A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias*

by

Orley Ashenfelter
Princeton University
&
NBER

Colm Harmon
University College Dublin
&
CEPR, London

Hessel Oosterbeek
Tinbergen Institute
&
University of Amsterdam

Draft 6.03 October 3rd, 1999

ABSTRACT

In this paper we provide an analytical review of previous estimates of the rate of return on schooling investments and measure how these estimates vary by country, over time, and by estimation method. We find evidence of reporting (or "file drawer") bias in the estimates and, after due account is taken of this bias, we find that differences due to estimation method are much smaller than is sometimes reported, although some are statistically significant. We also find that estimated returns are higher in the U.S. and they have increased in the last two decades.

* There are a number of persons to thank for commenting on early drafts of this paper including Joshua Angrist, Kevin Denny, Henry Farber, Alan Krueger and David Madden, seminar participants at the University of Essex and MIT, and Roope Uusitalo, who first suggested a meta-study. The work also draws on the research of many authors with whom we have benefited from discussions, especially David Card. Harmon acknowledges the support of a research award under the terms of the President's Award Scheme in University College Dublin and the hospitality of the Industrial Relations Section at Princeton, which facilitated this collaboration. Thanks to Emma Barron for assistance in preparing the data on rates of return for the meta-study in this paper. The usual disclaimer applies. The computer code for the procedures used in this paper can be obtained on request from the authors.

1. Introduction

In recent years there has been considerable interest in whether measured correlations between schooling and earnings reflect the causal impact of schooling on earnings. This interest has led to some innovative methods of estimation that take advantage of either exogenous determinants of schooling decisions (instrumental variables) or comparisons between genetically identical workers (twins). In this paper we provide an analytical survey of estimates of the rate of return to schooling that is designed to determine the extent to which rates of return differ by country, over time, and with the method of estimation.

In providing this survey we have taken especial care to study the role that "reporting" or "file drawer" bias may have played in the studies we observe. Reporting bias may arise because of the tendency in virtually all scientific fields to report statistical results that tend to reject the hypothesis of no effect. The result is that estimated effects of schooling may be correlated with sampling errors and, if these are in turn correlated with other variables, conclusions about the determinants of rates of return may be seriously biased. The existence of any such bias is no reflection on any individual scholar, but is instead the natural working of a scientific process designed to discover important new results. We implement a generalized method for testing for reporting bias, and for adjusting reported estimates for reporting bias, that is due to Hedges (1992).

The results of our study provide some evidence that estimated rates of return do suffer from reporting bias, especially those based on instrumental variables or within-twin estimators. After adjustment for reporting bias we find strong evidence of a considerable payoff to schooling that differs less with the estimation method used to measure that payoff than is sometimes reported.

2. Simple Earnings Functions & the Problems of OLS-Estimation

Ignoring other covariates, Mincer's (1974) specification for the determinants of earnings is

$$y_i = \alpha + \beta S_i + u_i \quad (1)$$

where y_i is the log of earnings of individual i, S_i is a measure of their schooling, u_i is a statistical error term and α and β are parameters to be estimated, with the parameter β being the *return* to schooling. In the early literature following Mincer's approach, equation (1) - extended with linear and quadratic experience terms to account for on-the-job training - was commonly estimated by means of ordinary least squares (OLS). This estimation technique assumes that the explanatory variables are uncorrelated with the unobserved disturbance in the equation, which for various reasons might not be fulfilled.

The coefficient β *i*s biased if an individual's 'ability' or motivation affects earnings but is omitted from the earnings equation, with the extent of the bias determined by the correlation between education and ability. The concern about the formulation of an estimate of the return to schooling β is that ability may be associated with both wages and schooling. Three approaches have been used to try to deal with this potential problem. One approach deals with the issue of ability bias by including explicit measures that proxy for unobserved ability. IQ and related tests are an example of such proxies (Griliches (1977), Griliches and Mason (1972)). The results of these studies have sometimes suggested that there is an upward bias in results that lack an ability measure. The method of adding ability proxies has been criticised, however, because it is extremely difficult to develop ability measures that are not themselves determined by schooling. When the ability measure is itself influenced by schooling, the use of ability proxies will, in fact, bias estimated rates of return downward.

The ‘siblings’ or ‘twins’ approach exploits a belief that siblings are more alike then a randomly selected pair of individuals, given that they share common heredity, financial support, peer influences, and geographic influences. The approach attempts to eliminate omitted ability bias by estimating the return to schooling from differences between siblings or twins in levels of schooling and earnings. If the omitted variable, say ability (A), is such that siblings have the same level of A, then any estimate of β from within family data will eliminate this bias.

Studies based on sibling or twin comparisons have suffered from two primary criticisms. First, if ability has an individual component as well as a family component, which is not independent of the schooling variable, the within-family approach may not yield estimates that are less biased than OLS estimates. Second, if schooling is measured with error, this will account for a larger fraction of the differences between the twins than across the population as a whole. This would imply that the bias from measurement error in schooling is likely to increase by forming differences between twins, which means the within-twin estimates will be biased downwards. Recent contributions to the twins and siblings literature have attempted to deal with the measurement error problem by collecting multiple measures of schooling by questioning the siblings about each other or by using independent measures of error variances to adjust the estimates. Many of the within-twin studies suggest that ability bias is relatively small, although this is only the case when measurement error has been controlled.

