Applied Economics Letters, 2003, **10**, 281–286

Skill, parental income, and IV estimation of the returns to schooling

ROBERT J. LEMKE* and ISAAC C. RISCHALL‡

Department of Economics and Business, Lake Forest College, 555 N. Sheridan Road, Lake Forest, IL 60045, USA and ‡Decision Science, Citigroup, 7th Floor – Wingren, 290 Carpenter Fwy, Irving, TX 75062, USA

Recently, attention has moved away from using parental background variables, such as parental education, in favour of using institutional features of the education system as instruments when estimating the return to schooling. In this paper, these possible instruments are revisited. Using the National Longitudinal Survey of Youth, several specifications of the wage equation are estimated and three types of instruments used – parental education, quarter of birth, and college proximity. It is shown that under some specifications – in particular, by including parental income and individual skill in the wage equation – parental education appears to be a valid and useful instrument. On the other hand, when using the institutional instruments, the weak correlation between the instruments and years of schooling produces imprecise and likely biased estimates of the return to schooling.

I. INTRODUCTION

Estimating the returns to schooling with instrumental variables is somewhat at a crossroads. Parental education has fallen out of favour as an instrument for the child's education, because of its correlation with the child's wage. In its place, institutional features of the education system – such as mandatory school attendance and living close to a college or university when in high school – have been used as instruments, but these suffer from being weakly correlated with the child's education. Additionally, all of these produce estimates above the ordinary least squares (OLS) estimates, which runs counter to the selection story in which more highly motivated people obtain more education and earn higher wages.

We contend that much can be learned by using a single data set containing all of these instruments and by broadening the wage specification to include skill and parental income. Using the National Longitudinal Survey of Youth, four findings are demonstrated. First, parental education is uncorrelated with the wage after controlling for skill and parental income. Second, when the wage equation includes skill and parental income, instrumenting with parental education produces lower instrumental variable (IV) estimates than OLS. Third, the institutional instruments produce imprecise estimates that likely contain at least as much bias as OLS estimates. And fourth, skill and parental income are predicted to positively affect the wage when instrumenting with parental education but to have no predictive power when using institutional instruments. These findings are viewed as lending support for needing to include skill and parental income in the wage equation and for using parental education as an instrument.

II. ESTIMATING THE RETURNS TO SCHOOLING

The wage equation specifies wages, w, as a function of their personal characteristics, X, years of schooling, s, and a random component, ε, so that $\ln(w) = \beta X + \alpha s + \varepsilon$. The central task is then to estimate α. Problems arise, however, because schooling is not exogenous. Rather, $s = \gamma Z + \nu$,

*Corresponding author: E-mail: lemke@lfc.edu

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/1350485032000078653

where Z is a set of personal characteristics and ν is an idiosyncratic term. If ε and ν are uncorrelated, then OLS produces unbiased and consistent estimates of α. If ε and ν are correlated, but some items in Z are not correlated with ε (and therefore are not in X), then IV produces unbiased and consistent estimates of α, where the elements in Z not in X serve as instruments for schooling.

Willis and Rosen (1979) remains a leading paper in estimating the returns to schooling by meticulously capturing the school selection issue and instrumenting for schooling with parental schooling. Ashenfelter *et al.* (1999) and Card (1999) summarize many other empirical studies that instrument for schooling similarly. Whereas OLS estimates of the returns to schooling are typically in the 5 to 10% range, IV estimates typically range from 5 to 15%.

Instrumenting with parental education, however, has been questioned on the grounds that it directly affects the wage or is at least correlated with ε. In the pursuit of new instruments, attention has turned to institutional features of the schooling system that affect the schooling decision while not affecting wages. Angrist and Krueger (1991) focus on school attendance policies, noting that people born earlier in the academic year can drop out of school at a younger age than those born later.[1] Card (1995) focuses on whether high school students have a nearby university to attend, arguing that students living close to a university face a lower cost of acquiring post-secondary education.[2] The IV estimates when using these institutional instruments tend between 9 and 15%, but their usefulness is suspect because of the imprecision of the estimates due to their weak correlation with schooling (Bound *et al.*, 1995; Staiger and Stock, 1997).

