Misreported Schooling and Returns to Education: Evidence from the UK*

Erich Battistin
University of Padova and Institute for Fiscal Studies

Barbara Sianesi
Institute for Fiscal Studies

first draft September 2004
this version 30th March 2005

* Preliminary and incomplete *

Abstract

This paper aims to provide a number of contributions of policy, practical and methodological interest to the study of the returns to educational qualifications. First, we derive estimates of the returns to different qualifications for the UK that allow for the possibility of misreported attainment using data from the British National Child Development Survey. To date, any major empirical evidence on the importance of this issue is restricted to the US, where it was shown that errors might indeed play a non-negligible role. Second, we aim to provide the academic and policy community with estimates of the accuracy of commonly used types of data on educational attainment: administrative files, self-reported information close to the date of completion of the qualification and recall information ten years after completion. Third, by using the unique nature of our data, we assess how the biases from measurement error and from omitted ability and family background variables interact in the estimation of returns. We intend to produce simple calibration rules to allow policymakers to use nationally representative data, such as the Labour Force Survey, that totally rely on self-reported information on qualifications and contain little or no information on individual ability and family background. We rely throughout on a unified framework of general applicability to study the impact of misreported treatment status on the estimation of causal treatment effects.

JEL Codes: C10, I20, J31.

Keywords: Measurement Error, Misclassification, Programme Evaluation, Returns to Educational Qualifications, Treatment Effects.

*This paper benefited from helpful discussion with Andrew Chesher and comments by audiences at Padova (July 2004), "XIX Italian Conference of Labour Economics" (Modena, September 2004) and "Cemmap Metrics Lunch Seminar" (London, December 2004). Financial support from the ESRC under the research grant RES-000-22-1163 is gratefully acknowledged. Address for correspondence: Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE - UK and Department of Statistics, Via Cesare Battisti 243-5, 35123 Padova - Italy. E-mail: erich_b@ifs.org.uk and barbara_s@ifs.org.uk.

Contents

1 Introduction

This paper considers the identification and estimation of the returns to educational qualifications when educational attainment is potentially misreported, with an application to UK data.

The measurement of the return to education, that is of the individual wage gain from investing in more education, has become probably the most explored and prolific area in labour economics. Policymakers too have shown increasing interest, with estimated returns feeding into debates on national economic performance, educational policies, or the public funding of education. Reliable measures of the impact of education on individual earnings are in fact needed to establish whether it is worthwhile for individuals to invest in more education (and in which type), to compare private and social returns to education, or to assess the relative value that different educational qualifications fetch on the labour market. For an extensive discussion of the policy interest of the individual wage return from education, see Blundell, Dearden and Sianesi (2004).

As to the measurement of education, a first issue is whether we can summarize it in the single, homogeneous measure of years of schooling. Although particularly convenient, this 'one-factor' model is *a priori* quite restrictive, in that it assumes that the returns increase linearly with each additional year, irrespective of the level and type of educational qualifications the years refer to. When interested in a wide range of education levels with potentially very different returns, a more adequate framework is the 'multiple-factor' model, in which different educational levels are allowed to have separate effects on earnings. An interesting sub-case is the single-treatment specification, which focuses on the return to a specific educational level, such as undertaking higher education compared to not doing so.[1]

In the schooling system in the US, grades generally follow years, and it has long been argued that the returns to an additional year are reasonably homogeneous (see for example Card, 1999). In the UK and other European countries, however, there are alternative nationally-based routes leading to quite different educational qualifications, and the importance of distinguishing between different types of qualifications is widely accepted. Blundell, Dearden and Sianesi

[1]Another limitation of using years of schooling as a measure of educational attainment is that whilst in the UK and the US students have increasingly been studying part-time, most surveys do not provide information on the mode of study, and only ask about years of full-time study or the age a person first left full-time education.

(2005) highlight the potential shortcoming of the 'one-factor' model when applied to the UK's educational system, in which individuals with the same number of years of schooling have quite different educational outcomes. Not only would this obfuscate the interpretation of the return to one additional year, but imposing equality of yearly returns across educational stages was found to be overly restrictive.

A second important issue as to the measurement of education – and the one object of this paper – is the possibility of errors in recorded education and its consequences on the estimated returns. Misrecorded education could arise from data transcript errors, as well as from misreporting: survey respondents may either lie, not know if the schooling they have had counts as a qualification or simply not remember.

With the continuous years-of-schooling measure of education, standard results based on classical measurement error show that OLS estimates are downward biased and that appropriate IV methods applied to the linear regression model provide consistent estimates. Indeed, the trade-off between attenuation bias due to measurement error and upward bias due to omitted variables correlated with both schooling and wages (the so-called 'ability bias') has been at the heart of the early studies on returns to years of schooling. The received wisdom has traditionally been that ability bias and measurement error bias largely cancel each other out (for a review see in particular Griliches, 1977, and Card, 1999; for a recent UK study see Bonjour *et al.*, 2003).

With the categorical qualification-based measure of education, however, any measurement error in educational qualifications will vary with the true level of education. Individuals in the lowest category can never under-report their education level and individuals in the top category cannot over-report, so that the assumption of classical measurement error cannot hold (see, for example, Aigner, 1973). In the presence of such non-classical measurement error, OLS estimates are not necessarily downward biased, so that the cancelling out of the ability and measurement error biases cannot be expected to hold in general. Moreover, the IV methodology cannot provide consistent estimates of the returns to qualifications (see, for example, Bound, Brown and Mathiowetz, 2001). The implications of this problem, although of longstanding concern amongst researchers, were recently discussed by Kane, Rouse and Staiger (1999) for the estimation of returns to education in the US.

Two approaches have been developed to overcome the bias induced by misreported educational qualifications. A first possibility is to derive bounds on the returns by making *a priori* assumptions on the misclassification probabilities (see e.g. Manski, 1990, and Molinari, 2004). In most instances, such restrictions on the nature of reporting errors can be obtained by looking at results from previous research and/or behavioural theories that seem reasonable for the phenomenon under investigation. This approach only allows *partial identification* of the parameters of interest, and the qualitative information this would provide for a study on returns to qualifications would need to be established on a case-by-case basis.

The alternative approach is more demanding in terms of data requirements but, if feasible, it allows *point identification* of the returns. An additional appealing feature is that it provides direct estimates of the measurement error in the educational measures, which may often be of independent interest. What is needed is (at least) two categorical reports of educational qualifications for the same individuals, both potentially affected by reporting error but independent of each other. Kane, Rouse and Staiger (1999) and Black, Berger and Scott (2000) have developed a procedure to simultaneously estimate the returns to qualifications and the distribution of reporting error in each educational measure. Repeated measurements on educational qualifications are typically obtained by matching complementary datasets, for example exploiting administrative records and information self-reported by individuals.

To date, empirical evidence on the importance of these issues is restricted to the US, where it was in fact shown that measurement error might play a non-negligible role, as we review in the section below. For the UK there are no estimates of the returns to educational qualifications that adequately correct for measurement error. This is of great concern, in view of the stronger emphasis on returns to discrete levels of educational qualifications in the UK and given the widespread belief amongst UK researchers and policymakers that ability and measurement error biases still cancel out (Dearden, 1999b, Dearden *et al.*, 2002, and McIntosh, 2004).

This paper provides a number of *new* contributions of considerable policy and practical relevance, as well as of methodological interest.

First, we provide reliable estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment. To this end we use an extremely rich dataset - the National Child Development Survey (NCDS) - which allows us both to directly control for ability and family background influences and to exploit a number of repeated measurements of individual educational attainment.

Second, we identify the extent of measurement error in *three* different types of data sources on educational qualifications: administrative school files, self-reported information very close to the dates of completion of the qualifications and self-reported recall information ten years later. We thus provide the academic and policy community with estimates of the relative reliability of commonly used types of data. This represents a *new* piece of evidence for the UK, which will allow one to check the robustness of current estimates of returns to the presence of misreported qualifications. Knowing the extent of misreporting also has obvious implications for the interpretation of other studies that use educational attainment as an outcome variable or for descriptive purposes.

Our third contribution is to explore how the biases from measurement error and from omitted variables interact in the estimation of returns to educational qualifications in the UK. The aim is to produce some simple calibration rules to allow policy makers to use nationally representative data sets such as the Labour Force Survey to estimate returns to qualifications. These data totally rely on recall about the qualifications individuals have and do not contain any information on individual ability and family background.

Throughout we explore a unified general framework for the study of the impact of misreported treatment status on evaluation methods widely used in the literature. Mahajan (2004), Lewbel (2004) and Molinari (2004) are the only examples we are aware of (see also Battistin and Chesher, 2004).

The evaluation problem, that is the measurement of the causal impact of a generic 'treatment' on an outcome of interest, can be fruitfully framed within the potential outcome framework (for a review see Heckman, LaLonde and Smith, 1999). This set-up is both non-parametric as well as of extremely general scope, and in particular it allows for arbitrarily heterogeneous individual returns. Our discussion of the methods used to address measurement error in educational qualifications is thus of far wider interest, since the same issues arise in any application looking at the effects of a binary or categorical variable. Examples include: the returns to work-related training, where the occurrence of training is typically self-reported by individuals who are asked to recall whether they have undertaken any course for work purposes; the effects of programmes (or policy schemes) in which participation (or eligibility) is not recorded in administrative data and the treatment status is obtained from survey respondents, who have been typically shown to have rather poor recall or awareness of the kind of schemes they are in; the effects of government schemes where the researcher cannot directly observe or measure actual take-up and has to 'impute' the treatment status; or a medical study of the effectiveness of a new drug when patients may fail to take it or else may follow alternative treatments.

In order to focus fully on the potential biases arising from measurement error, in this paper we use the uniquely rich data from the 1958 British NCDS cohort to avoid issues related to omitted variables bias. In particular, in this work we only consider evaluation methods based on the selection on observables (or conditional independence) assumption, and rely on Blundell, Dearden and Sianesi (2005) who could not find any strong evidence of remaining selection bias given the information available in that data.