A third approach to the problem of ability bias exploits natural variation in data caused by different influences on the schooling decision. The essence of this 'natural experiment' approach is to provide a suitable determinant (or instrument) for schooling that is not correlated with the earnings residual. In principle, natural experiments provide the closest equivalent to a randomised trial in a clinical study. In the context presented here the treatment group is chosen (albeit not randomly) independent of individual characteristics. The treatment and control groups should, in principle, be identical in other observed and unobserved characteristics that affect earnings except for schooling. By constructing instruments for schooling that are uncorrelated with the earnings residual the instrumental variables (IV) approach will generate a consistent estimator of the return to schooling. The basic idea of the IV estimator is to proceed in two stages. First, estimate the effect of the instrumental variable on schooling; then estimate the effect of the instrumental variable on earnings. Since, by assumption, the instrument is correlated with earnings only because it influences schooling, the ratio of the effect of the instrument on earnings to its effect on schooling provides an estimate of the causal effect of schooling on earnings. The primary criticism of IV estimates revolves around the concern that the instrument may not, in fact, be truly independent of the earnings residual. If, for example, the instrument is positively correlated with earnings, the IV estimator may be upward biased.

The results from IV studies are varied, but some point towards the presence of a downward bias in OLS estimates. Card (1998) has proposed an explanation for this phenomenon that is based on the hypothesis that the return to schooling is heterogeneous and declines at higher levels of schooling. IV estimates will differ from OLS estimates to the extent that the instrument influences schooling decisions at different levels. If the instrument influences decisions primarily at lower levels of schooling, the IV estimator may be higher than the OLS estimator because it reflects the payoff to schooling at lower rather than higher schooling levels.

It is apparent from this discussion that the estimates of returns to schooling may differ because of the estimation method. In what follows we systematically investigate the role of the estimation method - along with region and time period - as determinants of the payoff to schooling.

3. Meta-Analysis of the Returns to Schooling Literature

There already exist several extensive summaries of the payoff to schooling, including Psacharopolous (1994) and Card (1998). Here we use methods common among statisticians, and sometimes called “meta-analysis,” to test whether estimated payoffs are sensitive to estimation period or time period covered and to provide a framework to determine whether our inferences are sensitive to reporting (or “file drawer”) bias. As noted in Huque (1988), Hunter *et al.*(1990) and Egger and Smith (1997) a meta-analysis combines and integrates the results of several studies that share a common aspect so as to be 'combinable' in a statistical manner.

Although less common in economics[1], there has been considerable concern in the medical and statistical literature over whether the observed sample of published results was selected solely because they were "statistically significant". If they were, then any survey of these results suffers from the sample selection bias so well known in a different context in econometric analyses. In what follows we provide estimates of the extent of this kind of selection bias and also of its effect on estimates of the factors associated with differences across time, across econometric methods, and across regions in the return to schooling. We test for publication bias using a method due to Hedges (1992) that we have generalized to accommodate systematic heterogeneity in the payoff to education.

It is important to understand that “reporting bias” may exist even without the authors of individual studies being aware of it. The potential problem simply arises because of the desire to report useful findings. Except in unusual circumstances, evidence against the null hypothesis—

[1] Similar issues have been addressed extensively by financial economists; see Brown, Goetzmann, and Ross (1995) and Lo and MacKinlay (1990). Card and Krueger (1995) also comment on this problem in their survey of minimum wage studies.

that is, favorable to the finding of a treatment effect—is more valuable and more likely to be reported in any rational weighting of the usefulness of empirical evidence.

Table 1 ***Sample Statistics – Rates of Return Data (27 studies, 9 countries)***

Variable	**ALL**		**OLS**		**IV**		**TWINS**	
	Mean	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
Year	88.71	*5.93*	88.38	*5.90*	88.14	*6.46*	92.00	*3.00*
Sample Size/1000	32.70	*86.65*	35.67	*96.50*	38.54	*84.74*	0.631	*0.470*
Ability Controls? (1=Yes)	0.198	*0.401*	0.180	*0.388*	0.257	*0.443*	0.091	*0.302*
Estimated Rate	0.079	*0.036*	0.066	*0.026*	0.093	*0.041*	0.092	*0.037*
Standard Error	0.015	*0.017*	0.006	*0.007*	0.026	*0.022*	0.021	*0.010*
Published?	0.365	*0.484*	0.420	*0.500*	0.286	*0.458*	0.364	*0.505*
Measurement Error?	0.063	*0.243*	0.020	*0.141*	0.057	*0.236*	0.273	*0.467*
N	96		50		35		11	

For our analysis we created a data set of 96 different returns to schooling, obtained from 27 studies. Table 1 reports some descriptive statistics for this sample of both published and unpublished studies, broken down by estimation method, with the full listing of the studies reported in Appendix Table A1.[2] The year of estimation averages in the mid-80's but ranges from 1974 to 1995. Sample size is quite varied with the smallest sample naturally being observed in the twins' studies. Ability controls are quite common in the literature, with around 20% of the OLS estimated sample containing ability measures, and a somewhat higher representation in the IV estimated returns. Explicit control for the presence of measurement error is increasingly a feature of the literature and we see an average of 6% of the rates of return coming from models where this is the case, although control for measurement error is far more common in the twins studies. With respect to the returns and their corresponding standard errors/t-statistics we see the pattern emerging clearly - average returns of 6/7% with corresponding IV and twins study estimated returns of 9%. Precision is lost when we move from OLS however, as seen in the far larger standard errors among the IV and twins studies.