When estimating wage profiles, the importance of controlling for innate skill (Griliches and Mason, 1972; Griliches, 1977; Blackburn and Neumark, 1993, 1995; Neal and Johnson, 1996) and parental income (Solon, 1992; Zimmerman, 1992) has been well-documented. The standard wage specification in returns to schooling studies using IV, however, fails to include measures of skill or parental income. The main contribution of this study is to show that skill and parental income need to be included in the wage equation when using IV. When they are included, parental education appears to be a better instrument than the typical institutional instruments. The regression statistics are relied on to compare the results across specifications and across the set of instruments. Basmann (1960) provides a test using the over-identifying restriction(s).[3] We also report the partial R^2 and the F statistic from the first stage regression of the two-stage least squares regression, as Bound *et al.* (1995) have shown that the inconsistency of IV estimates diminishes as the partial R^2 increases and that $1/F$ captures the relative inconsistency of IV to OLS that stems from spurious correlation between the instruments and ε.

III. THE DATA – NLSY 79

The National Longitudinal Survey of Youth (NLSY) is used, which contains information on youths aged 14 to 22 in 1979. Using the original survey along with income and education information from the 1996 follow-up survey, males are selected at most 17 years-old who lived with at least one parent in 1979. From this, all observations are dropped with a reported 1995 hourly wage of less than \$1 or more than \$75. Observations with any missing data are also dropped. This leaves a sample of 944 individuals. Table 1 provides the summary statistics.

The measure of skill comes from the Armed Forces Qualification Test (AFQT).[4] AFQT has been normalized for the entire sample to have a mean of zero and a standard deviation of one. The normalized scores were then age-adjusted to more accurately measure skill.

IV. RESULTS

The first seven columns of Table 2 presents OLS ln(wage) regression results. The specifications each include the individual's age and three indicator variables for race being white, living in the South in 1995, and living in an MSA in 1995. The first four specifications document the importance of education, parental income, skill (as measured by AFQT), and parental education on the wage. Notice in column 5 that parental income and skill remain strong determinants of the wage even after controlling for schooling, and that the return to schooling falls by half (from 9.4 to 4.7%) when these are included in the wage regression. In contrast, columns 6 and 7 show that parental education has little effect on wages after controlling for education, skill, and parental income.

In columns 8–13, Table 2 then reports the IV coefficients and test statistics from using three sets of instruments – quarter of birth, college proximity, and parental education – to estimate two wage equations, one without skill or parental income and another with both.

[1] Bound and Jaeger (1996) show that birth date can be correlated with earnings. Harmon and Walker (1995) and Staiger and Stock (1997) use similar instruments.
[2] Conneely and Uusitalo (1997) and Kane and Rouse (1995) use similar instruments.
[3] The Basmann test assumes at least one of the instruments is valid and tests the validity of the remaining instruments, so that the Basmann test has low power whenever there is a high degree of colinearity between the instruments.
[4] Neal and Johnson (1996) argue AFQT is an exogenous measure of skill for people under 18.

Table 1. *Summary Statistics*

	Non-Missing Value NLSY Sub-sample ($N = 944$)			
	Mean	Stnd. Dev.	Minimum	Maximum
1995 hourly wage	14.45	9.37	1.27	70.83
ln(1995 wage)	2.492	0.670	0.242	4.260
Years of education	13.37	2.38	7	20
Age in May 1994	28.97	1.047	27.08	30.87
White	0.7119	0.4531	0	1
Standardized AFQT score	0.0984	0.3563	−0.6665	0.7485
Lives in MSA in 1995	0.7934	0.4051	0	1
Lives in South in 1995	0.3591	0.4800	0	1
1979 parental income (1995 $)	45734	30057	1169	175309
ln(parental income)	10.50	0.74	7.06	12.07
Dad's years of education	11.60	3.32	0	20
Mom's years of education	11.68	2.35	0	20
1979 county has a 2-yr college	0.8814	0.3235	0	1
1979 county has a 4-yr college	0.7479	0.4345	0	1
Born January–March	0.232	0.4223	0	1
Born April–June	0.2511	0.4339	0	1
Born July–September	0.2934	0.4556	0	1
Born October–December	0.2235	0.4168	0	1

Note: The NLSY sub-sample is for all males aged 14 to 17 in 1979 who report an hourly wage of between $1 and $75 in 1995. AFQT has been normalized to have a mean of zero and standard deviation of one and then age-adjusted to more accurately measure skill.

