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

5 April 2006

Abstract

In this paper we study the impact of misreported treatment status on the estimation of causal treatment effects. We characterise the bias introduced by misclassification on the average treatment effect on the treated under the assumption of selection on observables. Although the bias of matching-type estimators computed from misclassified data cannot in general be signed, we show that the bias is most likely to be downward if misclassification does not depend on variables entering the selection-on-observables assumption, or only depends on such variables via the propensity score index. We extend the framework to multiple treatments. We provide results to bound the returns to a number of educational qualifications in the UK semi-parametrically, and by using the unique nature of our data we assess the plausibility for the two biases from measurement error and from omitted variables to cancel out.

JEL Codes: C10, I20, J31.

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

*First draft September 2004. This paper benefited from helpful discussions with Richard Blundell, Andrew Chesher, Francesca Molinari, Peter Mueser, Enrico Rettore and Yu Xie and comments by audiences at University of Padova (July 2004), "XIX Italian Conference of Labour Economics" (September 2004), "Cemmap Metrics Lunch Seminar" (November 2004), Bank of Italy (April 2005), Second World Conference SOLE/EALE (June 2005), "The Empirical Evaluation of Labour Market Programmes" (Nuremberg, June 2005), ESWC (August 2005), Policy Studies Institute (September 2005), ADRES Conference on "Econometric Evaluation of Public Policies: Methods and Applications" (December 2005), Franco Modigliani Fellowship Workshop (February 2006), CEE (February 2006), Michigan (March 2006), Kentucky (March 2006) and Maryland (March 2006). 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.battistin@unipd.it and barbara_s@ifs.org.uk.

1 Introduction

Countless theoretical and applied work has addressed itself to the evaluation problem, that is to the measurement of the causal impact of a generic 'treatment' on one or more outcomes of interest (for a review see Heckman, LaLonde and Smith, 1999, and Imbens, 2004). Especially in applied work, interest mostly lies in recovering the average effect of the treatment on the sub-population of participants - the average treatment effect on the treated (ATT in the following).

While endogeneity of treatment status has been the main preoccupation of both theoretical and empirical research, measurement error in recorded treatment status and its consequences for the estimation of causal effects has received far less attention (Molinari, 2004, Lewbel, 2005, and Mahajan, 2006, are the only examples we are aware of). However, in empirical applications the possibility of misrecorded treatment status is far from negligible. 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 (e.g. eligibility to some means-tested benefits); or a randomised study where the extent of actual compliance (in terms of participants failing to take the treatment and/or controls taking an alternative one) is not recorded in the data. A more general class of example applications comprises those situations in which the treatment status is derived by splitting the sample based on an underlying continuous variable which is itself potentially measured with error (e.g. income or consumption to define poverty status, or firm size to define some form of eligibility).

Since the treatment indicator is a binary or categorical variable, any measurement error will by construction vary with the true treatment status (e.g. Aigner, 1973, or Card, 1996). In the presence of such non-classical errors, OLS estimates are biased, though not necessarily downward, and so would standard IV (e.g. Bound, Brown and Mathiowetz, 2001).[1]

[1]By contrast, if it were a continuous variable to be affected by measurement error, standard results show that OLS estimates would be downward biased, while appropriate IV methods applied to the linear regression model would provide consistent estimates.

In this paper we study the impact of misreported treatment status on the estimation of causal treatment effects. In particular, we characterise the bias introduced by misclassification on the ATT under the assumption of selection on observables, and derive results for partial identification that do not rely on additional information (e.g. multiple measures or instrument-like variables) and that can be implemented in a non/semi-parametric fashion. We further extend the framework to multiple treatments.

The causal effects that motivated this paper are the wage returns to educational qualifications in the UK. While the estimation of the return to education is amongst the most explored and prolific areas in labour economics and attracts constant policy interest (for a discussion, see Blundell, Dearden and Sianesi, 2004), there is a real possibility of errors in education data: in addition to data transcript errors, survey respondents may either over-report their attainment, simply not remember, or just not know if the schooling they have had counts as a qualification. As to the latter in particular, the British education system is remarkably complex, with a plethora of different - and changing - sub-qualifications classified in broader levels, often based on obtained grades.

The received wisdom from the studies on the returns to years of education has traditionally been that the upward bias from omitted 'ability' variables and the downward bias from (classical) measurement error 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). Such a continuous years-of-schooling measure, although particularly convenient, imposes the restriction that the returns increase linearly with each additional year, irrespective of the level and type of educational qualifications the years refer to. 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 and allowing each to have a separate effect on earnings is widely accepted (see the discussion in Blundell, Dearden and Sianesi, 2005). With a categorical qualification-based measure of education, however, as mentioned above the assumption of classical measurement error cannot hold, as individuals in the lowest category can never under-report their education level and individuals in the top category cannot over-report. Given that OLS estimates are not necessarily downward biased, 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.

To date, empirical evidence on the importance of these issues for the estimation of returns to education 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).

When an instrument-like variable, or an additional, independent measure of the treatment of interest becomes available, point identification of returns can be achieved (Lewbel, 2005, and Mahajan, 2006). In our companion paper (Battistin and Sianesi, 2006) we exploit the characterisation of the bias developed in this paper together with repeated measurements of individual educational qualifications to arrive at point estimates of the returns to educational qualifications that allow for measurement error. However, multiple measures are far from being the norm in practical applications; this paper thus suggests a bounding approach and corresponding sensitivity analysis that while easy to implement can provide an often quite informative robustness check. In most instances, 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; alternatively, bounds can be calculated for a range of plausible values of reporting errors to assess the robustness of the evaluation inference to the presence of misclassification.

This paper provides a number of new contributions of considerable policy and practical relevance, as well as of more general methodological interest. We start by characterising the bias introduced by misclassification on the average treatment effect on the treated under the assumption of selection on observables (or conditional independence). To the best of our knowledge, all papers dealing with this issue only consider the treatment effect for a given value of the observable characteristics. Although the resulting bias of matching-type estimators computed from misclassified data cannot in general be signed, we show that the bias is most likely to be downward if misclassification does not depend on variables entering the selection-on-observables assumption, or if misclassification only depends on such variables via the propensity score index.

We further extend our approach to a multiple-treatment framework, which becomes necessary if interest lies in estimating the incremental impacts of multiple treatments and in fact the impacts of binary but more narrowly-defined treatments that do not split up the entire population. In either of such cases, account needs to be taken of the potential misclassification in the reporting of *all* treatment levels, not just in the two ones being considered. Furthermore, as we argue in more detail in Section 6, the move to a multiple treatment framework is often necessary just to be able to justify the non-differential misclassification assumption widely invoked in the literature (see Bound, Brown and Mathiowetz, 2001).

The characterisation of the bias we provide straightforwardly suggests the derivation of bounds for the ATT(s) by making *a priori* assumptions on the extent of misclassification. Such bounds can be derived by exploiting the observed propensity score in a non/semi-parametric way, in particular allowing for arbitrarily heterogeneous individual returns, for arbitrary non-linearities in the no-treatment outcome equation and for misreporting probabilities to depend on individual characteristics, albeit only through the propensity score.

In our empirical application we implement this approach and provide bounds to the returns to a number of educational qualifications in the UK, both in a binary and multiple-treatment setting. To motivate the conditional independence assumption we 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.

Further, by using the uniquely rich National Child Development Survey (NCDS) data we assess the plausibility that the biases from measurement error and from omitted variables cancel out in the estimation of returns in the UK. If this were the case, to estimate up-to-date returns to qualifications policy-makers could simply rely on Labour Force Survey-type datasets, which totally rely on recall about individuals' and do not contain any information on individual ability and family background.

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 introduces 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 causal treatment effects. In Section 6 we discuss how to extend our identification strategy to deal with misreporting of categorical treatments. 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, and defines our parameters of interest. In Section 8 we first sketch our strategy for partial identification of causal effects in the presence of misclassification, before presenting and discussing our results. 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, 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 misreporting might have on estimated returns to education.[3] 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.

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

[3]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.

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 percent of those with a degree reported so accurately and less than 1 percent 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 percent of those with no completed years of college misreported to have completed some, and 6 percent of those who had completed some college misreported to have completed none. Estimates of returns that ignore such misclassification were found to be severely biased, either upwards or downwards depending on the educational level of interest. Similarly, the application in Lewbel (2005) 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.

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.[4] 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 we start by considering the *binary treatment* setting, with treatment defined by having achieved an educational outcome or not, denoted by $D^* = 1$ and $D^* = 0$. Examples include completing college compared to not doing so, or attaining any qualification compared to dropping out of high-school with none. The generalization to the multiple-treatment

[4]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).

case, though notationally more demanding, proceeds along the same lines and will be considered in Section 6. 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 were to achieve 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, $Y_1 - Y_0$. The observed individual wage can then be written as $Y = Y_0 + D^*(Y_1 - Y_0)$, 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.[5]

Since no individual can be in two different educational states at the same time, either Y_1 or Y_0 is missing, which makes it impossible to ever observe the individual return. A more modest though still challenging aim is to identify the average return in some population of interest. The 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 qualification of interest:

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

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 (post-compulsory) educational qualifications is voluntary, in this paper we shall focus on the ATT, capturing the average payoff to individuals' own educational choices.[6]

[5]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.

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

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:*

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

This assumption 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.

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

$$\Delta^*(x) \equiv E(Y_1 - Y_0|x) = E(Y|D^* = 1, x) - E(Y|D^* = 0, x)$$

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 in the population. This effect is integrated with respect to the distribution of X for participants.[7]

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.

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), and analogue estimators of (2) will therefore be constructed from (Y, D, X).

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

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

where:

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

[7] In its bare essentials, estimation proceeds by considering the empirical analogues of the quantities on the right-hand-side of (2). In particular one can perform any type of semi/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 methods (see Imbens, 2004, for a review).

