



ELSEVIER

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Misreported schooling, multiple measures and returns to educational qualifications[☆]

Erich Battistin^{a,b,c}, Michele De Nadai^d, Barbara Sianesi^{e,*}

^a Queen Mary University, UK

^b IRVAPP, Italy

^c IZA, Germany

^d University of Padova, Italy

^e Institute for Fiscal Studies, UK

ARTICLE INFO

Article history:

Received 17 September 2012

Received in revised form

3 March 2014

Accepted 8 March 2014

Available online xxx

JEL classification:

C10

I20

J31

Keywords:

Misclassification

Mixture models

Returns to educational qualifications

Treatment effects

ABSTRACT

We consider the identification and estimation of the average wage return to attaining educational qualifications when attainment is potentially measured with error. By exploiting two independent measures of qualifications, we identify the extent of misclassification in administrative and self-reported data on educational attainment. The availability of multiple self-reported educational measures additionally allows us to identify the temporal patterns of individual misreporting errors across survey waves. We provide the first reliable estimate of a highly policy relevant parameter for the UK, namely the return from attaining any academic qualification compared to leaving school at the minimum age without any formal qualification. We identify returns to qualifications under two alternative settings: a strong ignorability assumption and an exclusion restriction. All these results are obtained by casting the identification problem in terms of a mixture model, and using a semi-parametric estimation approach based on balancing scores, which allows for arbitrarily heterogeneous individual returns.

© 2014 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

The estimation of the return to education has probably become the most explored and prolific area in labour economics.¹ If a

continuous years-of-schooling measure of education is affected by recording errors, standard results based on classical measurement error show that OLS estimates of the return to an additional year of education are downward biased, while appropriate IV methods applied to the linear regression model provide consistent estimates. If however a categorical qualification-based measure of education is affected by errors, any such error will necessarily vary with the true level of education, so that the assumption of classical measurement error cannot hold (see, for example Aigner, 1973). In this case, OLS estimates of the returns to qualifications are no longer necessarily downward biased, and the IV methodology cannot provide consistent estimates (see, for example Bound et al., 2001).

To date, empirical evidence on the importance of misreporting and returns to discrete educational levels is restricted to higher education in the US, where it was in fact shown that measurement error might play a non-negligible role (see Kane et al., 1999; Black

[☆] This paper benefited from helpful discussions with Enrico Rettore and comments by audiences at Policy Studies Institute (London, September 2005), ADRES Conference on “Econometric Evaluation of Public Policies: Methods and Applications” (Paris, December 2005), Franco Modigliani Fellowship Workshop (Rome, February 2006), the 5th ESRC Research Methods Festival (Oxford, 2012) and European University Institute (Florence, 2013). Two editors and three referees further provided many insightful suggestions. Financial support from the ESRC under the research grant ES/I02574X/1 is gratefully acknowledged. Address for correspondence: Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK; School of Economics and Finance, Queen Mary, University of London (Mile End Campus), Mile End Road, London E1 4NS, UK and Department of Economics and Management, Via del Santo 33, 35123 Padova, Italy.

* Correspondence to: Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK. Tel.: +44 20 7291 48 00.

E-mail addresses: e.battistin@qmul.ac.uk (E. Battistin), denadai@stat.unipd.it (M. De Nadai), barbara_s@ifs.org.uk (B. Sianesi).

¹ Policymakers too have shown increasing interest, with estimated returns feeding into debates on national economic performance, educational policies, or

the public funding of education. For an extensive discussion of the policy interest of the individual wage return from education, see Blundell et al. (2005a).

<http://dx.doi.org/10.1016/j.jeconom.2014.03.002>

0304-4076/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

et al., 2003; Lewbel, 2007). For the UK there are no estimates of the returns to educational qualifications that adequately correct for measurement error. This is of great concern, given the importance of focusing on discrete levels of educational qualifications² and given the widespread misconception amongst UK researchers and policymakers that the bias from measurement error (believed to be downward) and the so-called “ability bias” largely cancel each other out (Dearden, 1999; Dearden et al., 2002; McIntosh, 2006; Department for Business, Innovation and Skills, 2011).

This paper focuses on the returns to educational qualifications when attainment is potentially misrecorded³ and offers a two-fold contribution. First, it provides the first reliable estimates of a highly policy relevant parameter for the UK, namely the return from attaining any academic qualification compared to leaving school at the minimum age without any formal qualification. Secondly, it estimates misclassification probabilities and patterns of misclassification, including the temporal correlations in misreporting by individuals across survey waves.