When using quarter of birth as instruments, the return to schooling is greater under IV than OLS, however, a 95% confidence interval of the IV estimate contains the OLS point estimate (and zero).[5] The estimated coefficients on parental income and AFQT are not statistically different from zero, which is explained by the weak correlation between quarter of birth and schooling. The Basmann test rejects the hypothesis that all of the quarter of birth instruments are valid. And the extremely low partial R^2 and F statistics suggest that the IV bias is on par with the OLS bias.

When using college proximity to instrument for schooling, the IV point estimates of the return to schooling are less than the OLS estimates, but they remain imprecisely estimated.[6] The estimated effects of parental income and AFQT on wages are in line with the OLS estimates, but they remain statistically insignificant. The partial R^2 and F statistics remain low suggesting that the IV estimates retain much of the bias associated with OLS, however, the Basmann test fails to reject the validity of the college proximity instruments.

Unlike the previous two sets of results, the wage specification matters when using parental education to instrument for schooling. When neither parental income nor AFQT are controlled for (column 12), the return to education is 13.4%, which exceeds the comparable OLS estimate of 9.4%. When parental income and skill are included in the wage equation (column 13), however, the estimated return to schooling falls to 3.8% which is not statistically different from zero or from the OLS point estimate of 4.7%.[7] The estimates on parental income and AFQT are also statistically different from zero and roughly identical to the OLS estimates. And even after controlling for parental income and skill, parental education explains 27.7% of the remaining variation in schooling whereas only 11.4% is explained when parental income and AFQT are omitted from the specification. The Basmann statistic fails to reject the validity of the parental education instruments (though high correlation between father's and mother's education makes the power of the test very low), and the large F statistics suggest little bias remains in the estimates.

It is interesting to compare our results to those from the returns to schooling literature that uses sibling/twin data (Ashenfelter and Krueger, 1994; Ashenfelter and Zimmerman, 1997; Ashenfelter and Rouse, 1998). In those papers, the return to education is estimated by regressing the difference in earnings on the difference in schooling, which nets out any family specific effect present in earnings or education. The general consensus, in that

[5]Instrumenting similarly, Angrist and Krueger report OLS and IV estimates of 0.0521 (0.0003) and 0.0779 (0.0239).
[6]Instrumenting similarly, Card (1995) reports OLS and IV estimates of 0.073 (0.004) and 0.132 (0.049).
[7]Similar wage equations have been estimated using the NLS: Class of 1972. The OLS estimates range between 6.5 to 7.6%. When instrumenting with parental education, the returns increase to 11.7% if parental income and skill are omitted but falls to 5.8% when they are included.