Table 2 Meta-Analysis – OLS Regression of Estimates of Returns to Schooling

	ALL		US		Non-US	
	Estimate	*Std.Err*	Estimate	*Std.Err*	Estimate	*Std.Err*
(A)						
Non-US Data	-0.003	*0.009*				
Year of Sample	0.002	*0.001*	0.002	*0.001*	0.001	*0.001*
Estimated by IV	0.031	*0.007*	0.033	*0.010*	0.031	*0.009*
Estimated by Twins	0.016	*0.012*	0.026	*0.014*	-0.001	*0.017*
Sample Size/1,000,000	0.001	*0.004*	-0.002	*0.004*	0.379	*0.233*
Ability Controls	-0.002	*0.009*	-0.054	*0.015*	0.024	*0.011*
Published	0.018	*0.010*	0.016	*0.011*	0.037	*0.015*
Measurement Error	0.013	*0.015*	0.000	*0.017*	-0.002	*0.025*
Constant	0.030	*0.016*	0.038	*0.017*	0.034	*0.020*
Adjusted R^2	0.179		0.419		0.251	
(B)						
Non-US Data	-0.004	*0.008*				
Year of Sample	0.002	*0.001*	0.002	*0.001*	0.000	*0.001*
Estimated by IV	0.007	*0.008*	0.005	*0.010*	0.009	*0.010*
Estimated by Twins	0.003	*0.011*	0.005	*0.012*	-0.009	*0.015*
Sample Size /1,000,000	0.005	*0.004*	0.003	*0.003*	0.510	*0.209*
Ability Controls	0.001	*0.008*	-0.040	*0.012*	0.027	*0.009*
Published	0.006	*0.009*	-0.000	*0.009*	0.028	*0.014*
Measurement Error	0.005	*0.013*	-0.000	*0.013*	-0.020	*0.022*
Standard Error	1.103	*0.215*	1.082	*0.225*	1.181	*0.313*
Constant	0.031	*0.014*	0.040	*0.013*	0.034	*0.017*
Adjusted R^2	0.364		0.652		0.416	
N	96		41		55	

[2] The criteria for study inclusion are not very stringent. The starting point was the list originally produced by David Card (1998), reported in Appendix Table A1 and the comprehensive review of Cohn and Addison (1997) where the details about the original data were provided. The deadline for entries in this study is June 1998.

The top half of Table 2 provides some regression results whereby the estimated return is related to a range of other variables that may influence the estimated return. The omitted category is important to note - here we use an unpublished return estimated via OLS without ability or measurement error corrections as our specification. The year variable is re-scaled so that 1974=0, so that the constant in the regression measures the rate of return in 1974. For results pooled across countries the omitted category is the US. The dependent variable is the level of the estimated return. The pooled results suggest little difference in the estimated returns by geographical region - countries in this non-US grouping include Finland, Honduras, Indonesia, Ireland, Netherlands, Portugal and the United Kingdom. Estimation methods have significant effects throughout with rates of return by IV and fixed effects/twins some 3% and 1.6% higher than the OLS default category. Sample size in itself has no effect on the estimated return but controlling for ability lowers the OLS estimate for the US studies in line with conventional wisdom, but raises the OLS estimate in the non-US studies. Controlling for measurement error has no significant effect on estimated rates of return, but published papers do tend to report higher rates of return (although this is only significant for the non-US studies). Looking at the results for the US we see that the more recent estimates are higher for the US studies in line with recent suggestions of a general shift upwards in the returns to schooling in the US. However this result is not apparent for the non-US studies, confirming the findings of Harmon and Walker (1995) of a relatively stable pattern of returns over time in the UK, which would be the largest grouping in the non-US block.

The bottom half of Table 2 incorporates all of the earlier elements but in addition controls explicitly for the standard error of the regression coefficient estimated for the rate of return in the model. The results here are rather startling. Unlike the top half of the table we no longer find any evidence of differences in the returns estimated by different estimation procedures, nor do we observe the pattern of higher estimates in published studies (although the upward drift of returns over time is still observed). This is a very important result, for in the absence of any bias in the reporting of results the estimates should not be correlated with the standard error. This leads us in the next section to a more formal consideration of the results in the context of reporting bias.