The remainder of the paper is organized as follows. First, in Section 2 we start by reviewing the evidence on measurement error and returns to educational qualifications. Section 3 sets out the general evaluation framework, while Section 4 allows for the possibility of misclassification in the treatment status. In Section 5 we show the consequences that such reporting errors might have for the estimation of returns. The estimation strategies that we exploit to correct for this bias are described in Section 6. Section 7 discusses how information in the NCDS will allow us to implement this strategy under fairly weak assumptions on the nature of the data collected. Estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment are presented in Section 8, while Section 9 concludes.

2 The evidence so far

Whilst use of years of completed education has a long history in the US, for the UK most authors prefer qualification-based measures of educational attainment. Recent examples include Robinson (1997), Dearden (1999a,b), Blundell *et al.* (2000), Gosling, Machin and Meghir (2000), Conlon (2001), Blanden *et al.* (2002), Dearden *et al.* (2002), Galindo-Rueda and Vignoles (2003), McIntosh (2004) and Blundell, Dearden and Sianesi (2005).[2]

However despite the importance of schooling both as an outcome and as an explanatory variable in applied research[3], hardly any effort has been devoted to assessing either the accuracy of widely used survey reports of educational attainment in the UK, or the impact that

[2]For a review and summary of some recent work on returns to qualifications, see Sianesi (2003).

[3]Note, however, that the effect of misreporting when educational qualifications enter the right-hand-side or the left-hand-side of a regression equation can be considerably different (see, amongst others, Hausman *et al.*, 1998). As we are interested in the estimation of returns, in what follows we will focus on the former case.

misreporting might have on estimated returns to education.[4] To date, the only work in the latter direction is Dearden (1999b) and Dearden *et al.* (2000 and 2002), who however ignore the non-classical nature of measurement error caused by misreporting of discrete qualifications and conclude that measurement error bias and omitted ability bias largely cancel out in the estimation of returns. Indeed, some recent work based on the UK Labour Force Survey (e.g. McIntosh, 2004) at times appeals to this result.

As a starting point and a benchmark it is thus worth considering the evidence on categorical education measures available for the US, most of which being provided by the study by Kane, Rouse and Staiger (1999) (see also the work referred to by Card, 1999). Overall, misreporting was found to be more likely to happen for low levels of qualification, with over-reporting being more likely than under-reporting (see also Black, Sanders and Taylor, 2003) and events such as degree completion being more accurately reported than completed years of college. Interestingly, transcript measures were often found to be subject to at least as much – and at times even more! – measurement error as self-reported survey measures.

With regard to their more specific findings, extensive measurement error was found in self-reported measures for those completing less than 12 years of schooling (i.e. the high-school drop-outs). As to bachelor's degree attainment, they found that 95% of those with a degree reported so accurately and less than 1% of those without a degree misreported having one, with self-reported information being actually more accurate than information from administrative data. As to years of college completed, however, both measures were found to be often inaccurate. In particular, 6% of those with no completed years of college misreported to have completed some, and 6% of those who had completed some college misreported to have completed none. Estimates of returns that ignore such misclassification were found to be severly biased, either upwards or downwards depending on the educational level of interest. Similarly, the application in Lewbel (2004) points to seriously inaccurate transcript information as to degree attainment and finds that allowing for misclassification has a considerable impact on estimated returns to college, leading to around a 5-fold increase in the return to a degree.

[4]Ives (1984) only offers a descriptive study of the mismatch between self-reported and administrative information on qualifications in the NCDS, finding serious discrepancies particularly for the lower-level academic qualifications.

3 The evaluation set-up

3.1 Potential outcomes framework

The measurement of the causal impact of a generic 'treatment' can be fruitfully framed within the potential outcome framework.[5] In the next section we extend such a framework to study the consequences for the identification of causal effects of allowing measurement error in recorded treatment status.

The specific evaluation problem we have in mind is the measurement of the returns to educational qualifications – that is of the causal effects of qualifications on individual (log) wages in the population of interest – when measurement error affects the reporting of education.

To ease the exposition, throughout this paper we consider the single treatment setting with treatments defined by *two* different educational qualifications, denoted by $D^* = 1$ and $D^* = 0$, so that the treatment status is binary. The generalization to the multiple-treatment case proceeds along the same lines, but it is notationally more demanding. We thus study the identification of the wage return of obtaining a qualification of interest ($D^* = 1$ say for college), relative to another qualification, where $D^* = 0$ can denote a specific alternative (e.g. high-school) or just the non-attainment of the qualification of interest (non-college in this example). To ease the comparison with the general evaluation literature, we will often refer to individuals with $D^* = 1$ as the group of 'participants' (in the educational qualification of interest) and to those with $D^* = 0$ as the group of 'non-participants'.

Letting Y_1 be the wage if the individual achieved the qualification of interest and Y_0 the wage if the individual were not to achieve the qualification, the individual causal effect (or return) of achieving the qualification is defined as the difference between the two potential outcomes, $\beta \equiv Y_1 - Y_0$. The realised individual wage can then be writtten as $Y = Y_0 + D^*\beta$, with $Y = Y_1$ if the individual is a participant and $Y = Y_0$ if the individual is a non-participant. This set-up is extremely general, in particular it does not assume that the returns to a given qualification are homogeneous across individuals.[6]

Since no individual can be in two different educational states at the same time, either Y_1

[5]For reviews of the evaluation problem see Heckman, LaLonde and Smith (1999) and Imbens (2004). For the potential outcome framework, the main references are Fisher (1935), Neyman (1935), Roy (1951), Quandt (1972) and Rubin (1974).

[6]Note however that for this representation to be meaningful, the stable unit-treatment value assumption needs to be satisfied (Rubin, 1980), requiring that an individual's potential outcomes as well as the chosen education level are independent from the schooling choices of other individuals in the population.

or Y_0 is missing, which makes it impossible to ever observe the individual return β. Our more modest though still challenging aim is to identify the average return in some population of interest. A group which has traditionally received most attention in the evaluation literature is the group of treated. In our case, the *average effect of treatment on the treated* (ATT) represents the average return to education for those individuals who have chosen to undertake the educational qualification of interest:

$$\Delta^* \equiv E(\beta|D^* = 1) = E(Y_1 - Y_0|D^* = 1) = E(Y_1|D^* = 1) - E(Y_0|D^* = 1). \tag{1}$$

This is the parameter of interest when the 'treatment' is voluntary, and is the one needed for a cost-benefit analysis. Given that achievement of educational qualifications is voluntary, in this paper we shall focus on the ATT, capturing the average payoff to individuals' own educational choices.[7]

3.2 Identification in the absence of misclassification

As to the identification of the ATT, the first term in (1) is observed, since $E(Y_1|D^* = 1) = E(Y|D^* = 1)$ for individuals acquiring the qualification. The average unobserved counterfactual $E(Y_0|D^* = 1)$ needs however to be somehow constructed on the basis of some usually untestable identifying assumptions.

As we aim to characterize the impact of measurement error in the reporting of D^*, in what follows we will assume that the outcome-relevant differences in the composition of participants and non-participants can purely be attributed to *observable* characteristics (*selection on observables* or *conditional independence assumption*), or, in other words, that D^* is exogenous given X:

Assumption 1 ***(Conditional Independence Assumption)*** *Conditional on a set of observable variables X, the educational choice D^* is mean independent of the no-education outcome Y_0:*

$$E(Y_0|D^*, X) = E(Y_0|X). \qquad \blacksquare$$

[7]An additional reason to focus on this parameter relates to the relative ease of its identification in the available data. Identification of the average effect of treatment on the non-treated, or of the average treatment effect requires in fact more restrictive assumptions and was found to be too demanding on the data we use (see Blundell, Dearden and Sianesi, 2005).

This assumption thus requires the evaluator to observe *all* those characteristics that jointly affect the decision to acquire the qualification of interest and potential wages in the absence of that educational investment. Its plausibility for our empirical application, and in particular the issue of 'ability bias', will be addressed in the data section. As discussed by Blundell, Dearden and Sianesi (2005), the set of regressors available from the NCDS data seems rich enough to assume that the selection problem can be dealt with by conditioning on observable characteristics.

In the absence of measurement error, the *true* participation status D^* is observed. Under Assumption 1, one could always define a correspondence which identifies parameters of interest as functionals of the joint distribution (Y, D^*, X). Let F_{YD^*X} denote this distribution. The parameter Δ^* is identified by a correspondence $\Delta^* \longleftarrow \mathcal{H}(F_{YD^*X})$, and $\mathcal{H}$ is termed an identifying functional. Matching, and other estimators employed in practice, are *analogue estimators* obtained by applying the identifying functional to an estimate of the distribution of observable outcomes and covariates, that is $\hat{\Delta}^* \equiv \mathcal{H}(\hat{F}_{YD^*X})$.

To give empirical content to Assumption 1, we also require the following condition on the support of the X variables:

Assumption 2 ***(Common Support)*** *For all values X, there are both participants and non-participants, that is*

$$0 < e^*(x) \equiv Pr(D^* = 1|X = x) < 1, \qquad \forall x$$

where $e^*(x)$ is the propensity score. ■

Under Assumptions 1 and 2, the causal effect of education for those who participated in education – that is the ATT parameter (1) – is identified as:

$$\Delta^* = \int \Delta^*(x) f(x|D^* = 1)dx, \tag{2}$$

where

$$\Delta^*(x) \equiv E(Y|D^* = 1, x) - E(Y|D^* = 0, x)$$

is the *conditional treatment effect*, that is the average treatment effect (or average return) for individuals with characteristics $X = x$. Note that, because of Assumption 2, the conditional effect is well defined for all values X. This effect is integrated with respect to the distribution of X for participants.

In its bare essentials, estimation proceeds by considering the empirical analogues of the quantities on the right-hand-side of (2). In particular, without invoking any functional form assumption one can perform any type of non-parametric estimation of the conditional expectation function in the non-participation group, $E(Y|D^* = 0, x)$, and then average it over the distribution of X in the participants' group (within the common support).

