Economics
The Open-Access, Open-Assessment E-Journal

Vol. 5, 2011-12 | September 2, 2011 | http://dx.doi.org/10.5018/economics-ejournal.ja.2011-12

Labor Market Returns to Higher Education in Vietnam

Tinh Doan and Philip Stevens
Ministry of Economic Development, New Zealand

Abstract This paper employs the Ordinary Least Squares, Instrumental Variables and Treatment Effect models to a new dataset from the Vietnam Household Living Standards Survey (VHLSS) to estimate return to four-year university education in 2008. Our estimates reveal that income premium of four-year university education is about 97 percent above that of high school education, and robust to the various estimators.

Special Issue
Quasi Markets in Education

JEL C31, J31, O15
Keywords Economic transition; returns to higher education; IV model; Vietnam

Correspondence Tinh Doan, Department of Economics, Waikato Management School, The University of Waikato, Hamilton, New Zealand; e-mail: thanhtinhdoan@gmail.com

Citation Tinh Doan and Philip Stevens (2011). Labor Market Returns to Higher Education in Vietnam. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 5, 2011-12. doi:10.5018/economics-ejournal.ja.2011-12. http://dx.doi.org/10.5018/economics-ejournal.ja.2011-12

1 Introduction

The most challenging task of estimating return to higher education is that one does not have sufficient information about studied subjects. Observationally identical individuals make different choices; we do not know why some people decide to take university education, while some do not, so that difference in their earnings may be due to education participation or unobservable attributes. To estimate return to university education, we should measure how much people would have earned if they did not have university degrees (Heckman and Li 2004), but we are unable to measure this counterfactual earnings.

The ordinary least squares (OLS) do not account for the factors affecting the higher education decision. This is particular true in Vietnam where potential students faced with liquidity constraints (Glewwe and Jacoby 2004; Glewwe and Patrinos 1999). Furthermore, university intake places are limited due to the government's quota and capacity of education providers; about three fourths of high school leavers are unable to go to university.[1]

According to a population census in 2009, only 5 percent of the Vietnam population (86 million people) hold higher education degrees (GSO 2010). This is much lower than other countries both in the Southeast Asia and indeed across the world. Given the fact that university candidates have to take highly competitive entrance examinations as well as face liquidity constraints,[2] factors such as individual ability, family resources and motivation may play important roles in their pursuing higher education. Therefore, university students self-select into higher education on both family background and student ability (observed and unobservable attributes).

In such a case like this and without experimental data, one may employ instrumental variable or fixed effect methods. However, IV method is preferred method as instrumental variable estimator provides a consistent estimate of the return to education (Ashenfelter et al. 1999), and is less prone to mis-specification

[1] See at http://www.business-in-asia.com/vietnam/education_system_in_vietnam.html; and http://www.gso.gov.vn/default_en.aspx?tabid=474&idmid=3&ItemID=10220

[2] The Vietnam government annually allocates a certain quota of student intakes for each university depending on their facility and staff capacity, so demand for higher education is always higher than the supply

than FE estimator (Belzil 2007; Keane 2010) because the fixed effect is highly sensitive to measurement error in schooling (Ashenfelter and Zimmerman 1997).

Our results show that the income premium for four-year university education in Vietnam in 2008 is 97 percent based on IV model and about 101 percent based on the OLS and Treatment Effect models. Thus, the bias by OLS model is not too large to be concerned in the context of Vietnam higher education.

This paper is structured as follows. Section 2 presents empirical models and data. Section 3 discusses estimation results. Concluding remarks are presented in Section 4.

2 Empirical Models and Data

As noted in the previous section, we employ IV estimator in this study. What we need to do is to search for instruments that affect schooling choices but not earnings. In reality, there are two groups of instruments relating to either supply side or demand side of schooling. On the supply side, many studies make use of institutional sources of schooling variation, such as minimum school leaving age (Harmon and Walker 1995), proximity to school (Card 1995). On the demand side, variables such as quarter of birth (Angrist and Krueger 1991; Staiger and Stock 1997), and family background such as parental education, year of birth, brother's education, sibling composition (Ashenfelter and Zimmerman 1997; Butcher and Case 1994; Card 1995, 1999; Staiger and Stock 1997) are used. Hogan and Rigobon (2010) use both sides of the market to exploit the heterogeneity in education attainment caused by differences between regions resulting from different population density, variation in the proximity to school, parental income, and income distribution, demographics, school quality, and weather etc. across regions.

In our case, we look at the return to four-year university education using demand side factors such as parental education, assets and ratio of the university and post-graduated members in family (called ratio of higher education members hereinafter) as instruments. Family information such as parental education is often utilized to either directly control for unmeasured ability or as an instrument for children's schooling (Ashenfelter and Zimmerman 1997; Card 1995; Heckman and Li 2004; Griliches 1977). This is because children's education is highly correlated with their parents' characteristics especially education and economic conditions (income and assets) (Card 1999). Moreover, education in Vietnam is not free of charge (Glewwe and Jacoby 2004; Glewwe and Patrinos 1999). We utilize household assets and parental education to proxy for permanent household income (Musgrove 1979) which is believed to be correlated with children's education.