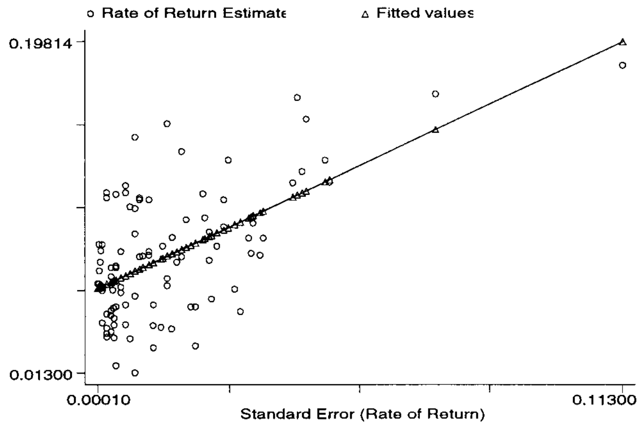
4. Model of Reporting Bias

One straightforward interpretation of the contrast between the result in panels A and B in Table 2 is that estimated payoffs that are significantly different from zero are more likely to be reported in journals and, since the twins studies and IV studies tend to have larger sampling errors in general, a less representative sample of these studies is typically reported.

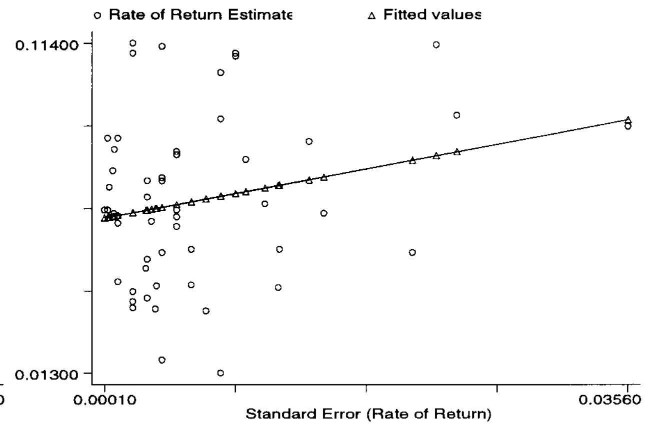
The graphs in Figure 1 examine this issue more closely and are related to the material presented in Table 2 where we related the return to schooling to the estimated standard error. These graphs show a plot of the estimated return against the standard error, together with the estimated regression line. In the absence of any selective reporting this line should be horizontal, as the return to schooling should not vary in proportion to its standard error. However if the tendency is to only report where the t-ratio is greater than 2 the estimated return will increase as the standard error increase in order to maintain the t-ratio at or above 2. Over all of the estimates in our meta-analysis, shown in panel (a) we find a positive slope which is significant (t = 7.11) but in the case of the OLS returns in panel (b) the slightly positive slope is not statistically significant. However the estimated returns in panel (c) and (d) for the IV and Twins estimates respectively show far steeper slopes which are statistically significant (with estimated t-ratios of 3.92 and 5.42 respectively).

Figure 1 Estimated Returns Plotted Against Estimated Standard Error

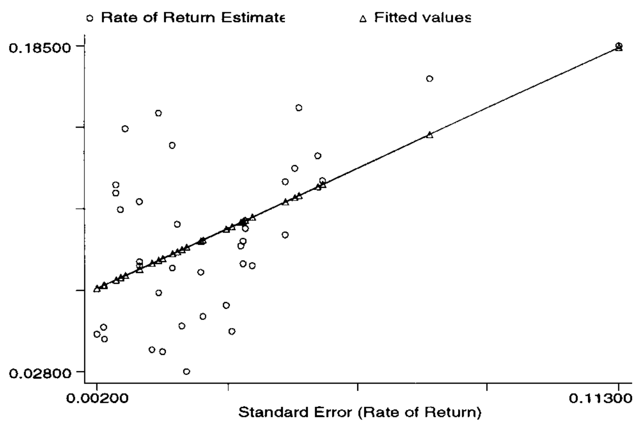
(a) All Estimates



(b) OLS Estimates



(c) IV Estimates



(d) Twins Estimates

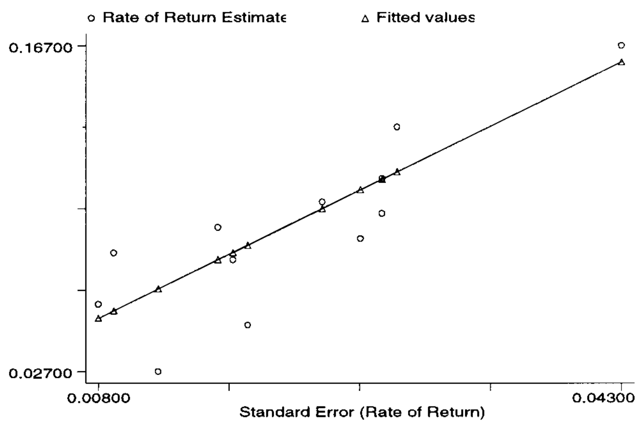
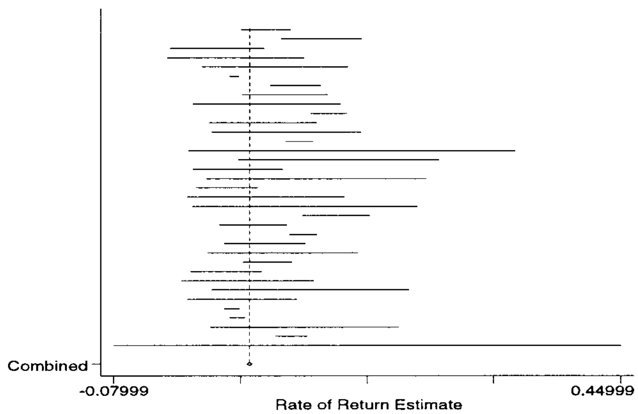
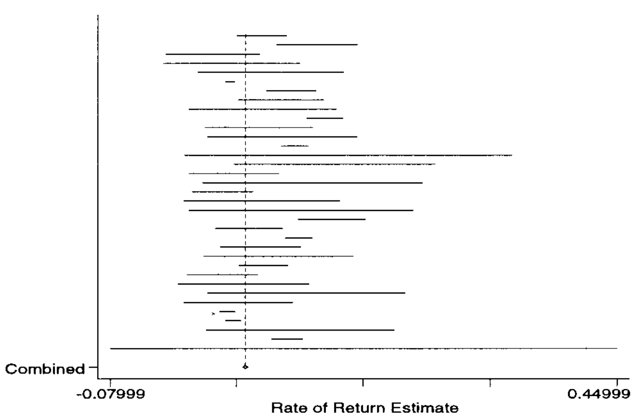