One way of implementing this non-parametric regression is via *matching*, whereby participants are matched with respect to their observable characteristics X to non-participants and the difference in the average outcomes of the two matched groups is then taken as an estimate of the ATT. More generally, the outcomes of non-participants are appropriately re-weighted so as to realign their distribution of X to the one of the group of participants (see Imbens, 2004, for a review).[8]

4 Misclassified treatment status

4.1 General formulation of the problem

Either because individuals are left to self-report their qualifications or because of transript errors, the treatment status D which is recorded in the data may differ from the actual status D^*. By analogy to the definition of D^*, let $D = 1$ be the group of individuals who self-report to have attained the educational qualification of interest, and $D = 0$ the group of individuals reporting not to have attained it.[9]

In the absence of measurement error, data are informative about (Y, D^*, X); as seen above, estimators based on Assumptions 1 and 2 establish a correspondence between this triple and the parameter of interest in (1). By contrast when qualifications are misreported, data are informative about the distribution of *measurement-error contaminated* variables. If measurement error is ignored, or not perceived, causal effects will thus be inferred using realizations of (Y, D, X) as if they were realizations of (Y, D^*, X).

[8]Note that although both matching and simple OLS regression rely on Assumption 1, matching is not subject to several potential misspecification biases for the ATT compared with standard parametric methods like OLS. In particular, OLS may suffer from misspecification bias for the non-education outcome equation; it may use this imposed functional form to extrapolate outside the common support, if need be; and in the presence of heterogeneous effects it does not in general identify the ATT (see Angrist, 1998, and Blundell, Dearden and Sianesi, 2005).

[9]Note that the extension to the multiple treatments setup is rather simple, as (2) and (3) could be thought as one of the possible pairwise comparisons for treatments $D^* = i$ and $D^* = j$ under a suitably extended Assumption 1.

In particular, the object that can be computed from the observed data is therefore:

$$\int_{\mathcal{S}} \Delta(x) f(x|D=1)dx \equiv \Delta, \tag{3}$$

where:

$$\begin{aligned}\Delta(x) &\equiv E(Y|D=1,x) - E(Y|D=0,x),\\ \mathcal{S} &\equiv \{x : 0 < e(x) \equiv Pr(D=1|X=x) < 1\}.\end{aligned}$$

Notice that the expression above is simply the analogue of Δ^* when D^* is replaced by D. In particular $\mathcal{S}$ is the observed (common) support for the self-reported participants in education and $e(x)$ is the propensity score calculated from the mismeasured qualification D. It is worth noting that, as we will discuss in the next section, although Assumption 2 implies that the *true* score $e^*(x)$ is strictly between zero and one, misclassification can cause the *observed* score $e(x)$ to take on values at the boundaries.

Since some individuals with $D^* = 0$ will erroneously be misclassified as participants on the basis of the error-affected indicator D and only part of those individuals reporting $D = 1$ have actually got the qualification of interest, the estimation of causal effects based on (Y, D, X) will in general be biased for treatment effects, with the magnitude of this bias depending on the extent of misclassification. This is shown in Section 5.3, where we derive the difference between the causal parameter of interest that would consistently be estimated if we observed the correct triple (Y, D^*, X) – equation (2) – and the parameter that would instead be estimated from the observable triple (Y, D, X) – equation (3).

4.2 The misclassification probabilities

In what follows we build on Molinari (2004) to introduce the notation required to study this problem, as well as the assumption on the classification errors we will maintain throughout (Assumption 3).

We start by defining the *(mis)classification probabilities* as

$$\lambda_{ji}(x) \equiv Pr(D^* = j|D = i, x), \qquad i, j \in \{0, 1\},$$

which may in general depend on X. In the binary case, there are two types of misclassification: $\lambda_{10}(x)$, the proportion of true participants amongst those reporting $D = 0$; and $\lambda_{01}(x)$, the proportion of true non-participants amongst those with $D = 1$.

Of recurrent use will be the probabilities of *exact* classification, that is one minus the probability of misclassification (or for the case of multiple treatments, one minus the *sum* of the misclassification probabilities):

$$\begin{aligned}
\lambda_{00}(x) &\equiv \lambda_0(x) = Pr(D^* = 0|D = 0, x), \\
\lambda_{11}(x) &\equiv \lambda_1(x) = Pr(D^* = 1|D = 0, x),
\end{aligned}$$

where for ease of notation only one subscript is retained.[10] It is convenient to collect the (mis)classification probabilities into the *matrix of (mis)classification probabilities*:

$$\Pi(x) = \begin{bmatrix} \lambda_0(x) & 1 - \lambda_0(x) \\ 1 - \lambda_1(x) & \lambda_1(x) \end{bmatrix}.$$

Throughout our discussion, we will assume that the classification error is *non-differential*, as this can help us write down relatively detailed but still manageable models (see Bound, Brown and Mathiowetz, 2001). Accordingly, we will maintain the assumption that, conditional on a person's actual qualification and on other covariates, reporting errors are independent of earnings.[11]

Assumption 3 ***(Non-Differential Misclassification given*** X*) Any variable* D *which proxies* D^* *does not contain information to predict the outcome of interest* Y *conditional on* D^* *and* X*:*

$$E(Y|D^*, D, X) \;=\; E(Y|D^*, X),$$

namely the following two conditions are verified

$$\begin{aligned}
&(a)\;\; E(Y_0|D^* = 0, D = 1, X) &= \;& E(Y_0|D^* = 0, D = 0, X), \\
&(b)\;\; E(Y_1|D^* = 1, D = 1, X) &= \;& E(Y_1|D^* = 1, D = 0, X). \qquad \blacksquare
\end{aligned}$$

These two conditions highlight how this assumption would not hold if an individual's propensity to misreport treatment status is related to outcomes. In particular, note that (b) is implied by:

$$E(Y_0|D^* = 1, D = 1, X) = E(Y_0|D^* = 1, D = 0, X)$$

[10] The (mis)classification probabilities are often defined conditional on the true treatment status: $\gamma_1 = Pr(D = 1|D^* = 1)$ and $\gamma_0 = Pr(D = 0|D^* = 0)$. These γ's are linked to our λ's via Bayes' Theorem.

[11] It is worth noting that the set of covariates X considered in Assumption 3 below coincides with (or is included in) the set of covariates already considered in Assumption 1. Our discussion could be extended to deal with the potentially interesting case in which some observables are known to affect only the (mis)classification probabilities and not to enter the strong ignorability assumption.

and

$$E(\beta|D^* = 1, D = 1, X) = E(\beta|D^* = 1, D = 0, X).$$

Thus Assumption 3 would be violated if those graduates ($D^* = 1$) who experience a very low Y_1 - either because they have received a negative productivity shock to their no-education earnings Y_0 and/or because they have reaped a very low return from the degree β - are more inclined to deny possessing the qualification (this violation would be even more likely if respondents are asked by the interviewer about their education and earnings at the same time!).

Under Assumption 3, we have that

$$\begin{aligned} E(Y|D = 1, x) &= \lambda_1(x)E(Y|D^* = 1, x) + [1 - \lambda_1(x)]E(Y|D^* = 0, x), \\ E(Y|D = 0, x) &= [1 - \lambda_0(x)]E(Y|D^* = 1, x) + \lambda_0(x)E(Y|D^* = 0, x). \end{aligned}$$

It can be seen from the last two expressions that individuals for whom we observe $D = d$ are in fact a *mixture* of participants ($D^* = 1$) and non participants ($D^* = 0$), with mixing weights given by the (mis)classification probabilities. This system of equations can be written more compactly in matrix algebra notation as

$$\begin{bmatrix} E(Y|D = 0, x) \\ E(Y|D = 1, x) \end{bmatrix} = \Pi(x) \begin{bmatrix} E(Y|D^* = 0, x) \\ E(Y|D^* = 1, x) \end{bmatrix},$$

from which we have that

$$\Pi^{-1}(x) \begin{bmatrix} E(Y|D = 0, x) \\ E(Y|D = 1, x) \end{bmatrix} = \begin{bmatrix} E(Y|D^* = 0, x) \\ E(Y|D^* = 1, x) \end{bmatrix}. \tag{4}$$

provided that $det[\Pi(x)] = \lambda_0(x) + \lambda_1(x) - 1 \neq 0$:

Assumption 4 ***(Informative Recorded Treatment Status)*** *Misclassification is such that*

$$\lambda_1(x) + \lambda_0(x) - 1 \neq 0,$$

namely

$$Pr(D^* = 1|D = 1, X) \neq Pr(D^* = 1|D = 0, X)$$

for all values X. ■

Assumption 4 appears reasonable. It requires that conditional on X, the proportion of true graduates among those who self-report having a degree to be different from the proportion of true graduates among those who self-report not having a degree; or in other words, that the marginal effect of recorded status D on true status D^* conditional on X is non-zero. Assumption 4 only requires inequality; it is however convenient to spell out here the two possible cases:

$$\text{4-(a)} \quad \lambda_1(x) + \lambda_0(x) > 1 \quad \Leftrightarrow \quad Pr(D^* = 1|D = 1, X) > Pr(D^* = 1|D = 0, X),$$

$$\text{4-(b)} \quad \lambda_1(x) + \lambda_0(x) < 1 \quad \Leftrightarrow \quad Pr(D^* = 1|D = 1, X) < Pr(D^* = 1|D = 0, X).$$

Case 4-(a) is a situation of limited misclassification in the sense that, given X, the proportion of true graduates among those reporting to have a degree is *higher* than the proportion of true graduates among those reporting not to have a degree. By contrast, 4-(b) represents a case of such extensive misclassification for it to be more likely to randomly draw a true graduate from the group reporting no degree than from the group reporting a degree.

Another way to look at this is to note that $Cov(D, D^*|x) = (\lambda_1(x) + \lambda_0(x) - 1)Var(D|x)$; so that the sign of $\lambda_1(x) + \lambda_0(x) - 1$ determines the sign of the correlation between D and D^*. Case 4-(a) can then be seen as preventing measurement error to be so severe as to reverse the (positive) correlation between the observed and the true treatment measures.

5 The bias introduced by misclassification

5.1 Bias on the conditional treatment effect

In deriving how the parameter that can be recovered from the observed data (3) compares to the causal parameter of interest (2), we start by considering the bias introduced in the estimation of the causal treatment effect *conditional* on X, that is on $\Delta^*(x)$. This bias can be straightforwardly characterized using (4). The result in (5) coincides with the result in Lewbel (2004; see Proof of Theorem 1), and more in general follows from Aigner (1973). The proof is reported in the Appendix.