Hence $\mathcal{S}$ 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 further below, 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. It is also worth noting that estimators of the ATT based on $e(x)$ (e.g. estimators based on propensity score matching or re-weighting) are equivalent to the estimator defined by the empirical analogue of (3), as we have that:

$$\Delta = \int_{\mathcal{S}} \Delta[e(x)] f[e(x)|D=1] de. \tag{4}$$

The result straightforwardly follows from x being finer than $e(x)$ and by the balancing property of the propensity scores (see Rosenbaum and Rubin, 1983).

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) and the parameter that would instead be estimated from the observable triple (Y, D, X) – namely Δ.

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

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

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:

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

where for ease of notation only one subscript is retained.[8] 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.

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*:*

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

For the binary treatment case, this amounts to the following two conditions:

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

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:

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

and

$$E(Y_1 - Y_0|D^*=1, D=1, X) = E(Y_1 - Y_0|D^*=1, D=0, X).$$

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 and/or because they have reaped a very low if not negative return from their degree - are more inclined to deny possessing the qualification. In addition to such type of behaviour by respondents, there is a more technical consideration that would in fact guarantee a violation

[8] The (mis)classification probabilities can also be 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.

of this assumption. If in defining the treatment indicator one were to ignore a feature of the treatment that affects both its effect and recall precision, Assumption 3 would by construction break down. The obvious solution to this is to refine the treatment to fully reflect the feature causing the violation, thus extending the framework to look at the treatment components separately. We further elaborate on this issue in Section 6, where we turn to multiple treatments.

Under Assumption 3, individuals for whom we observe $D = d$ are a *mixture* of participants ($D^* = 1$) and non participants ($D^* = 0$), with mixing weights given by the (mis)classification probabilities. This result can be written 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

$$\Pi^{-1}(x) \begin{bmatrix} E(Y|D=0,x) \\ E(Y|D=1,x) \end{bmatrix} = \begin{bmatrix} E(Y|D^*=0,x) \\ E(Y|D^*=1,x) \end{bmatrix}. \tag{5}$$

provided that $det[\Pi(x)] = \lambda_0(x) + \lambda_1(x) - 1 \neq 0$. We formalise this condition as:

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 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-(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. By contrast, 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. 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, i.e. $\lambda_1(x) > 0.5$ and $\lambda_0(x) > 0.5$ (see e.g. Bollinger, 1996).[9]

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 (5). The result in (6) coincides with the result in Lewbel (2005), 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*

$$\Delta^*(x) = \frac{\Delta(x)}{\lambda_0(x) + \lambda_1(x) - 1}. \tag{6}$$

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, an attenuation bias result still holds. An interesting implication of (6) 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 $\Delta(x) = -\Delta^*(x)$.

[9] Another way to look at this is to note that $Cov(D, D^*|x) = (\lambda_1(x) + \lambda_0(x) - 1)Var(D|x)$. Hence 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.

5.2 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 there are both individuals with $D^* = 1$ and with $D^* = 0$, 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:

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

It follows that the parameter (3) estimated from the triple (Y, D, X) could in general refer to a different population than the one implied by (Y, D^*, X).

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

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. 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 were interested in the average treatment effect (ATE), that is the average return for an individual irrespective of whether the qualification of interest has been acquired or not:

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

the discussion could stop here.[10] 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 would keep holding unconditional on X, so that ignoring measurement error in treatment status would lead to a downward-biased estimate of the ATE. The correspondence between a zero raw average effect and a zero true 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 average return to education for those who invested in that qualification (ATT), 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.[11] 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.

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

$$\begin{aligned} \Delta^* &= \int \omega(x)\Delta(x) f(x|D=1) dx, \\ &= \Delta + \int [\omega(x) - 1]\Delta(x) f(x|D=1) dx, \end{aligned} \tag{7}$$

[10] 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.

[11] By contrast, this distribution could be directly inferred if information on the true treatment status D^* and X were available from an external survey. In this case, it can be easily shown that Δ would underestimate Δ^*.

where

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

and

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

This result 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, $\Delta(x)$, with weights defined by $\omega(x)$. Note that, as it should be, $\omega(x) = 1$ for all individuals if there is no measurement error. Moreover, weights cannot be signed in general, implying that Δ^* can be over- or under-estimated depending on the unknown probabilities $\lambda_1(x)$ and $\lambda_0(x)$. A notable exception under limited mislcassification (4-(a)) is when the true incidence of treatment in the population, $P(D^*=1)$, is smaller than the one observed from raw data, $P(D=1)$. This is a sufficient condition for Δ to provide a downward-biased estimate of Δ^*, as all weights would be larger than one. Although $P(D^*=1)$ is in general unobserved, one could gauge the relative size of the two probabilities if external validation data (e.g. government statistics on educational attainment in our application) were available.

Furthermore, Δ being zero no longer implies the absence of a treatment effect, as was the case when conditioning on X.

Finally, it is worth noting that, because of (4), the result derived in Proposition 2 also applies to the bias induced by misclassification when estimation is carried out with respect to the observed propensity score $e(x)$.

5.4 Special cases

In what follows we discuss two sets of restrictions that can be imposed on the probabilities $\lambda_1(x)$ and $\lambda_0(x)$ to sign the bias induced by misclassification.

Special Case 1 *(**Only over-reporting of qualifications**) The misclassification probabilities are such that only over-reporting can happen:*

$$\lambda_0(x) = 1$$

for all values X.■

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. By setting $\lambda_0(x)$ equal to one in (6) we get that the conditional treatment effect is always right-signed for all X, but biased towards zero. Furthermore, it follows from Proposition 2 that

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

for all X, so that the estimated effect Δ is always biased towards zero for Δ^*.[12]

Special Case 2 *(**Misclassification independent of** X) The percentage of correct classification is independent of the characteristics X of respondents:*

$$\lambda_1(x) = \lambda_1 \quad \textit{and} \quad \lambda_0(x) = \lambda_0$$

for all values X.■

Although this assumption is clearly only made here for convenience, it could be weakened by assuming constant probabilities within cells defined by X. Alternatively, the same arguments made below would still apply if one allowed misclassification to vary with X but only through the propensity score index, $e(X)$.

Using Proposition 2 it follows that

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

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 (and testable) condition for $\omega(x)$ to be larger than one 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

[12]The 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 that $\lambda_1(x) < \lambda_0(x)$ 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.

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 one. However we found that even in this case the distribution of weights is skewed towards values (often much) larger than one, so that in most empirical applications the 'raw' estimate is most likely to be a lower bound.[13]

6 Extension to multiple treatments

So far we have considered treatments within the binary (single) treatment framework, and in fact treatments where the specific educational level of interest cuts right through the entire educational spectrum (e.g. any qualification versus none, or degree versus non-degree). This does not of course rule out interest in the incremental returns to sequential multiple treatments, or in the returns to binary treatments for a more narrowly defined educational split, such as the return to college vis-à-vis stopping with high school diploma, or the return to college vis-à-vis dropping out from school without qualifications, or the return to finishing school with some qualifications vis-à-vis nothing.

It is important to note that even for considering such types of *binary* treatments, the analysis needs to be extended to a multiple-treatment framework, since account needs to be taken of the potential misclassification in the reporting of *all* educational levels, not just in the two ones being considered. So for instance, even if one only wanted to compare college to high school, the other categories would still need to be considered, since, first, individuals reporting no qualifications might in reality have a high-school diploma or a college degree, and, second, individuals reporting college or high school diploma might in reality have neither of the two qualifications of interest.

In addition to the policy relevance of estimating returns to multiple educational qualifications and/or to more disaggregated ones, another important reason to turn to the multiple treatment framework relates to the validity of the non-differential misclassification assumption. As anticipated in Section 4, this assumption would by construction be violated if in defining the treatment indicator one were to lump across features of the treatment that affect both its effect and the precision of its recall. To see this, consider the treatment being defined as having

[13] More detailed results are available upon request.

any qualification as opposed to none.[14] In such a situation, an individual with a degree will be more likely both to correctly report to have any qualification *and* to have higher earnings than an observationally-equivalent individual who has only completed high school.

More generally, a need to consider an extended framework arises in any situation where underlying the binary treatment indicator is a dose-response framework. In our application we consider educational categories, which are inherently ordered and sequential, but such a set-up can occur much more generally. For instance, when considering completion of college for those who enrolled, or participation in a programme for the unemployed, the underlying treatment – college or the programme – has itself a duration, which is likely to affect both recall of the event and outcomes. Another example relates to treatments taken more or less recently; recall is likely to depend on how long ago the treatment was received, and the treatment effects themselves might evolve, in particular depreciate, over time.

Assumption 3 would thus appear to be most defensible when the treatment is disaggregated into multiple treatments that fully embody the feature that if ignored would cause the violation (sequential categories, duration, how long ago taken, etc.). In conclusion, extending the evaluation framework to look at the treatment components/features as separate treatments not only is often policy-relevant, but in the presence of misclassification it may often become a necessity for justifying the widely invoked non-differential misclassification assumption.

With our educational application in mind, in the following we extend our framework to consider *three* levels of qualifications (or more generally, of exposure), which we assume to be of increasing intensity. Let these levels be defined by $D^* = 0$, $D^* = 1$ and $D^* = 2$, denoting, for example, high-school drop-outs, high-school graduates and college graduates, respectively. We are interested in the estimation of pairwise incremental returns, that is the wage return of obtaining a qualification of interest (e.g. college) relative to a lower qualification (e.g. high-school), when the only available measure of educational attainment D is potentially affected by error. We focus on the following three ATT's:

$$\begin{aligned}
\Delta^*_{10} &\equiv E(Y_1 - Y_0 | D^* = 1), \\
\Delta^*_{21} &\equiv E(Y_2 - Y_1 | D^* = 2), \\
\Delta^*_{20} &\equiv E(Y_2 - Y_0 | D^* = 2).
\end{aligned}$$

[14] We thank Peter Mueser for pointing this out to us.