Figure 2 Predicted Interval Estimates of Returns Based on Random Effects Meta-Analysis

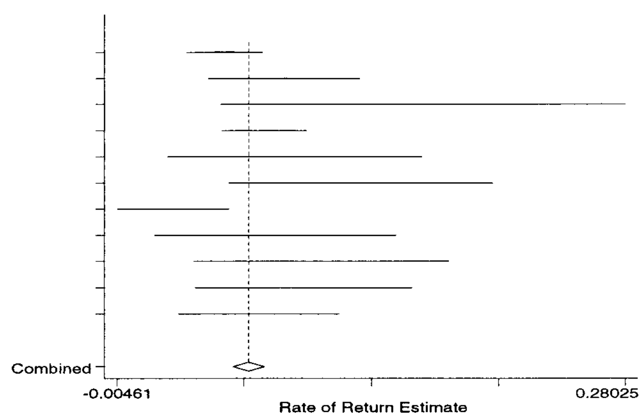
(a) OLS Estimates



(b) IV Estimates



(c) Twins Estimates



An alternative way to consider this problem is to estimate the study returns from the information collected via a simple meta regression. We might expect some degree of heterogeneity in the returns compared to some overall pooled 'average' return if what we are seeing in the reported studies is not exhibiting some degree of non-randomness. The graphs in Figure 2 represent pooled meta-analysis estimates of the returns based on the relationship between the estimate and the standard error of the estimate. Note that in the case of the OLS estimates the pooled estimate (represented in the graph by the vertical line) bypasses many of the individual interval estimates. On the contrary the results for the IV and Twins estimates suggest a clustering of the results around the pooled estimate in almost every instance.

Hedges (1992), in a review of over seven hundred studies on the effectiveness of aptitude tests as a predictor of employment outcomes, proposes a formal model of publication bias based on the assumption that there is a weight function (based on outcome p-values) that determines the probability a study is observed. Full details of this weight function are outlined in an appendix to this paper but the estimation procedure generates parameters that determine the increasing or decreasing probability of observing a study. We have specified different probabilities of observation of a study according to whether the p-value for that study is $0.01<p<0.05$ (denoted ω_2) or $p>0.05$ (denoted ω_3), relative to a default category of $0<p<0.01$. This default category's weight ω_1 is normalized to unity expressing the assumption that results with p-values in this bracket are reported with probability one. In the absence of reporting bias ω_2 and ω_3 should equal unity as well, indicating the equality of outcome probabilities when significance of results is accounted for. In addition to these ω_i parameters the overall pooled estimate for the return to schooling, denoted Δ, is provided based on the observed studies. Finally, the heterogeneity (measured by the standard deviation) in rates of return is estimated, and denoted σ.

Table 3 Hedges' publication bias model; all studies

All Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	1.273	0.497	--	--
ω_3	0.146	0.089	--	--
Δ ('true' rate of return)	0.068	0.004	0.073	0.003
σ	0.030	0.003	0.028	0.002
Log-Likelihood	288.67		279.73	
N	96		96	

OLS Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	1.692	1.426	--	--
ω_3	0.409	0.470	--	--
Δ ('true' rate of return)	0.064	0.004	0.065	0.004
σ	0.025	0.003	0.025	0.003
Log-Likelihood	158.19		157.46	
N	50		50	

IV Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	2.134	1.130	--	--
ω_3	0.265	0.214	--	--
Δ ('true' rate of return)	0.081	0.009	0.086	0.007
σ	0.034	0.006	0.032	0.005
Log-Likelihood	100.74		95.83	
N	35		35	

Table 3 presents the results for all studies together, for the studies using OLS and for the IV-studies, based on the modified Hedges' procedure. The first column gives the results of the full model, while the results in the second column are based on the model assuming no publication bias. The presence of publication bias is determined by examining the parameters ω_i

and testing the restriction of $\omega_2 = \omega_3 = 1$. Rejection of this restriction indicates the presence of publication bias. The test statistic is the difference of the log-likelihood values of the two models times 2. This statistic has a chi-square distribution with 2 degrees of freedom. Hence for the OLS studies equality of ω_2 and ω_3 to unity cannot be rejected, while for the IV studies equality of ω_2 and ω_3 to unity has to be rejected at the 1% level (test statistic equals 9.82 with a critical value of 9.21). Thus the IV studies appear to exhibit publication bias.

