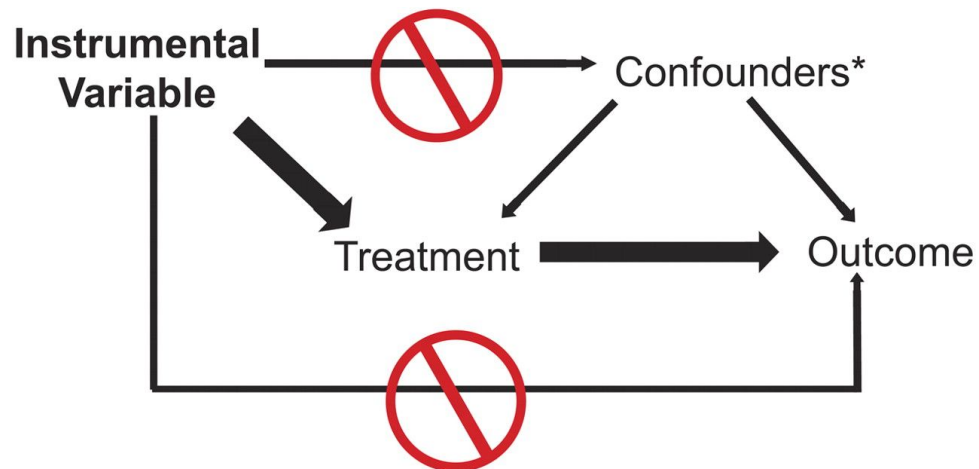# LATE TO THE PARTY: INVESTIGATING INSTRUMENT VARIABLES FOR EDUCATION

Cassandra Pengelly

20346212

Econometrics 871: Cross Section Project

Stellenbosch University

July 2021

Table of Contents

1. Introduction

The return to education is a cornerstone topic in labour economics and is of particular interest to policymakers. This is no different in South Africa, especially given the significant income inequality, which is often claimed to be strongly linked to differences in education. However, wages cannot simply be regressed on education because there is likely endogeneity present. This arises from the problem of there being an omitted variable, where education and wages are both correlated with a variable in the error term. One variable of this nature that has been extensively studied is innate ability (Hertz (2003)). Returns to schooling could be biased upwards if ability is positively correlated with

both income and education. However, as Lang (1993) assesses, the overall findings of research on the impact of ability bias are inconclusive. In fact, Lang (1993: 1) notes that several papers find that returns to education are *downwardly* biased.

This paper will investigate whether OLS estimates are biased in the South African case by estimating the return to education using four different instrumental variable estimators: the first two exploit parents' education, and the other two use parents' occupations. These instruments estimate a 'local average treatment effect', which - this paper will argue - is more appropriate than OLS estimators for analysing the returns to education for South Africa. These instruments are tested for strength and validity and then implemented on the NIDS Wave 5 data set. This essay[1] is structured as follows: section 2 details the data set used and discusses the descriptive statistics. Section 3 outlines the methodology and argues that the LATE assumptions for the instrumental variables hold. Following this, section 4 presents the regression results and evaluates the robustness of the estimators used to obtain a causal effect. The final section, 5, concludes.

2. Data

mention survey design for NIDS *International standard classification of occupations* (2012) for occupation classification
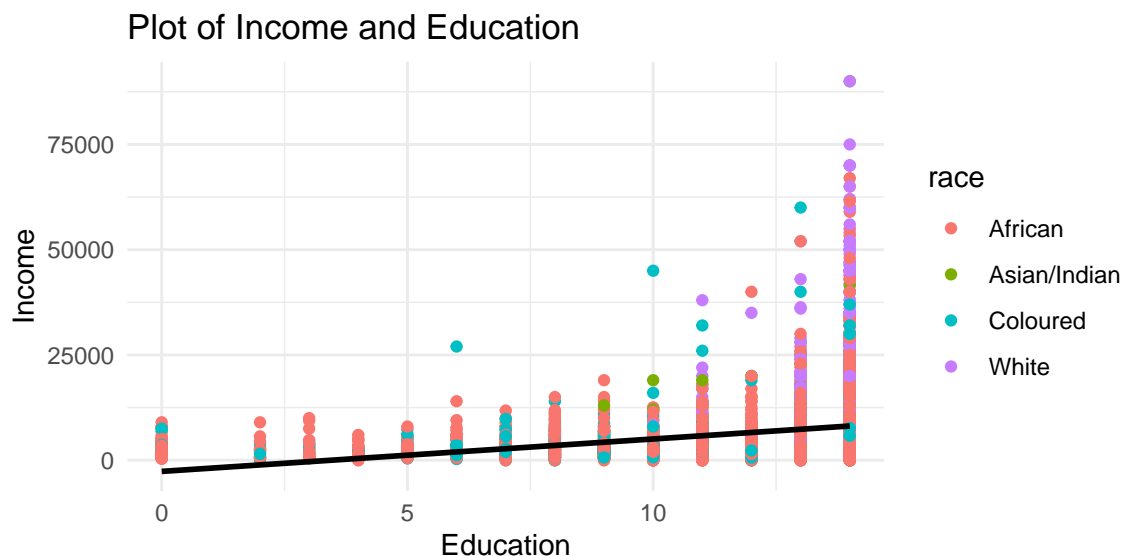