Our empirical models are as follows:

OLS model: $\text{Log } y_i = \alpha + \beta S_i + \lambda X_i + \delta . Z_i + \varepsilon_i$ (1)

IV model: $\text{Log } y_i = \alpha + \hat{\beta} S_i + \lambda X_i + u_i$ and $S_i = \gamma Z_i + v_i$ (2)

where $Z_i u_i$ are independent or $\text{cov}(Z_i, u_i) = 0$, and $\text{cov}(Z_i, S_i) \neq 0$. The variable S is a 0/1 variable that equals to one if an individual has a bachelor's degree (four-year university graduate) and 0 if an individual has a high school diploma. We remove post-graduate degree holders, three-year-college and vocational-diploma holders,[3] and below-high-school educated individuals. The variable X_i is a set of controlling variables including experience, experience squared, gender, ethnicity, urban, economic sectors, and eight geographical regions in Vietnam. The estimated coefficient β in equations 1 and 2 reflects a percentage difference in earnings between individuals with a bachelor's degree and ones with high-school graduation degree. This coefficient is referred as the four-year university premium. The variable Z_i is a set of family background such as mother's education, father's education, ratio of higher education members, and household assets (durable, fixed assets and houses) which was acquired at least one year prior to the survey.[4]

On the one hand, the use of family back ground variables such as parental education is controversial since they often violate the validity assumption. For example, father's education may affect both children's schooling and children's incomes (Psacharopoulos and Patrinos 2004). On the other hand, Hoogerheide et al. (2010) show that using the family background variables as instruments is a practical option to deal with the endogeneity problem of education when it is hard to find good instruments as they are often available in many household surveyed

[3] Because the number of years for achieving these educational levels is greater or fewer than four-years relative to university education, adding them into the four-year university graduates will bias the estimated return to the four-year university education.

[4] This is to avoid the reversal causality effect of current earnings on the assets.

datasets and the possible size of bias is smaller than the standard error of education's estimated coefficient in the IV model.

In any case, one should carefully conduct the necessary IV tests to make sure that two assumptions of instruments (relevance and validity) hold, or at least to avoid using invalid and weak instruments leading to imprecise estimates and conclusions. Instrument (Z) needs to be valid and strongly correlated with endogenous variable of education. In other words, instruments should be determinants of schooling decision, but uncorrelated with earnings residual (error term). IV method first estimates effect of instrumental variables (Z) on schooling (S), then estimates the effect of the schooling (S) on earnings (y). By this procedure, instruments affect the dependent variable (earnings) only through schooling but do not have a direct effect on the dependent variable (earnings). The relevance assumption implies that an instrument should be strongly correlated with endogenous variable of education. If the instrument does not meet this condition, we have a weak instrument problem that makes it difficult to provide correct estimates (Hoogerheide et al. 2010). If an instrument is also correlated with the earnings residual, the estimates will be biased as IV violates the validity or exclusive restriction assumption (Angrist et al. 1996; Staiger and Stock 1997), especially if Z_i are weakly correlated with schooling (S_i) and positively correlated with earnings, the estimates would be highly upward biased (Murray 2006; Stock and Yogo 2002). The lower the correlation between the instruments and treatment participation, the more sensitive the IV estimate is to violations of the exclusion restriction assumption (Angrist et al. 1996, p. 451).

Family background (Z_i) can also be used to check the robustness of the OLS estimates (Yakusheva 2010). Even though family background variables may not be legitimate instruments for education, controlling for these variables in OLS may reduce the bias in estimated return to schooling because they are often correlated with unobserved children's ability (Card 1999). In a review of many studies that also controlled for ethnicity, region and age, the family background explain up to about 0.30 of the variance of observed schooling, and expected attenuation of the education coefficient could be as high as 15 percent. This attenuation is almost as high as the attenuation bias by measurement error in measured schooling (Card 1994). Thus, controlling for these variables also is as important as correcting for measurement error in reported education.

Given the fact that to enter universities, in Vietnam the candidates have to take highly competitive entrance examinations, factors such as individual ability, and family resources and parental motivation play important roles in entering university education. These factors can be reflected through family background since individuals are more likely to have similar innate ability and family background than randomly selected (Ashenfelter and Zimmerman 1997). On the supply side of the university education, some studies use proximity to college as an instrument to predict schooling in Vietnam (e.g. Arcand et al. 2004). We do not use this information because the data of distance to university from each household measured in the current survey do not reflect properly the distance to university when the surveyed wage-earners were at ages for the university entry. This is because there is a high rate of migration, especially of wage-earners, in Vietnam since the economic reform was introduced in late 1980s (International Organization of Migration;[5] GSO 2010). In other words, distance from wage-earners' current homes to the nearest university may not be exactly the distance to the nearest university when they were students.