The conditional ATT's based on *true* and *observed* attainment levels are respectively defined as follows:

$$\begin{aligned}\Delta^*_{ij}(x) &= E(Y|D^*=i,x) - E(Y|D^*=j,x),\\ \Delta_{ij}(x) &= E(Y|D=i,x) - E(Y|D=j,x),\end{aligned}$$

where $i>j$ and $(i,j)\in\{0,1,2\}$, for which the adding up conditions $\Delta^*_{20}(x)=\Delta^*_{10}(x)+\Delta^*_{21}(x)$ and $\Delta_{20}(x)=\Delta_{10}(x)+\Delta_{21}(x)$ hold by definition. Along the lines of what discussed in Section 4, the relationship between true quantities and quantities observed from raw data depends on the 3×3 matrix of misclassification probabilities through the following expression:

$$\begin{pmatrix} E(Y|D=0,x)\\ E(Y|D=1,x)\\ E(Y|D=2,x)\end{pmatrix} = \Pi(x)\begin{pmatrix} E(Y|D^*=0,x)\\ E(Y|D^*=1,x)\\ E(Y|D^*=2,x)\end{pmatrix}.$$

If this matrix is invertible, each $\Delta^*_{ij}(x)$ can be written as a function of the $\Delta_{ij}(x)$'s, thus providing an extension of the result in Proposition 1.

The ATT's of interest can then be written as:

$$\begin{aligned}\Delta^*_{10} &= \int \Delta^*_{10}(x)f(x|D^*=1)dx,\\ \Delta^*_{21} &= \int \Delta^*_{21}(x)f(x|D^*=2)dx,\\ \Delta^*_{20} &= \int \Delta^*_{20}(x)f(x|D^*=2)dx = \Delta^*_{21} + \int \Delta^*_{10}(x)f(x|D^*=2)dx,\end{aligned}$$

which depend on the conditional distributions of X given $D^*=1$ and $D^*=2$.

For the above quantities to have a causal interpretation we need suitably extended versions of Assumptions 1 and 2 (see e.g. Imbens, 2000). Furthermore, to move to the unconditional effects we need to invoke an extended version of Assumption 5. We combine all of these Assumptions in:

Assumption 6 ***(Extended Conditional Independence, Common Support and Extent of Misclassification)***

$$\begin{aligned}E(Y_j|D^* &= i,X) = E(Y_j|D^*=j,X), \qquad (i,j)\in\{0,1,2\}, i>j\\ 0 &< e^*_i(x)\equiv Pr(D^*=i|X=x)<1 \qquad i\in\{0,1\},\quad \forall x.\end{aligned}$$

Moreover, misclassification is such that the observed common support coincides with the true one for each pairwise comparison.

Restrictions imposed on the misclassification probabilities can help simplify the relationship between moments involving D^* and moments involving D, and therefore the analytical tractability of the problem. With our application in mind we thus impose a particular structure for the misclassification problem by assuming that misclassification can occur only for *adjacent* categories of education. This corresponds to assuming the following form for the misclassification matrix:

$$\Pi(x) = \begin{bmatrix} \lambda_{00}(x) & \lambda_{10}(x) & 0 \\ \lambda_{01}(x) & \lambda_{11}(x) & \lambda_{21}(x) \\ 0 & \lambda_{12}(x) & \lambda_{22}(x) \end{bmatrix}, \tag{8}$$

which is a function of the *four* unknown probabilities (because of three adding up conditions): $\lambda_{00}(x)$,$\lambda_{11}(x)$, $\lambda_{22}(x)$ and $\lambda_{21}(x)$. The following proposition extends Proposition 1 to the case of multiple treatments. The proof is reported in the Appendix.

Proposition 3 ***(Bias on Treatment Effects given X)*** *Provided that the determinant of* $\Pi(x)$ *is different from zero:*

$$\delta(x) \equiv \lambda_{00}(x)[\lambda_{22}(x) - \lambda_{21}(x)] - \lambda_{22}(x)[1 - \lambda_{11}(x) - \lambda_{21}(x)] \neq 0$$

and if Assumptions 3 and 6 hold, we have that:

$$\begin{aligned}
\Delta^*_{10}(x) &= \frac{\lambda_{22}(x) - \lambda_{21}(x)}{\delta(x)}\Delta_{10}(x) - \frac{\lambda_{21}(x)}{\delta(x)}\Delta_{21}(x), \\
\Delta^*_{21}(x) &= \frac{\lambda_{11}(x) + \lambda_{21}(x) - 1}{\delta(x)}\Delta_{10}(x) + \frac{\lambda_{00}(x) + \lambda_{11}(x) + \lambda_{21}(x) - 1}{\delta(x)}\Delta_{21}(x), \\
\Delta^*_{20}(x) &= \frac{\lambda_{11}(x) + \lambda_{22}(x) - 1}{\delta(x)}\Delta_{10}(x) + \frac{\lambda_{00}(x) + \lambda_{11}(x) - 1}{\delta(x)}\Delta_{21}(x).
\end{aligned}$$

A few comments are worth mentioning. The true conditional effects can be expressed as weighted sums of the raw effects $\Delta_{10}(x)$ and $\Delta_{21}(x)$ (remember the adding up condition). Weights are such that in the absence of misclassification raw effects coincide with true effects. Most importantly, in sharp contrast to the binary treatment case (see Proposition 1), the effects of misclassification on the relationship between $\Delta^*_{ij}(x)$ and $\Delta_{ij}(x)$ is not easily pinned down; depending on the extent and nature of misreporting across *all* categories, as well as on the sign and magnitude of both $\Delta_{10}(x)$ and $\Delta_{21}(x)$, $\Delta_{ij}(x)$ could be upward or downward biased for $\Delta^*_{ij}(x)$. Nonetheless, by assuming that the mean response is monotonically increasing with D^* (which corresponds to assuming non-negative wage returns) and that the misclassification error is non-differential, one can derive conditions on the extent of misclassification under which

$\Delta^*_{ij}(x)$ and $\Delta_{ij}(x)$ have at least the same sign. We show in the Appendix that if $\lambda_{ii}(x) > 0.5$ for all $i \in \{0, 1, 2\}$ sign reversal is avoided.

The following proposition extends Proposition 2 to the case of multiple treatments.

Proposition 4 *(**Bias on Treatment Effects**) If Assumptions 3, 6 are satisfied and* $\Pi(x)$ *has a non-zero determinant, the relationship between the true incremental ATT's and the effects estimated from raw data can be written as follows*

$$\Delta^*_{ij} = \int \omega_{1,ij}(x)\Delta_{10}(x)f(x|D=1)dx + \int \omega_{2,ij}(x)\Delta_{21}(x)f(x|D=2)dx,$$

where $i > j$ *and* $(i, j) \in \{0, 1, 2\}$*, and* $\omega_{1,ij}(x)$ *and* $\omega_{2,ij}(x)$ *are defined in the Appendix.*

7 Data and educational qualifications of interest

7.1 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, and our measure of educational qualifications is the one self-reported by respondents at age 33. Our sample of 3,642 is obtained by focusing on males only and restricting attention to those in work in 1991 with non-missing wage and education information.[15]

[15]Of the 5,606 males interviewed, the loss being mainly driven by non-response - less than 10% were not currently in work.

7.2 Educational qualifications of interest

We start by briefly outlining the educational system in Britain to put into context the educational qualifications to which we estimate the returns. Those students deciding to stay on past the minimum school leaving age of 16 can either continue along an academic route or else undertake a vocational qualification before entering the labour market. Until 1986, pupils choosing the former route 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. A levels still represent the primary route into higher education (HE). The vocational path is much more heterogeneous, 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, and Higher Education to College.[16]

In our binary framework we consider the following two parameters in turn:

1. the return to achieving **any academic qualification** compared to none[17];

   In the UK system, this translates into acquiring at least O levels compared to leaving school at the minimum age of 16 without any formal qualification, the counterfactual being thus akin to high-school drop-out status in the US. This parameters reflects a very well defined and homogenous qualification, and it 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. Additional policy relevance of the returns to O levels arises from the finding that reforms raising the minimum school leaving age in the UK have impacted on individuals achieving low academic qualifications, in particular O levels (Chevalier *et al.*, 2003, Del Bono and Galindo-Rueda, 2004).

[16] 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.

[17] The 'None' category also includes the level 1 academic qualification Certificates of Secondary Education (CSE) at grade 2 to 5. (Students at 16 could take the lower-level 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.)

2. the return from undertaking **some form of higher education** (HE) compared to anything less. This considers both the academic route and its vocational equivalent (levels 4 and 5).

When we extend our framework to multiple treatments we consider incremental returns to the following three broad and sequential education levels:

1. **no qualifications** (neither academic nor vocational);

   This treatment level basically reflects dropping out of school with no qualifications without later undertaking any vocational studies or formally recognised practice.[18]

2. **intermediate qualifications** (level 2 – O levels or their vocational equivalent);

   In addition to the academic O level exams held at age 16, this educational category includes their level-2 vocational qualifications (e.g. intermediate City and Guilds and Royal Society of Arts).

3. **advanced qualifications** (level 3 or above – at least A levels or their vocational equivalent).

   This collection of qualifications goes from high-school diploma (A levels) or level-3 vocational qualifications (e.g. advanced City and Guilds and Royal Society of Arts, or Ordinary National Diplomas, generally taken at age 16-18 and mostly in practical subjects such as hairdressing, catering, building techniques, or computing), all the way up to university and postgraduate studies or level-4 vocational (e.g. higher City and Guilds and Royal Society of Arts, or Higher National Diplomas) and level-5 (professional degree) qualifications.