Proposition 1 ***(Bias on Treatment Effects given X)*** *If Assumptions 1 to 4 are satisfied, it follows that*

$$\Delta^*(x) = \frac{\Delta(x)}{\lambda_0(x) + \lambda_1(x) - 1}. \qquad \blacksquare \tag{5}$$

Accordingly, the estimates of $\Delta^*(x)$ based on the triple (Y, D, X) are always biased towards zero, but possibly with the opposite sign if the measurement error is very strong (the denominator being negative in case 4-(b)). In terms of the conditional treatment effect, therefore, the classical attenuation bias result still holds.

An interesting implication of (5) is that $\Delta(x) = 0 \Leftrightarrow \Delta^*(x) = 0$, so that the raw difference in observed outcomes given X being zero actually implies that the true conditional treatment effect is zero. Finally, if there is no misclassification (that is, $\lambda_0(x) = \lambda_1(x) = 1$), then of course $\Delta(x) = \Delta^*(x)$; and if there is complete reversal in the classification (that is, $\lambda_0(x) = \lambda_1(x) = 0$), then $\Delta(x) = -\Delta^*(x)$.

5.2 Support condition

We have thus far defined the bias for the treatment effect conditional on a given value of the vector X. In order to characterize the bias for the ATT, we have to integrate over the distribution of X in the treated group, which brings us to discuss support issues. As we pointed out earlier in this paper, although Assumption 2 ensures that at each point in the support of the X distribution both individuals with $D^* = 1$ and individuals with $D^* = 0$ are observed, the extent of misclassification can be such that the same condition does not hold for individuals with $D = 1$ and $D = 0$.

To see this, we use the law of iterated expectations to write $e^*(x)$ in terms of $e(x)$:

$$e^*(x) = [1 - \lambda_0(x)] + e(x)[\lambda_0(x) + \lambda_1(x) - 1],$$

so that, solving for $e(x)$ and using Assumption 4:

$$e(x) = \frac{e^*(x) - [1 - \lambda_0(x)]}{\lambda_0(x) + \lambda_1(x) - 1}.$$

from which we see that $e(x)$ will take on values at the boundaries according to:

$$\begin{aligned} e(x) = 0 \quad &\Leftrightarrow \quad \lambda_0(x) = 1 - e^*(x), \\ e(x) = 1 \quad &\Leftrightarrow \quad \lambda_1(x) = e^*(x). \end{aligned}$$

It follows that the parameter (3) estimated from the triple (Y, D, X) – that is the ATT for observed participants in the observed common support $\mathcal{S}$ – could in general refer to a different population than the one implied by (Y, D^*, X).[12]

To avoid this, we ensure that $e(x)$ is strictly between 0 and 1 for all values of X by assuming:

[12]Note incidentally that without any specific assumptions, the fact that $e(x) \in [0, 1]$ implies the following additional restrictions on the extent of misclassification under case 4-(a) (the inequality signs being reversed

Assumption 5 ***(Restriction on the extent of misclassification)*** *Misclassification is such that at each value X, at least one holds among*

$$\begin{aligned} \lambda_0(x) &\neq 1 - e^*(x) \\ \lambda_1(x) &\neq e^*(x) \end{aligned}$$

(the other one holding automatically given Assumption 4). ■

We make this assumption for formal convenience, in that it allows us to treat the common support in the presence of measurement error as the true common support. If this were not the case, the integrals in the following would be defined over a different subset of the truly treated than the full one. More specifically, if Assumption 5 does not hold and we imposed common support based on $e(x)$, true participants not belonging to the observed $\mathcal{S}$ may be discarded so that the ATT estimated from (Y, D, X) would refer to a *different* population of participants than the population of participants the true ATT refers to.

5.3 Bias on the treatment effect on the treated

If one is interested, as is Lewbel (2004), in the average treatment effect (ATE) – the average return for an individual irrespective of whether the qualification of interest has been acquired or not:

$$E(\beta) \equiv E(Y_1 - Y_0) = \int \Delta^*(x) f(x) dx,$$

the discussion can stop here.[13] In particular, one would only need to integrate the conditional average treatment effect $\Delta^*(x)$ over the distribution of X in the population, the latter being observed in the data.

Note also that the attenuation-bias result from Proposition 1 keeps holding unconditional on X, in other words, ignoring measurement error in treatment status leads to a downward-biased estimate of the ATE. The correspondence between a zero raw average effect and a zero true

under case 4-(b)):

$$\begin{aligned} \lambda_0(x) &\geq 1 - e^*(x), \\ \lambda_1(x) &\geq e^*(x), \end{aligned}$$

for all values $X = x$. Note further that if one were willing to assume that the misclassification error did not depend on X, the restrictions above would become $\lambda_0 \geq 1 - \min_x\{e^*(x)\}$ and $\lambda_1 \geq \max_x\{e^*(x)\}$.

[13] Note that identification of ATE requires a strengthened Assumption 1, implying in particular homogeneous returns (given X) or the absence of selection into education based on unobserved returns β.

average effect, however, no longer holds, unless the misclassification probabilities are assumed not to depend on X.

By contrast, if interest lies in recovering the ATT – the average return to education for those who invested in that qualification – the next step is not straightforward, since, as shown in equation (2), the conditional effect $\Delta^*(x)$ needs to be integrated over the distribution of X in the (truly) treated group, $f(x|D^*=1)$, which is not observed.

The following proposition provides a characterization of the bias introduced by measurement error for the estimation of (1), that is the relationship between Δ^* and Δ. The proof is reported in the Appendix.[14]

Proposition 2 ***(Bias on Treatment Effects)*** *If Assumptions 1 to 5 are satisfied, the effect estimated from raw data can be written as follows*

$$\Delta^* = \frac{Pr(D=1)}{Pr(D^*=1)}\Delta + \frac{1}{Pr(D^*=1)}\int \frac{1-\lambda_0(x)}{\lambda_0(x)+\lambda_1(x)-1}\Delta(x)f(x)dx, \tag{6}$$

where

$$Pr(D^*=1) = \int [1-\lambda_0(x)]f(x)dx + \int [\lambda_0(x)+\lambda_1(x)-1]e(x)f(x)dx.$$

■

Apparently, misclassification is such that the effect Δ^* can be over- or under-estimated depending on the unknown probabilities $\lambda_1(x)$ and $\lambda_0(x)$. Furthermore, Δ being zero no longer implies the absence of a treatment effect, as was the case when conditioning on X.

Note that we could solve for Δ in the last expression if these probabilities were known. This implies that the causal effect of education can always be expressed as a *known* functional of the triple (Y, D, X) and $\lambda_1(x)$ and $\lambda_0(x)$, that is

$$\Delta^* = \Delta^*[Y, D, X, \lambda_0(X), \lambda_1(X)], \tag{7}$$

[14]To the best of our knowledge, this has not been considered in the literature on treatment effects although it mimics well known results from the linear regression theory. In a linear regression setting, treatment effects can be inferred from the following regression

$$y = \beta d + \gamma x + \varepsilon$$

if Assumption 1 is satisfied. It is well known that, if the variable D is affected by classical measurement error, the coefficient β is biased toward zero. However, since the treatment status is binary, the measurement error is negatively correlated with D^*, so that the sign of the bias is not determined in general (see Aigner, 1973, and Bound, Brown and Mathiowetz, 2001).

but cannot be identified without the knowledge of the misclassification probabilities. If both probabilities are one, then we get standard identification of treatment effects. Note that these probabilities may depend on X in general.[15]

To conclude this section, we show that any estimator of the ATT based on the observed propensity score $e(x)$ (e.g. propensity score matching or re-weighting)

$$\int \Delta[e(x)]f[e(x)|D=1]de$$

is equivalent to the estimator defined by (3). Since we have that

$$\begin{aligned} E[Y|D,e(x)] &= \int E[Y|D,e(x),x]f[x|D,e(x)]dx, \\ &= \int E[Y|D,x]f[x|D=1,e(x)]dx, \end{aligned}$$

where the last equality follows from x being finer than $e(x)$ and by the balancing property of the propensity score (see Rosenbaum and Rubin, 1983), we also have

$$\Delta[e(x)] = \int \Delta[x]f[x|D=1,e(x)]dx.$$

Therefore

$$\begin{aligned} \int \Delta[e(x)]f[e(x)|D=1]de &= \int\int \Delta[x]f[x,e(x)|D=1]dxde, \\ &= \int \Delta[x]f[x|D=1]dx \equiv \Delta, \end{aligned}$$

which is enough to conclude that the bias induced by misclassification when the estimation is carried out with respect to the observed propensity score is equivalent to the bias derived in Proposition 2.

5.4 Special cases

CASE I: Limited Misclassification

Misclassification is such that 4-(a) applies, that is:

$$\lambda_1(x) + \lambda_0(x) > 1$$

[15]The extension to the multiple treatments setting proceeds along the same lines, but requires more work. For example, for the case of three treatments, the matrix of misclassification probabilities has dimension 3×3, and therefore the expressions in (5) and (6) become less tractable although conceptually identical to the ones derived for the 2×2 case. As expected, the analytical tractability of the problem worsens as the number of treatments increases. Further assumptions on the misclassification probabilities can help simplify the resulting expressions (as we discuss below).

for all values X. $\blacksquare$

As discussed above, this represents the most likely case, and is also implied by the assumption that observations on D are more accurate than pure guesses once X is corrected for (see for example Bollinger, 1996), that is:

$$\lambda_1(x) > 0.5, \qquad \lambda_0(x) > 0.5.$$

An implication of this assumption which is worth stressing again is that $\Delta(x)$ – though biased toward zero – is always right signed for all values of X. This can be seen from (5) by noting that the scaling factor defined by the misclassification probabilities is always between zero and one.