Table 2. ln(*Wage*) *regression results*

	Ordinary Least Squares							Instrumental Variables					
								Instruments: Quarter of Birth		Instruments: College Proximity		Instruments: Parental Education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Years of education	0.094*				0.047*	0.086*	0.047*	0.181	0.133	0.051	0.041	0.134*	0.038
	0.008				0.010	0.008	0.010	0.109	0.114	0.150	0.199	0.018	0.055
ln(parental income)		0.202*			0.103*		0.104*		0.073		0.105		0.106*
		0.028			0.027		0.028		0.048		0.074		0.033
AFQT			0.734*		0.461*		0.465*		0.078		0.487		0.502*
			0.055		0.070		0.072		0.510		0.890		0.251
Dad's years of education				0.016*		0.003	−0.006						
				0.007		0.007	0.007						
Mom's years of education				0.042*		0.019*	0.006						
				0.010		0.010	0.009						
Age	0.037*	0.039*	0.030	0.042*	0.028	0.036*	0.028	0.029	0.029	0.041	0.028	0.033	0.028
	0.017	0.018	0.017	0.018	0.016	0.017	0.016	0.021	0.017	0.022	0.016	0.017	0.016
White	0.284*	0.210*	0.104*	0.289*	0.089	0.269 *	0.094*	0.220*	0.177	0.316*	0.084	0.254*	0.080
	0.041	0.047	0.044	0.044	0.046	0.042	0.047	0.092	0.125	0.118	0.208	0.043	0.072
South in 1995	−0.020	−0.022	0.002	0.001	0.005	−0.007	0.004	0.066	0.088	0.195	0.122	0.112*	0.123*
	0.038	0.040	0.038	0.040	0.037	0.038	0.037	0.119	0.061	0.156	0.084	0.048	0.048
MSA in 1995	0.152*	0.203*	0.154*	0.189*	0.120*	0.142*	0.124*	−0.003	−0.002	−0.029	0.005	−0.012	0.005
	0.045	0.047	0.044	0.047	0.043	0.045	0.044	0.046	0.039	0.049	0.040	0.039	0.037
Constant	−0.149	−1.071	1.356*	0.228	−0.236	−0.268	−0.260	−0.970	−1.101	0.256	−0.179	−0.531	−0.145
	0.496	0.573	0.488	0.521	0.548	0.497	0.549	1.158	1.274	1.504	2.077	0.527	0.772
R-squared	0.223	0.144	0.240	0.142	0.274	0.229	0.275	0.115	0.211	0.197	0.274	0.192	0.273
Basmann *p*-value								0.0379	0.0168	0.6398	0.8434	0.4226	0.3362
Partial *R*-squared								0.001	0.002	0.047	0.004	0.114	0.277
First Stage F-St.								1.718	2.410	1.238	1.077	99.201	14.598

Note: Number of observations is 944. The Basmann statistic tests for over identifying restrictions. The *p*-value lists the probability with which one cannot reject that all instruments are valid. Partial *R*-squared is the amount of variation of education explained by the instruments after controlling for the exogenous variables. The First Stage *F*-Statistic is the test statistic that the coefficients on all instruments equal zero in the first stage regression of education regressed on the exogenous variables and the instruments.
*Indicates the estimate is significantly different from zero with 95% confidence.

literature, is that the IV and OLS estimates are of about the same magnitude. By including skill and parental income in the wage and education equations, the results in column 13 essentially control for a family specific effect, and in so doing, the IV point estimate is found to come into line with the OLS point estimate.

V. DISCUSSION

It has been long established that innate skill and parental income affect wages. Until now, however, there has been no systematic effort to include both of these in the wage equation when estimating the returns to schooling with instrumental variables. This omission is particularly striking given that omitting variables from the wage equation that are positively related to schooling and the wage will not only be unbiased but the IV procedure will exacerbate, not diminish, the bias associated with the OLS estimates (Card, 1999). The results confirm this pattern and lend support for using parental education to instrument for schooling in place of institutional instruments.

In Table 2 (as well as throughout the literature), IV estimates of the return to schooling exceed the OLS estimates when skill and parental income are omitted from the wage specification regardless of which instruments are used. This runs counter to the typical selection story, but is expected given the correlation between skill and parental income with the wage. Only when skill and parental education are included and schooling is instrumented for with parental education, however, do the IV estimate falls below the OLS estimate. Moreover, the test statistics and estimated coefficients overwhelmingly support using parental education as an instrument.

When instrumenting with parental education, the return from an additional year of schooling in the NLSY is estimated to be 3.8% and not statistically different from zero. At face value, there is no statistically significant return to an additional year of schooling. Taking a step back, however, and assuming that productivity is rewarded in the work place, the issue becomes one of measuring an individual's productivity. To this end, *quality of schooling* rather than years of schooling may be the better measure of productivity. It is quite possible that AFQT and parental income better capture this productivity (or quality of education) than does years of education.