Table 1 shows our sample split in our multiple treatment case, that is when we distinguish between those who stopped education with no formal qualification, those who stopped after completing level 2, and those who stopped after completing at least level 3. Note that this educational split encompasses academic school-based qualifications as well as their vocational equivalents. Next, the table shows the sample split into our two binary treatments: HE versus anything less, and any academic qualification versus none.

[18]The 'None' category also includes very low-level qualifications at NVQ level 1 or less, i.e. CSE grade 2 to 5 qualifications, other business qualifications, other qualifications not specified and Royal Society of Arts level 1 qualifications.

8 Partial identification of causal effects in the presence of misclassification

8.1 Estimation issues

The aim of this section is to discuss how we derived estimates of the ATT for *known* values of the misclassification probabilities and confidence intervals for the partially identified ATT. In our empirical application we will implement the approach outlined in this section to provide bounds to the returns to a number of educational qualifications in the UK. Such bounds can be derived by exploiting the *observed* propensity score in a non/semi-parametric way, in particular allowing for arbitrarily heterogeneous individual returns and leaving the no-treatment outcome equation unspecified.[19] Furthermore, both for the binary and the multiple treatment case, we allow misreporting to depend on X, albeit only through the propensity score index. To the best of our knowledge, we are the first ones to use the observed propensity score as a solution to deal with the curse of dimensionality arising from all these sources. In fact, most applications concerned with the estimation of the returns to educational attainment in the presence of misreporting either ignore the presence of X (Black, Berger and Scott, 2000); assume linearity, homogeneity in returns and misclassification independent of X (Kane, Rouse and Staiger, 1999); or impose a parametric structure together with possibly doubtful restrictions to ease estimation (Lewbel, 2005). The issue is further considered in our companion paper (see Battistin and Sianesi, 2006b).

The idea is most simply put across by considering the case of binary treatments, though it can be trivially extended to the case of multiple treatments. If the misclassification probabilities are constant with respect to X, we have shown that the weights in Proposition 2 vary with X only through the observed score $e(x)$. By applying the law of iterated expectations to the term on the right-hand-side of (7) we get

$$\begin{aligned} E\{\omega(x)\Delta(x)|D=1\} &= E\{\omega(x)E\{\Delta(x)|D=1,e(x)\}|D=1\}, \\ &= E\{\omega(x)\Delta[e(x)]|D=1\}, \end{aligned} \tag{9}$$

the last expression following since x is finer than $e(x)$ and from the observed propensity score

[19] Note that although both simple OLS regression and non-parametric methods such as matching rely on Assumption 1, matching is not subject to several potential misspecification biases for the ATT. In particular, OLS may suffer from misspecification bias for the no-treatment 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).

being a balancing score for the distribution of X for individuals with $D = 1$ and $D = 0$. Known values of the probabilities of correct classification uniquely define the weights $\omega(x)$, and alternative estimators of the ATT result from considering the empirical analogue of (9). For example, one could match individuals with $D = 1$ to individuals with $D = 0$ and then integrate the outcome difference in the two groups weighted by $\omega(x)$ with respect to the distribution of the score for $D = 1$.

Note that semi-parametric estimation is also feasible if weights are constant within cells defined by the propensity score $e(x)$ or, alternatively, if the misclassification probabilities are left to vary with X through $e(x)$. In this case we have

$$P(D^* = 1) = E\{1 - \lambda_0[e(x)]\} + E\{(\lambda_0[e(x)] + \lambda_1[e(x)] - 1)e(x)\}$$

and weights in Proposition 2 can be used.

By using Proposition 2, bounds on the true ATT can be derived by taking 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 $[0, 1] \times [0, 1]$. Indeed, leaving the misclassification probabilities to vary between zero and one is most likely to imply unreasonably high misclassification rates for the problem under consideration. One possibility is to use *a priori* restrictions on these probabilities 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 in Section 5.4.

In our application, we consider the empirical analogue of (9) by stratifying observations on the value of the propensity score and by allowing the λ's be *stratum specific*. Partial identification of the relevant ATT is then obtained through Proposition 2 (in the binary case) or Proposition 4 (in the multiple-treatment case) by considering the *maximum* and the *minimum* values of the Δ^*'s over the set defined by the sum of the exact classification probabilities exceeding a given value and by imposing that overreporting is more likely than underreporting.

As to the implementation details for the binary case, we first account for the high dimensionality of X by using stratification matching (see e.g Dehejia and Wahba, 1999) and regression adjustment within each stratum.[20] Specifically, we stratify the sample on 5 values

[20]Given our relatively small sample size, we performed the latter to account for residual imbalance of X

of the estimated propensity score. Within each stratum k we run a regression of the outcome Y on $e(x)$, separately for the $D = 1$ and $D = 0$ groups, and calculate the corresponding raw conditional treatment effect $\Delta(e_k)$ as the difference in predicted outcomes at the mean stratum propensity score. We then specify a grid of step 0.05 between 0.5 and 1 for $\lambda_0(x)$ and $\lambda_1(x)$, and consider only the 25 points that are consistent with the identification region defined by $\{\lambda_1(x) + \lambda_0(x) \geq 1.6)\} \cap \{\lambda_1(x) \leq \lambda_0(x)\}$, where the latter restriction reflects overreporting being more likely than underreporting (see Section 5.4). We thus end up with 25^5 possible combinations of $\{\lambda_1(x), \lambda_0(x)\}$ across strata. Finally, we compute the values of Δ^* using Proposition 2 for all admissibile combinations, and take the *maximum* and the *minimum* over the sets defined by $\lambda_0(x) + \lambda_1(x) \geq k$, where $k \in [1.6, 2]$.

In the multiple treatment case, we deal with the high dimensionality of X by defining strata on the two scores $e_1(x) \equiv P(D = 1|x)$ and $e_2(x) \equiv P(D = 2|x)$. To ensure cells of reasonable sample size, we stratify observations on 2 values of each score, thus obtaining 4 strata. We again perform regression adjustment within each stratum to account for residual imbalance, this time regressing Y on $e_1(x)$, $e_2(x)$ and $e_1(x) \cdot e_2(x)$ separately for each subgroup defined by D. For given values of λ_{22} and λ_{21}, we then construct a grid of step 0.25 between 0.5 and 1 for $\lambda_{00}(x)$ and $\lambda_{11}(x)$, and impose that $\lambda_{11}(x) \leq 1 - \lambda_{21}(x)$. This restriction is needed to ensure that the resulting $\Pi(x)$ is indeed a probability matrix. Note that, as in the binary case, the assumption we maintain that observations on (each value of) D are more accurate than pure guesses, $\lambda_{ii}(x) \geq 0.5$ for $i = 0, 1, 2$, implies that sign reversal is avoided and that misclassification is limited, in the sense that the probability of picking someone who truly has a given qualification is higher when drawing randomly from the group that claims to have that qualification rather than from the group claiming to have a different one. We finally further ensure that overreporting is more likely than underreporting, which in the multiple-treatment case translates into the condition that $\lambda_{00}(x) \geq \max\{\lambda_{11}(x) + \lambda_{21}(x), \lambda_{22}(x)\}$.[21] We finally derive bounds on the incremental ATT's using Proposition 4 and by proceeding as in the binary case.

A final issue concerns the significance of our estimates. A growing body of research in the

within stratum. Both in the case of any academic qualification and HE we could reject joint balancing of the observables at 10% for only one stratum.

[21] This translates into the assumption that all elements of $\Pi(x)$ above the diagonal are smaller than those below.

last years has looked into the problem of constructing confidence intervals for partially identified parameters. In our application, we follow Horowitz and Manski (2000) and derive confidence intervals for bounds that cover the entire *identification region* with 95 percent probability. By denoting with $\hat{L}$ and $\hat{U}$ the lower and the upper bounds, respectively, we report confidence intervals of the form $[\hat{L} - \zeta, \hat{U} + \zeta]$, where ζ is a positive constant obtained by bootstrapping the distribution of bounds so to ensure the required probability of coverage. As stated in Imbens and Manski (2004; see Lemma 1), the probability that the interval considered covers the *true* ATT is *at least* 95 percent (thus leading to conservative inference).

8.2 Results

8.2.1 Binary levels of attainment

The return Δ from attaining any academic qualification at 16 for those who chose to do so is estimated at 23.4 percent in the raw data (using the full set of controls outlined in Section 7.1). The average return to higher education for graduates is estimated at 23.1 percent (see Table 2).

We investigated the sensitivity of these estimates to the presence of misclassification by performing several types of analyses. We first used Proposition 2 to bound the true returns Δ^* by considering the case of misclassification independent of X when λ_0 and λ_1 are left to vary between 70 percent (very severe misclassification) and 100 percent (exact reporting). The first panel in Figure 1 plots how the value of the true return to any academic qualification varies as a function of the extent of misclassification. The minimum and the maximum values of the true return are 23.4 percent and 55.7 percent and are achieved for $\lambda_0 = 1$ and $\lambda_1 = 1$ and for $\lambda_0 = 0.7$ and $\lambda_1 = 0.7$. We thus find that, in line with our previous discussion, returns estimated from raw data represent a lower bound for the true return. A similar result holds for the return to higher education, for which the identification region corresponds to $(23.1, 60.1)$.

The second panel in Figure 1 considers the identifying power of assuming that over-reporting is more likely than under-reporting. Such restriction embodies the finding of cognitive studies that respondents tend to over-report desirable features and behaviours. As we have pointed out in Section 5.4, this corresponds to finding the maximum and the minimum in the region for which $\lambda_0 \geq \lambda_1$. It is evident from the figure that such a restriction on the nature of misclassification does not help improve the identification power, as the lower and upper bounds

coincide with those defined from the unconstrained region.