CASE II: Only over-reporting of qualifications

The misclassification probabilities are such that only over-reporting can happen:

$$\lambda_0(x) = 1$$

for all values X. $\blacksquare$

To see why this condition represents a situation where only over-reporting of qualifications can occur, note that it corresponds to $P(D^* = 1|D = 0) = 0$, which rules out that true graduates may be found among those reporting not to have a degree, in other words, ruling out under-reporting. As in this case we have that:

$$\Delta^*(x) = \frac{\Delta(x)}{\lambda_1(x)},$$

the conditional treatment effect is always right-signed for all X, but biased towards zero. Furthermore, setting $\lambda_0(x)$ equal to 1 in Proposition 2 yields

$$\Delta^* = \frac{Pr(D=1)}{\int \lambda_1(x)e(x)f(x)dx}\Delta = \frac{\int e(x)f(x)dx}{\int \lambda_1(x)e(x)f(x)dx}\Delta$$

where the factor multiplying Δ is greater than 1 since $\lambda_1(x) < 1$. It thus follows that the estimated effect Δ is always biased towards zero for Δ^*, and with the right sign.[16]

[16] As already pointed out for the case of multiple treatments, the restrictions imposed on the misclassification probabilities can simplify the relationship between moments involving D^* and moments involving D, and therefore the analytical tractability of the problem. For example, by assuming that the misclassification problem only arises because of over-reporting of qualifications, for the 3×3 case we have

$$\Pi(x) = \begin{bmatrix} \lambda_0(x) & 0 & 0 \\ \lambda_{01}(x) & \lambda_1(x) & 0 \\ \lambda_{02}(x) & \lambda_{12}(x) & \lambda_2(x) \end{bmatrix},$$

which is function of four unknown probabilities.

CASE III: Misclassification independent of X

A last interesting special case where a tractable expression for (7) can be derived is when the percentage of correct classification is assumed to be independent of the characteristics X of respondents:

$$\begin{aligned}\lambda_1(x) &= \lambda_1, \\ \lambda_0(x) &= \lambda_0,\end{aligned}$$

for all values X. $\blacksquare$

Although this assumption is clearly only made here for convenience, it could be weakened by assuming constant probabilities within cells defined by X.[17] By using Proposition 2 we have

$$\Delta^* = \frac{1}{Pr(D^*=1)}\left(Pr(D=1)\Delta - \frac{\lambda_0 - 1}{\lambda_0+\lambda_1-1}\int \Delta(x)f(x)dx\right)$$

and

$$Pr(D^*=1) = 1-\lambda_0 + [\lambda_0+\lambda_1-1]Pr(D=1).$$

Using (3) it follows that

$$\begin{aligned}\Delta^* &= \frac{\int \Delta(x)\left[1 - \frac{1}{Pr(D=1)}\frac{f(x)}{f(x|D=1)}\frac{\lambda_0-1}{\lambda_0+\lambda_1-1}\right]f(x|D=1)dx}{\frac{1-\lambda_0}{Pr(D=1)} + (\lambda_0+\lambda_1-1)}, \\ &= \frac{\int \Delta(x)\left[1 - \frac{1}{e(x)}\frac{\lambda_0-1}{\lambda_0+\lambda_1-1}\right]f(x|D=1)dx}{\frac{1-\lambda_0}{Pr(D=1)} + (\lambda_0+\lambda_1-1)},\end{aligned}$$

where the last expression is derived using the definition of $e(x)$ and Bayes theorem. It follows that

$$\Delta^* = \int \omega(x)\Delta(x)f(x|D=1)dx, \tag{8}$$

where

$$\omega(x) = \frac{1+\frac{1}{e(x)}\frac{1-\lambda_0}{\lambda_0+\lambda_1-1}}{\frac{1-\lambda_0}{Pr(D=1)} + (\lambda_0+\lambda_1-1)}.$$

This last expression, which represents the *exact* calculation of (7) for this particular example, shows that *if* the two λ's were known, the ATT could be estimated by appropriately re-weighting the conditional differences in outcomes based on recorded treatment data, $\Delta(x)$,

[17] As we will show in Section 6.2, this can be tested against real data if additional measurements of D^* become available. To a certain extent, this is possible using NCDS data.

with weights defined by $\omega(x)$. Note that, as it should be, $\omega(x) = 1$ for all individuals if there is no measurement error. If there is only over-reporting ($\lambda_0 = 1$), we have that $\omega(x) = \frac{1}{\lambda_1} \geq 1$ for all x. In fact, if all the weights are larger than 1, the 'raw' effect Δ will provide a lower bound on the true treatment effect, so that the classical attenuation-bias result applies.

Under the likely scenario of limited misclassification (assumption (4)-a), all the weights are positive and a first-order approximation to $\omega(x)$ around ($\lambda_0 = 1, \lambda_1 = 1$) yields

$$\omega(x) \simeq 1 + (1 - \lambda_0) + (1 - \lambda_1)[\frac{1}{e(x)} - \frac{1 - Pr(D = 1)}{Pr(D = 1)}],$$

from which it can be seen that a sufficient (but not necessary) and testable condition for $\omega(x)$ to be larger than 1 is that the propensity score at x be smaller than the odds ratio, i.e. $e(x) \leq \frac{Pr(D=1)}{1-Pr(D=1)}$. From a study of $\omega(x)$ as a function of the λ's, it can be shown that only for values of the parameter $P(D = 1)$ smaller than 0.3 is there the possibility that, depending on the value of $e(x)$, the corresponding weight at x is positive but smaller than 1. However we found that even in this case the distribution of weights is skewed towards values (often much) larger than 1, so that in most empirical applications the 'raw' estimate is most likely to be a lower bound.

6 Identification in the presence of misclassification

6.1 Partial identification of causal effects

Because of (7), bounds can be derived by looking at the *maximum* and the *minimum* value of the estimate of Δ^* when the probabilities $\lambda_0(x)$ and $\lambda_1(x)$ vary over the unit interval, or on a suitably chosen subset of $[0, 1] \times [0, 1]$. Thus, without additional information, only partial identification of treatment effects can be achieved.

Note however that misclassification probabilities left to vary between zero and one are likely to imply unreasonably high misclassification rates. One possibility often exploited in the literature is to use *a priori* restrictions on these probabilities, most of the times derived from previous studies or from knowledge of the economic context under investigation. For example, results from validation studies and behavioral theories developed in the social sciences often suggest restrictions on misclassification. Some fairly general restrictions that can be applied to the study of returns to education include the three cases considered above.[18]

[18]The assumptions below are spelled out only for the case of binary treatments, but can be generalized to

- **CASE I: Limited misclassification**: $\lambda_1(x) + \lambda_0(x) > 1$ for all values X.

- **CASE II: Only over-reporting of qualifications**: $\lambda_0(x) = 1$ for all values X. This assumption that individuals never under-report qualifications they have obtained can be weakened by assuming that over-reporting is just more likely than under-reporting. This case of **monotone misclassification** imposes $\lambda_1(x) < \lambda_0(x)$ for all values X, or, in a more intuitive form, $P(D^* = 0|D = 1) > P(D^* = 1|D = 0)$. Monotone misclassification reflects the idea supported by cognitive studies that when respondents are asked questions about socially and personally sensitive topics, they tend to under-report undesirable behaviours and attitudes, and over-report desirable ones.

- **CASE III: Misclassification independent of X**. As a specific illustration on how to obtain bounds for the ATT parameter, let us go back to a situation where misclassification is independent of respondents' characteristics. This simplifying assumption allowed us to obtain the exact expression (8) for Δ^*. For *given* values of the two misclassification probabilities λ_0 and λ_1, this expression can be estimated to get 'base case' bounds by calculating the empirical analogues of the quantities involved: recorded participants and non-participants can be matched according to their characteristics X and the resulting conditional raw differences in outcomes $\Delta(x)$ averaged with weights $\omega(x)$. Varying the two λ's over their support defines different weights $\omega(x)$. Choosing the maximum and the minimum from the resulting range of estimates for Δ^* finally provides lower and upper bounds for the ATT. Tighter bounds on the returns to educational qualifications can then be obtained by restricting further the values the two λ's can take. In particular, one can exploit the limited-misclassification condition $\lambda_1 + \lambda_0 > 1$ for different values of the sum of misclassification probabilities, or tighten the bounds even further by assuming monotone misclassification ($\lambda_1 < \lambda_0$) or that misclassification arises only because of over-reporting ($\lambda_0 = 1$).

the case of multiple treatments. See Molinari (2004) for a discussion of additional restrictions that could be imposed on the matrix of misclassification probabilities.

6.2 Point identification of causal effects

6.2.1 General idea

Throughout this section, we will assume than an additional measurement of D^* becomes available, so that we observe two measurements of educational qualifications. To fix ideas, in line with our application to returns to education, let us call the two measures D_S (Self-reported by the respondent, previously simply denoted by D) and D_T (from Transcript files).[19] Multiple reports of D^* can solve for misclassification, provided that errors are independent across reports (see Kane *et al.*, 1999, and Black *et al.*, 2000, Lewbel, 2004, in particular Proposition 2 and Proposition 3). In our discussion we will maintain the following assumption.

Assumption 6 ***(Multiple Independent Reports)*** *The qualification measures D_S and D_T are conditionally independent given D^* and X, that is*

$$D_S \perp D_T | D^*, X. \qquad \blacksquare$$

Moreover, in what follows we extend the non-differential error assumption 3 to hold for both measures separately:

$$\begin{aligned} E(Y|D^*, D_S, X) &= E(Y|D^*, X), \\ E(Y|D^*, D_T, X) &= E(Y|D^*, X). \end{aligned} \tag{9}$$

Alternatively, one could let the non-differential error Assumption 3 hold jointly:

$$E(Y|D^*, D_S, D_T, X) = E(Y|D^*, X),$$

which together with Assumption 6 implies (9).

Under these assumptions, Kane *et al.* (1999), Black *et al.* (2000) and Lewbel (2004) show that information on the number of individuals classified differently by D_S and D_T can be combined with information on the difference of their earnings to estimate the distribution of reporting errors (i.e. the misclassification probabilities) *in both measures.* It follows from the discussion above that error corrected estimates of the returns to qualifications can be obtained.[20]

[19]The extension to the case of multiple treatments / multiple measurements is notationally more demanding but proceeds along the same lines (see Lewbel, 2004). Frazis and Loewenstein (2002) extend the analysis to the case where the second measure is replaced by one or more instruments.

[20]One might be tempted to use the second measurement D_T to instrument the first one, D_S. Albeit similar, the approach suggested in this section is different from IV estimation: actually, it can be shown that instrumenting one report with the other tend to produce upward biased estimates of treatment effects because of the non-classical measurement error (see Bound, Brown and Mathiowetz, 2001).

