RESEARCH INSTITUTE OF
INDUSTRIAL ECONOMICS

IFN Working Paper No. 762, 2008

Comment on Education Returns of Wage Earners and Self-employed Workers

Henrik Jordahl, Panu Poutvaara and Juha Tuomala

Comment on Education Returns of

Wage Earners and Self-employed Workers[*]

Henrik Jordahl,[a] Panu Poutvaara,[b] and Juha Tuomala[c]

August 26, 2008

Abstract: In a recent paper, García-Mainar and Montuenga-Gómez (2005) apply the generalized IV model of Hausman and Taylor to estimate education returns of wage earners and the self-employed in Portugal and in Spain. Our examination reveals several problems which relate to the validity and documentation of the instrumental variables, as well as the robustness of the results.

Keywords: Education; Entrepreneurship; Human capital; EGIV estimator

JEL Classification: C23; I21; J31

1. Introduction

In a recent article, García-Mainar and Montuenga-Gómez (2005) estimate the returns to education for wage earners and the self-employed in Portugal and in Spain. Since this can be quite relevant for governments that seek to promote education and entrepreneurship, we provide a close examination of these findings in this note. Our investigation suggests that the results presented by García-Mainar and Montuenga-Gómez should be interpreted with great caution. The choice of instruments appears dubitable a priori and the authors have not provided the information necessary to replicate their study. An attempt to carry out a corresponding investigation with Finnish data illustrates that the results could be sensitive to the choice of instruments.

2. Brief summary of García-Mainar and Montuenga-Gómez (2005)

García-Mainar and Montuenga-Gómez (2005) report several findings, among them a lower return to (years of) education for self-employed relative to wage earners in Spain, and a negative return to a university education (relative to secondary education) for the self-employed in both countries.[1] The data set is obtained from the European Commission Household Panel for the period 1994–2000. For Spain there are 6,652 observations of 2,240 self-employed persons, and 28,651 observations of 8,365 wage earners. For Portugal there are 7,181 observations on 2,225 self-employed and 28,914 observations on 6,968 wage earners. The empirical investigation focuses on the Efficient Generalized Instrumental Variables (EGIV) estimator of Hausman and Taylor (1981).

The EGIV estimator divides the explanatory variables into one exogenous and one endogenous set. In a series of emails, we have asked the authors to specify the variables that they included as exogenous and which they included as endogenous. The corresponding author has replied that they cannot find this information, and that they used different instruments (determined by tests) in each regression. The only available information in the article is that experience and education "and some other variables" (p 167) are considered endogenous, and that a Hausman test is used to identify "the regressors that are strictly exogenous" (p 168).

3. Discussion of the instruments

The EGIV estimator is designed for situations where none of the explanatory variables are correlated with the idiosyncratic error, but some of them are correlated with the individual-level random effects. Although Hausman and Taylor (1981: 1378) emphasize that their method allows for testing for regressors that are exogenous to the individual random effects, additional strategies are needed to support the economic plausibility of this assumption.[2] García-Mainar and Montuenga-Gómez (2005) rely exclusively on the Hausman test to

---

[1] Especially the latter finding is surprising. Van der Sluis and van Praag (2004) find the opposite, i.e. higher returns to (years of) education for the self-employed than for wage earners. Van der Sluis and van Praag note that when it comes to empirical strategies, the literature on the returns to education for employees has been technically more sophisticated than the literature on returns to education for the self-employed. For example, the selection into self-employment is often not accounted for.
[2] Cf. Angrist and Krueger (2001) for a general discussion, or the strategies employed by Levitt (1996) to determine the validity of his instruments.

establish that some of their regressors are exogenous and can be used as instruments. This procedure is problematic, especially when several potential instruments share a mutual vulnerability to being invalid. As an example, which we will return to, one may be led to classify marriage as exogenous but divorce as endogenous.

In general, it is far from clear that the potential instruments considered by García-Mainar and Montuenga-Gómez (gender, marital status, seniority, occupation, sector of activity, and whether the worker has taken some training course) are uncorrelated with the unobserved individual effect (which in this literature is often interpreted as innate ability). In a related study, van der Sluis and van Praag (2004: 195) bring up Card's (1999) doubts about the validity of parental background variables as identifying instruments for education. According to Wooldridge (2002: 328), the identifying assumptions which have been made when applying the EGIV estimator to the returns to education are not especially convincing.

4. Data and empirical investigation

In order to test the robustness of the findings with respect to the choice of instruments, we have made our best efforts to follow the empirical specifications in García-Mainar and Montuenga-Gómez (2005). Using a large administrative data base from Finland, we show that statistical testing alone does not identify a unique set of exogenous variables that can be used as instruments. And depending on the choice of instruments, we obtain different estimates of the returns to education.

Our empirical investigation is based on micro-level panel data constructed by Statistics Finland (Employment Statistics) for the period 1993−1997. The data cover all 350,000 individuals aged 12–75 in 1997 with permanent residence in Finland. Statistics Finland has constructed the data by combining information from several administrative registers.

Statistics Finland defines occupational status in the following way. First of all, students, retirees and people in military or civil service during the last week of the year are defined to be outside of the labor force. Of the remaining population, people registered as unemployed job searchers during the last working day of the year are counted as unemployed. Among the labor force, people with an employment contract or who contribute towards a self-employed pension during the last week of the year are counted as employed. Based on their source of income, employed persons are classified either as workers or as self-employed. The distinction between the two types of income is based on mandatory pension contributions, which distinguish between income earned as a worker and income from self-employment. Workers have higher income from employment than from self-employment. The self-employed have higher income from self-employment than from employment. Using this definition, the sample contains 280,480 workers and 22,470 self-employed (where the members of both groups are classified either as a worker or as self-employed during the whole period 1993−1997). Our variables are described in Table 1.