A further sensible restriction that can be imposed to obtain tighter bounds is $\lambda_0+\lambda_1 \geq k > 1$ for increasing values of k. This corresponds to consider the intersection of the second panel in Figure 1 with two-dimensional planes that increasingly shift to the right of the figure for higher values of k. We implemented this idea for the case of misclassification probabilities dependent on x as described in the previous section. Table 2 reports the lower and upper bounds for the return to attaining any academic qualification and for the return to completing higher education for a number of values of k. Each set of lower and upper bounds thus relates to a different value of the sum of the misclassification probabilities. The table also reports the 95 percent confidence intervals for the identification region, which as mentioned lead to conservative inference for the parameter of interest.

As expected, the bounds become more informative (narrower) as sum of λ's approaches 2. An interesting result is that even when we allow for a non-negligible extent of misclassification - up to 10 percent of individuals misreporting their attainment in either direction - the point estimates of the lower and upper bounds for the returns are quite close (i.e. around 23-26 percent for both any academic qualification and HE).

By using the uniquely rich NCDS data we can assess the plausibility that the biases from measurement error and from omitted variables cancel out in the estimation of returns in the UK. If this were the case, to estimate up-to-date returns to qualifications policy-makers could simply rely on Labour Force Survey-type datasets, which totally rely on recall about individuals' educational attainment and do not contain any information on individual ability and family background.

To perform this exercise, we have estimated the return from the raw data controlling only for the Labour Force Survey-style variables of gender, age, ethnicity and region (gender and age implicitly via sample choice), Δ_{LFS}. For academic qualifications, Δ_{LFS} is 32.9 percent and for HE 35.9 percent. As to returns to any academic qualification, for the sum of the λ's roughly larger than 1.8, ignoring both measurement error and ability biases yields an upper bound. More generally, we find that there is a *chance* for the two biases to cancel out only if measurement error is rather severe, in particular when at least 20 percent of individuals misreport their qualifications in either direction. In fact, for the case where $\lambda_0 + \lambda_1 \geq 1.9$, which is in line with the little available evidence so far (see in particular Kane, Rouse and

Staiger, 1999), even the conservative confidence intervals do not contain the point estimate of Δ_{LFS}. As to the returns to HE, it is even harder to appeal to the cancelling of the two biases.[22]

We also derived the bounds for the case of constant λ's. As expected, they turned out to be sharper; they however led to the same inferential conclusions about the true effect and its relationship with Δ_{LFS}.

8.2.2 Multiple levels of attainment

We now turn to a more disaggregated analysis in a multiple treatment framework that focuses on the incremental returns to three broad education levels. We consider a sequence starting with no (or extremely low-level) qualifications, O levels or vocational equivalent ("intermediate") qualifications and at least A levels or vocational equivalent ("advanced") qualifications. Ignoring potential misclassification, the incremental ATT's to acquire intermediate qualifications is 10.6 percent, to move from intermediate to advanced qualifications is 18.4 percent, and to acquire advanced qualifications compared to remaining with none is 28.3 percent.

As done for the binary case, we proceeded to assess the robustness of these estimates to the presence of misclassification in recorded educational attainment. Starting with the case of misclassification independent of X, we considered the bounds which arise when letting the exact classification probabilities (λ_{00}, λ_{11} and λ_{22}) vary between 70 percent (very severe misclassification) and 100 percent (exact classification), and the probability of 'forgetting' their advanced qualifications by those stating to have only intermediate ones (λ_{21}) between 0 and 15 percent.

As was the case for our binary treatments, the ensuing identification regions are quite wide: $(5.8, 27.1)$ for Δ^*_{10}, $(11.3, 33.0)$ for Δ^*_{21}, and $(27.9, 39.9)$ for Δ^*_{20}. In our multiple-treatment application, imposing the restriction that over-reporting is more likely than underreporting does however improve identification power, albeit only for Δ^*_{10} (a 5.6 percentage reduction in width by lowering the upper bound) and Δ^*_{21} (a 14.1 percentage reduction by raising the lower bound).

To gain some further insight we move to a graphical analysis. This is similar to what done for the binary treatment; however to overcome the additional difficulty that we now are working in

[22]This discussion is heuristic in that it takes no account of the uncertainty of the estimated Δ_{LFS}. A more formal indicator would be obtained by considering the proportion of bootstrapped values of Δ_{LFS} that fall within the corresponding bootstrapped upper and lower bounds.

a 5-dimensional space, we fix two dimensions, λ_{22} and λ_{21}. Specifically, we consider two rather extreme but still plausible profiles: one of severe misclassification in these two dimensions ($\lambda_{22} = 0.90, \lambda_{21} = 0.05$) and one of little misclassification ($\lambda_{22} = 0.95, \lambda_{21} = 0.01$).

In contrast to the binary case, one can visually appreciate from Figures 2-3 that the restriction in terms of over/under-reporting significantly increases identifying power. As can be evinced from the graphs, the percentage gain is in fact larger in the case of little as opposed to severe misclassification in terms of λ_{22} and λ_{21}. For Δ^*_{10} and Δ^*_{20}, the gain from the restriction arises from 'cropping' the top, i.e. lowering the upper bound; for Δ^*_{21} , from raising the lower bound instead. Comparing the three ATT's, the percentage reduction in bounds width is largest for Δ^*_{20} for either $\{\lambda_{22}, \lambda_{21}\}$ profile. The identification regions for the three parameters under the restriction are in fact quite similar for the two profiles, and are roughly 10 to 15 percent for for Δ^*_{10}, 16 to 20 percent for Δ^*_{21}, and 29 to 30 percent for Δ^*_{20}. Indeed, under our restriction, Δ^*_{20} is very tightly bound under either profile.

Next, we allow for heterogeneity in the two misclassification probabilities $\lambda_{00}(x)$ and $\lambda_{11}(x)$ and investigate the identifying power of considering bounds that, whilst keeping the values of $\{\lambda_{22}, \lambda_{21}\}$ constant to those of either of the two profiles and meeting the over/under-reporting restriction, further satisfy $\lambda_{00}(x) + \lambda_{11}(x) \geq k$, for increasing values of k. The results are reported in Table 3, together with the 95 percent confidence intervals for the identification region, which as mentioned leads to conservative inference for the ATT of interest.

As expected, for given $\{\lambda_{22}, \lambda_{21}\}$ profile, the bounds monotonically narrow as $\lambda_{00}(x)+\lambda_{11}(x)$ approaches 2. For Δ^*_{10}, bounds become more informative because of a lowering of the upper bound, for Δ^*_{21} because of a raising of the lower bound. For Δ^*_{20}, however, the bounds remain unchanged for $k = 1.6$ and $k = 1.7$, and for higher levels of k both the lower and upper bounds move closer. Interestingly, the result that bounds either stay the same or become narrower as we get closer to exact reporting in terms of λ_{00} and λ_{11} no longer applies in terms of λ_{22} and λ_{21}. Specifically, bounds do not always become more informative when moving from severe to little misclassification in terms of λ_{22} and λ_{21} (see $k = 1.6$ for Δ^*_{10} and $k = 1.6, 1.7, 1.8$ for Δ^*_{20}).

As to how the 'raw' incremental effects Δ_{ij}'s relate to the bounds, in contrast to the binary case no clear pattern applies. In fact, although the Δ_{ij}'s are mostly downward biased, in terms of point estimate, at times the estimated lower bound proves sharper than the raw effect. Furthermore, the raw effects could at times be upward biased, even though in our

application this could happen only for Δ^*_{21} and, interestingly, only for a situation of more severe misclassification in all dimensions (λ_{00}, λ_{11}, λ_{22} and λ_{21}).

Finally, we consider the relationship between the bounds for the true incremental ATT's and the *naive* estimates based on the raw data and LFS-style controls. For the two sets of values of $\{\lambda_{22}, \lambda_{21}\}$ we have considered, ignoring misreporting and ability biases (Δ_{LFS}) yields an upward biased estimate, in the sense that the Δ_{LFS}'s always lie well outside of the bounds. Whatever the sum of λ_{00} and λ_{11}, we thus find no evidence that there is a chance for the two biases to cancel out. For Δ^*_{20}, this conclusion keeps holding in terms of the (conservative) CI for the identification region - estimating returns to advanced qualifications compared to none ignoring misclassification and selection yields severely upward biased estimates. For Δ^*_{10}, and especially Δ^*_{21}, this conclusion is more likely to hold, as one would expect, in the case of less severe misclassification ($k = 1.8, 1.9$, and $\lambda_{00} = 0.95$,$\lambda_{21} = 0.01$).

In contrast to the 'traditional' binary treatment case, when account is taken of potential misclassification across multiple treatments, the results in this section show that the patterns that emerge can be exceptionally varied. However, the conclusion that in our application we cannot expect measurement error to cancel out all ability bias keeps holding in our multiple-treatment extension. More generally, our results show that under relatively mild restrictions we can obtain strong conclusions regarding our question of interest, although more assumptions are needed to obtain statistical significance.

9 Conclusions

In this paper we have looked at the bias introduced by misclassification of the treatment indicator on the average treatment effect on the treated (ATT) under the assumption of selection on observables. Matching-type estimators of the treatment effect computed from misclassified data are in general biased for the true ATT. The true ATT can be over- or under-estimated depending on the amount of misclassification. However, if misclassification does not depend on variables entering the selection on observables assumption, or only depends on them via the propensity score, the attenuation bias result is still most likely to hold.

We have extended the framework to multiple treatments.

We have provided results to bound the returns to a number of important educational qualifications in the UK semi-parametrically, and by using the unique nature of our data we have

assessed the plausibility for the two biases from measurement error and from omitted variables cancel out.

In future work we plan to investigate the extent to which 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 selection-on-observables assumption.