For the single-treatment case, the intuition behind the identification strategy goes as follows. To begin with, assume that the relationship stated in Assumption 6 holds without conditioning on X (or let the following hold given X). We can then compute the *four* sample means of the outcome for the four groups defined by $D_S = 1$, $D_S = 0$, $D_T = 1$ and $D_T = 0$:

$$\begin{aligned} E(Y|D_S = i) &= E(Y|D^* = 0) + \Delta^* Pr(D^* = 1|D_S = i), \\ E(Y|D_T = i) &= E(Y|D^* = 0) + \Delta^* Pr(D^* = 1|D_T = i) \end{aligned} \tag{10}$$

Accordingly, wages in the four cells for $i = 0$ and $i = 1$ are function of *six* unknowns: $E(Y|D^* = 0)$, Δ^*, $\lambda_1^S \equiv Pr(D^* = 1|D_S = 1)$, $\lambda_0^S \equiv Pr(D^* = 0|D_S = 0)$, $\lambda_1^T \equiv Pr(D^* = 1|D_T = 1)$ and $\lambda_0^T \equiv Pr(D^* = 0|D_T = 0)$. On the other hand, the sample proportions of individuals in each cell provide *three* usable equations (one equation is lost because of the adding-up condition), which rearranged yield:

$$\begin{aligned} \frac{Pr(D_S = i, D_T = j)}{Pr(D_S = i)Pr(D_T = j)} &= \frac{Pr(D^* = 1|D_S = i)Pr(D^* = 1|D_T = j)}{E(D^*)} \\ &+ \frac{(1 - Pr(D^* = 1|D_S = i))\,(1 - Pr(D^* = 1|D_T = j))}{1 - E(D^*)} \end{aligned} \tag{11}$$

in *one* additional unknown, $E(D^*)$.

It follows that the expressions in (10) and (11) define a system of *seven* equations which is non-linear in *seven* unknowns. By denoting with θ the vector of the unknown parameters, this system can be written more compactly as

$$\Psi(\theta) = 0, \tag{12}$$

with $\Psi : A \subset R^k \to R^h$ and $k = h = 7$. In particular, although the estimating equations are non-linear in the parameter θ, it can be shown that they define a unique (analytical) solution and thereby identify the misclassification probabilities and the treatment effect (see Kane *et al.*, 1999, Black *et al.*, 2000 and Lewbel, 2004).[21]

When additional regressors become available, the set of equations in (10) and (11) can be defined conditional on X. To fix ideas, if the support of X is discrete (e.g. if X contains only categorical variables), seven equations are defined *for each* point of the support and the number of unknown parameters to be estimated increases with the cardinality of the support.

[21] The solution to the system can be derived iteratively by replacing Ψ with a linear approximation, and then solve the linear problem to generate the next guess using Newton's methods.

Yet exact identification is obtained, as in (12) the number of equations h equals the number of unknowns k.

If one is willing to make the assumption that some of the unknown parameters θ do *not* depend on X, the system (12) is *over-identified* as there are more equations than unknowns ($k < h$). For example, one could assume that the misclassification probabilities do not depend on X, as in Case III of Section 5.4, or are constant within cells defined by X. One common way to define the solution of an over-identified system corresponds to choosing that θ that solves the following problem

$$\min_{\theta} \Psi(\theta)^{\top} \Sigma \Psi(\theta),$$

where Σ is a matrix of weights suitably defined. The solution to the previous problem coincides with the (non-linear) GMM estimator of the parameter θ, once Σ is chosen to minimize the variance of estimation (Matyas, 1999).[22]

7 Data and educational qualifications of interest

In order to put into context the educational qualifications to which we estimate the returns, this section starts by briefly outlining the educational system in Britain. It then describes the data we use in the paper and concludes by specifying the qualifications – hence parameters – of interest.

7.1 The British educational system

Education in the United Kingdom is compulsory for everyone between the ages of 5 and 16. Those deciding to stay on past the minimum can either continue along an *academic* route or else undertake a *vocational* qualification before entering the labour market.

The former route is based on a series of national public examinations marked by independent assessors. Until 1986, pupils could take Ordinary Levels (O level) at 16 and then possibly move on to attain Advanced Levels (A levels) at the end of secondary school at 18.[23] A levels still represent the primary route into higher education.

[22]It is possible to exploit more than two measures of educational qualifications in the GMM estimation, as we do in our application.

[23]Less academically-oriented pupils could go for the lower-level Certificates of Secondary Education (CSE) option at 16 before they left school. In 1986 the CSEs and O levels exams were replaced by General Certificates of Secondary Education (GCSEs).

The vocational path is much more heterogeneous, with a plethora of options ranging from job-specific, competence-based qualifications often delivered within a work environment to more generic work-related qualifications. The academic and wide range of vocational qualifications have been classified into equivalent National Vocational Qualification (NVQ) levels, ranging from level 1 to level 5.

The British system is thus quite distinct from the one in the US; nevertheless, forcing some comparisons, one could regard the no-qualifications group as akin to the group of high-school drop-outs, A levels to High School, Higher Education to College.[24]

7.2 Data

In this paper we only consider methods relying on Assumption 1, and we thus require very rich background information capturing all those factors that jointly determine the attainment of educational qualifications and wages. We use the uniquely rich data from the British National Child Development Survey (NCDS), a detailed longitudinal cohort study of all children born in a week in March 1958. There are extensive and commonly administered ability tests at early ages (mathematics and reading ability at ages 7 and 11), as well as accurately measured family background (parental education and social class) and school type variables, all ideal for methods relying on the assumption of selection on observables. In fact, Blundell, Dearden and Sianesi (2005) could not find evidence of remaining selection bias for the higher education versus anything less decision once controlling for the same variables we use in this paper. We thus invoke this conclusion in assuming that there are enough variables to be able to control directly for selection.

Our outcome is real gross hourly wages at age 33.

Specifically, we use the sweeps from 1958 (at birth), 1965 (aged 7) and 1969 (aged 11) for family background and individual characteristics, from 1974 (aged 16) and 1981 (aged 23) for educational attainment and from 1991 (aged 33) for the wage outcome as well as for further measures of educational attainment. As to the latter, note that cohort members were asked to report the qualifications they had obtained as of March 1981 not only in 1981, but also in 1991. In other words, we can construct a separate measure of the qualifications the individuals

[24]In such a comparison the group with O levels as highest qualification is quite atypical, being made up of individuals who stop at the minimum leaving age with formal qualifications.

obtained up to 1981 based solely on responses in the 1991 survey.[25] Furthermore, the 1991 survey also asked respondents about all the qualifications they had achieved till then.

In addition to the main sweeps, in 1978 the schools cohort members attended when aged 16 provided information on the results of public academic examinations entered up to 1978 (i.e. by age 20). Similar details were collected from other institutions if pupils had taken such examinations elsewhere.[26]

For each individual our data thus contains four measurements of educational qualifications; these are summarised in Table 1 together with the corresponding broad categories we consider.[27] Accordingly, returns can be derived by considering any of the pairwise comparisons arising from these categories once measurement error in the reporting of qualifications is taken into account.[28]

Table 4 summarises the criteria according to which our final sample was selected.

7.3 Educational qualifications of interest

For our partial identification analysis we focus on the self-reported measure of educational qualifications achieved by age 33. Parameters that can be bound thus include the return to achieving at least O levels or their vocational equivalent compared to remaining without qualifications; the return to achieving at least A levels or their vocational equivalent compared to stopping with O levels or with no qualifications; or the return from undertaking some form of higher education compared to anything less.

[25] In the 1981 survey itself, respondents were asked whether they had any O or A levels. Furthermore, they were asked what their highest post-school qualification was, irrespective of whether it was an academic or a vocational one. In addition to an educational measure of academic qualifications up to A levels, one can thus only construct their highest educational qualification in terms of NVQ level only, i.e. without being able to separate out the highest academic and the highest vocational qualification.

[26] Results were obtained for approximately 95% of those whose secondary school could be identified.

[27] The 'None' category also includes very low-level qualifications at NVQ level 1 or less, in particular the academic CSE grade 2 to 5 qualifications. (Students at 16 could take the lower-level Certificates of Secondary Education (CSE) or the more academically demanding O levels. The top grade (grade 1) achieved on a CSE was considered equivalent to an O level grade C. Most CSE students tended to leave school at 16.) For full details of the various educational groupings, see the Appendix.

[28] We might not expect too large an extent of measurement error in such so broadly defined educational categories. The questionnaires as well as school files are very detailed in their questions, listing all possible types of individual qualifications. We then aggregate the responses to categories such as 'any A levels or vocational equivalent'. Also, we would expect problems especially at the lower end, in discriminating between no qualifications - or very low CSEs and other NVQ level 1 or less qualifications, and O levels. In fact, returns to moving from the lowest to the O level group are of particular policy interest. Still, these aggregations have been and are being used, so confirming the actual extent of their reliability is of value. There would be in any case surely many more errors in reporting 1 versus 2 or more A levels of in discriminating between good and bad O levels. Corresponding returns have in fact been shown to vary considerably.

Table 1: Measurements of highest educational qualifications

Obtained by age 33, self-reported at age 33 (1991 sweep)	
(separately identifies highest academic and highest vocational qualifications)	
None (level 0-1)	
O level or vocational equivalent	
A level or vocational equivalent	
HE or vocational equivalent	
Obtained by age 23, self-reported at age 33 (1991 sweep)	
(separately identifies highest academic and highest vocational qualifications)	
None (level 0-1)	
O level or vocational equivalent	
A level or vocational equivalent	
HE or vocational equivalent	
Obtained by age 23, self-reported at age 23 (1981 sweep)	
(identifies academic O and A levels, as well as highest NVQ level)	
None academic (level 0-1)	NVQ level 0-1
O level	NVQ level 2
A level or above	NVQ level 3
	NVQ level 4-5
Obtained by age 20, administrative information (1978 School Files)	
(identifies academic O and A levels)	
None academic (level 0-1)	
O level	
A level or above	

All of the preceding examples fall into the 'single treatment' framework, and in fact in one where the specific educational level of interest cuts right through the entire educational spectrum. This does not of course rule out interest in 'multiple treatments', or in single treatments for a more narrowly defined educational split, such as the return to higher education *vis-à-vis* stopping with A levels, or the return to higher education *vis-à-vis* dropping out at 16 without qualifications, or the return to finishing school with O levels *vis-à-vis* nothing.