Figure 2.1: Income and Education Relationship

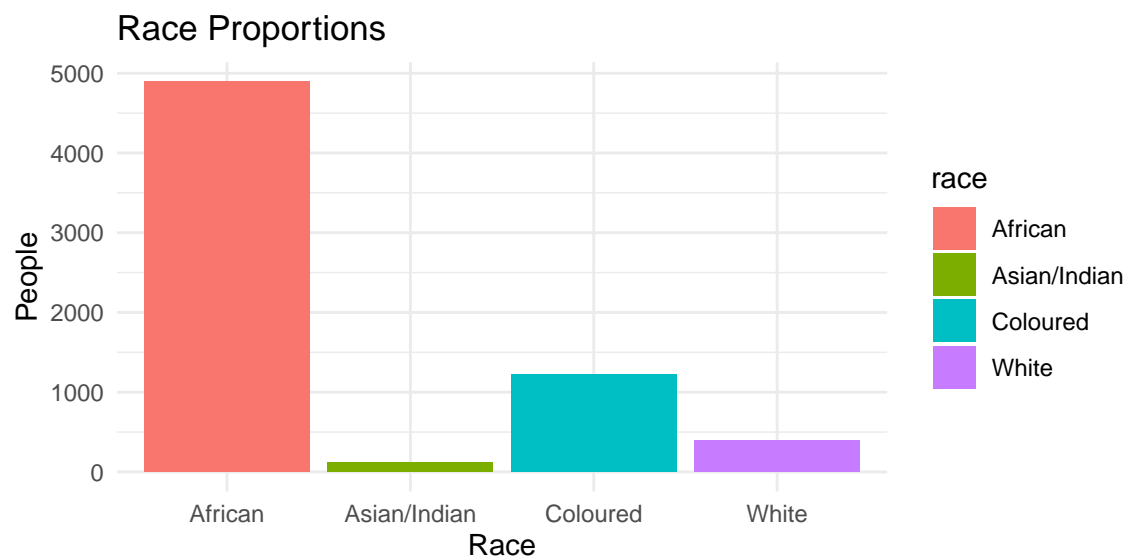[1]This essay was written in R using the Texevier package by Katzke (2017)

Figure 2.2: Income and Education Relationship

3. Methodology

This project makes use of 4 different instrument variables: fathers occupation, mothers occupation, mothers education and fathers education

4. Results

4.1. Specification Tests

Table 4.1: Regressions: Father's Education as Instrument

	OLS	2SLS FE	2SLS ME	2SLS FO	2SLS MO
age	0.047 ***	0.040 ***	0.049 ***	0.056 ***	0.072 ***
	(0.009)	(0.009)	(0.009)	(0.012)	(0.015)
age2	-0.000 **	-0.000	-0.000 *	-0.000 *	-0.000 **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
education	0.194 ***	0.309 ***	0.280 ***	0.327 ***	0.387 ***
	(0.005)	(0.016)	(0.015)	(0.027)	(0.028)
Coloured	0.072	0.135 **	0.124 **	0.146 **	0.248 ***
	(0.039)	(0.043)	(0.040)	(0.053)	(0.062)
Asian/Indian	0.388 ***	0.214 *	0.205	0.174	0.003
	(0.095)	(0.105)	(0.112)	(0.130)	(0.211)
White	0.689 ***	0.303 ***	0.408 ***	0.263 **	0.163
	(0.050)	(0.074)	(0.071)	(0.098)	(0.105)
Male	0.464 ***	0.490 ***	0.497 ***	0.526 ***	0.474 ***
	(0.027)	(0.029)	(0.028)	(0.039)	(0.045)
N	3376	3376	3613	2178	1747
R2	0.456	0.364	0.368	0.337	0.247

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

wn possibly long tables. Note that the following will fit on one page if it can, but cleanly spreads over multiple pages:

5. Conclusion

You should NOT write an extensive essay about the topic of interest, nor should you conduct an extensive literature review behind it: you should only reference articles that support the motivation for your econometric strategy (in the quest of finding a causal interpretation of a relationship that you are modeling). Provide descriptive statistics and graphs to aid in your discussion. The assignment should be between 3-6 pages, including your tables and figures.

You are required to go beyond simply estimating and presenting your results, but to convince the reader of their robustness by presenting alternative specifications. You should apply different estimators and specifications where possible. Discuss the shortcomings of the estimators in obtaining a causal effect and argue why your strategy is the best available to obtain a causal effect that satisfies relevant assumptions.

Determine an effect of interest and find instrumental variables to estimate it causally. If possible, use more than one instrument. Given the LATE assumptions, try and explain why your results differ and which is likely to represent the causal effect you are looking for. Conduct sufficient specification tests to establish whether you are overidentifying your instrument set or whether the IVs shift the estimates enough to indicate that OLS would be inconsistent.

Econometri
Students

References

10 Hertz, T. 2003. Upward bias in the estimated returns to education: Evidence from south africa. *American Economic Review.* 93(4):1354–1368.

International standard classification of occupations: Structure, group definitions and correspondence tables. 2012. (ISCO - 08). International Labour Office, Geneva; International Labour Organization.

Katzke, N.F. 2017. *Texevier: Package to create elsevier templates for rmarkdown.* Stellenbosch, South Africa: Bureau for Economic Research.

Lang, K. 1993. *Ability bias, discount rate bias and the return to education.* (MPRA Paper). University Library of Munich, Germany. [Online], Available: https://EconPapers.repec.org/RePEc:pra:mprapa:24651.

Appendix

Appendix A

Some appendix information here

Appendix B