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 Sianesi, B. (2006), *Misreported Schooling, Multiple Measures and Returns to Educational Qualifications*, 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. (1996), *The Effect of Unions on the Structure of Wages: A Longitudinal Analysis*, Econometrica, Vol.64, No.4, pp.957-979.

[14] 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.

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

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

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

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

[19] 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.

[20] 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.

[21] Dehejia, R.H and Wahba, S. (1999), *Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programmes*, Journal of the American Statistical Association 94, 1053-1062.

[22] 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.

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

[24] 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.

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

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

[27] 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.

[28] Horowitz, J.L. and Manski, C.F. (2000), *Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data*, Journal of the American Statistical Association, Vol. 95, No. 449, pp. 77-84.

[29] Imbens, G.W. (2000), *The Role of the Propensity Score in Estimating Dose-Response Functions*, Biometrika, 87, 3, 706-710.

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

[31] Imbens, G.W. and Manski, C.F. (2004), Confidence Intervals for Partially Identified Parameters, Econometrica, Vol. 72, No. 6, pp. 1845-1857.

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

[33] 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.

[34] Lewbel, A. (2005), *Estimation of Average Treatment Effects With Misclassification*, unpublished manusctipt, Department of Economics, Boston College.

[35] Mahajan, A. (2006), *Identification and Estimation of Regression Models with Misclassification*, forthcoming Econometrica.

[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.

Table 1: Educational sample split (N=3,642)

	Multiple treatment	**HE vs Less**	**Any Academic vs None**
None	895 (25%)		
O/eq.	941 (26%)		
A/eq.+	1,806 (50%)		
No Acad			1,243 (34%)
Any Acad			2,399 (66%)
Below HE		2,472 (68%)	
HE		1,170 (32%)	

See main text for the definition of the educational categories of interest.

Table 2: Bounds on returns from Any Academic Qualification and from Higher Education

	Any Acad Qual				**Higher Educ**			
Δ^*	Estimate		95% CI		Estimate		95% CI	
	lower	upper	lower	upper	lower	upper	lower	upper
$\lambda_0(x) + \lambda_1(x) \geq 1.6$	0.2336	0.3975	0.1421	0.4890	0.2310	0.4032	0.1926	0.4416
$\lambda_0(x) + \lambda_1(x) \geq 1.7$	0.2336	0.3340	0.1517	0.4159	0.2310	0.3378	0.1923	0.3765
$\lambda_0(x) + \lambda_1(x) \geq 1.8$	0.2336	0.2947	0.1595	0.3688	0.2310	0.2956	0.1947	0.3319
$\lambda_0(x) + \lambda_1(x) \geq 1.9$	0.2336	0.2607	0.1646	0.3297	0.2310	0.2597	0.1941	0.2966
Raw data:								
Full controls:	$\Delta = 0.2336$				$\Delta = 0.2310$			
LFS controls:	$\Delta_{LFS} = 0.3286$				$\Delta_{LFS} = 0.3588$			

Bounds are derived from Proposition 2 by allowing the misclassification probabilities to depend on x through the propensity score and by imposing that over-reporting is more likely than under-reporting. Confidence intervals covering the identification region with 95 percent probability have been derived from 500 bootstrap replications following Horowitz and Manski (2000).

Table 3: Bounds on returns to incremental levels of attainment

Δ^*_{10}	$\lambda_{22}(x)=0.95$ and $\lambda_{21}(x)=0.01$ Estimate		95% CI		$\lambda_{22}(x)=0.90$ and $\lambda_{21}(x)=0.05$ Estimate		95% CI	
	lower	upper	lower	upper	lower	upper	lower	upper
$\lambda_0(x)+\lambda_1(x)\geq 1.6$	0.1057	0.1591	0.0331	0.2317	0.0981	0.1566	0.0249	0.2298
$\lambda_0(x)+\lambda_1(x)\geq 1.7$	0.1057	0.1504	0.0352	0.2209	0.0981	0.1392	0.0261	0.2112
$\lambda_0(x)+\lambda_1(x)\geq 1.8$	0.1057	0.1314	0.0358	0.2013	0.0981	0.1200	0.0285	0.1896
$\lambda_0(x)+\lambda_1(x)\geq 1.9$	0.1057	0.1156	0.0361	0.1852	0.0981	0.1046	0.0285	0.1742

Raw data:
Full controls: $\Delta_{10}=0.1060$, LFS controls: $\Delta_{10,LFS}=0.1922$

Δ^*_{21}	$\lambda_{22}(x)=0.95$ and $\lambda_{21}(x)=0.01$ Estimate		95% CI		$\lambda_{22}(x)=0.90$ and $\lambda_{21}(x)=0.05$ Estimate		95% CI	
	lower	upper	lower	upper	lower	upper	lower	upper
$\lambda_0(x)+\lambda_1(x)\geq 1.6$	0.1519	0.1944	0.0886	0.2577	0.1766	0.2159	0.1100	0.2825
$\lambda_0(x)+\lambda_1(x)\geq 1.7$	0.1550	0.1944	0.0926	0.2568	0.1820	0.2159	0.1166	0.2813
$\lambda_0(x)+\lambda_1(x)\geq 1.8$	0.1725	0.1944	0.1176	0.2493	0.1981	0.2159	0.1399	0.2741
$\lambda_0(x)+\lambda_1(x)\geq 1.9$	0.1860	0.1944	0.1341	0.2463	0.2107	0.2159	0.1549	0.2717

Raw data:
Full controls: $\Delta_{21}=0.1843$, LFS controls: $\Delta_{21,LFS}=0.2432$

Δ^*_{20}	$\lambda_{22}(x)=0.95$ and $\lambda_{21}(x)=0.01$ Estimate		95% CI		$\lambda_{22}(x)=0.90$ and $\lambda_{21}(x)=0.05$ Estimate		95% CI	
	lower	upper	lower	upper	lower	upper	lower	upper
$\lambda_0(x)+\lambda_1(x)\geq 1.6$	0.2874	0.2944	0.1899	0.3919	0.2975	0.3104	0.2015	0.4064
$\lambda_0(x)+\lambda_1(x)\geq 1.7$	0.2874	0.2944	0.1905	0.3913	0.2975	0.3104	0.2015	0.4064
$\lambda_0(x)+\lambda_1(x)\geq 1.8$	0.2883	0.2944	0.1908	0.3919	0.2991	0.3076	0.2016	0.4051
$\lambda_0(x)+\lambda_1(x)\geq 1.9$	0.2889	0.2917	0.1908	0.3898	0.3004	0.3029	0.2017	0.4016

Raw data:
Full controls: $\Delta_{20}=0.2825$, LFS controls: $\Delta_{20,LFS}=0.4339$

See main text for the definition of Δ^*_{10}, Δ^*_{21} and Δ^*_{20}. Bounds are derived from Proposition 4 by allowing the misclassification probabilities to depend on x through the propensity scores and by imposing that over-reporting is more likely than under-reporting. Confidence intervals covering the identification region with 95 percent probability have been derived from 500 bootstrap replications following Horowitz and Manski (2000).

Proof of Proposition 1

By using (5) we have that

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

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

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

it also follows that

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

Since

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

we can use (6) to write

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

where

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

and

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

Proof of Proposition 3

Conditioning on $X = x$ is left implicit throughout. Start from

$$\begin{aligned} \begin{pmatrix} E[Y|D=0] \\ E[Y|D=1] \\ E[Y|D=2] \end{pmatrix} &= \begin{pmatrix} \lambda_{00} & 1-\lambda_{00} & 0 \\ 1-\lambda_{11}-\lambda_{21} & \lambda_{11} & \lambda_{21} \\ 0 & 1-\lambda_{22} & \lambda_{22} \end{pmatrix} \begin{pmatrix} E[Y|D^*=0] \\ E[Y|D^*=1] \\ E[Y|D^*=2] \end{pmatrix}, \\ &= \Pi \begin{pmatrix} E[Y|D^*=0] \\ E[Y|D^*=1] \\ E[Y|D^*=2] \end{pmatrix}, \end{aligned}$$

where we need

$$det(\Pi) = \lambda_{00}(\lambda_{22} - \lambda_{21}) - \lambda_{22}(1 - \lambda_{11} - \lambda_{21}) \neq 0.$$

$$
\begin{aligned}
\Delta_{10} &= E[Y|D^*=0](1-\lambda_{11}-\lambda_{21}-\lambda_{00}) \\
&+ E[Y|D^*=1](\lambda_{11}+\lambda_{00}-1) \\
&+ E[Y|D^*=2](\lambda_{21}), \\
&= E[Y|D^*=0](1-\lambda_{11}-\lambda_{21}-\lambda_{00}) \\
&+ E[Y|D^*=0](\lambda_{11}+\lambda_{00}-1) \\
&+ \Delta^*_{10}(\lambda_{11}+\lambda_{00}-1) \\
&+ E[Y|D^*=2](\lambda_{21}), \\
&= \Delta^*_{10}(\lambda_{11}+\lambda_{00}-1)+\Delta^*_{20}\lambda_{21}, \\
&= \Delta^*_{10}(\lambda_{11}+\lambda_{00}+\lambda_{21}-1)+\Delta^*_{21}\lambda_{21},
\end{aligned}
$$

where the last equality follows (under CIA) from $\Delta^*_{20} = \Delta^*_{21} + \Delta^*_{10}$. By assuming $\Delta^*_{10} \geq 0$ and $\Delta^*_{21} \geq 0$, restrictions on the λ's can be defined to avoid sign reversal. For example, a sufficient condition for this is $\lambda_{11} + \lambda_{00} \geq 1$.