Table 1. Descriptive statistics

	Workers				Self-employed			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Income	24467.35	16210.51	167.94	1984189	28126.60	69720.30	167.94	8808356
Experience	22.73	5.78	4	35	24.34	5.36	7	35
Primary education	0.22	0.41	0	1	0.30	0.46	0	1
Secondary education	0.55	0.50	0	1	0.59	0.49	0	1
Higher education	0.23	0.42	0	1	0.11	0.31	0	1
Female	0.51	0.50	0	1	0.32	0.46	0	1
Married	0.66	0.47	0	1	0.72	0.45	0	1
Divorced	0.11	0.31	0	1	0.10	0.30	0	1
Single parent	0.06	0.24	0	1	0.04	0.19	0	1
Disabled	0.001	0.03	0	1	0.001	0.02	0	1
Finnish language	0.94	0.24	0	1	0.94	0.24	0	1
Swedish language	0.05	0.23	0	1	0.05	0.23	0	1
Foreign language	0.01	0.07	0	1	0.004	0.07	0	1
Uusimaa region	0.31	0.46	0	1	0.21	0.41	0	1
Southern Finland	0.35	0.48	0	1	0.37	0.48	0	1
Middle Finland	0.23	0.42	0	1	0.29	0.45	0	1
Northern Finland	0.11	0.31	0	1	0.13	0.34	0	1
Working months	11.41	2.33	0	12	11.48	2.35	0	12

Source: Statistics Finland, Employment Statistics.

We report results from EGIV specifications that closely resemble the specifications preferred by García-Mainar and Montuenga-Gómez. According to the Hausman test, different sets of regressors can be used as instruments. In Table 2 we present estimates from two specifications for workers that pass the Hausman test and which therefore − according to García-Mainar and Montuenga-Gómez − should give reliable estimates of the returns to education. The example demonstrates that the estimated coefficients for secondary and higher education (the educational categories that García-Mainar and Montuenga-Gómez use) are both sensitive to the choice among similar instruments (married or divorced). Relative to primary education, the estimated returns to secondary education can be either positive or negative, and the estimated coefficient for higher education varies from 0.31 to 2.78. We do not obtain as striking differences for the self-employed as for workers, but the general problem is the same, casting doubt on any finer comparison of estimates.

Table 2. Returns to education for workers

Variable type	Variable name	Specification 1	Specification 2
Time-invariant endogenous	Secondary education	3.180***	–0.579***
		(1.034)	(0.161)
	Higher education	2.769***	0.306***
		(0.659)	(0.112)
Time-invariant exogenous	Female	–0.278***	–0.292***
		(0.017)	(0.006)
	Swedish language	0.186**	–0.040**
		(0.073)	(0.016)
	Foreign language	1.612***	–0.386***
		(0.557)	(0.089)
Time-varying exogenous	Married	–0.031***	
		(0.004)	
	Divorced		–0.009**
			(0.005)
	Single parent	0.001	–0.000
		(0.004)	(0.004)
	Disabled	–0.186***	–0.203***
		(0.021)	(0.020)
Time-varying endogenous	Married		–0.032***
			(0.004)
	Divorced	–0.009	
		(0.005)	
	Experience	0.015***	0.014***
		(0.003)	(0.004)
	Experience squared	–0.063***	–0.063***
		(0.003)	(0.003)
Observations		234,214	234,214
Hausman test		4.75	7.94

Notes: EGIV estimator for workers. The dependent variable is logarithm of annual gross income. Standard errors in parentheses. Primary education is the reference category for education. Finnish is the reference language category. The specifications also include unreported variables: a constant term, dummies for year and region, and number of working months. The Hausman test tests the column's EGIV specification against a corresponding fixed-effects model. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

5. Conclusions

Estimating the returns to education is a tough econometric challenge, which we don't claim to have overcome in this note. A major problem is that individual educational choices are correlated with individual characteristics which affect earnings directly. At the same time, reliable estimates of the returns to education would be a highly relevant input to education policy. Depending on the relative returns to education for workers and the self-employed, governments may want to promote the education of people with entrepreneurial talent or make entrepreneurial incentives contingent on education. Considering this policy relevance, we have made our best efforts to examine some of the surprising findings of García-Mainar and Montuenga-Gómez (2005), including the negative returns to a university education for the self-employed (relative to secondary education).

García-Mainar and Montuenga-Gómez (2005: 165) state that their application of the EGIV estimator makes it possible to "avoid the insecurity associated with the choice of suitable instruments." In this comment we have demonstrated that this claim is overly optimistic. The identifying assumptions are not especially convincing, and García-Mainar and Montuenga-

Gómez have not specified which variables they classify as exogenous. Since our own estimates based on Finnish data are sensitive to this classification, the lack of information appears worrisome and empirically relevant.

References

Angrist, J. D. & Krueger, A. B. (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4), 69–85.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics*, Vol. 3A. Amsterdam, North-Holland.

García-Mainar, I. & Monteuenga-Gómez, V. M. (2005). Education returns of wage earners and self-employed workers: Portugal vs. Spain. *Economics of Education Review,* 24(2), 161−170.

Hausman, J. A. & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica* 49(6), 1377−1398.

Levitt, S. D. (1996). The effect of prison population size on crime rates: evidence from prison overcrowding litigation. *Quarterly Journal of Economics*, 111(2), 319–351.

Van der Sluis, J. & van Praag, M. (2004). Economic returns to education for entrepreneurs: the development of a neglected child in the economics of education. *Swedish Economic Policy Review,* 11(2), 183−225.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.