Misreported Schooling, Multiple Measures and Returns to Educational Qualifications*

Erich Battistin
University of Padova and Institute for Fiscal Studies

Barbara Sianesi
Institute for Fiscal Studies

* **Preliminary** *

Abstract

In this paper we provide a number of contributions of policy, practical and methodological interest to the study of the returns to educational qualifications in the presence of misreporting. First, we derive estimates of the returns to different qualifications for the UK that allow for the possibility of misreported attainment using data from the British National Child Development Survey. To date, any major empirical evidence on the importance of this issue is restricted to the US, where it was shown that errors might indeed play a non-negligible role. Second, we aim to provide the academic and policy community with estimates of the accuracy of commonly used types of data on educational attainment: administrative files, self-reported information close to the date of completion of the qualification and recall information ten years after completion. Third, by using the unique nature of our data, we assess how the biases from measurement error and from omitted ability and family background variables interact in the estimation of returns. We intend to produce simple calibration rules to allow policymakers to use nationally representative data, such as the Labour Force Survey, that totally rely on self-reported information on qualifications and contain little or no information on individual ability and family background. On the methodological front, we propose a semi-parametric estimation approach based on balancing scores and mixture models, in particular allowing for arbitrarily heterogeneous individual returns. Since our proposed approach is embedded in the general evaluation framework, it is applicable to the estimation of any causal treatment effects in the presence of potentially misreported treatment status. Our approach also deals with cases where the analyst does not want *a priori* to commit on the two educational measures being independent.

JEL Codes: C10, I20, J31.

Keywords: Misclassification, Mixture Models, Returns to Educational Qualifications, Treatment Effects.

*This draft 15th February 2006. This paper benefited from helpful discussions with Enrico Rettore and comments by audiences at Policy Studies Institute (London, September 2005), ADRES Conference on “Econometric Evaluation of Public Policies: Methods and Applications” (Paris, December 2005) and Franco Modigliani Fellowship Workshop (Rome, February 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

This paper considers the identification and estimation of the returns to educational qualifications for the UK allowing for misreporting of educational attainment.

The measurement of the return to education, that is of the individual wage gain from investing in more education, has become probably the most explored and prolific area in labour economics. Policymakers too have shown increasing interest, with estimated returns feeding into debates on national economic performance, educational policies, or the public funding of education. Reliable measures of the impact of education on individual earnings are in fact needed to establish whether it is worthwhile for individuals to invest in more education (and in which type), to compare private and social returns to education, or to assess the relative value that different educational qualifications fetch on the labour market.[1]

As to the measurement of education, a first issue is whether we can summarize it in the single, homogeneous measure of years of schooling. Although particularly convenient, this 'one-factor' model is a priori quite restrictive, in that it assumes that the returns increase linearly with each additional year, irrespective of the level and type of educational qualifications the years refer to. When interested in a wide range of education levels with potentially very different returns, a more adequate framework is the 'multiple-factor' model, in which different educational levels are allowed to have separate effects on earnings.[2]

In the schooling system in the US, grades generally follow years, and it has long been argued that the returns to an additional year are reasonably homogeneous (see for example Card, 1999). In the UK and other European countries, however, there are alternative nationally-based routes leading to quite different educational qualifications, and the importance of distinguishing between different types of qualifications is widely accepted. Blundell, Dearden and Sianesi (2005) highlight the potential shortcoming of the 'one-factor' model when applied to the UK's educational system, in which individuals with the same number of years of schooling have quite different educational outcomes. Not only would this obfuscate the interpretation of the return to one additional year, but imposing equality of yearly returns across educational stages was found to be overly restrictive.

A second important issue as to the measurement of education - and the one object of this paper - is the possibility of errors in recorded education and its consequences on the estimated returns. Misrecorded education could arise from data transcript errors, as well as from misreporting: survey

[1] For an extensive discussion of the policy interest of the individual wage return from education, see Blundell, Dearden and Sianesi (2004).

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

respondents may either over-report their attainment, not know if the schooling they have had counts as a qualification or simply not remember.