Data used in this study come from Vietnam Household Living Standards Survey conducted by the Vietnam General Statistical Office in 2008 (VHLSS 2008).[6] The survey interviewed 9,186 households that makes up about 40,000 members covering all provinces and eight geographical regions of Vietnam. The survey is representative for national level of Vietnam. The survey collected a rich set of variables such as demographics, education, healthcare, employment, income sources, expenditures, assets, housing, borrowing, participation in governmental supporting programs.

The probability of being wage-earners in Vietnam was low. Most people earned from self-employment such as small family businesses, farmers and other unofficial economic activities (Doan and Gibson 2010). Because of this the subsample of individuals who have either four-year university degrees or high school diplomas and earned from wages and salaries is rather small relative to the entire sample. Further, the IV model used in the current paper needs more

---

5 http://www.iom.int.vn/joomla/index.php

6 More information about the survey, see at http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1181743055198/3877319-1207149468624/BINFO_VHLSS_02_04.pdf

information such as mother's education, father's education, household assets, ratio of university and higher educated members within family; this generates more missing observations and offers a subsample of relatively young individuals (see Table 1 in the next section). Therefore, the final sample shrinks and provides only 651 individuals who have either high school diplomas or four-year university degrees and also have earnings from wages and salaries to estimate the return to university education.

Specifically, our subsample was obtained as follows. The entire VHLSS2008 sample makes up about 40,000 observations of which 22,723 observations are in labor force and have finished high school (whose age ranges from 18 to 60). There are 7,760 wage earners (34 percent) in this subsample of which 778 are high school graduates and 705 are four-year university graduates. However, instrument method using parental background use only observations with parental education, household assets, and siblings' education information. Therefore, subsample of 7,760 wage-earners continues to shrink to 2,608 observations of which 360 are high school graduates and 294 are four-year university graduates. We also removed three extreme observations. Eventually, we have a subsample of 651 observations to estimate the return to university education.

3 Estimation results

In this section we estimate series of models from OLS with basic controls, then with further controls of family background such as father's education, mother's education, ratio of higher education members, and the log of household assets. Next, we estimate an IV model and conduct tests of instruments accordingly. Finally, Treatment Effect model estimation will be run to corroborate the IV estimates.

The unconditional mean of hourly wage of four-year university wage-earners is almost twice as high as that of high school graduated wage-earners (Table 1). University graduates are more likely to work in state sector but less likely to work in private sector. They also have better family background such as parental education, assets, and have more siblings who obtained university and post-graduate degrees. They are also more likely to be in the major ethnic group (Kinh and Chinese) and living in urban areas. The university graduated wage-earners are about 3 years older but have about one year of experience fewer than the high school graduated wage-earners since they stayed at school longer (Table 1). This also indicates that the subsample used in this paper covers younger individuals

Table 1: Summary Statistics

Variables	High school graduates (n=360)		The university graduates (n=291)		*t*-value for equal mean
	Mean	Std. error	Mean	Std. error	
Hourly wage (VND 1,000)	9.146	0.478	17.571	0.696	9.98*
Log of hourly wage	1.956	0.036	2.685	0.036	14.36*
Worked in state sector	0.222	0.022	0.704	0.027	13.93*
Worked in foreign sector	0.114	0.017	0.082	0.016	1.35
Worked in private sector	0.664	0.025	0.213	0.024	13.02*
Age (year)	26.706	0.288	29.793	0.368	6.61*
Experience (year)	8.706	0.288	7.801	0.367	1.94***
Gender (male=1)	0.597	0.026	0.550	0.029	1.21
Majority (Kinh and Chinese=1)	0.922	0.014	0.979	0.008	3.48*
Urban (yes=1)	0.339	0.025	0.718	0.026	10.43*
Region 1—Red River	0.286	0.024	0.268	0.026	0.51
Region 2—North East	0.097	0.016	0.107	0.018	0.39
Region 3—North West	0.039	0.010	0.014	0.007	2.05**
Region 4—North Central	0.050	0.012	0.058	0.014	0.47
Region 5—South Central	0.114	0.017	0.117	0.019	0.12
Region 6—Central Highlands	0.025	0.008	0.027	0.010	0.20
Region 7—South East	0.217	0.022	0.271	0.026	1.61
Region 8—Mekong Delta	0.172	0.020	0.137	0.020	1.22
Instruments					
Mother's education (year)	5.778	0.236	9.646	0.327	9.59*
Father's education (year)	6.331	0.264	9.405	0.385	6.58*
Ratio of higher education members	0.017	0.004	0.432	0.013	30.58*
Log total assets acquired before 2007	12.599	0.063	13.606	0.062	11.32*

*, **, *** denote significance at the level of 1, 5, and 10 percent, respectively.

than the entire sample in the same year of 2008 (see Doan and Gibson 2010), so the estimate may not be representative for the entire sample in Vietnam.