The parameter Δ in this context can be interpreted as the true mean effect corrected for publication bias. Correcting for this bias the IV estimated rate of return to a year of schooling is equal to 0.081, higher than the OLS rate of return of 0.064. There is evidence of considerable heterogeneity in estimated returns, with an estimated standard deviation of around .03. Notice also that the corrected IV return is somewhat below the uncorrected IV return (0.081 vs. 0.086). The point estimates for ω_2 exceed 1 so it is expected that studies with p-values below 0.01 have a lower probability to be observed than studies with a p-values between 0.01 and 0.05, but these estimates are never significantly higher than one. On the other hand, the probability of observing a study with a p-value numerically larger than 0.05 is much smaller than that of observing a p-value smaller than 0.01, and this difference is statistically significant (Hedges found a similar pattern).

The interpretation of Δ as the true mean effect corrected for publication bias (with only random heterogeneity) is a sensible interpretation in applications where it is indeed reasonable to expect that there is one uniform global effect. With medical interventions this might indeed be the case. When returns to schooling are considered, however, we have presented evidence that the returns vary across, for instance, countries and periods. A natural extension of Hedges' likelihood function is to parameterize Δ, thereby allowing the true return to schooling to vary with some of these characteristics. Thus the restriction, present in Table 3, that the estimate of the true rate of return is a constant is removed which allows us to show how the 'reporting-bias-corrected' return to schooling varies with the studies' characteristics.

In Table 4 we supplement the parameter Δ with interactions between it and the various regressors in Table 2, thus combining the OLS regressions in Table 2 with the results in Table 3.

Table 4 Extended Publication Bias Model: All Studies

All Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	1.598	0.655	--	--
ω_3	0.226	0.144	--	--
Δ_0	0.035	0.010	0.038	0.010
Δ_1(IV)	0.018	0.008	0.022	0.007
Δ_2(US)	0.013	0.007	0.013	0.006
Δ_3(Year)	0.002	0.001	0.001	0.0005
Δ_4(Twins)	0.009	0.012	0.012	0.011
Δ_5(Ability)	0.002	0.008	0.002	0.008
σ	0.026	0.003	0.025	0.002
Log-Likelihood	295.29		289.1	
N	96		96	

OLS Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	2.114	1.824	--	--
ω_3	0.558	0.658	--	--
Δ_0	0.033	0.011	0.033	0.011
Δ_1(IV)	--	--		
Δ_2(US)	0.016	0.007	0.015	0.007
Δ_3(Year)	0.002	0.0006	0.002	0.0006
Δ_4(Twins)	--	--		
Δ_5(Ability)	0.003	0.009	0.003	0.009
σ	0.023	0.003	0.023	0.002
Log-Likelihood	162.56		161.92	
N	50		50	

IV Studies	Unrestricted		Restricted ($\omega_2 = \omega_3 = 1$)	
Parameter	Coefficient	Standard error	Coefficient	Standard error
ω_2	2.052	1.091		
ω_3	0.251	0.204		
Δ_0	0.072	0.025	0.081	0.020
Δ_1(IV)				
Δ_2(US)	-0.005	0.018	-0.006	0.015
Δ_3(Year)	0.001	0.001	0.000	0.001
Δ_4(Twins)				
Δ_5(Ability)	-0.002	0.018	-0.001	0.015
σ	0.034	0.006	0.031	0.005
Log-Likelihood	101.03		96.07	
N	35		35	

Overall the results are fairly similar. Again by a likelihood ratio test of restricted versus unrestricted models, the IV studies appear to suffer from reporting bias whereas the OLS studies do not. The results in the top panel of Table 4 provide perhaps the most general summary of our analysis. The results in this panel reject the hypothesis that there is no publication bias. The benchmark estimates of the overall average return to schooling is 3.5% in 1974, but increasing at a rate of about 2 percentage points per decade. The estimated unexplained heterogeneity in rates of return, which is reduced with the use of the covariates in Table 4, has a standard deviation of around 2.6 percentage points. The IV and within-twins estimates of the return are 1.8 and 0.9 percentage points higher than the OLS estimates, and the difference between the IV and OLS estimates is statistically significant while the difference between the within-twins and OLS estimates are not. However, these "reporting-bias-corrected" differences in returns due to estimation method are much smaller than the uncorrected differences of 3.1 and 1.6 percentage points reported in Panel A of Table 2. Ability control have no effect on the estimated rates of return, but there remains evidence that returns have increased over time, at a rate of about 2 percentage points per decade.

5. Conclusion

There appears little controversy in the general principle underpinning the theory of schooling and earnings - schooling adds considerably to the earnings of individuals. What is at the centre of the debate is that in any context schooling is a choice variable and may not be independent of other factors that affect earnings. This raises the possibility that the observed correlation between schooling and earnings is not a causal relationship, but merely masks a correlation between other factors, such as ability, and earnings.