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

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

To date, empirical evidence on the importance of these issues is restricted to the US, where it was in fact shown that measurement error might play a non-negligible role (see the results in Kane, Rouse and Staiger, 1999, Black, Sanders and Taylor, 2003, and Lewbel, 2005). For the UK there are no estimates of the returns to educational qualifications that adequately correct for measurement error.[3] 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, 1999, Dearden *et al.*, 2002, and McIntosh, 2004).

Two approaches have been developed to overcome the bias induced by misreported educational qualifications. A first possibility is to derive bounds on the returns by making a priori assumptions on the misclassification probabilities (see e.g. Manski, 1990, and Molinari, 2004). This approach only allows partial identification of returns. In our companion paper (Battistin and Sianesi, 2006) we suggest bounds that can be derived exploiting the observed propensity score in a non-parametric way, in particular allowing for arbitrarily heterogeneous individual returns. The corresponding sensitivity

[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.

analysis is easy to implement and can provide an often quite informative robustness check.

The alternative approach is more demanding in terms of data requirements but, if feasible, it allows point identification of the returns. An additional appealing feature is that it provides direct estimates of the measurement error in the educational measures, which may often be of independent interest. What is needed is (at least) two categorical reports of educational qualifications for the same individuals, both potentially affected by reporting error but independent of each other (for the proof of non-parametric identification, see Lewbel, 2005, and Mahajan, 2006). Repeated measurements on educational qualifications are typically obtained by combining complementary datasets, for example exploiting administrative records and information self-reported by individuals. In this paper, we build on this approach.

We provide a number of new contributions of considerable policy and practical relevance, as well as of methodological interest.

On the methodological front, we propose a semi-parametric estimation approach based on balancing scores and mixture models. As far as we are aware, we are the first ones to cast this problem in terms of a mixture model, which combined with the propensity score allows us to proceed in a semi-parametric way, allowing in particular for arbitrarily heterogeneous individual returns. A further advantage of our approach is that it does not require the two educational measures to be independent. Independence can be ruled out, for example, if the same information is repeatedly asked to the same individuals over time, as done by longitudinal studies for educational qualifications and other information. Other less clear-cut cases are those where selection/assortative matching may lead to questioning the independence assumption, e.g. when the same question is asked to a child and its parents, to twins, to husband and wife, to employee and employer, or to a child and its school. We thus offer an approach that deals with cases where the analyst does not want *a priori* to commit on the two measures being independent. This feature of the approach we propose will be particularly useful in our application, as for certain important educational categories such as higher education we can only recover repeated measurements self-reported by the same individuals over time.

Note that throughout we explore a unified general framework for the study of the impact of misreported treatment status on the estimation of causal treatment effects (Molinari, 2004, Lewbel, 2005, and Mahajan, 2006, are the only examples we are aware of). Our proposal of a new estimation method to address measurement error in educational qualifications is thus of far wider interest, since the same issues arise in any application looking at the effect of a potentially misrecorded binary or categorical variable, such as eligibility to policy schemes, participation in (multiple) government programmes or work-related training.

Second, we provide reliable estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment. In order to focus fully on the potential biases arising from measurement error, we use the uniquely rich data from the British National Child Development Survey (NCDS) to avoid issues related to omitted variables bias. In particular, in this work we only consider evaluation methods based on the selection on observables (or conditional independence) assumption, and rely on Blundell, Dearden and Sianesi (2005) who could not find any strong evidence of remaining selection bias given the information available in that data. The NCDS data is crucial also in that it provides us with a number of repeated measurements of individual educational attainment.

Third, we identify the extent of measurement error in three different types of data sources on educational qualifications: administrative school files, self-reported information very close to the dates of completion of the qualifications and self-reported recall information ten years later. We thus provide the academic and policy community with estimates of the relative reliability of commonly used types of data. Results in this paper together with those in Battistin and Sianesi (2006) represent a new piece of evidence for the UK, which will allow one to check the robustness of current estimates of returns to the presence of misreported qualifications. Knowing the extent of misreporting also has obvious implications for the interpretation of other studies that use educational attainment as an outcome variable or for descriptive purposes.

