

EDUCATION POLICY REFORM AND THE RETURN TO SCHOOLING FROM INSTRUMENTAL VARIABLES

Kevin Denny
Colm Harmon

Education Policy Reform and the Return to Schooling from Instrumental Variables *

KEVIN J. DENNY

University College Dublin & Institute for Fiscal Studies, London

COLM P. HARMON

University College Dublin, University of Warwick & CEPR, London

Version 2.10, April 7th 2000

Abstract:

This paper exploits an unusual policy reform that had the effect of reducing the direct cost of schooling in Ireland in the late 1960's. This gave rise to an increased level of schooling but with effects that vary substantially across family background. This interaction of educational reform and family background generates a set of instrumental variables that are used to estimate the return to schooling allowing for the endogeneity of schooling. Using a standard Mincer type model we find a large and well-determined rate of return of around 12% which are substantially higher than the OLS estimates of around 7%.

* Material from the Survey of Income Distribution, Poverty and Usage of State Services Survey is used with permission of the Economic and Social Research Institute in Dublin. The providers bear no responsibility for the calculations contained herein nor for any interpretation made by the authors. Our thanks to Ian Walker for encouraging us to write this paper, to Gauthier Lanot for useful suggestions on the econometrics, to Karen Conneely and seminar participants in Athens, IFS and NUI Maynooth for comments and to Sandra Redmond for research assistance. The usual disclaimer applies. The research was part funded by the EU Commission under the auspices of the TSER programme (contract SOE2-CT98-2044) and by the award of a Marie Curie fellowship to Colm Harmon that facilitated the completion of this paper. Corresponding author: Dr. Colm Harmon, Department of Economics, University College Dublin, Belfield, Dublin 4, Ireland. Phone (+353 1) 7068318. Fax (+353 1) 2830068. E-mail colm.harmon@ucd.ie.

I. Introduction

The rate of return to schooling is an important parameter in labour economics and having a good estimate is clearly important for public policy as well as for individuals. In recent years considerable research effort has been invested in refining the estimation techniques used to estimate this parameter. This is due to a variety of sources of bias associated with OLS estimates of the return to schooling making recovery of a consistent estimate of the return to schooling problematic.

Recent research on the returns to schooling (along with other areas of labor research such as the impact of training) has drawn analogies with 'treatment' and 'control' group concepts in the medical/psychology literature. Since random assignment of schooling is not possible, the literature on returns to schooling is increasingly focused on finding natural experiments that exogenously assign individuals to different treatments, thus creating 'natural' variation in data. For example Angrist and Krueger (1991) explore how an individual's season of birth may imply that some students reach school leaving age after fewer months of compulsory education than others, allowing for the creation of suitable instruments to exploit in an Instrumental Variables (IV) approach. Harmon and Walker (1997) use the change in compulsory schooling law, which raised the minimum schooling age in Britain to generate an exogenous change in education. In both of these approaches the key variable will affect the education decisions of a subset of the population, those who leave school as soon as they can, so one interpretation of these results is that the IV estimates identify the rate of return to the marginal or "treated" group only. As argued by Card (1999) "IV estimation based on an intervention that affects a narrow sub-group may lead to an estimated return to schooling above or below an OLS estimator for the same sample".

A somewhat different approach is used in the paper by Duflo (1999). In that paper estimation is based on the exposure of individuals to a massive investment program in education in Indonesia in the early 1970's. Individuals were assigned to the treatment on the basis of their date of birth (pre and post reform) and the district they lived in (as investment was a function of local level needs assessment). Meghir and Palme (1999) pursue a similar strategy in their analysis of reforms in Sweden in the 1950's that was intended to extend the schooling level nationally. This was piloted in a number of school districts prior to its adoption nationally and it is from this pre-trial experiment that the variation in attainment comes. Both these papers rely on large-scale reforms or "natural experiments" whose effect differed across individuals. This paper follows a similar approach in looking at a fundamental change in the educational system in 1960's Ireland, which not only affected the entire population of school going individuals but in a way which differed across socio-economic backgrounds. This allows one to generate Instrumental Variables that permit consistent estimates of the return to schooling.