$$
\begin{aligned}
\Delta_{21} &= E[Y|D^*=0](-1+\lambda_{11}+\lambda_{21}) \\
&+ E[Y|D^*=1](1-\lambda_{22}-\lambda_{11}) \\
&+ E[Y|D^*=2](\lambda_{22}-\lambda_{21}), \\
&= E[Y|D^*=0](-1+\lambda_{11}+\lambda_{21}) \\
&+ E[Y|D^*=1](1-\lambda_{22}-\lambda_{11}) \\
&+ E[Y|D^*=1](\lambda_{22}-\lambda_{21}) \\
&+ \Delta^*_{21}(\lambda_{22}-\lambda_{21}), \\
&= \Delta^*_{21}(\lambda_{22}-\lambda_{21})+\Delta^*_{10}(1-\lambda_{11}-\lambda_{21}).
\end{aligned}
$$

By assuming $\Delta^*_{21} \geq 0$ and $\Delta^*_{10} \geq 0$, a sufficient condition for not having sign reversal is $\lambda_{22} \geq \lambda_{21}$. It therefore follows that

$$
\begin{aligned}
\Delta_{20} &= \Delta_{21} + \Delta_{10}, \\
&= \Delta^*_{10}(\lambda_{00}-\lambda_{21})+\Delta^*_{20}\lambda_{21}+\Delta^*_{21}(\lambda_{22}-\lambda_{21}), \\
&= \Delta^*_{10}\lambda_{00}+\Delta^*_{21}\lambda_{22}.
\end{aligned}
$$

The previous expressions can be written with matrix notation by

$$\begin{pmatrix} \Delta_{10} \\ \Delta_{21} \end{pmatrix} = \begin{pmatrix} \lambda_{00} - (1 - \lambda_{11} - \lambda_{21}) & \lambda_{21} \\ 1 - \lambda_{11} - \lambda_{21} & \lambda_{22} - \lambda_{21} \end{pmatrix} \begin{pmatrix} \Delta^*_{10} \\ \Delta^*_{21} \end{pmatrix},$$

so that

$$\begin{pmatrix} \lambda_{00} - (1 - \lambda_{11} - \lambda_{21}) & \lambda_{21} \\ 1 - \lambda_{11} - \lambda_{21} & \lambda_{22} - \lambda_{21} \end{pmatrix}^{-1} \begin{pmatrix} \Delta_{10} \\ \Delta_{21} \end{pmatrix} = \begin{pmatrix} \Delta^*_{10} \\ \Delta^*_{21} \end{pmatrix}.$$

By solving the last expression we get

$$\begin{aligned} \Delta^*_{10} &= \frac{\lambda_{22} - \lambda_{21}}{det}\Delta_{10} - \frac{\lambda_{21}}{det}\Delta_{21}, \\ \Delta^*_{21} &= \frac{\lambda_{11} + \lambda_{21} - 1}{det}\Delta_{10} + \frac{\lambda_{00} + \lambda_{11} + \lambda_{21} - 1}{det}\Delta_{21}. \blacksquare \end{aligned}$$

Proof of Proposition 4

Start from the parameters of interest

$$\begin{aligned} \Delta^*_{10} &= \int \Delta^*_{10}[x] f[x|D^* = 1] dx, \\ \Delta^*_{21} &= \int \Delta^*_{21}[x] f[x|D^* = 2] dx, \\ \Delta^*_{20} &= \int \Delta^*_{20}[x] f[x|D^* = 2] dx = \Delta^*_{21} + \int \Delta^*_{10}[x] f[x|D^* = 2] dx, \end{aligned}$$

where we have

$$\begin{aligned} f[x|D^* = 1] &= \frac{P[D^* = 1|x] f[x]}{P[D^* = 1]}, \\ f[x|D^* = 2] &= \frac{P[D^* = 2|x] f[x]}{P[D^* = 2]}. \end{aligned}$$

Computation of $f[x|D^* = 1]$

By using the adding up condition $P[D = 0|x] = 1 - P[D = 1|x] - P[D = 2|x]$ we can write

$$\begin{aligned} P[D^* = 1|x] &= (1 - \lambda_{00}) P[D = 0|x] + \lambda_{11} P[D = 1|x] + (1 - \lambda_{22}) P[D = 2|x], \\ &= (1 - \lambda_{00}) + (\lambda_{00} + \lambda_{11} - 1) P[D = 1|x] + (\lambda_{00} - \lambda_{22}) P[D = 2|x], \end{aligned}$$

so that

$$\begin{aligned} f[x|D^* = 1] &= (1 - \lambda_{00}) \frac{f[x]}{P[D^* = 1]} + (\lambda_{00} + \lambda_{11} - 1) \frac{P[D = 1]}{P[D^* = 1]} f[x|D = 1] \\ &+ (\lambda_{00} - \lambda_{22}) \frac{P[D = 2]}{P[D^* = 1]} f[x|D = 2]. \end{aligned}$$

The following two expressions follow. First, we have

$$\begin{aligned} f[x|D^*=1] &= (1-\lambda_{00})\frac{P[D=1]}{P[D^*=1]}\frac{1}{P[D=1|x]}f[x|D=1] \\ &+ (\lambda_{00}+\lambda_{11}-1)\frac{P[D=1]}{P[D^*=1]}f[x|D=1] \\ &+ (\lambda_{00}-\lambda_{22})\frac{P[D=1]}{P[D^*=1]}\frac{P[D=2|x]}{P[D=1|x]}f[x|D=1], \end{aligned}$$

so that

$$\begin{aligned} f[x|D^*=1] &= \eta_1(x)f[x|D=1], \\ \eta_1(x) &= \frac{P[D=1]}{P[D^*=1]}[\frac{1-\lambda_{00}}{P[D=1|x]}+(\lambda_{00}+\lambda_{11}-1)+\frac{(\lambda_{00}-\lambda_{22})P[D=2|x]}{P[D=1|x]}]. \end{aligned}$$

Second, we have

$$\begin{aligned} f[x|D^*=1] &= (1-\lambda_{00})\frac{P[D=2]}{P[D^*=1]}\frac{1}{P[D=2|x]}f[x|D=2] \\ &+ (\lambda_{00}+\lambda_{11}-1)\frac{P[D=2]}{P[D^*=1]}\frac{P[D=1|x]}{P[D=2|x]}f[x|D=2] \\ &+ (\lambda_{00}-\lambda_{22})\frac{P[D=2]}{P[D^*=1]}f[x|D=2], \end{aligned}$$

so that

$$\begin{aligned} f[x|D^*=1] &= \eta_2(x)f[x|D=2], \\ \eta_2(x) &= \frac{P[D=2]}{P[D^*=1]}[\frac{1-\lambda_{00}}{P[D=2|x]}+\frac{(\lambda_{00}+\lambda_{11}-1)P[D=1|x]}{P[D=2|x]}+(\lambda_{00}-\lambda_{22})]. \end{aligned}$$

Computation of $f[x|D^*=2]$

By using

$$P[D^*=2|x]=\lambda_{21}P[D=1|x]+\lambda_{22}P[D=2|x],$$

we have

$$\begin{aligned} f[x|D^*=2] &= \lambda_{21}\frac{P[D=1]}{P[D^*=2]}f[x|D=1]+\lambda_{22}\frac{P[D=2]}{P[D^*=2]}f[x|D=2], \\ &= \lambda_{21}\frac{P[D=1]}{P[D^*=2]}f[x|D=1]+\lambda_{22}\frac{P[D=1]}{P[D^*=2]}\frac{P[D=2|x]}{P[D=1|x]}f[x|D=1], \\ &= \lambda_{21}\frac{P[D=2]}{P[D^*=2]}\frac{P[D=1|x]}{P[D=2|x]}f[x|D=2]+\lambda_{22}\frac{P[D=2]}{P[D^*=2]}f[x|D=2], \end{aligned}$$

so that

$$\begin{aligned} f[x|D^* = 2] &= \eta_3(x) f[x|D = 1], \\ \eta_3(x) &= \frac{P[D=1]}{P[D^*=2]}[\lambda_{21} + \lambda_{22}\frac{P[D=2|x]}{P[D=1|x]}], \end{aligned}$$

or

$$\begin{aligned} f[x|D^* = 2] &= \eta_4(x) f[x|D = 2], \\ \eta_4(x) &= \frac{P[D=2]}{P[D^*=2]}[\lambda_{21}\frac{P[D=1|x]}{P[D=2|x]} + \lambda_{22}]. \end{aligned}$$

Computation of Δ^*_{10}

$$\begin{aligned} \Delta^*_{10} &= \int \Delta^*_{10}[x] f[x|D^* = 1] dx, \\ &= \int \frac{\lambda_{22} - \lambda_{21}}{det} \Delta_{10}[x] f[x|D^* = 1] dx - \int \frac{\lambda_{21}}{det} \Delta_{21}[x] f[x|D^* = 1] dx, \\ &= \int \omega_{1,10}(x) \Delta_{10}[x] f[x|D = 1] dx + \int \omega_{1,10}(x) \Delta_{21}[x] f[x|D = 2] dx, \\ \omega_{1,10}(x) &= \frac{\lambda_{22} - \lambda_{21}}{det} \eta_1(x), \\ &= \frac{P[D=1](\lambda_{22} - \lambda_{21})(\lambda_{11} + \lambda_{00} - 1 + \frac{P[D=2|x](\lambda_{00} - \lambda_{22}) + 1 - \lambda_{00}}{P[D=1|x]})}{det(1 - \lambda_{00} + (\lambda_{11} + \lambda_{00} - 1)P[D=1] + (\lambda_{00} - \lambda_{22})P[D=2])}, \end{aligned}$$

the last expression following from having misclassification independent of X. Moreover we have

$$\begin{aligned} \omega_{2,10}(x) &= -\frac{\lambda_{21}}{det} \eta_2(x), \\ &= -\frac{P[D=2]\lambda_{21}(\lambda_{00} - \lambda_{22} + \frac{P[D=1|x](\lambda_{00} + \lambda_{11} - 1) + 1 - \lambda_{00}}{P[D=2|x]})}{det(1 - \lambda_{00} + (\lambda_{11} + \lambda_{00} - 1)P[D=1] + (\lambda_{00} - \lambda_{22})P[D=2])}, \end{aligned}$$

where the last expression follows under misclassification constant with respect to X. In what follows, we report similar expressions for Δ^*_{21} and Δ^*_{20}, where weights are derived with and without imposing constant misclassification probabilities.