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

The remainder of the paper is organized as follows. In Section 2 we allow for the possibility of misclassification in the treatment status in the general evaluation framework. The identification problem is discussed in Section 3. The estimation strategy that we exploit to correct for the bias introduced by misclassification is presented in Section 4. Section 5 discusses how information in the NCDS will allow us to implement this strategy under fairly weak assumptions on the nature of the data collected. Estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment are in Section 6, while Section 7 concludes.

2 General formulation of the problem

In the potential-outcomes framework, interest lies in the causal impact of a given 'treatment' D^* on an outcome of interest Y.[4] To fix ideas and with our application in mind, in the following let the treatment be the qualification of interest and let the outcome be individual (log) wages. Let Y_1 and Y_0 denote the *potential* wages from having and not having the qualification of interest, respectively.[5] Let D^* be a binary indicator for the qualification of interest, which we will later allow to be potentially observed with error amongst individuals. The individual causal effect of (or return to) achieving the qualification is defined as the difference between the two potential outcomes, $\beta \equiv Y_1 - Y_0$. The observed individual wage can then be written as $Y = Y_0 + D^*\beta$. We are interested in recovering the average return for those individuals who have chosen to undertake the qualification of interest, that is the *average effect of treatment on the treated* (ATT):[6]

$$\Delta^* \equiv E_{Y_1|D^*}[Y_1|1] - E_{Y_0|D^*}[Y_0|1].$$

In the absence of misreporting of D^*, identification of the unobserved counterfactual term $E_{Y_0|D^*}[Y_0|1]$ follows straightforwardly from the following two assumptions, which we will maintain throughout:

Assumption 1 ***(Unconfoundness)*** *Conditional on a set of observable variables X, the educational choice D^* is independent of the two potential outcomes:*

$$f_{Y_0,Y_1|D^*,X}[y_0, y_1|d^*, x] = f_{Y_0,Y_1|X}[y_0, y_1|x].$$

For the plausibility of this assumption, which allows one to focus on the impact of measurement error in the reporting of D^*, we draw on Blundell, Dearden and Sianesi (2005), who find the set of regressors X available in our NCDS data to be rich enough to control for the endogeneity of educational choices. To give empirical content to Assumption 1, we also require the following condition on the support of the X variables:

Assumption 2 ***(Common Support)*** *Individuals with and without the qualification of interest can be found at all values of X, that is*

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

[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).

[5]For this representation to be meaningful, the stable unit-treatment value assumption needs to be satisfied (Rubin, 1980), requiring that an individual's potential wages and the chosen qualification are independent from the qualification choices of other individuals in the population.

[6]In the remainder of this paper, $f_{Y|X}[y|x]$ and $E_{Y|X}[Y|x]$ will denote the conditional distribution and the conditional mean of Y given $X = x$, respectively. Also, we will use upper-case letters for random variables and lower-case letters for their realisations.

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

Under these two assumptions, but without invoking any functional form or constant-effect assumption, one can perform any type of semi-non-parametric estimation of the conditional expectation function in the non-participation group, $E_{Y_0|D^*,X}[Y_0|0,x]$, and then average it over the distribution of X in the participants' group (within the common support) to get the counterfactual term of interest (see Heckman, LaLonde and Smith, 1999, and Imbens, 2004). Conditions 1-2 together make the *strong ignorability* condition of Rosenbaum and Rubin (1983).

When qualifications are misreported, either because individuals are left to self-report or because of transcript errors, the treatment information recorded in the data may differ from the actual status D^*. From now on, we assume that we observe *two* measurements of educational qualifications for each individual, D_A and D_B. Note that they need not be independent of each other; for example, they can be measures of educational qualifications reported by the same individuals in two different interview waves. Neither of these measures need correspond with D^*, that is either of them can potentially be affected by misclassification. Throughout our discussion we will however assume that the classification error in either measure is *non-differential*, that is conditional on a person's actual qualification and on other covariates, reporting errors are independent of wages (see Battistin and Sianesi, 2006, for a more detailed discussion of the implications of this assumption).