II. Reforms as Instruments and the Identification Strategy

The model is given by the following standard two-equation system describing log earnings, (y_i), and years of schooling, (S_i):

$$(1) \quad y_i = \mathbf{X}'_i \boldsymbol{\delta} + \beta S_i + u_i$$

$$(2) \quad S_i = \mathbf{Z}'_i \boldsymbol{\alpha} + v_i$$

where \mathbf{X} , \mathbf{Z} are a vectors of observed attributes and β is interpreted as the return to schooling. Estimation of equation (1) by OLS will yield an unbiased estimate of β only if the S_i is exogenous that is $E(\mathbf{X}_i u_i) = 0$. If this is not the case alternative estimation methods have to be used.

This paper uses an exogenous change in the educational distribution of individuals caused by a policy innovation in Ireland in the late 1960's whereby secondary schooling was made free for all school-age youths. The fee-paying aspect to secondary education prevailing prior to reform was a major hurdle for families. Annual fees per pupil at the time of introduction of the policy were approximately two weeks wages for the average manual worker. Taking into account the large family sizes that prevailed at that time it is entirely feasible for secondary school fees to represent up to one-sixth of total household income. These two reforms had a significant effect on the participation rate in education. Archer (1998) notes that contemporaneous accounts of the policy change recorded a 33% increase in the number of children participating in a school transport scheme. In an econometric evaluation analysis of time series data on participation rates, Tussing (1978) found increases in enrolment for age groups that were significantly in excess of the trend in participation rates, which had been increasing in the period prior to the policy announcement. The long run increase in school participation due to the reform, abstracting from other trends present in the data, has been estimated to be about 20%.

What is most important for present purposes is not the aggregate change in education but the distribution effect. Prior to the reform those that received secondary (and by implication third level) education came from a wealthier socio-economic background. Thus the elimination of fees for secondary schooling had a differential effect, with larger increases in participation for those from less well off backgrounds. It is this interaction that we exploit to generate exogenous variation in schooling. In particular a dummy variable ("No fees") is defined for individuals who were born after 1955 and hence faced a regime of no fees at secondary school. This dummy variable is then interacted with the provided information on socio-economic background of parents, a six-point classification based on the occupation of

the main earner, ranging from Professional (the omitted category) through to Unskilled Manual.

The dependent variable is the gross hourly wage rate. We estimate a standard "Mincer" earnings function using years of full time education. As with other studies we include a quadratic in age to proxy for experience given the possible endogeneity of labour market experience. We also include the participation rate in education when the individual was 15 to proxy for trends in participation and to eliminate the potential interpretation of the free fees dummy as a cohort effect alone. The direct effects of family background are included in the earnings equation since they may be correlated with unobserved characteristics such as motivation; see Rischall (1999). Card (1999) has argued that family background may not be a valid instrument even if it is not directly a determinant of earnings because it may not completely absorb the effect of omitted measures of ability. In an application to Finnish data, Conneely and Uusitalo (1999) experiment with family background as an instrumental variable but reject the hypothesis that it is uncorrelated with the error term in the earnings equation. The remaining variables in the earnings function are a dummy variable for those resident in urban areas, for union membership and for married status.

III. Data and Results

The data used in this study is a sample of employed males from a household survey conducted by the Economic and Social Research Institute in 1987 (hereafter ESRI87)¹. A total of 3,300 households were interviewed, generating information on over 6,500 adults. The total number

¹ More recent data at the individual level is not publicly available. In addition the problems of sample selection for women have been found in this dataset (Callan and Wren, 1994) so we focus in this paper on the male sample.

of male employees aged 18-64 in year of interview for whom all the necessary information is available is 1,158. More complete information on this data can be found in Callan and Harmon (1999). Descriptive statistics are provided in Table 3 at the end of the paper.

Table 1 presents the results from the estimation of equation (1) and (2). The first column gives the OLS estimates of the earnings equation with an estimated rate of return equal to 7.9%, an estimate that is consistent with the existing literature. Returns to experience are 12.2%. Changes in participation rates have no direct effect on earnings but there is a large premium to urban residence of around 12% (which may simply reflect a higher cost of living). Being married and being in a trade union also capture large premia – almost 20% each. Finally parental class is jointly significant.

The second and third set of estimates are the schooling and earnings equation respectively for the IV estimates. What is notable here is the importance of parent's socio-economic background and how it changes with the introduction of the reform. The coefficients on parental class show that those from poorer (manual or 'blue collar') backgrounds have lower education *ceteris paribus* than those from non-manual (or 'white-collar') backgrounds, in keeping with a wide body of other, mostly sociological, research. The policy reform has the effect of reducing these socio-economic penalties, approximately by a half. For example whereas an individual from an unskilled manual background has 3.7 years *less* education than a professional background individual, after the reform this penalty is reduced to about 2 years. This effect is consistent with studies of the reform discussed earlier such as Tussing (1978).