Computation of Δ_{21}^*

$$
\begin{aligned}
\Delta_{21}^* &= \int \Delta_{21}^*[x] f[x|D^* = 2] dx, \\
&= \int \frac{\lambda_{11} + \lambda_{21} - 1}{det} \Delta_{10}[x] f[x|D^* = 2] dx + \int \frac{\lambda_{00} + \lambda_{11} + \lambda_{21} - 1}{det} \Delta_{21}[x] f[x|D^* = 2] dx, \\
&= \int \omega_{1,21}(x) \Delta_{10}[x] f[x|D = 1] dx + \int \omega_{2,21}(x) \Delta_{21}[x] f[x|D = 2] dx, \\
\omega_{1,21}(x) &= \frac{\lambda_{11} + \lambda_{21} - 1}{det} \eta_3(x), \\
&= \frac{P[D = 1](\lambda_{11} + \lambda_{21} - 1)(\lambda_{21} + \lambda_{22} \frac{P[D=2|x]}{P[D=1|x]})}{det(\lambda_{21} P[D = 1] + \lambda_{22} P[D = 2])}, \\
\omega_{2,21}(x) &= \frac{\lambda_{00} + \lambda_{11} + \lambda_{21} - 1}{det} \eta_4(x), \\
&= \frac{P[D = 2](\lambda_{00} + \lambda_{11} + \lambda_{21} - 1)(\lambda_{22} + \lambda_{21} \frac{P[D=1|x]}{P[D=2|x]})}{det(\lambda_{21} P[D = 1] + \lambda_{22} P[D = 2])}.
\end{aligned}
$$

Computation of Δ_{20}^*

$$
\begin{aligned}
\Delta_{20}^* &= \Delta_{21}^* + \int \Delta_{10}^*[x] f[x|D^* = 2] dx, \\
&= \Delta_{21}^* + \int \frac{\lambda_{22} - \lambda_{21}}{det} \Delta_{10}[x] f[x|D^* = 2] dx - \int \frac{\lambda_{21}}{det} \Delta_{21}[x] f[x|D^* = 2] dx, \\
&= \int \omega_{1,20}(x) \Delta_{10}[x] f[x|D = 1] dx + \int \omega_{2,20}(x) \Delta_{21}[x] f[x|D = 2] dx, \\
\omega_{1,20}(x) &= \frac{\lambda_{11} + \lambda_{22} - 1}{det} \eta_3(x), \\
&= \frac{P[D = 1](\lambda_{11} + \lambda_{22} - 1)(\lambda_{21} + \lambda_{22} \frac{P[D=2|x]}{P[D=1|x]})}{det(\lambda_{21} P[D = 1] + \lambda_{22} P[D = 2])}, \\
\omega_{2,20}(x) &= \frac{\lambda_{00} + \lambda_{11} - 1}{det} \eta_4(x), \\
&= \frac{P[D = 2](\lambda_{00} + \lambda_{11} - 1)(\lambda_{22} + \lambda_{21} \frac{P[D=1|x]}{P[D=2|x]})}{det(\lambda_{21} P[D = 1] + \lambda_{22} P[D = 2])}. \blacksquare
\end{aligned}
$$

Figure 1: Identification region for the return to any academic qualification, assuming constant misclassification probabilities and, in the right-hand-side panel, that overreporting is more likely than underreporting

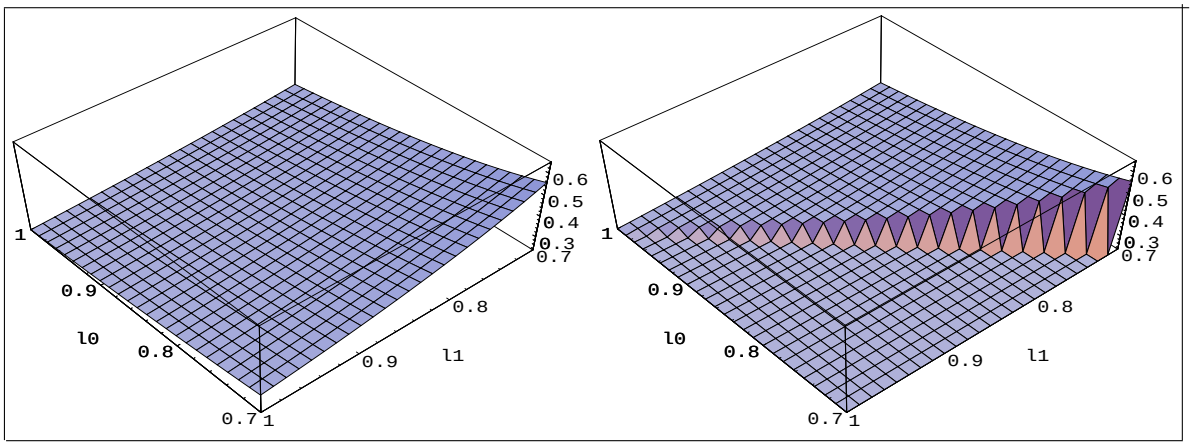


Figure 2: Identification region for returns to incremental levels of attainment by imposing $\lambda_{22}(x) = 0.90$ and $\lambda_{21}(x) = 0.05$

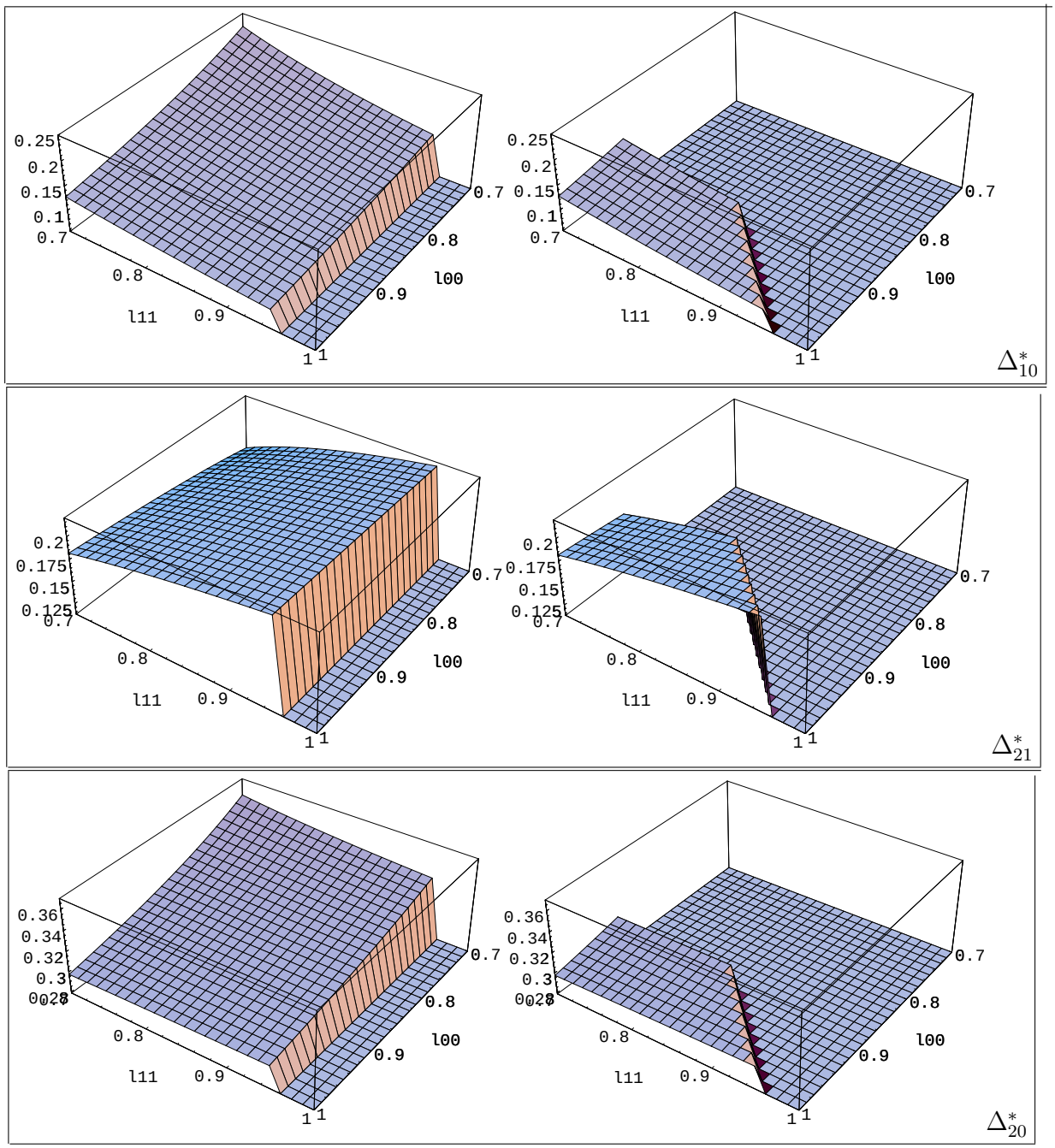


Figure 3: Identification region for returns to incremental levels of attainment by imposing $\lambda_{22}(x) = 0.95$ and $\lambda_{21}(x) = 0.01$