Assumption 3 ***(Non-Differential Misclassification given** X**)*** *Any variables D_A and D_B which proxy D^* do not contain information to predict Y conditional on the true measure D^* and X:*

$$f_{Y|D^*,D_A,D_B,X}[y|d^*,a,b,x] = f_{Y|D^*,X}[y|d^*,x].$$

As shown in Battistin and Sianesi (2006), even under Assumptions 1-3, causal inference drawn from the triples (Y, D_A, X) or (Y, D_B, X) will in general be invalid for the ATT, with the magnitude of this bias depending on the extent of misclassification. The extent of misclassification can be conveniently summarised by the *misclassification probabilities.* In particular, the probabilities of *exact* classification relative to D_A and D_B are defined as:

$$\begin{aligned} \lambda_0^j(x) &\equiv Pr[D^* = 0|D_j = 0, x], \\ \lambda_1^j(x) &\equiv Pr[D^* = 1|D_j = 1, x], \end{aligned}$$

for $j \in \{A, B\}$, and may in general depend on X. The $\lambda_1^j(x)$'s represent the proportion of individuals actually having the qualification of interest amongst those reporting so, separately for the two reported measures D_A and D_B. A similar interpretation applies to the $\lambda_0^j(x)$'s, summarising the extent of exact reporting amongst those reporting not to have the qualification.

Under the assumption of non-differential misclassification (Assumption 3), the distribution of observed wages conditional on X for the 2×2 groups defined by $D_A \times D_B$ can be written as a *mixture* of *two* latent distributions: the distribution of wages in the presence of the qualification, i.e. Y_1, and the distribution of wages in the absence of the qualification, i.e. Y_0:

$$f_{Y|D_A,D_B,X}[y|a,b,x] = [1 - p_{a,b}(x)]f_{Y_0|X}[y|x] + p_{a,b}(x)f_{Y_1|X}[y|x], \tag{1}$$

where the equality follows from Assumption 1 and the probability

$$p_{a,b}(x) \equiv E_{D^*|D_A,D_B,X}[D^*|a,b,x]$$

denotes the true proportion of individuals with the qualification of interest amongst those with $D_A = a$ and $D_B = b$ within cells defined by X.

Knowledge of the $p_{a,b}(x)$'s suffices to identify the extent of misreporting in either measure of educational qualifications, since

$$\begin{aligned}
\lambda_0^A(x) &= 1 - \sum_b p_{0,b}(x)P[D_B = b|D_A = 0, x], \\
\lambda_1^A(x) &= \sum_b p_{1,b}(x)P[D_B = b|D_A = 1, x],
\end{aligned}$$

for D_A, and

$$\begin{aligned}
\lambda_0^B(x) &= 1 - \sum_a p_{a,0}(x)P[D_A = a|D_B = 0, x], \\
\lambda_1^B(x) &= \sum_a p_{a,1}(x)P[D_A = a|D_B = 1, x],
\end{aligned}$$

for D_B. Knowing the probabilities of exact classification in turn allows identification of

$$\Delta^*(x) \equiv E_{Y_1|D^*,X}[Y_1|1,x] - E_{Y_0|D^*,X}[Y_0|1,x]$$

(see Battistin and Sianesi, 2006), corresponding – under Assumption 1 – to the causal effect of having the qualification of interest for individuals with $X = x$. Alternatively, note that $\Delta^*(x)$ can be obtained from the difference of the first moments of the mixture components in (1). Moreover, as the ATT is obtained by integrating $\Delta^*(x)$ with respect to:

$$f_{X|D^*}[x|1] = \frac{Pr[D^* = 1|x]f_X[x]}{\int Pr[D^* = 1|x]f_X[x]dx},$$

it follows from

$$\begin{aligned}
Pr[D^* = 1|x] &= [1 - \lambda_0^A(x)] + Pr[D_A = 1|x][\lambda_0^A(x) + \lambda_1^A(x) - 1], \\
&= [1 - \lambda_0^B(x)] + Pr[D_B = 1|x][\lambda_0^B(x) + \lambda_1^B(x) - 1],
\end{aligned}$$

that the ATT is also identified by knowing the $p_{a,b}(x)$'s (see Battistin and Sianesi, 2006, for the exact characterisation of the relationship between the true ATT, the effect estimated using either misrecorded measure and the latter's misclassification probabilities).[7]