Table 1: Estimated Schooling and Earnings Functions

	OLS		OLS		IV	
	Earnings		Schooling		Earnings	
	Co-eff.	Std. Err.	Co-eff.	Std. Err.	Co-eff.	Std. Err.
Years of Schooling	.0793	.0057			.1360	.0251
Age	.1224	.0125	.1928	.0699	.1155	.0127
Age ²	-.0011	.0001	-.0016	.00061	-.0010	.0001
Trend in Secondary Participation	.0128	.0024	.0654	.0203	.0114	.00247
Resident in urban area	.1177	.0228	.0617	.1281	.1178	.02357
Married	.1950	.0335	-.2072	.1875	.2057	.03483
Union member	.1916	.0239	-.1529	.1333	.2032	.02506
Parental Class 2 (= Admin/Clerical)	-.0191	.0641	-.0393	.3747	.0028	.0579
Parental Class 3 (= Other Non-Manual)	-.0867	.0572	-.8463	.3258	-.0280	.0552
Parental Class 4 (= Skilled Manual)	-.1338	.0537	-2.6109	.2995	-.0099	.0706
Parental Class 5 (= Semi-Skilled Manual)	-.1239	.0569	-3.1785	.3271	.0277	.0829
Parental Class 6 (= Unskilled Manual)	-.1478	.0576	-3.6738	.3071	.0287	.0908
No Fees (= 1 if born after 1955)	--	--	-1.8225	.5949	--	--
Parental Class 2 * No Fees	--	--	-1.1727	.6633	--	--
Parental Class 3 * No Fees	--	--	-0.3128	.5497	--	--
Parental Class 4 * No Fees	--	--	1.2414	.5059	--	--
Parental Class 5 * No Fees	--	--	1.4783	.5528	--	--
Parental Class 6 * No Fees	--	--	1.6750	.5306	--	--
Constant	-3.425	.4046	6.3923	2.5097	-3.5694	.4702
N	1158	1158	1158	1158	1158	1158
$\overline{R^2}$	0.5141		0.2247		0.4554	
Over-identification: $\chi^2_{(5)}$ (p-value)					1.989(5) (.8507)	
Exogeneity: F test (1, 1144 d.f.'s) (p-value)					5.97 (.0147)	

What is striking in these results is the large and well determined increase in the return to schooling in the IV earnings equation in the marginal return to schooling. The estimated return to education is close to 13%. This is at the high end of estimated returns though not as

high as the 15% estimated by Harmon and Walker (1995) who also use a "natural experiment" with British data².

Using the test proposed by Davidson and McKinnon (1993) we are unable to reject the over-identifying restrictions. The instrument set used is jointly significant in the schooling equation in keeping with the informal test suggested by Bound *et al.* (1995). We also report a test for endogeneity of schooling, which allows us to reject the null that the OLS estimates are consistent³.

Table 2 reports the key results from a number of alternative specifications. Since marital status and union membership are potentially endogenous and there are no obvious instruments, column 1 shows the implications of dropping both from the system. As before the IV estimates of schooling returns are about 5% higher with IV than OLS. Column 2 restores the marital status dummy while still excluding union membership but this has virtually no impact on the parameters of interest. So far the estimates have included the direct or "main effects" of parental background as covariates in the earnings equation since they may be correlated with unobservables such as motivation and it is the *interaction* of the reform (as well as the reform itself) that has generated the Instruments. These direct effects are not jointly significant in the earnings equation (Table 1) so column 3 in table 2 excludes them from the earnings equation and hence treats them as *additional* instruments. The gap between the OLS and IV rate of return is narrowed because the OLS estimate rises to almost 9%. The over-

² There are other explanations for the IV estimate being higher than the OLS one. Ashenfelter *et al.* (1999) presents an alternative explanation for the dominance of IV results that are higher than OLS. By a meta-analysis of some 100 estimated rates of return they find that the average premium of around 3% of IV over OLS may be partly (1.8%) explained by selective reporting of results by researchers.

³ This test, see Davidson & McKinnon (1993) pp237-240 is similar to the familiar (Durbin-Wu) Hausman test but has the advantage that it always yield a computable test statistic whereas the former may not since it requires the difference in the estimated covariance matrices to be positive definite.

identifying restrictions continue not to be rejected. Column 4, rather than relying on parents socio-economic background, uses dummy variables for mothers education as a set of covariates in the earnings equations, on the basis that these may be a better proxy for motivation (or whatever mechanism that individuals are influenced by their families) while using the main effects of socio-economic background as well as the interactions with the reforms as instruments. The impact of using IV is still a large increase in the rate of return.

