percent went to the same institution. Thus while it is not clear exactly why schooling years differ from these data, it does not seem that ability is a large determinant. Nor does it seem that differences in school or teacher quality underlie the earnings differences since the vast majority of twins went to the same school and class.

[19] This was a subsample of 134 individuals, 67 pairs, who attended the unit for medical tests.

TABLE 6—SCHOOL EXPERIENCE OF THE TWINS (SAMPLE OF 134 INDIVIDUALS, 67 IDENTICAL PAIRS)

	Percent	Number
Primary school		
Same school	100	134
Same class	90	120
Reason for different class:		
Apart	4	5
Ability	4	5
Other	2	4
Secondary school		
Same school	96	129
Reason for different school:		
Apart	1	1
Ability	3	4
Other		
Same class	62	79
Reason for different class:		
Apart	17	22
Ability	16	20
Other	5	7
Higher education		
Both no higher education	60	79
Same institution	16	21
Reason for different institution:		
Only one stayed	7	9
Apart	2	2
Different courses	9	12
Other	6	8

Notes: "Apart": the twins were put into separate classes/schools either because of school policy or parental preferences. "Ability": twins reported that it was due to different ability that they were in separate classes/schools.
Source: Authors' calculations from follow-up survey.

Second, we asked the twins why they had different years of schooling up to age 21. In fact, 80 percent reported that they had had the same years. The reasons given by the remaining 20 percent were very mixed; 4 percent reporting differences in "interest" and 4 percent "being less academic."[20]

Finally, we identified a number of other possible reasons for schooling differences. We asked twins if they had taken a career break, the age when they first got married, and if they worked in the public sector. The only positive correlation between within-twin-pair differences in education and differences in these characteristics was the age at which first married (a correlation of −0.21, significant at 10 percent). This suggests putting this age into the earnings regression, or possibly using this as an instrument, but the sample sizes are rather too small to do this effectively.

In sum, the only correlate of within-twin-pair differences in schooling is within-twin-pair differences in age at first marriage. No other characteristic is correlated with this and there is no evidence from schooling records that differences in abilities are marked (at least differences in abilities wide enough for different classes).

IV. Conclusions

We have used a new sample of identical U.K. twins to estimate returns to education using the within-twin-pair method to correct for measurement error. Our findings consist of (i) those arising from replication of method on new data and (ii) new findings. Concerning replication, we have four main findings. The point estimates from our twins sample confirm the theoretical prediction that, first, measurement error biases estimated returns to education down and, second, omitted ability biases estimates up. Third, in fact these effects roughly cancel each other out indicating a private return to education for women of 7.7 percent. Fourth, using similar correlates of ability to Ashenfelter and Rouse (1998), such as tenure, partner's occupation, etc., we find no correlation between differences in these measures within twin pairs and differences in their education, but a strong correlation between average family measures and average family education. This pattern is repeated using data on birthweight (which Ashenfelter and Rouse, 1998, did not have). Thus we find no evidence that ability bias is likely to bias our within-twin-pair results by more than the pooled results. Thus we expect ability biases to be less for within-pair estimators than for estimators not controlling for ability. Therefore, conditional on positive ability bias, which we find, our estimates at least tighten the upper bound for the returns to education.

Our new findings are threefold. First, for our

[20] The remaining scores were: got married, 1 percent; inspired by teacher, 0 percent; no access to course, 2 percent; financial, 0 percent; other, 4 percent; and missing, 1 percent.

whole sample we have data on early smoking behavior. Our results suggest that smoking is more likely to reflect family background than individual discount rates. Therefore, smoking used as an instrument for education is likely to exacerbate ability bias. Second, for a smaller sample we have data on twins' exam and reading scores. Like the other characteristics, we find no correlation between differences in these scores within twin pairs and differences in their education. To the extent that these measure ability differences, this again suggests that ability differences within twin pairs are less than those between families. Third, we have information on differences in schools and school classes attended. Not only do the vast majority of twins in our data attend the same school but they also are in the same class. This suggests that our findings are not affected by different school or class qualities.

How do our results compare with other twins studies? First, concerning point estimates of returns, of the seven studies we are aware of (see footnote 3), IV estimates for men and women range from 0.167 (Ashenfelter and Krueger, 1994, for U.S. twins) to 0.042 (Isacsson, 1999, for Swedish twins). Our results are about in the midpoint of these estimates. Second, the comparisons of findings with other twins studies using the education of the other twin as an instrument show a consistent pattern of results. First, in all studies the OLS within-pair estimates are less than the IV within-pair estimates. This suggests that measurement error bias in within-pair studies is negative, in some cases severely. Second, in all studies the pooled IV estimates are larger than the within-pair IV estimates.[21] This suggests that ability bias is positive. Third, what is the overall effect on schooling returns? Comparing the pooled OLS and the within-twin-pair IV results, in two studies (Isacsson and Rouse) the pooled results are larger, but by under 10 percent, in two (Miller et al., 1995; Ashenfelter and Rouse, 1998) they are larger by 20–25 percent, and in one (Ashenfelter and Krueger, 1994) smaller by 99 percent (in this study they are the same). The point estimate therefore suggests that in most cases the positive ability bias is slightly larger than the negative measurement error bias.

