INSTRUMENTATION VERSUS ORDINARY LEAST SQUARE ESTIMATES OF RETURNS TO EDUCATION

Ochalibe, Alexander Ibu
Department of Agricultural Economics, Federal University of Agriculture, Makurdi

Mgbebu, Ezekiel Sunday
Department of Agricultural Economics, University of Nigeria, Nsukka
&
Onyia, Chiebonam Chukwuemeka
Department of Agricultural Economics, University of Nigeria, Nsukka

ABSTRACT

This work estimates the returns to education using 5% of data drawn from a 20% random sample of working-age men in England from the UK Quarterly Labour Force Survey (QLFS). From the result, the coefficient for rosla is 0.064, statistically significant at 5% level, which means a one-year increase in the leaving age raises educational qualification by 6.4%. This implied that the instrument is relevant. All the variables selected are statistically significant at 5% level with p-value of less than 0.05. This justifies the inclusion of the variables. The result shows that the OLS estimate of returns to an additional qualification (controlling for demographic characteristics) is 19% with standard error of 0.008 (3d.p), while that IV instrument gives 10.6% with a larger warfare error *of* 0.267 (3d.p). Additionally, the F-static in first stage is roughly 21 which is greater than 10, the 'rule of thumb' this means the instrument is not weak. The result showed that OLS estimates of the return to schooling are smaller than their IV counterpart. It was concluded that OLS bias downward and the IV estimates obtained are a better indicator of the population average than OLS estimates. Additionally, Hausman test clearly justifies the use of IV instrument hence instrument is both relevant and consistent. It was recommended that IV regression was more suitable and should be utilized in the presence of endogeneity.

Keywords: OLS, IV, instrument, regression and education.

INTRODUCTION

Instrument variable (IV) is used to estimate causal relationship when controlled experiments are not feasible. Usually an estimator obtained by OLS is BLUE (Best Linear Unbiased Estimator) if it has minimum variance and is unbiased. This is the case when if all classical linear assumptions hold however, OLS estimate are unlikely to be best if there is violation in any of the assumptions hence the need for other methods with better estimators. The wage premium for education is probably one parameter that is of interest in modern economics. According to Christian (2005) education measure are potentially affected by unobserved individual skills which correlated with individual wages hence the need to instrument on them to correct the "ability bias".

A major complication that is emphasized in micro econometrics is the possibility of inconsistent parameter estimation due to endogenous regressors. Usually, OLS regression estimates is a measure of only the magnitude of association, rather than the magnitude and direction of causation which is needed for policy analysis. The instrumental variables estimator provides a way to obtain consistent parameter estimates. This method, widely used

in econometrics and rarely used elsewhere, is conceptually difficult and easily misused. When there is a presence of endogeneity, a treatment approach is still possible using observational data, provided there exists an instrument z that has the property that changes in z are associated with changes in x but do not led to change in the dependent variable (y)

The second assumption requires that there is some association between the instrument and the variable being instrumented. Examples of an Instrument in many econometric applications abound though it is difficult to obtain a legitimate instruments. For instance, if we want to estimate the response of market demand to exogenous changes in market price: Quantity demanded clearly depends on price, but prices are not exogenously given since they are determined in part by market demand. A suitable instrument for price is a variable that is correlated with price but does not directly affect quantity demanded. An obvious candidate is a variable that effects market supply, since this also effect prices, but is not a direct determinant of demand.

In the field of agriculture to examine the favorable growing conditions of a product, The choice of instrument here is uncontroversial, provided favorable growing conditions do not directly affect demand, and is helped greatly by the formal economic model of supply and demand. Additionally if one is interested in estimating the returns to exogenous changes in schooling, most observational data sets lack measures of individual ability, so regression of earnings on schooling has error that includes unobserved ability and hence is correlated with the regressor schooling. We need an instrument z that is correlated with schooling, uncorrelated with ability and more generally is uncorrelated with the error term which means that it cannot directly determine earnings. According to Card, (1995) one popular candidate for z is proximity to college or university This clearly satisfies condition 2 as, for example, people whose home is a long way from a community college or state university are less likely to attend college. It most likely satisfies 1, though since it can be argued that people who live a long way from a college are more likely to be in low-wage labor markets one needs to estimate a multiple regression for y that includes as additional regressors controls such as indicators for non-metropolitan area.

A second candidate for the instrument is month of birth (Angrist and Krueger,1991). This clearly satisfies condition 1 as there is no reason to believe that month of birth has a direct effect on earnings if the regression includes age in years. Surprisingly condition 2 may also be satisfied, as birth month determines age of first entry into school which in turn may affect years of schooling due to laws that specify a minimum school leaving age. Bound et al. (1995) provide a critique of this instrument.

According to Doherty (2007) Instrumental variables are generally inconsistent if the instruments are correlated with the error term in the equation of interest; if weak instrument that are poor predictor in the first stage equation are selected. This may result in poor prediction of the endogenous explanatory variable by the instrument and the predicted value will have little variation. The pertinent question is whether the instrument is valid and if they are, are they consistent and relevant?