Studies of twins and other siblings and studies that use instrumental variables have been a major focus of research in the last decade in sophisticated attempts to measure the causal effect of schooling on earnings. Our survey of these studies suggests that, once the impact of the likelihood that a study result will be reported is controlled, there are relatively small differences among the estimates produced by the different estimation methods although some of these differences are statistically significant. Estimated rates of return to schooling appear to be higher in the U.S. than elsewhere, in part because of increased returns in the U.S. in the last two decades. However apart from this difference the estimates of the returns are considerably closer to each other than a simple glance at the range of estimates would provide. The evidence that schooling investments have a significant economic payoff is therefore very strong.

A number of future directions exist for this research. For many purposes it is often more useful to know the returns to specific types of schooling (by level and field) or the payoff to increased quality of schooling. It appears that the current methodology to estimate returns to years of schooling should be applied to these other topics as well. Likewise, studies of the returns to work-related training (firm training) should be subject to similar analyses. These, and related empirical studies of human capital investments, are essential to making wise public and private choices.

References

Angrist, J.D. and Krueger, A.B. (1991) "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" Quarterly Journal of Economics, 106, pp. 979-1014.

_____. (1995) "Split Sample Instrumental Variable Estimates of the Returns to Schooling." *Journal of Business and Economic Studies*, 13(2) pp.225-235.

Angrist, J.D. and Newey, W.K. (1991). "Over-Identification Tests in Earnings Functions with Fixed Effects." *Journal of Business and Economic Statistics*, 9, pp.317-323.

Ashenfelter, O. and Rouse, C. (1998). "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics*, 113, pp. 253-84.

Bedi, A. and Gaston, N. (1999). "Using Variation in Schooling Availability to Estimate Educational Returns for Honduras." *Economics of Education Review*, 18(1), pp. 107-16.

Blackburn, M. and Neumark, D. (1993). "Omitted Ability Bias and the Increase in the Return to Schooling." *Journal of Labor Economics*, 11, pp.521-543.

_____. (1995). "Are OLS Estimates of the Return to Schooling Biased Downward? Another Look." *Review of Economics and Statistics*, 77, pp.217-229.

Blanchflower, D. and Elias, P. (1993). "Ability, Schooling and Earnings: Are Twins Different?" Mimeo, Dartmouth College, New Hampshire.

Butcher, K.F. and Case, A. (1994). "The Effect of Sibling Composition on Women's Education and Earnings." *Quarterly Journal of Economics*, 109, pp. 531-563.

Brown, S., Goetzmann, W. and Ross, S. (1995). "Survival." *Journal of Finance*, L, pp. 853-873.

Card, D. (1993) "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." National Bureau of Economic Research (Cambridge MA.) Paper No. 4483.

_____. (1998) "The Causal Effect of Education." Mimeo, Center for Labor Economics, University of California at Berkeley.

Card, D., and Krueger, A. (1995). "Time-Series Minimum-Wage Studies: A Meta-analysis, *American Economic Review*, 85, pp. 238-243.

Cohn, E. and Addison, J.T. (1997), "The Economic Returns to Lifelong Earnings." Mimeo (College of Business Administration, University of South Carolina, South Carolina, USA).

Conneely, K. and Uusitalo, R. (1998). "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model." Mimeo, Princeton University.

Dearden, L. (1995). "The Returns to Education and Training for the United Kingdom." Unpublished Ph.D. Dissertation, University College London.

_____. (1997). Ability Families Education and Earnings in Britain. Mimeo. (Institute for Fiscal Studies, London, UK).

Duflo, E. (1998). "Evaluating the Schooling and Labor Market Consequences of a School Construction Program." Mimeo, MIT.

Egger and Smith, G.D. (1997). "Meta-Analysis: Potentials and Problems." *British Medical Journal*, Vol. 315, pp. 1371-1374.

Griliches, Z. (1977) "Estimating the Return to Schooling: Some Econometric Problems." *Econometrica*, 45(1), pp.1-22.

Griliches, Z. and Mason, W. (1972) "Education, Income and Ability." *Journal of Political Economy*, 80(2), pp. 74-103.

Hanson, J. and Wahlberg, R. (1998). "Endogenous Schooling and the Distribution of the Gender Wage Gap." Mimeo, Concordia University.

Harmon, C. and Walker, I. (1995). "Estimates of the Economic Return to Schooling for the UK, *American Economic Review*, 85, pp. 1279-1286.

_____. (1999). "The Marginal and Average Effect of Schooling." *European Economic Review*, 43, pp. 879-887.

_____. (1999). "The Returns to the Quantity and Quality of Education – Evidence for Men from England and Wales." *Economica*, forthcoming.

Hedges, Larry V. (1992). "Modeling Publication Selection Effects in Meta-Analysis." *Statistical Science*, Vol. 7, pp. 246-255.

Hunter, John E. and Schmidt, Frank L. (1990). Methods of Meta-Analysis – Correcting Error and Bias in Research Findings. London: Sage.

Huque, M.F. (1988). "Experiences with Meta-Analysis in NDA Submissions." *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, vol. 2, pp. 28-33.

Isaacson, G. (1997) "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins." Mimeo. (Centre for Research on Transportation and Society, Borlange, Sweden).

Lo, A. and MacKinley, C. (1990). "Data Snooping Biases in Tests of Financial Asset Pricing Models." *Review of Financial Studies*, 3, pp. 431-467.