These differences highlight the importance of either controlling for the family background variables in the OLS wage equation or using them as instruments for schooling. Table 2 shows the results of estimation of the probabilities of going to university as controlling for the family background variables. We observe significant effects of these variables on the probability, this suggests that they meet

Table 2: Probability of Going to University

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.1131	0.1113	0.0660	0.1303	0.1152	0.0736
	(3.27)*	(3.10)*	(1.55)	(3.63)*	(3.41)*	(1.71)***
Age squared	–0.0014	–0.0012	–0.0008	–0.0015	–0.0015	–0.0010
	(2.67)*	(2.26)**	(1.31)	(2.83)*	(2.84)*	(1.57)
Gender (male=1)	–0.0093	–0.0095	0.0732	–0.0405	–0.0227	0.0465
	(0.18)	(0.18)	(0.93)	(0.77)	(0.44)	(0.57)
Majority	0.1888	0.1837	0.0892	0.1908	0.0644	0.0045
	(1.41)	(1.31)	(0.73)	(1.51)	(0.46)	(0.04)
Urban	0.3422	0.2436	–0.0026	0.3064	0.1904	–0.0749
	(6.87)*	(4.39)*	(0.03)	(5.79)*	(3.18)*	(1.00)
Mother's education		0.0465				–0.0152
		(8.43)*				(1.59)
Ratio of higher education members			4.0215			4.1493
			(7.03)*			(7.37)*
Father's education				0.0296		0.0096
				(6.63)*		(0.87)
Log total assets					0.1880	0.1116
					(6.18)*	(2.68)*
8 region dummies controlled?	Yes	Yes	Yes	Yes	Yes	Yes
Wald χ^2	104.77	146.01	73.20	131.52	145.85	114.09
Prob > χ^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Pseudo R^2	0.1582	0.2624	0.7290	0.2208	0.2274	0.7426
Observations	651	651	651	651	651	651

Robust *z*-statistics in parentheses; *, **, *** denote significance at the level of 1, 5, and 10 percent, respectively.

the "*relevance*" condition, cov(Z_i, S)≠0. When all the family background variables are included in the model, however, father and mother's education turn out to be insignificant. This is because of high correlations between these variables (the last column of Table 2). This also suggests utilizing either of them as an IV at a time.

3.1 Ordinary Least Squares Estimates

Estimates of return to education using OLS estimator show that university graduated wage-earners earned 103 percent above that of the high school wage-earners (Table 3).[7] When the family background is further controlled for, the return slightly declines. Interestingly, only father's education and household assets have a direct effect on individual earnings, while mother's education and the ratio of higher education members do not have such effects on earnings. This sheds some light on the validity of mother's education and the ratio of higher education members, but casts doubt on father's education and assets when they are used as IVs. We shall come back to the test of IV validity in the following paragraphs.

Table 3: Return to Schooling Using OLS with and without Family Background Controls

Variables	(1)	(2)	(3)	(4)	(5)	(6)
University education	0.7134	0.6954	0.7116	0.6927	0.6561	0.6738
	(11.11)*	(10.65)*	(8.21)*	(10.72)*	(10.39)*	(7.96)*
Experience	0.0490	0.0500	0.0490	0.0550	0.0495	0.0535
	(4.11)*	(4.21)*	(4.12)*	(4.66)*	(4.27)*	(4.71)*
Experience squared	–0.0012	–0.0012	–0.0012	–0.0013	–0.0012	–0.0013
	(2.99)*	(2.94)*	(2.98)*	(3.25)*	(3.16)*	(3.35)*
Gender (male=1)	0.2386	0.2379	0.2387	0.2264	0.2274	0.2189
	(4.37)*	(4.35)*	(4.38)*	(4.15)*	(4.32)*	(4.12)*
Majority	0.5764	0.5743	0.5764	0.5784	0.4941	0.5013
	(2.44)**	(2.40)**	(2.43)**	(2.46)**	(2.05)**	(2.09)**
Urban	0.0935	0.0816	0.0932	0.0813	–0.0139	–0.0114

[7] The percentage is calculated for dummy variable in a semi-logarithmic regression as 100 x (e^{β}–1).