The returns to education using LFS data has been computed by Chevalier et al (2004), Walker and Zhu (2003). Similarly the correlation between education and wages has been analyzed by (Becker 1964, 1967 and Mincer 1958, 1974 quoted in Christian 2005).However little or none has been done to compare instrumental variable regreesion and ordinary least

square method using post estimation techniques. This work attempts to use instrument variable approach to obtain a consistent estimator of returns to education when it is suspected to correlate with the disturbance term as result of omitted variable (ability bias) and make comparison with result from OLS estimates. This is significant because it establishes the right model and may spur policy makers to make right policies that has to do with returns to education.

METHODOLOGY

The data used is drawn from a 20% random sample of working-age men in England from the UK Quarterly Labour Force Survey (QLFS). The Choice of this country arises from the opportunity available in getting complete enormous dataset. The dataset contain measures of returns to education and other relevant variables of interest. The logarithm of real hourly wage (*logwage*) was used as a measure of returns while nvqlv2 was used as a measure of education. In order to avoid inconsistencies and outliers the data was checked before running the regression. After the regression a diagnostic test of misspecification and heteroscedasticty were carried out to assess the validity of the empirical model.

The approach to selecting an instrument following the trend of previous experiment, Hermon and Walker (1995) use changes in school leaving age to provide instrument for education in the estimation of returns to schooling. Given the composition of the dataset *rosla* is likely to be suitable instrument because according to Webster (2010) Good instruments are often created by policy changes. Additionally, using the national vocational qualification level 2 (nqvlv2) as a measure for education to determine wage, raising of school leaving age (rosla) will be used as an instrument variable for nvqlv2. This is because it is suspected to correlate with nvqlv2, unlikely to be correlated with the disturbance term and unlikely to be a direct determinant of returns.

Instrumenting *nvqlv2* on rosla, it will be a valid instrument based on the following requirements which is sometimes described as instrument relevance: (a) it must be correlated with the endogenous explanatory variable that is, increases in compulsory schooling has effect on educational qualification (b) instrument variable must not be correlated with the error term which implies that, increase in compulsory schooling are uncorrelated with the ability. The approach used here is instrumenting raising of school living age (*rosla*) on *nvqlv2* which is suspected to correlate with the unobserved ability (ability bias). Because ability cannot be observed, the assumption for zero correlation which is also called an orthoganality assumption will be difficult to test directly. However according to Baum (2006) such test could be constructed in the presence of multiple instruments.

Model Specification

In the first instance, OLS returns to education was estimated without correcting for ability bias. This involves estimating the regression:

$$In(Y_i) = \pi + \beta nqvlv2_i + \delta Z_i + \varepsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Where the dependent variable (Y) is the log real gross hourly wage, *nqvl2* is the individual's qualification and Z is a vector of demographic characteristics (cohab, lim_dis age se London), β describes rate of return to an additional qualification.

To use rosla as an Instrument for educational qualification, we estimate the following first stage regression:

$$In(nqvlv2_i) = \pi + \beta(rosla2)_i + \delta_i + \varepsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Where rosla is an indicator variable for born after September 1957 (affected by rosla)
In both cases, the second stage equation is:

$$In(Y_t) = \theta + \alpha nqvlv2_t + nZ_t + V_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

Table 1: Summary Statistics for key variables of LFS survey

VARIABLES	MEAN	SD	Observation
Log real gross hourly wage	2.249	0.398	11839
National vocational qualifications	0.262	0.440	11839
Married	0.555	0.475	11839
Cohabiting	0.136	0.343	11839
Age	39.555	6.451	11839
Nonwhite	0.20	0.138	11839
Health problem	0.084	0.277	11839
LFS year of survey	100,343	3.031	11839
Greater London	0.079	0.270	11839
Outside London	0.244	0.430	11839

Source: LFS sample

RESULTS AND DISCUSSION

Interpretation of stage 1

The regression output in table 2 is the predicted values from the first stage IV regression where rosla was used as an instrument. The idea is to decompose *nqlv2* into free and problematic component (that is, to eliminate the endogeneity problem) as well as looking at the relevance of instrument.

From the result, the coefficient for rosla is 0.064, statistically significant at 5% level, which means a one-year increase in the leaving age raises educational qualification by 6.4%. This satisfies one of the conditions of a valid instrument (instrument relevance).

Additionally, the F-static in first stage is roughly 21 which is greater than 10, the 'rule of thumb' this means the instrument is not weak. According to Doherty (2007) Instrumental variables are generally inconsistent if the instruments are correlated with the error term in the equation of interest; if weak instrument that are poor predictor in the first stage equation are selected. This may result in poor prediction of the endogenous explanatory variable by the instrument and the predicted value will have little variation.