The extension of the bounds analysis to such types of treatments – though conceptually quite straightforward – is however computationally (extremely) complex, since account would need to be taken of the potential misclassification in the reporting of *all* educational levels, not just in the two being considered. So for instance, even if one only wanted to compare higher education to A levels, the other categories would still need to be considered, since, first, individuals reporting no qualifications and individuals reporting O levels might in reality have higher education or A levels, and, second, individuals reporting higher education or A levels might in reality have neither of the two qualifications of interest. The price to pay in order to allay the computational burden is to impose some a priori restrictions on the matrix of misclassification probabilities.

Such extension, by contrast, is not only conceptually but also computationally straightforward in the point identification (GMM) approach. As it has been highlighted before, this approach is however more demanding, in that it needs an additional, independent measure of educational attainment. In the NCDS data, such a measure is offered by the School Files, which however only record academic qualifications (i.e. O and A levels), and only those achieved by age 20.

Although driven by the availability of an independent school measure for O and A levels only, focusing on academic qualifications does offer some advantages, and allows one to estimate highly policy relevant parameters.

First, academic qualifications are well defined and homogenous, with the central government traditionally determining their content and assessment. By contrast, the provision of vocational qualifications is much more varied and ill-defined, with a variety of private institutions shaping their content and assessment. In fact, as mentioned in Section 7.1 there is a wide assortment of options ranging from job-specific, competence-based qualifications to more generic work-related qualifications, providing a blend of capabilities and competences in the most disparate fields.

A second advantage of focusing on O and A levels is that they are almost universally taken through mostly uninterrupted education, whereas vocational qualifications are often taken after having entered the labour market. It is thus more difficult to control for selection into post-school (vocational) qualifications, since one would ideally want to control also for the labour market history preceding the acquisition of the qualification.

Additional interest in O levels arises from the finding that in the UK, reforms raising the minimum school leaving age have impacted on individuals achieving low academic qualifications, in particular O levels. Chevalier *et al.* (2003) show that the main effect of the reform was to induce individuals to take O levels. Del Bono and Galindo-Rueda (2004) similarly show that changes in features of compulsory schooling have been biased towards the path of academic attainment; the main effect of the policy was not to increase the length of schooling, but rather to induce individuals to leave school with an academic certification. In such a context it is of great policy interest to estimate the returns to finishing school with O levels compared to leaving with no qualifications. Indeed, Blundell, Dearden and Sianesi (2005) found a non-negligible return of 18% for those who did leave with O levels and of 13% for those who dropped out at 16 without any qualifications.

Another interesting parameter is the return to acquiring at least O levels compared to nothing; this parameter captures all the channels in which the attainment of O levels can impact on wages later on in life, in particular the potential contribution that attaining O levels may give to the attainment of A levels and then of higher education.

As to A levels, special interest arises from the fact that they were – and still are – the almost exclusive way to get the chance of a university education. Thus, the return to obtaining at least A levels compared to stopping at 16 without qualifications again captures the effect that the attainment of A levels may have on progression to university.[29]

For academic qualifications there are clearly defined targets and requirements for progression to the next level. Thus another advantage of focusing on academic qualifications is that one can look at incremental returns, since those who have an A level or university qualification also have the preceding lower qualifications (O level or O level and A level, respectively). An interesting parameter is thus the return from attaining at least A levels compared to stopping at O levels, capturing the full incremental return to A levels.

It is important to highlight that since we compare O and A level attainment recorded by the schools by the time the individuals were aged 20 to O and A level attainment self-reported by individuals by the time they were aged 23, we need to further assume:

Assumption 7 ***(Age-20 completion)*** *O Level and A Level qualifications are completed by age* 20. ■

We can however safely consider this assumption to be met, at the very least for our NCDS cohort. From the school files, we could verify that only a negligible fraction of O levels achieved by age 20 had in fact been achieved after the typical age of 16 and similarly hardly any of the A levels achieved by age 20 had been achieved after the typical age of 18.

[29] We can see how likely A-level or above includes HE, that is $P(HE|A+)$.

Tables 5-7 present cross tabulations between the various educational measurements described in Table 1.[30] Before briefly discussing the tables, it is reassuring to note that the patterns that emerge from them are the same irrespective of the samples selected on the basis of non-missing educational information ever or non-missing wage information in 1991 (the latter obviously also restricting attention to those employed in 1991).

If we were to believe the school files, almost 5% of those students who did achieve O-levels reported to have no academic qualifications at age 23, while only a much smaller proportion (1.4%) incorrectly denied having taken their A-levels. At age 33, when asked to recall the qualifications they had attained by age 23, individuals are observed to make more mistakes, with over 10% of O-level achievers and 5% of A-level achievers 'forgetting' their attainment. Conversely, still taking the school files at face value, it appears that almost one fifth of those with no formal qualifications over-report their achievement, mostly stating that they have obtained O levels. A smaller but still sizeable fraction of 13% of those who according to the school files have only achieved O levels maintain to have in fact A levels. As was the case with under-reporting, over-reporting behaviour seems to worsen when moving further away from the time the qualification was achieved. When relying on recall information, almost one fourth of individuals with no formal qualifications state to have some, while almost 15% of individuals with O levels as their highest qualification according to the administrative files affirm to have A levels or even HE. This discussion on over-reporting crucially relies on assumption 7 that O and A level be completed by age 20. Although there is indirect support for this from the data and anecdotal evidence, in future research we will explore the sensitivity of our estimates to violations of this assumption.

8 Results

Based on the assumptions of misreporting being both independent of X and limited (respectively Cases III and I in Section 5.4), Table 2 reports the lower and upper bounds for the return to higher education and for the return to attaining any qualification that arise from different values of the sum of the misclassification probabilities. Educational attainment is based on achievement by age 33 self-reported at age 33.

[30]Note that our measures are in terms of highest achievement, and are thus not directly comparable to the ones used by Ives (1984).

Table 2: Base case bounds on treatment effects

	$\lambda_0 + \lambda_1 \geq 1.4$		$\lambda_0 + \lambda_1 \geq 1.5$		$\lambda_0 + \lambda_1 \geq 1.6$	
	lower	upper	lower	upper	lower	upper
HE vs less	0.222	0.634	0.222	0.505	0.222	0.416
Any vs None	0.278	0.680	0.278	0.555	0.278	0.463
	$\lambda_0 + \lambda_1 \geq 1.7$		$\lambda_0 + \lambda_1 \geq 1.8$		$\lambda_0 + \lambda_1 \geq 1.9$	
	lower	upper	lower	upper	lower	upper
HE vs less	0.222	0.349	0.222	0.300	0.222	0.257
Any vs None	0.278	0.396	0.278	0.347	0.278	0.308

In Table 3 we will present some GMM results for the return to achieving any academic qualification and for the return to achieving at least A levels qualifications. Returns and misclassification probabilities will be calculated using administrative information on academic qualifications obtained by age 20 together with academic qualifications obtained by age 23 as self-reported either at age 23 or, alternatively, at age 33.

9 Conclusions

This paper provides reliable estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment. We additionally identify the extent of misreporting in different types of commonly used data sources on educational qualifications and thus provide estimates of their relative reliability. We also intend to produce some simple rules as to how to correct returns estimated on data that rely on recall about individual qualifications and contain limited or no information on individual ability and family background characteristics (such as the Labour Force Survey). We also plan to explore the advantages of combining repeated measurements of education reported by the same individuals over time. Finally, we will consider the impact of misclassification on evaluation methods based on *propensity score matching*.

Table 3: Academic qualifications: point estimates

	age 23, self-reported at 23		age 23, self-reported at 33	
	Any vs None	**A+ vs Less**	**Any vs None**	**A+ vs Less**
	Raw and true returns			
Δ				
Δ^*				
	Transcript information: probabilities of exact classification			
λ_1^T				
λ_0^T				
γ_1^T				
γ_0^T				
	Self-reported information: probabilities of exact classification			
λ_1^S				
λ_0^S				
γ_1^S				
γ_0^S				

$\lambda_1^j \equiv Pr(D^* = 1|D_j = 1)$ and $\gamma_1^j \equiv Pr(D_j = 1|D^* = 1)$, $j = T, S$
$\lambda_0^j \equiv Pr(D^* = 0|D_j = 0)$ and $\gamma_0^j \equiv Pr(D_j = 0|D^* = 0)$, $j = T, S$

References

[1] Aigner, D. (1973), *Regression with a Binary Independent Variable Subject to Errors of Observation*, Journal of Econometrics, 1, 49-60.

[2] Angrist, J. (1998), *Estimating the labour market impact of voluntary military service using social security data on military applicants*, Econometrica, 66, 249–88.

[3] Battistin, E. and Chesher, A. (2004), *The Impact of Measurement Error on Evaluation Methods Based on Strong Ignorability*, unpublished manuscript, Institute for Fiscal Studies.

[4] Black, D., M. Berger, and F. Scott (2000), *Bounding Parameter Estimates with Non-Classical Measurement Error*, Journal of the American Statistical Association, 95, 451, 739-48.

[5] Black, D., Sanders, S. and Taylor, L. (2003), *Measurement of Higher Education in the Census and Current Population Survey*, Journal of the American Statistical Association, 98, 463, 545-554.

[6] Blanden, J., Goodman, A., Gregg, P. and Machin, S. (2002), *Changes in Intergenerational Mobility In Britain*, Centre for Economics of Education Discussion Paper No. 26.

[7] Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000), *The Returns to Higher Education in Britain: Evidence from a British Cohort*, Economic Journal, 110, F82–F99.

[8] Blundell, R. Dearden, L. Sianesi, B. (2004), *Measuring the Returns to Education*, chapter 6 in Machin, S. and Vignoles, A. (eds), The Economics of Education in the UK, Princeton University Press, forthcoming.

[9] Blundell, R. Dearden, L. Sianesi, B. (2005), *Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS*, forthcoming, Journal of the Royal Statistical Society A.

[10] Bollinger, C.R. (1996), *Bounding Mean Regressions When a Binary Regressor is Mismeasured*, Journal of Econometrics, 73, 387-399.

[11] Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D., and Spector, T., (2003), *Education and Earnings: Evidence from UK Twins*, American Economic Review, December, 1799-1812.

[12] Bound, J., Brown, C. and Mathiowetz, N. (2001), *Measurement error in survey data*, in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics. Vol. 5*, Amsterdam: North-Holland, 3705-3843.