ACKNOWLEDGEMENTS

Lemke thanks Dimitrios Thomakos for several helpful suggestions on an earlier draft. Rischall thanks the Canadian International Labour Network (CILN) for financial support. CILN is a major research initiative of the Social Sciences and Humanities Research Council of Canada (SSHRCC) and McMaster University. Thanks also to Martin Browning, John Kennan, Aloysius Siow, Jim Walker, and seminar participants at the University of Wisconsin-Madison, McMaster University, University of Western Ontario, Brock University, York University, and the University of Toronto for their valuable comments.

REFERENCES

Angrist, J. A. and Krueger, A. B. (1991) Does compulsory school attendance affect schooling and earnings?, *Quarterly Journal of Economics*, **106**, 979–1014.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999) A review of estimates of the schooling/earnings relationship, with tests for publication bias, *Labour Economics*, **6**, 453–70.

Ashenfelter, O. and Krueger, A. (1994) Estimates of the economic return to schooling from a new sample of twins, *American Economic Review*, **84**, 1157–73.

Ashenfelter, O. and Rouse, C. (1998) Income, schooling, and ability: evidence from a new sample of identical twins, *Quarterly Journal of Economics*, **113**, 253–84.

Ashenfelter, O. and Zimmerman, D. J. (1997) Estimates of the returns to schooling from sibling data: fathers, sons, and brothers, *Review of Economics and Statistics*, **79**, 1–9.

Basmann, R. L. (1960) On finite sample distributions of generalized classical linear identifiability test statistics, *Journal of the American Statistical Association*, **55**, 650–59.

Blackburn, M. L. and Neumark, D. (1993) Omitted-ability bias and the increase in the return to schooling, *Journal of Labor Economics*, **11**, 521–44.

Blackburn, M. L. and Neumark, D. (1995) Are OLS estimates of the return to schooling biased downward? Another look, *Review of Economics and Statistics*, **77**, 217–30.

Bound, J. and Jaeger, D. A. (1996) On the validity of season of birth as an instrument in wage equations: a comment on Angrist and Krueger's 'Does compulsory school attendance affect schooling and earnings?', National Bureau of Economic Research working paper 5835.

Bound, J. Jaeger, D. A. and Baker, R. M. (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.

Card, D. (1995) Using geographic variation in college proximity to estimate the return to schooling, in *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp* (Ed.) L. N. Christofides, E. Kenneth Grant, Robert Swidinsky, and John Vanderkamp, University of Toronto Press Incorporated, Toronto, pp. 201–22.

Card, D. (1999) The causal effect of education on earnings, in *Handbook of Labor Economics*, Volume 3, (Ed.) O. Ashenfelter and D. Card. North-Holland, New York.

Conneely, K. and Uusitalo, R. (1997) Estimating heterogeneous treatment effects in the Becker schooling model, Unpublished discussion paper, Princeton University Industrial Relations Section.

Griliches, Z. (1977) Estimating the returns to schooling: some econometric problems, *Econometrica*, **45**, 1–22.

Griliches, Z. and Mason, W. M. (1972) Education, income, and ability, *Journal of Political Economy*, **80**, S74–S103.

Harmon, C. and Walker, I. (1995) Estimates of the economic return to schooling for the United Kingdom, *American Economic Review*, **85**, 1278–86.

Kane, T. K. and Rouse, C. E. (1995) Labor-market returns to two- and four-year college, *American Economic Review*, **85**, 600–14.

Neal, D. A. and Johnson, W. R. (1996) The role of premarket factors in black-white wage differences, *Journal of Political Economy*, **104**, 869–95.

Solon, G. (1992) Intergenerational income mobility in the United States, *American Economic Review*, **82**, 393–408.

Solon, G., Corcoran, M., Gordon, R. and Laren, D. (1991) A longitudinal analysis of sibling correlations in economic status, *Journal of Human Resources*, **26**, 509–34.

Staiger, D. and Stock, J. H. (1997) Instrumental variables regression with weak instruments, *Econometrica*, **65**, 557–86.

Willis, R. J. and Rosen, S. (1979) Education and self-selection, *Journal of Political Economy*, **87**, S7–S36.

Zimmerman, D. J. (1992) Regression toward mediocrity in economic stature, *American Economic Review*, **82**, 409–29.