Table 2: First stage IV regression

A: Dependent variable: national vocational qualification	
School leaving age(rosla)	0.064
	(0.014)
Married	0.026
	(0.010)
Cohabiting	0.007
	(0.014)
Age	-0.019
	(0.007)
Non-white	-0.056
	(0.029)
Hearth problem	-0.028
	(0.015)
London	0.038
	(0.015)
Outside London	0.036
	(0.010)
Observations	11839
R-squared	0.0073

Source: LFS data; *Note the figures in parentheses () represent the standard error

Table 3: Results comparing OLS with IV instrument

B: Dependent variable: log real wage	OLS	IV Rosla
Qualification	0.200	1.062
	(0.008)	(0.267)
Married	0.143	0.121
	(0.009)	(0.014)
Cohabiting	0.096	0.089
	(0.012)	(0.017)
Age	0.048	0.057
	(0.006)	(0.009)
Non-white	-0.056	-0.106
	(0.025)	(0.039)
Health problem	-0.125	-0.101
	(0.012)	(0.019)
London	0.022	0.187
	(0.013)	(0.021)
Outside London	0.145	0.113
	(0.008)	(0.015)
F-test for excluded instrument	-	21.56
Observations	11839	11839
R-squared	0.13	

Source: LFS data; *Note the figures in parentheses () represent the standard error

Comparing OLS IV instrument

Table 3 compares the OLS estimates with the stage 2 of IV regression where the logarithm of real gross hourly wage was regress on the predicted values from the first stage. The idea here is to exploit the free component of endogenous nvqlv2 in order to have an unbiased as well as consistent estimator. The *result* from various studies have generally found increase returns to education, Leigh and Ryan (2005); Harmon and Walker (1995) and (Meghir and Palme 2003) in a-discussion paper observed a 10% higher returns to education corrected for-ability bias.
All the variables selected are statistically significant at 5% level with p-value of less than 0.05. This justifies the inclusion of the variables. The result shows that the OLS estimate of returns to an additional qualification (controlling for demographic characteristics) is 19% with standard error of 0.008 (3d.p), while that IV instrument gives 10.6% with a larger warfare error *of* 0.267 (3d.p). The coefficient of nvqlv2 is larger in the IV regression suggesting that omitted variable bias has led to a downwards bias in its coefficient in the OLS regression.

The standard error in IV regression is however much larger than in OLS regression which implies loss in efficiency. This IV estimation can be improve by drawing on other variables instead of just nvqlv2 to instrument for nvq then carryout a formal test of the difference in coefficients. Similarly the IV approach predicts approximately 12.1% and 8.8% higher wages for married and cohabiting respectively ceteris paribus. While living in London or the rest Southeast region is associated with approximately 18.7% and 11.3% higher wages respectively, other things being equal. Age is associated with 5.6% higher wages. However health problem and non-white means 10.1% and 10.6% lower wages respectively. This implies that individuals with health problem and non-whites are disadvantage.

Hausman test rejects the null of no systematic difference between the OLS and IV instrument this implies that nvqlv2 is endogenous and maintaining that rosla is exogenous. However it should be taken cautiously as this test could be misleading; it has low power and assumes instruments are valid.

CONCLUSION

This work estimates the returns to education and found that OLS estimates of the return to schooling are smaller than their IV counterpart, OLS bias downward. The IV estimates obtained are a better indicator of the population average than OLS estimates. Additionally, Hausman test clearly justifies the use of IV instrument hence instrument is both relevant and consistent which proves the focus of this work. For policy makers the implication here is that, increasing the level of education in the population is rewarding and raising of school leaving age is likely to affect some set of individuals. In the case of health problem, policy makers may wish to mitigate the effect of such employee wages through tax rebate and other measures. The wage differential observed in race relations could be addressed by way of a balanced legislation

REFERENCES

Chevalier, A, Harmon, C. Walker, L and Zhu, Y. (2004). Does Education Raise Productivity, or just Reflect re?, Economic journal 214, F499-F517.

Baum, C. F. 2006. Introduction to Modern Econometrics *Using* STATA, *STATA Press, ISBN-10: 1-5S723-013-C.*

Christian B. (2005). The Return to Schooling and Experience and the Ability Bias in Structure Models: Centre National de Recherche Scientifique (GATE, Ecully, France) Institute for the study of Labor (IZA), CIRANO and CREQ.May 3.

Dougherty, C., 2007 Introduction to Econometrics, Oxford University Press.

Harmon C. and Ian Walker, (1995), "Estimates of the economic return to schooling for the UK, "IFS Working Papers W95/12, institute for Fiscal Studies.

Leigh, A. and Ryan, C. (2005), Estimating Returns to Education: Three Natural Experiments Compared.

Meghir, C. and Palme, M. (2003), "Ability, parental background and educational policy: Empirical Evidence from a social experiment," IFS Working Papers W03/05, Institute for Fiscal Studies.

Wa*lker* I. and Zhu, Y. (2003), Education, Earnings and Productivity—Recent UK Evidence, Labour market Trends, March, 145-152.

A Dictionary *Definition of instrumental Variables*- http:/www.websters-online-dictionary.org Mekonnen, A., and Kohlin, G., 2009. Determinants of Household Fuel Choice