Table 2 : Some alternative specifications

	1	2	3	4	5	6
	Without marital status or union	Without trade union	Parental background as instruments	Mother's education as covariates	Participation rate as extra instrument	IALS 1994
OLS rate of return	.0766	.0774	.0878	.0818	.0805	.0898
s.e.	(.0059)	(.0058)	(.0049)	(.0056)	(.0057)	(.0126)
IV rate of return	.1274	.1256	.1251	.1304	.1648	.1680
s.e.	(.0256)	(.0251)	(.0101)	(.0148)	(.0259)	(.1052)
N	1158	1158	1158	1158	1158	518
Over-identification : χ^2 (d.f.) (p-value)	2.46 (5) .7825	1.81(5) .875	4.37(10) .929	4.65(10) .9133	20.544(6) .0022	
Exogeneity :F test d.f. (p value)	4.47 1,1146 .0346	4.14 1,1146 .0421	19.29 1,1149 .0000	14.37 1,1141 .0000	13.76 1,1145 .0000	

Throughout the analysis so far the aggregate rate of secondary school participation (based on that prevailing when the individual was 12) has been used as a covariate (that is, it is in both schooling and earnings equations). This is to ensure that our reform interactions are not simply reflecting aggregate trends in participation. Treating the participation rate as an additional instrument (relative to the specification in Table 1) however leads to an immediate

rejection of the over-identifying restrictions as shown by the χ^2 test in column 5 above. This illustrates how careful specification testing is required when using IV.

Finally specification (6) updates the analysis based on a more recent alternative data source – the International Adult Literacy Survey (IALS) of 1994. The earnings data from this survey is banded instead of being continuous so one cannot directly use IV. However it is possible to control for the endogeneity of schooling in the earnings equation in a similar two-step method. One estimates a reduced form earnings equation by maximum likelihood with a corresponding schooling equation estimated by OLS. Hence one can compute the implied rate of return to schooling by Minimum Distance methods with bootstrapped standard errors (based on 100 replications). This "IV" return is considerably less precise than in the rest of the table 2 and table 1. Nonetheless controlling for endogeneity of schooling with comparable specifications of the two equation system leads to a large increase in the rate of return from 9% to almost 17%⁴. Generalising across the results in table 2 one finds that despite the different data specifications and sources the principal finding remains the same – OLS returns in Ireland are somewhat higher than normal for industrialised regions and controlling for the endogeneity generates a substantial large increase in the rate of return.

Much of the recent literature on schooling returns has emphasised the possibility of heterogeneous returns typically using a variant of a model due to Gary Becker (e.g. Conneely & Uusitalo (1999), Kling (1999)). One of the appealing aspects of this approach is that as Card shows, with heterogeneous returns and a binary treatment the IV return can be written as a weighted average of the returns of the sub groups. One advantage of the standard IV

⁴ This estimate is taken from Denny *et al*(2000).

estimator with a binary treatment is that in a random coefficients framework given certain conditions, it may identify the Average Causal Response to the treatment (Angrist & Imbens (1995)). However a necessary condition is monotonicity, that in this case the reform does not decrease educational attainment for any individual. As the schooling equation shows this does *not* hold since those from better off backgrounds actually had lower attainments than before as a result of the reform. It appears that those from better off background were partially "crowded out" by the higher attainment of those at the bottom end of the socio-economic distribution. Given the presence of binding supply side constraints it is not surprising that a fall in relative prices to one sub-group should have this impact on the others. Equally it may reflect not so much a physical crowding out as the possibility that higher participation by the less well off reduces the premium to the better off. The results here are similar in spirit to that of Card (1995) and subsequently Kling (1999) who use an interaction between family background and proximity to college as an instrumental variable with bigger effects on schooling for those from poorer backgrounds.

IV. Conclusion

In a standard model of education and earnings we exploit an unusual policy reform, which had the effect of reducing the direct cost of schooling. This gave rise to an increased aggregate level of schooling but with effects that vary across family background. This interaction generates a set of instrumental variables that we use to estimate the return to schooling allowing for the endogeneity of schooling. We control for any direct effects that family background may have on earnings caused by correlation with unobserved ability or motivation. Crucially, we also control for aggregate changes in school participation to eliminate pure cohort effects. The results show a large and well-determined increase in the rate

of return of the order of 12% to 13%, substantially higher than the OLS estimates of around 7%.