Table 3 continued

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	(1.79)***	(1.56)	(1.77)***	(1.57)	(0.24)	(0.20)
State sector	0.1351	0.1320	0.1350	0.1252	0.0846	0.0817
	(2.00)**	(1.93)***	(1.99)**	(1.85)***	(1.31)	(1.25)
Foreign sector	0.3243	0.3162	0.3243	0.3143	0.2846	0.2791
	(3.78)*	(3.56)*	(3.78)*	(3.67)*	(3.42)*	(3.26)*
Mother's education		0.0061				0.0010
		(1.06)				(0.15)
Ratio of higher education members			0.0050			–0.0843
			(0.03)			(0.54)
Father's education				0.0113		0.0078
				(2.47)**		(1.67)***
Log total assets					0.1242	0.1155
					(4.51)*	(4.20)*
Constant	0.6217	0.5832	0.6219	0.5018	–0.8156	–0.8080
	(3.05)*	(2.88)*	(3.04)*	(2.44)**	(2.14)**	(2.16)**
8 region dummies controlled	Yes	Yes	Yes	Yes	Yes	Yes
Observations	651	651	651	651	651	651
R-squared	0.48	0.48	0.48	0.49	0.50	0.50
F-value	30.73	29.23	29.88	30.24	30.22	27.24
Prob>*F*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Robust *t*-statistics in parentheses; *, **, *** denote significance at the level of 1, 5, and 10 percent, respectively. Private sector is set as a comparison base group for state and foreign sector.

3.2 Instrumental Variable Estimates

Estimates of return to four-year education using the Maximum Likelihood IV estimator (a joint estimation procedure) are presented in Table 4. Before presenting the estimated coefficients, we discuss the IV tests. The test results are presented in the bottom panel of Table 4. We emphasize the tests for the weak

Table 4: Return to schooling using IV estimator (LIML estimation)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
University education	1.0661 (4.10)*	0.6819 (9.16)*	2.0061 (3.25)*	1.9510 (5.27)*	1.5469 (5.95)*	0.7250 (9.41)*	0.6780 (9.09)*
Experience (year)	0.0693 (3.89)*	0.0495 (4.24)*	0.1179 (2.99)*	0.1150 (4.28)*	0.0942 (4.59)*	0.0517 (4.38)*	0.0493 (4.22)*
Experience Square	–0.0017 (2.87)*	–0.0012 (2.82)*	–0.0032 (2.39)**	–0.0031 (3.25)*	–0.0025 (3.32)*	–0.0012 (2.93)*	–0.0012 (2.80)*
Constant	0.6951 (3.45)*	0.7337 (3.97)*	0.6006 (2.24)**	0.6061 (2.34)**	0.6467 (2.82)*	0.7293 (3.92)*	0.7341 (3.98)*
F-value	22.09	30.83	13.05	12.68	17.69	30.65	30.64
Prob > *F*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Uncentered R^2	0.94	0.95	0.90	0.90	0.92	0.95	0.95
Root MSE	0.5841	0.5598	0.7731	0.7585	0.6626	0.5606	0.5598
Observations	651	651	651	651	651	651	651
Excluded instruments	Mother's education	Ratio of higher education members	Father's education	Log total assets	Mother's education and log total assets	Ratio of higher education members and log total assets	Ratio of higher education members and mother's education
Test for instruments jointly equal zero in the first stage, *F*-value [*p*-value in bracket]	29.08 [0.0000]	355.80 [0.0000]	9.97 [0.0017]	26.07 [0.0000]	22.64 [0.0000]	196.01 [0.0000]	178.19 [0.0000]
Partial R^2 of excluded instruments	0.0475	0.4837	0.0171	0.038	0.0722	0.4889	0.4843
Weak identification test (Kleibergen–Paap Wald rk *F*-statistic) [Stock–Yogo weak id test critical value at 10 percent maximal LIML size in bracket]	29.08 [16.38]	355.80 [16.38]	9.97 [16.38]	26.07 [16.38]	22.64 [8.68]	196.01 [8.68]	178.19 [8.68]
Hansen J statistic (overid test) [*p*-value in bracket]	Just-identified	Just-identified	Just-identified	Just-identified	4.696 [0.0302]	21.735 [0.0000]	2.731 [0.0984]
Endogeneity test of university education (*p*-value)	0.0742	0.3719	0.0019	0.0000	0.0000	0.2032	0.4240

Robust *z*-statistics in parentheses; *, **, *** denote significance at the level of 1, 5, and 10 percent, respectively. All the models controlled for gender, ethnicity, urban, economic sectors, and 8 geographical regions in Vietnam.

identification, the exclusion restriction or over-identification assumption, and endogeneity in the last three rows.

First, we consider models with only one instrument at a time in columns 1, 2, 3 and 4 of Table 4. The weak identification test accepts the hypothesis that the father's education variable is a weak instrument since the Kleibergen–Paap rank F statistic (9.97) is much smaller than the Stock–Yogo's weak identification critical value at 10 percent maximal LIML size (Stock and Yogo 2002). Furthermore, the *F*-statistic on the excluded instrument in the first stage is smaller than 10. This casts doubt on the validity of the father's education as an instrument, and suggests that this instrument is weak. The point estimates in this case are very biased and seriously inconsistent, thus, it is unable to predict the magnitude of the effects accurately when applying father's education as an instrument in IV models. In column 6, the Hansen test for exclusion restriction or over-identification rejects the validity of a combination of two instruments (the ratio of higher education members and total household assets). This implies at least one instrument in this combination is not valid, while in column 7 of Table 4 the combination of two instruments (the ratio of higher education members and mother's education) is accepted. This means at least one instrument in the combination is exogenous (Wooldridge 2002). The test results of these two combinations of instruments suggest that the ratio of higher education members (instrument) is exogenous.