[13] Card, D. (1999), *The Causal Effect of Education on Earnings*, Handbook of Labor Economics, Volume 3, Ashenfelter, A. and Card, D. (eds.), Amsterdam: Elsevier Science.

[14] Chevalier, A., Harmon, C., Walker, I. and Zhu, Y. (2003), *Does education raise productivity?*, University College Dublin Working Paper, ISSC, WP2003/01, Dublin.

[15] Conlon, G. (2001), *The Differential in Earnings Premia between Academically and Vocationally Trained Males in the United Kingdom*, Centre for Economics of Education Discussion Paper No. 11.

[16] Dearden, L. (1999a), *The Effects of Families and Ability on Men's Education and Earnings in Britain*, Labour Economics, 6, 551–67.

[17] Dearden, L. (1999b), *Qualifications and earnings in Britain: how reliable are conventional OLS estimates of the returns to education?*, IFS working paper W99/7.

[18] Dearden, L., McIntosh, S., Myck, M. and Vignoles, A. (2000), *The Returns to Academic and Vocational Qualifications in Britain*, Centre for the Economics of Education Discussion Paper No. 04.

[19] Dearden, L., McIntosh, S., Myck, M. and Vignoles, A. (2002), *The Returns to Academic and Vocational Qualifications in Britain*, Bulletin of Economic Research, 54, 249-274.

[20] Del Bono, E. and Galindo-Rueda, F. (2004), *Do a Few Months of Compulsory Schooling Matter? The Education and Labour Market Impact of School Leaving Rules*, IZA Discussion Paper No. 1233

[21] Fisher, R.A. (1935), *The Design of Experiments*, Edinburgh: Oliver&Boyd.

[22] Frazis, H. and Loewenstein, M.A. (2002), *Estimating Linear Regressions with Mismeasured, Possibly Endogenous, Binary Explanatory Variables*, Working Paper 355, U.S. Bureau of Labor Statistics.

[23] Galindo-Rueda, F. and Vignoles, A. (2003), *Class-Ridden or Meritocratic? An Economic Analysis of Recent Changes in Britain*, Centre for Economics of Education Discussion Paper No. 32.

[24] Gosling, A., Machin, S. and Meghir, C. (2000), *The changing distribution of male wages, 1966–93*, Review of Economic Studies, 67, 635-666.

[25] Griliches, Z. (1977), *Estimating the returns to schooling: some econometric problems*, Econometrica, 45, 1–22.

[26] Harmon, C. and Walker, I. (1995), *Estimates of the Economic Return to Schooling for the UK*, American Economic Review, 85, 1278–86.

[27] Hausman, J.A., Abrevaya, J. and Scott Morton, F. (1998), *Misclassification of the dependent variable in a discrete response setting*, Journal of Econometrics, 87, 239–269.

[28] Heckman, J.J. Lalonde, R. and Smith, J. (1999), *The Economics and Econometrics of Active Labor Market Programs*, Handbook of Labor Economics, Volume 3, Ashenfelter, A. and Card, D. (eds.), Amsterdam: Elsevier Science.

[29] Imbens, G.W. (2004), *Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review*, Review of Economics and Statistics, 86, 4-29.

[30] Ives, R. (1984), *School reports and self-reports of examination results*, Survey Methods Newsletter, Winter 1984/85, 4-5.

[31] Kane, T.J., Rouse, C. and Staiger, D. (1999), *Estimating Returns to Schooling when Schooling is Mismeasured*, National Bureau of Economic Research Working Paper No. 7235.

[32] Lewbel, A. (2004), *Estimation of Average Treatment Effects With Misclassification*, unpublished manusctipt, Department of Economics, Boston College.

[33] Mahajan, A. (2003), *Misclassified Regressors in Binary Choice Models*, mimeo, Department of Economics, Stanford University, October

[34] Manski, C. F. (1990), *Nonparametric Bounds on Treatment Effects*, American Economic Review Papers and Proceedings, 80, 319-323.

[35] Matyas, L. (ed.) (1999), *Generalized Method of Moments Estimation*, Cambridge University Press.

[36] McIntosh, S. (2004), *Further Analysis of the Returns to Academic and Vocational Qualifications*, Centre for the Economics of Education Discussion Paper No. 35.

[37] Molinari, F. (2004), *Partial Identification of Probability Distributions with Misclassified Data*, unpublished manusctipt, Department of Economics, Northwestern University.

[38] Neyman, J. (with co-operation by Iwaszkiewicz, K. and Kolodziejczyk, S.) (1935), *Statistical Problems in Agricultural Experimentation*, Supplement of the Journal of the Royal Statistical Society, 2, 107-180.

[39] Quandt, R. (1972), *Methods for Estimating Switching Regressions*, Journal of the American Statistical Association, 67, 306-310.

[40] Robinson, P. (1997), *The Myth of Parity of Esteem: Earnings and Qualifications*, London School of Economics, Centre for Economic Performance, Discussion Paper No. 354.

[41] Roy, A. (1951), *Some Thoughts on the Distribution of Earnings*, Oxford Economic Papers, 3, 135-146.

[42] Rosenbaum, P.R. and Rubin, D.B. (1983), *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, Biometrika, Vol. 70, No. 1, 41-55.

[43] Rubin, D.B. (1974), *Estimating Causal Effects of Treatments in Randomised and Non-randomised Studies*, Journal of Educational Psychology, 66, 688-701.

[44] Rubin, D.B. (1980), *Discussion of 'Randomisation analysis of experimental data in the Fisher randomisation test' by Basu*, Journal of the American Statistical Association, 75, 591–3.

[45] Sianesi, B. (2003), *Returns to Education: A Non-Technical Summary of CEE Work and Policy Discussion*, mimeo, June.

Proof of Proposition 1

By using (4) we have that

$$\begin{aligned}
\Delta^*(x) &= \begin{bmatrix} -1 & 1 \end{bmatrix} \Pi^{-1}(x) \begin{bmatrix} E(Y|D=0) \\ E(Y|D=1) \end{bmatrix}, \\
&= \begin{bmatrix} -1 & 1 \end{bmatrix} \frac{1}{det[\Pi(x)]} \begin{bmatrix} \lambda_1(x) & \lambda_0(x)-1 \\ \lambda_1(x)-1 & \lambda_0(x) \end{bmatrix} \begin{bmatrix} E(Y|D=0) \\ E(Y|D=1) \end{bmatrix}, \\
&= \frac{\Delta(x)}{\lambda_0(x)+\lambda_1(x)-1}.
\end{aligned}$$

The same result can be derived by noting that Assumption 3 implies

$$\begin{aligned}
E(Y|D=1,x) &= E(Y|D^*=0,x) + \Delta^*(x)\lambda_1(x), \\
E(Y|D=0,x) &= E(Y|D^*=0,x) + \Delta^*(x)\lambda_{10}(x),
\end{aligned}$$

so that by taking the difference of the last two expressions

$$\Delta(x) = \Delta^*(x)[\lambda_1(x) - \lambda_{10}(x)],$$

so that the result follows since $\lambda_{10}(x) = 1 - \lambda_0(x)$.■

Proof of Proposition 2

Using Bayes theorem we get

$$\begin{aligned}
f(x|D=1) &= \frac{e(x)f(x)}{Pr(D=1)} \\
f(x|D^*=1) &= \frac{e^*(x)f(x)}{Pr(D^*=1)}
\end{aligned}$$

where $e(x)$ is the propensity score calculated from D. Since by the law of iterated expectations we have

$$e^*(x) = [1 - \lambda_0(x)] + e(x)[\lambda_0(x) + \lambda_1(x) - 1] \tag{13}$$

it follows that

$$\begin{aligned}
f(x|D=1) &= f(x|D^*=1)\frac{Pr(D^*=1)}{[\lambda_0(x)+\lambda_1(x)-1]Pr(D=1)} \\
&+ \frac{[\lambda_0(x)-1]f(x)}{[\lambda_0(x)+\lambda_1(x)-1]Pr(D=1)}.
\end{aligned}$$

By combining the results above to (5) we get the first relationship stated in Proposition 2. Moreover, using (13) and the fact that

$$Pr(D^* = 1) = \int e^*(x)f(x)dx,$$

we have

$$\frac{Pr(D^* = 1)}{Pr(D = 1)} = \frac{\int [1 - \lambda_0(x)]f(x)dx}{\int e(x)f(x)dx} + \frac{\int [\lambda_0(x) + \lambda_1(x) - 1]e(x)f(x)dx}{\int e(x)f(x)dx}. \blacksquare$$

Data appendix

Table 4: Sample selection

	sample size
NCDS birth cohort	17,000
Non-missing education	
1978 Exam Files	14,331
1981 Survey	12,537
1991 Survey	11,407
None missing	8,504
Males with non-missing wage in 1991	3,639
Non-missing wage in 1991 and education ever	2,713

Table 5: Academic (and vocational) qualifications

	School Files		1981 Survey		1991 Survey	
	Percent	**Sample**	**Percent**	**Sample**	**Percent**	**Sample**
None	41.2	1,118	35.0	950	23.2	628
O (or eq)	37.1	1,006	38.4	1,042	25.7	698
A (or eq)	21.7	589	26.6	721	18.0	489
HE or eq					33.1	898

Only academic qualifications (below HE) for School files and the 1981 Survey

Table 6: Academic qualifications: conditional probabilities by qualifications in the School files

age 20	age 23, at 23			age 23, at 33				age 33, at 33			
	None	**O**	**A**	**None**	**O**	**A**	**HE**	**None**	**O**	**A**	**HE**
None	80.6	18.3	1.1	76.1	22.5	1.1	0.3	72.8	24.5	1.3	1.3
O	4.8	82.5	12.7	10.5	74.8	10.1	4.6	6.1	72.9	10.6	10.4
A+	0.2	1.2	98.6	3.7	1.2	48.2	46.9	1.2	1.0	33.5	64.4

Table 7: Academic qualifications: sample size

age 20	age 23, at 23			age 23, at 33				age 33, at 33			
	None	**O**	**A**	**None**	**O**	**A**	**HE**	**None**	**O**	**A**	**HE**
None	901	205	12	851	252	12	3	814	274	15	15
O	48	830	128	106	752	102	46	61	733	107	105
A+	1	7	581	22	7	284	276	7	6	197	379