In order to overcome the bias introduced by misreported educational qualifications and to achieve point identification of the returns, (at least) two categorical reports of qualifications need to be available for the same individuals, both potentially affected by reporting error but coming from independent sources (for the proof of non-parametric identification, see Mahajan, 2006; Lewbel, 2007; Hu, 2008). Repeated measurements on educational qualifications are typically obtained by combining complementary datasets, for example exploiting administrative records and information self-reported by individuals. An additional appealing feature of this approach is that it provides estimates of the extent of misclassification in each educational measures, which is often of independent interest. As we will see, our case study employs additional measures on top of the minimum number required to achieve identification, thus allowing us to shed light on features of the measurement error process not unveiled in past studies.⁴

We focus on the return from attaining any academic qualification compared to leaving school at the minimum age of 16 without any formal qualification (the latter being akin to dropping out of high-school in the US). This return captures all the channels in which the initial decision to attain academic qualifications at the school-leaving age impacts on wages later on in life, in particular the contribution that attaining such qualifications gives to subsequent educational attainment. The policy relevance of this parameter for the UK is additionally highlighted by the finding that the main effect of changes in compulsory schooling was not to increase the length of schooling, but rather to induce individuals to

leave school with an academic certification (Del Bono and Galindo-Rueda, 2004).

We rely on detailed longitudinal data for the male sample of the British National Child Development Survey (NCDS), which allows us to identify returns under two alternative settings. This data appears particularly suited to support the strong ignorability assumption that the observables are enough to control for the endogeneity of educational choices. This is because, in addition to detailed family background and school type variables, the NCDS contains extensive measures of both cognitive and non-cognitive traits at early ages. Under the strong ignorability assumption we also explore how the biases from measurement error and from omitted variables interact in the estimation of returns to educational qualifications, providing simple calibration rules that policy makers can apply to nationally representative datasets relying on self-reported qualifications and with no information on individual ability and family background (e.g. the Labour Force Survey). Alternatively, we identify returns for a specific group (the “compliers”) exploiting an exclusion restriction.

Using the unique nature of our data we identify the extent of misclassification in *three* different data sources on educational qualifications: administrative school files, self-reported information very close to the dates of completion of the qualification, and self-reported recall information ten years later. To this end, we combine multiple measurements self-reported by individuals in the NCDS with administrative data on qualifications coming from school records. Compared to the existing articles in the literature, the availability of multiple self-reported measurements introduces a certain degree of over-identification, which allows us to isolate the extent of misreporting in school files from that of individuals, while allowing for persistence in the propensity to misreport across self-reported measurements. Thus, our setup gives us the unique chance of assessing the *temporal patterns* of misreporting errors across survey instruments and of decomposing misreporting errors into a systematic component linked to individuals’ persistent behaviour and into a transitory part reflecting survey errors that occur independently of individuals in each cross-section survey wave.

On the methodological front we propose a semi-parametric estimation approach based on balancing scores and mixture models. We cast the estimation problem in terms of a mixture model, which combined with the propensity score defines a semi-parametric procedure that allows for arbitrarily heterogeneous individual returns. Given that the misclassification problem can be stated in terms of finite mixtures with a known number of components, we find this approach particularly suited for the case at hand (Hui and Walter, 1980, also propose an approach to misclassification of a dichotomous variable based on maximum likelihood). The general identification problem in the case of two reports has been considered, amongst others, by Kane et al. (1999), Black et al. (2000), Mahajan (2006), Lewbel (2007) and Hu (2008). We build upon these papers, and in particular upon Hu (2008), to show that the components of the mixture model are non-parametrically identified given the setup we consider. Specifically, we first show that all the quantities of interest are non-parametrically identified from the data through the availability of our repeated measurements on educational qualifications. The conditions required for this result are very general in nature, or at least are as restrictive as those commonly invoked in the relevant literature on misclassification. We then proceed with estimation, drawing from the statistical literature on finite mixtures to propose a flexible strategy based on Bayesian modelling.

We report a number of findings of substantive importance. All our results pertain to males only of the NCDS birth cohort. Individuals are found to be appreciably *less* accurate than transcript files when they do not have any academic qualification, but slightly

² In the UK educational system, individuals with the same number of years of schooling can have different educational outcomes; this not only obfuscates the interpretation of the return to one additional year, but imposing equality of yearly returns across educational stages was found to be overly restrictive (see Blundell et al., 2005b).

³ Misrecorded education can arise from data transcript errors, as well as from misreporting, whereby survey respondents may either over-report their attainment, not know if the schooling they have had counts as a qualification or simply not remember.

⁴ An alternative to dealing with misclassification bias which does not require repeated educational measures is to derive bounds on the causal effect of interest by making *a priori* assumptions on the misclassification probabilities (see, for example Kreider and Pepper, 2007; Molinari, 2008; Kreider et al., 2012). In previous work (Battistin and Sianesi, 2011) we based the analysis on one self-reported measure and could only provide partial identification of returns under strong ignorability. The bounds we suggested can be derived allowing for arbitrarily heterogeneous individual returns, while the corresponding sensitivity analysis is easy to implement and can provide an often quite informative robustness check. By contrast, in the current paper we achieve point identification not only of the returns, but also of the distribution of measurement error. Additionally, we discuss identification of policy effects under two scenarios that in our context seem appropriate (under strong ignorability and in the presence of a valid exclusion restriction).

Download English Version:

<https://daneshyari.com/en/article/5095964>

Download Persian Version:

<https://daneshyari.com/article/5095964>

[Daneshyari.com](https://daneshyari.com)