The endogeneity test in the last row of Table 4 indicates that the hypothesis of endogeneity of university education is rejected when father's education, assets, and a combination of mother and assets are used as instruments. The weak identification test statistic in column 2, which strongly rejects the hypothesis of weak instrument of the ratio of higher education members, but the *p*-value of Hansen test (column 7) is not high enough to eliminate the suspicion of a strong instrument of mother's education since power of the test is low in the presence of a weak instrument, so adding a weak instrument may result in accepting the null hypothesis of over-identification just by increasing degrees of freedom (Baum et al, 2003). These test results suggest that father's education and household asset are invalid instruments, while mother's education is not a very strong instrument, and the ratio of higher education members is a good instrument. This finding contrasts with Arcand et al. (2004) who used a combination of father's and mother's education (parental education) as an instrument in a study of return to education in Vietnam for period 1992–1998. Mixing father's education and mother's education together may not properly reveal whose education plays a more important role in children's schooling and IV modelling.

To choose which model in either column 2 or 7 of Table 4, we look at F-statistic on the excluded instrument in the first stage, the *F*-value (355.8) for the model with one variable of the ratio of higher education members doubles that (178.2) of the model with two instruments in column 7. Additionally, one should choose model having valid instruments which has a minimum mean-square error (MSE) (Donald and Newey 2001). Furthermore, all the estimated coefficients, their standard errors, partial R^2 of excluded instruments, and MSE of these two model specifications are almost the same. This suggests that we can use either the models in columns 2 or 7 of Table 4.

Estimated return to university education varies largely. Using weak/invalid instrument of father's education and total household assets yields very highly upward biased results (columns 3 and 4, Table 4). Because father's education and assets are both correlated with schooling S_i (treatment participation) and positively correlated with earnings (see Tables 2 and 3), the estimates are highly upward biased (Angrist et al. 1996; Murray 2006; Staiger and Stock 1997; Stock and Yogo 2002). However, using either a good instrument of the ratio of higher education members or a combination of the ratio of higher education members and mother's education, the income premium for four-year university education is 97 percent (columns 2 and 7).[8] Interestingly, the estimated return using IV models with good instruments is almost the same with that based on the OLS model with family background controls (column 6 of Table 3). This implies that controlling for family background could do the similar job as using IV models as discussed in Card (1999), and there is no a serious ability bias in the OLS estimated return to the university education in Vietnam.

3.3 Treatment Effect Model estimates

In the above IV estimation with the joint estimation procedure, the normal distribution assumption of the first stage dependent variable was ignored even though it is a binary variable. The joint estimation procedure may be acceptable since the OLS still remain unbiased (Wooldridge 2002). However, the estimates may be woefully inefficient (Nichols 2009).

[8] The percentage is calculated for dummy variable in a semi-logarithmic regression as 100 x (e^{β} –1).

Treatment effect model may be an alternative approach to the problem of non-fulfilment of the normality assumption of binary endogenous variable of university education in the first stage. The binary endogenous regressor of university education is viewed as a treatment indicator, hence this estimation is considered as the treatment effect model (Heckman and Li 2004). Error terms of main equation (wage), and instrumental equation (schooling) are assumed to be correlated, i.e. $cov(u_i, v_i) = \rho\sigma^2$ where $u_i \sim NID(0, \sigma^2)$ and $v_i \sim N(0,1)$. This model offers an estimator similar to IV estimator in the case of a single binary endogenous variable, but it improves efficiency of estimates (Nichols 2009, p. 56). For the treatment effect model, the Lambda or inverse Mills' ratio is estimated in the first stage and then is included in the second stage to correct for selection bias. The identification is obtained by including factors (as of the valid instruments in columns 2 and 7 of IV models in Table 4) that influence university education participation but not earnings. The estimates are presented in Table 5. The estimated return to the four-year university education relative to that of high school education (103 and 101 per year for model with the ratio of higher education members-column 1, and a combination of the ratio of higher education members and mother's education-column 2, respectively) seems to accord with the estimates based on the previous IV models.