As in other twins studies, there are of course a number of caveats to our results. First, we do not have a large enough sample to show statistically significant differences between the various different estimation methods. Thus it is important to stress that our conclusions are on the basis of our point estimates. Nonetheless, the pattern of our point estimates are in line with both that predicted by theory and with other twins studies. Second, our estimates assume that returns to education are linear. This assumption is forced on us by lack of data; we do not have enough identical twin pairs with education differences across different qualifications to estimate different returns.[22] Third, selection is often cited as a worry in estimates of female wage equations, if it yields a nonrandom sample of wages. The issue here is somewhat different: selection would be a problem if it yields a nonrandom sample of wage differences. We find only very weak evidence that there are smaller wage differences in more educated families, and since their average level of education is higher, our sample of differences consists, to some small extent, of smaller wage differences between better-educated families. Smaller differences decrease precision, and if there are diminishing marginal returns then we may underestimate the average marginal return to education.

In future work we hope to be able to extend the data set to consider male twins and, with an increased sample size, consider the issue of heterogeneous returns to education both with respect to different qualifications and to parental background.

[21] With the exception of Ashenfelter and Krueger (1994); but Rouse (1999) shows this result appears in that particular cross section only and not found in the subsequent Twinsburg studies.

[22] Hawkes (2003) adds nonidenticals to our identical sample and, using pooled regressions, finds some evidence of different returns to different qualifications (10 percent for O-levels, 8 percent for A-levels, and 7 percent for degree). Again, these estimates are not significantly different.

REFERENCES

Angrist, Joshua and Krueger, Alan B. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, November 1991, *106*(4), pp. 979–1014.

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

Ashenfelter, Orley and Rouse, Cecilia. "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics*, February 1998, *113*(1), pp. 253–84.

Ashenfelter, Orley and Zimmerman, David J. "Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers." *Review of Economics and Statistics*, February 1997, *79*(1), pp. 1–9.

Behrman, Jere R. and Rosenzweig, Mark. " 'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates." *Economics of Education Review*, April 1999, *18*(2), pp. 159–67.

Blanchflower, David G. and Elias, Peter. "Ability, Schooling and Earnings: Are Twins Different?" Unpublished manuscript, Warwick University, 1999.

Blundell, Richard; Dearden, Lorraine and Sianesi, Barbara. "Estimating the Returns to Education: Models, Methods and Results." Centre for the Economics of Education (CEE) Discussion Paper No. 16, London School of Economics, October 2001.

Bound, John and Solon, Gary. "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling." *Economics of Education Review*, April 1999, *18*(2), pp. 169–82.

Bryan, Elizabeth. "Twins and Higher Multiple Births—A Guide to their Nature and Nurture." Auckland: Edward Arnold, 1992.

Card, David. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in L. Christofides, E. Grant, and R. Swidinsky, eds., *Aspects of labour market behaviour: Essays in honour of John Vanderkamp*. Toronto: University of Toronto Press, 1997, pp. 201–22.

______. "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of labor economics*, Vol. 3A. Amsterdam: Elsevier Science, 1999, pp. 1801–63.

Chevalier, Arnaud and Walker, Ian. "Further Evidence on the Returns to Education in the UK." Mimeo, Keele University, 1999.

Crawford, John R.; Dearly, Ian J.; Starr, John M. and Whalley, Lawrence J. "The NART as an Index of Prior Intellectual Functioning: A Retrospective Validity Study Covering a 66-Year Interval." *Psychological Medicine*, 2001, *31*, pp. 451–58.

Dearden, Lorraine. "Ability, Families, Education and Earnings in the UK." IFS Working Paper No. W98/14, 1998.

______. "Qualifications and Earnings in Britain: How Reliable are Conventional OLS Estimates of the Returns to Education?" IFS Working Paper No. W99/7, 1999.

Evans, William N. and Montgomery, Edward. "Education and Health: Where There's Smoke There's an Instrument." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4949, December 1994.

Fuchs, Victor R. "Time Preference and Health," in Victor R. Fuchs, ed., *The health economy*. Cambridge, MA: Harvard University Press, 1986, pp. 214–42.

Griliches, Zvi. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy*, October 1979, Pt. 2, *87*(5), pp. S37–S65.

Hamermesh, Daniel. "The Craft of Labormetrics." *Industrial Labor Relations Review*, April 2000, *53*(3), pp. 363–80.

Harmon, Colm and Walker, Ian. "Estimates of the Economic Return to Schooling for the United Kingdom." *American Economic Review*, December 1995, *85*(5), pp. 1278–86.

Hawkes, Denise. "Education, Earnings, Ability and Early Child Bearing: Evidence from a Sample of U.K. Twins." Ph.D. dissertation, Queen Mary, University of London, in draft, 2003.

Herrnstein, Richard J. and Murray, Charles. *The bell curve*. New York: Free Press, 1994.

Isacsson, Gunnar. "Estimates of the Return to Schooling in Sweden from a Large Sample of

Twins." *Labour Economics*, November 1999, *6*(4), pp. 471–89.

Miller, Paul; Mulvey, Charles and Martin, Nick. "What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings." *American Economic Review*, June 1995, *85*(3), pp. 586–99.

Neumark, David. "Biases in Twin Estimates of the Return to Schooling." *Economics of Education Review*, April 1999, *18*(2), pp. 143–48.

Rouse, Cecilia. "Further Estimates of the Economic Return to Schooling from a New Sample of Twins." *Economics of Education Review*, April 1999, *18*(2), pp. 149–57.

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

Tunali, Insan. "A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Remigration," in R. Ehrenburg, ed., *Research in labor economics*. Greenwich, CT: JAI Press Inc., 1986, pp. 235–83.