Meghir, C and Palme, M. (1999). "Assessing the Effect of Schooling on Earnings Using a Social Experiment." Mimeo, Institute for Fiscal Studies, London.

Miller, P., Martin, N. and Mulvey, C. (1994) "The Effects of Education on Income: Estimates from a Sample of Australian Twins." Mimeo, University of Western Australia.

Mincer, J. Schooling, Experience and Earnings. (1974) New York: National Bureau of Economic Research.

Plug, E. (1997). "Season of Birth, Schooling and Earnings." Mimeo, University of Amsterdam.

Psacharopoulos, G. (1994). "Returns to Investment in Education: A Global Update." *World Development*, 22, pp. 1325-1343.

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

Uusitalo, R. (1997) "Returns to Education in Finland." Mimeo. (University of Helsinki, Helsinki).

Viera, J. (1997) Returns to Education in Portugal: OLS, IV and Selection Model Estimates. Mimeo. (University of the Azores, Azores).

APPENDIX – A GENERALIZED PUBLICATION BIAS MODEL

Hedges (1992) model of publication bias builds on the assumption that there is a weight function that determines the probability that a study is observed. He posits that the weight function depends on the p-value, whereby studies with a lower p-value are more likely to be observed. It is assumed that a random effects model generates the observed data. More precisely the observed data, $X_1,\ldots,X_n$ are such that

$$X_i \sim N(\delta, \sigma_i^2), \tag{A1}$$

where σ_i^2 is known and δ is an unknown parameter distributed as

$$\delta \sim N(\Delta, \sigma^2). \tag{A2}$$

Hence

$$X_i \sim N(\Delta, \eta_i) \tag{A3}$$

where

$\sigma_i^2 + \sigma^2 = \eta_i$. The observation associated with each study forms part of the weight function $w(X_i)$ which determines the probability of being observed, with the relationship with X_i coming via the p-value. In Hedges' formulation the weight function is a step function with the steps at points determined a priori. In our application we distinguish three steps: $0<p<0.01$, $0.01<p<0.05$

and p>0.05[3]. Given this data generating process and the weight function, Hedges derives the joint log-likelihood for the data **X**, which has the following form (Hedges 1992, p.250) for i observations over j steps in the weight function,

$$L = c + \sum_{i=1}^{n} \log w_i (X_i, \omega) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{X_i - \Delta}{\eta_i} \right)^2 - \sum_{i=1}^{n} \log(\eta_i) - \sum_{i=1}^{n} \log \left[\sum_{j=1}^{k} \omega_j B_{ij} (\Delta, \sigma) \right], \qquad \text{(A4)}$$

where $B_{ij}(\Delta, \sigma)$ is the probability that a normally distributed random variable with mean Δ and variance η_i will be assigned weight value ω_i.

The parameter Δ can be interpreted as the true effect corrected for publication bias. This is a sensible interpretation in applications where it is indeed reasonable to expect that there is one uniform global effect. With medical interventions this might indeed be the case. When returns to schooling are considered however, we have presented evidence that the returns vary between for instance countries and periods. A natural extension of Hedges' likelihood function is therefore to parameterize Δ thereby allowing the true return to schooling to vary with some of these characteristics. The extended likelihood function now reads:

$$L = c + \sum_{i=1}^{n} \log w_i (X_i, \omega) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{X_i - Z_i \Delta}{\eta_i} \right)^2 - \sum_{i=1}^{n} \log(\eta_i) - \sum_{i=1}^{n} \log \left[\sum_{j=1}^{k} \omega_j B_{ij} (Z_i \Delta, \sigma) \right], \qquad \text{(A5)}$$

where Z_i is a vector of characteristics of study i and Δ is (now) a vector of parameters to be estimated. In our application the vector Z includes four dummies equal to 1 if the study uses IV,

[3] Due to data limitations it is impossible to distinguish the additional step of $0.05<p<0.10$.

twins data, relates to the US and when an ability measure is included, as well as year to which the study relates.

Appendix Table A1 Meta-Analysis – Sources

STUDY	YEAR	COUNTRY
Angrist and Krueger	1991a	USA
Angrist and Krueger	1991b	USA
Angrist and Krueger	1995	USA
Angrist and Newey	1991	USA
Ashenfelter and Rouse	1997	USA
Bedi and Gaston	1998	HONDURAS
Blanchflower and Elias	1993	UK
Blackburn and Neumark	1993	USA
Blackburn and Neumark	1995	USA
Butcher and Case	1994	USA
Card	1993	USA
Card	1998	USA
Conneely and Uusitalo	1998	FINLAND
Dearden	1995	UK
Dearden	1997	UK
Duflo	1998	INDONESIA
Hansen and Wahlberg	1998	SWEDEN
Harmon and Walker	1995	UK
Harmon and Walker	1999	UK
Harmon and Walker	1999	UK
Isaacsson	1999	SWEDEN
Meghir and Palme	1997	SWEDEN
Miller, Martin & Mulvey	1995	AUSTRALIA
Plug	1997	NETHERLANDS
Rouse	1997	USA
Uusitalo	1997	FINLAND
Viera	1997	PORTUGAL