Table 5: Return to Schooling Using Treatment Effect Model

Controls in wage equation	(1)	(2)
University education (yes=1)	0.7113	0.7030
	(9.27)*	(9.07)*
Experience (year)	0.0489	0.0485
	(4.17)*	(4.12)*
Experience squared	–0.0012	–0.0011
	(3.02)*	(2.98)*
Gender (male=1)	0.2388	0.2392
	(4.42)*	(4.42)*
Majority (Kinh and Chinese=1)	0.5766	0.5774
	(2.46)**	(2.47)**
Urban (yes=1)	0.0941	0.0967
	(1.80)***	(1.84)***
State sector (yes=1)	0.1361	0.1401
	(1.92)***	(1.98)**
Foreign sector (yes=1)	0.3245	0.3257
	(3.82)*	(3.83)*
Constant	0.6220	0.6230
	(3.09)*	(3.10)*
8 region dummies controlled	Yes	Yes
Controls in selection equation (the first stage)		
Variables as of the wage equation	Yes	Yes
Mother's education		Yes
Ratio of higher education members	Yes	Yes
Wald χ^2	434.96	342.01
Prob > χ^2	0.0000	0.0000
Observations	651	651

Robust *z*-statistics in parentheses; *, **, *** denote significance at the level of 1, 5, and 10 percent, respectively. Private sector is set as a comparison base group for state and foreign sector.

4 Concluding Remarks

This paper utilizes a recent dataset to estimate the return to higher education in Vietnam. We demonstrate that controlling for individual ability (family background) in the wage equation slightly reduces the estimated return to higher education. This effect holds when a good instrument (ratio of higher education members) is used. Therefore, OLS estimates are upward-biased, but the bias is not too large to concern us. Additionally, the paper demonstrates that using invalid or weak instruments, such as father's education and household assets, leads to highly incorrect estimates of the return.

In 2008 income premium for university education in Vietnam is about 97 percent above the high school education. The return to higher education reached the average return (if annualized) of higher education in Asia (Psacharopoulos and Patrinos 2004). The estimated return seems to be robust to various estimators of OLS, IV and Treatment Effect. The return to university education that approximately equals that of Asia suggesting that labor market rewards higher-skilled workers more after a longer period of economic transition to a market economy.[9] The high premium for university education may be also attributed to university graduates' comparative advantage in the Vietnam labor market where only 5 percent of the country population hold university or post-graduate degrees (GSO 2010). However, the IV estimation in the current paper may provide the local average treatment effect (Imbens and Angrist 1994) for a younger cohort subsample (whose family background information is available in the sample) which may be higher than for older cohorts because of more up-to-date skills (Card and Lemieux 2001; Heckman and Li 2004). As a result, we do preserve a caution to interpret the finding for the entire sample of Vietnam.

Acknowledgements: We thank David Dapice, anonymous referee and the associate editor for their helpful comments and suggestions to previous versions, thanks also are due to participants of the 52nd New Zealand Association of Economists Annual Conference, Wellington 29th June to 1st July 2011 for their comments. Any remaining errors in the paper are those of the authors.

---

9 The returns to higher education in early economic transition were low in transitional economies but improved after a longer period of economic transition. This fact is observed in many transitional economies such as China and Eastern Europe (Heckman and Li 2004).

References

Angrist, J., and Krueger, A. (1991). Does compulsory schooling attendance affect schooling and earnings? *Quarterly Journal of Economics, 106*(4), 979–1014. http://ideas.repec.org/a/tpr/qjecon/v106y1991i4p979-1014.html.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444–455. http://www.jstor.org/stable/2291629

Arcand, J., d'Hombres, B., and Gyselinck, P. (2004). *Instrument choice and the returns to education: New evidence from Vietnam*. Retrieved from http://129.3.20.41/eps/lab/papers/0510/0510011.pdf

Ashenfelter, O., and Zimmerman, D. (1997). Estimates of the return to schooling from sibling data: Fathers, sons and brothers. *Review of Economics and Statistics, 79*(1), 1–9. http://ideas.repec.org/a/tpr/restat/v79y1997i1p1-9.html

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*(4), 453–470. http://ideas.repec.org/a/eee/labeco/v6y1999i4p453-470.html

Baum, C., Schaffer, M., and Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal, 3*(1), 1–31. http://ideas.repec.org/a/tsj/stataj/v3y2003i1p1-31.html

Belzil, C. (2007). The return to schooling in structural dynamic models: A survey. *European Economic Review, 51*(5), 1059–1105. http://ideas.repec.org/a/eee/eecrev/v51y2007i5p1059-1105.html

Butcher, K., and Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics, CIX*(3), 531–563. http://ideas.repec.org/a/tpr/qjecon/v109y1994i3p531-63.html

Card, D. (1994). *Earnings, schooling, and ability revisited* (Working Paper 4832). Cambridge, MA: National Bureau of Economic Research. http://ideas.repec.org/p/nbr/nberwo/4832.html

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 (p. 201–222). Toronto: University of Toronto Press. http://ideas.repec.org/p/nbr/nberwo/4483.html

Card, D. (1999). The casual effect of education on earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* (Vol 3A). Amsterdam and New York: North-Holland. http://ideas.repec.org/h/eee/labchp/3-30.html

Card, D., and Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *The Quarterly Journal of Economics*, *116*(2), 705–746. http://ideas.repec.org/a/tpr/qjecon/v116y2001i2p705-746.html