References

- Angrist, Joshua D. and Imbens, Guido (1995). "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." Journal of the American Statistical Association, 90, pp. 431-442.
- Angrist, Joshua D. and Krueger, Alan B. (1991) "Does Compulsory Schooling Attendance Affect Schooling and Earnings?" Quarterly Journal of Economics, November, 106(4), pp. 979-1014.
- Archer, Peter. (1988). Inequality in Schooling in Ireland. Dublin: CMRS.
- Ashenfelter, Orley, Harmon, Colm and Oosterbeek, H. (1999). "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias." Labour Economics, Vol. 6 (4), November, pp. 453-470.
- Bound, J., Jaeger, D.A., Baker, R. (1995). Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak. Journal of the American Statistical Association, 90, 443-50.
- Callan, Tim and Colm Harmon (1999), "The Economic Return to Schooling in Ireland", Labour Economics, December 1999.
- Callan, Tim and Anne Wren (1994), Male-Female Wage Differentials: Analysis and Policy Issues. Dublin: Economic and Social Research Institute.
- Card, D. (1995). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In Christofides, Louis; Grant, E. Kenneth; Swidinsky, Robert (eds.) Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. Toronto; Buffalo and London: University of Toronto Press, 1995, pages 201-22.
- _____. (1999) "The Causal Effect of Education on Earnings." In Ashenfelter, O. and Card, D. (eds) Handbook of Labor Economics Volume 3A. Amsterdam: Elsevier.
- Conneely, K. and Uusitalo, R. (1999). "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model." Mimeo, Princeton University.
- Davidson, Russell and McKinnon, James (1993). Estimation and Inference in Econometrics. Oxford: Oxford University Press.
- Denny K.J., C.P. Harmon and S.R. Redmond (2000) "Functional Literacy, Educational Attainment and Earnings: Evidence from the International Adult Literacy Survey" (*mimeo* Department of Economics, UCD)
- Duflo, Esther (1999). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *Mimeo*, Department of Economics, MIT.
- Harmon, C. and Walker, I. (1995). "Estimates of the Economic Return to Schooling for the UK, American Economic Review, 85, pp. 1279-1286.

- Kling, Jeffrey (1999). "Interpreting Instrumental Variables Estimates of the Returns to Schooling." Industrial Relations Section, Princeton University Working Paper 415, January.
- Lang, K. (1993). "Ability Bias, Discount Rate Bias and the Return to Education." *Mimeo*, Boston University, Boston.
- Meghir, Costas and Palme, Marten (1999). "Assessing the Effect of Schooling on Earnings Using a Social Experiment." Institute for Fiscal Studies Working Paper 99/10, November.
- Rischall, Isaac (1999). "The Roles of Education, Skill and Parental Income in Determining Wages." *Mimeo*, McMaster University.
- Tussing, Dale (1978). "Irish Educational Expenditures - Past, Present and Future." E.S.R.I. General Research Series no. 92.

Table 3: Descriptive Statistics

	All		Before		After	
	<i>Mean</i>	<i>Std.Dev.</i>	<i>Mean</i>	<i>Std.Dev.</i>	<i>Mean</i>	<i>Std.Dev.</i>
Wage (1987 £ per hour)	5.507	3.344	6.383	3.506	4.110	2.502
Log Wage	1.562	0.535	1.736	0.475	1.284	0.508
Age (years)	37.15	12.70	45.15	9.201	24.39	4.263
Schooling (years)	11.55	2.43	11.395	2.771	11.796	1.725
Married	0.661	0.473	0.893	0.309	0.291	0.455
Union member	0.531	0.499	0.624	0.485	0.383	0.487
Participation rate	0.491	0.235	0.320	.101	0.763	.082
Parental Class 1 (Professional)	.0907	.2873	.100	.300	.076	.266
Parental Class 2 (Admin/Clerical)	.0769	.2665	.086	.280	.063	.243
Parental Class 3 (Other Non-Manual)	.1623	.3689	.160	.370	.166	.372
Parental Class 4 (Skilled Manual)	.2945	.4560	.270	.444	.334	.472
Parental Class 5 (Semi-Manual)	.1598	.3665	.159	.366	.161	.368
Parental Class 6 (Unskilled Manual)	.2159	.4116	.226	.419	.199	.400
N	1158		712		446	

"**Before**" means born in or before 1955 and "**After**" otherwise.