Doan, T., and Gibson, J. (2009). *Do returns to schooling go up during transition? The not so contrary case of Vietnam.* (Working Papers in Economics No. 09/08). Retrieved from University of Waikato, Department of Economics. ftp://mngt.waikato.ac.nz/RePEc/wai/econwp/0908.pdf

Doan, T., and Gibson, J. (2010). *Return to schooling in Vietnam during economic transition: Does the return reach its peak?* (MPRA Paper 24984). http://mpra.ub.uni-muenchen.de/24984/1/MPRA_paper_24984.pdf

Donald, S., and Newey, W. (2001). Choosing the Number of Instruments. *Econometrica*, 69(5), 1161–1191. http://ideas.repec.org/a/ecm/emetrp/v69y2001i5p1161-91.html

Glewwe, P., and Jacoby, H. (2004). Economic growth and the demand for education: Is there a wealth effect? *Journal of Development Economics, 74*(1), 33–51. http://ideas.repec.org/a/eee/deveco/v74y2004i1p33-51.html

Glewwe, P., and Patrinos, H. (1999). The role of the private sector in education in Vietnam: Evidence from the Vietnam Living Standards Survey. *World Development, 27*(5), 887–902. http://ideas.repec.org/a/eee/wdevel/v27y1999i5p887-902.html

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica, 45*(1), 1–22. http://ideas.repec.org/a/ecm/emetrp/v45y1977i1p1-22.html

GSO (2010). *Population and housing census 2009, 1 April 2009* (Complete report and major findings, General Statistical Office). Hanoi, Vietnam. http://www.gso.gov.vn/default_en.aspx?tabid=515&idmid=5&ItemID=9813

Harmon, C., and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review, 85*(5), 1278–1286. http://ideas.repec.org/a/aea/aecrev/v85y1995i5p1278-86.html

Hausman, J., and Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica, 44*(6), 1377–1398. http://ideas.repec.org/a/ecm/emetrp/v49y1981i6p1377-98.html

Heckman, J., and Li, X. (2004). Selection bias, comparative advantage and heterogeneous returns to education: Evidence from China in 2000. *Pacific Economic Review, 9*(3), 155–171. http://ideas.repec.org/p/hhs/ifauwp/2003_017.html

Hogan, V., and Rigobon, R. (2010, June). *Using unobserved supply shocks to estimate the returns to education.* Paper presented at the European Association of Labour Economists Conference at University College London, England. http://www.eale.nl/Conference2010/Programme/PaperscontributedsessionsI/add127862_ElgN5Qi516.pdf

Hoogerheide, L., Block, J.H., and Thurik, R. (2010). *Family background variables as instruments for education in income regressions: A Bayesian analysis* (Discussion Papers 10-075/3). Tinbergen Institute. http://www.tinbergen.nl/ti-publications/discussion-papers.php?paper=1627

Imbens, G., and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467–476. http://ideas.repec.org/a/ecm/emetrp/v62y1994i2p467-75.html

Keane, M. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics, 156*(1), 3–20. http://ideas.repec.org/a/eee/econom/v156y2010i1p3-20.html

Murray, M. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives, 20*(4), 111–132. http://ideas.repec.org/a/aea/jecper/v20y2006i4p111-132.html

Musgrove, P. (1979). Permanent household income and consumption in urban South America. *American Economic Review, 69*(3), 355–368. http://ideas.repec.org/a/aea/aecrev/v69y1979i3p355-68.html

Nichols, A. (2009). *Casual inference with observational data: Regression discontinuity and related methods in Stata*. Retrieved from http://repec.org/nasug2007/causal.pdf

Psacharopoulos, G., and Patrinos, H. (2004). Returns to investment in education: A further update. *Education Economics, 12*(2), 111–134. http://ideas.repec.org/a/taf/edecon/v12y2004i2p111-134.html

Staiger, D., and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557–586. http://ideas.repec.org/a/ecm/emetrp/v65y1997i3p557-586.html

Stock, J., and Yogo, M. (2002). *Testing weak instruments in linear IV research regression* (NBER Technical Working Paper 284). Cambridge, MA: National Bureau of Economic Research. http://ideas.repec.org/p/nbr/nberte/0284.html

VHLSS (2008). *Vietnam Household Living Standards Survey 2008* [Datasets and basic results]. General Statistical Office (GSO), Hanoi, Vietnam. http://www.gso.gov.vn/default_en.aspx?tabid=483&idmid=4&ItemID=9713

Wooldridge, J. (2002). *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: MIT Press. http://books.google.de/books?id=cdBPOJUP4VsC&printsec=frontcover&dq=Econometric+Analysis+of+Cross-Section+and+Panel+Data&hl=de&ei=ar9rTfSUK4at8gO8z_jxBw&sa=X&oi=book_result&ct=result&resnum=1&ved=0CDAQ6AEwAA#v=onepage&q&f=false

Yakusheva, O. (2010). Return to college education revisited: is relevance relevant? *Economics of Education Review, 29*, 1125–1142. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1692483