When D_A and D_B are *independent* reports, Kane, Rouse and Staiger (1999) and Black, Berger and Scott (2000) have developed a procedure to simultaneously estimate the returns to qualifications and the distribution of reporting error in each educational measure in a parametric way. The identification problem in the case of independent reports has been further investigated in a more general setting by Mahajan (2006), whose results imply that the $p_{a,b}(x)$'s are non-parametrically identified from (Y, D_A, D_B, X). Information on the number of individuals classified differently by D_A and D_B can be combined with information on the difference in their earnings to estimate the distribution of reporting errors (i.e. the misclassification probabilities) in both measures.[8] To the best of our knowledge, for the case of *dependent* reports there are no results on non-parametric identification of the $p_{a,b}(x)$'s.

3 The curse of dimensionality

The main problem that hampers estimation of the ATT is the curse of dimensionality arising from a large number of regressors in X. While non-parametric identification has been extensively dealt with in the literature (see Lewbel, 2005, and Mahajan, 2006), to the best of our knowledge all applications to real data rely upon a fully parametric approach to model to relationship (Y, D_A, D_B, X).

In this section we propose a method to reduce the dimensionality of the problem based on the properties of *balancing scores* (see Theorem 1 by Rosembaum and Rubin, 1983, and Imbens, 2000). Let $\mathcal{S}(X)$ be a balancing score such that the conditional distribution of X within cells defined by $\mathcal{S}(x)$ is independent of (D_A, D_B):

$$f_{X|D_A,D_B,\mathcal{S}(X)}[x|a,b,s] = f_{X|\mathcal{S}(X)}[x|s]. \tag{2}$$

In what follows, we discuss under which conditions the mixture representation given X in (1) implies a mixture representation given $\mathcal{S}(X)$. The idea is most simply put across by assuming that the $p_{a,b}(x)$'s do not vary with X, i.e. by assuming $p_{a,b}(x) = p_{a,b}$.

By using (2) and from the fact that X is finer than $\mathcal{S}(X)$ we can write

$$f_{Y|D_A,D_B,\mathcal{S}(X)}[y|a,b,s] = \int f_{Y|D_A,D_B,X}[y|a,b,x] f_{X|\mathcal{S}(X)}[x|s]dx.$$

[7] It is worth noting that the ATT is actually *over*-identified, as $\Delta^*(x)$ and the distribution $f_{X|D^*}[x|1]$ can be recovered from either knowledge of $\lambda_0^A(x)$ and $\lambda_1^A(x)$ or knowledge of $\lambda_0^B(x)$ and $\lambda_1^B(x)$.

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

Using (1) and the fact that the $p_{a,b}(x)$'s do not vary with X, the term on the right-hand-side of the last expression can be written as

$$[1-p_{a,b}]\int f_{Y_0|X}[y|x]f_{X|\mathcal{S}(X)}[x|s]dx + p_{a,b}\int f_{Y_1|X}[y|x]f_{X|\mathcal{S}(X)}[x|s]dx,$$

so that

$$f_{Y|D_A,D_B,\mathcal{S}(X)}[y|a,b,s] = [1-p_{a,b}]f_{Y_0|\mathcal{S}(X)}[y|s] + p_{a,b}f_{Y_1|\mathcal{S}(X)}[y|s],$$

where the last relationship again follows from X being finer than $\mathcal{S}(x)$. Accordingly, the distribution of wages for the group $(D_A = a, D_B = b)$ conditional on $\mathcal{S}(X) = s$ is again a mixture of two latent distributions. The components of this mixture are weighted means of the components in (1) taken over individuals with $\mathcal{S}(X) = s$, with mixture weights given by $p_{a,b}$. Note that the mixture representation holds for all groups defined by $D_A \times D_B$; most importantly, because of the balancing property of $\mathcal{S}(X)$, the mixture components remain the same for all four groups, while the mixture weights are group specific.

Note also that the same representation would hold if the $p_{a,b}(x)$'s were left to vary with X through the index $\mathcal{S}(x)$ (or if they were constant within cells defined by the balancing score):

$$f_{Y|D_A,D_B,\mathcal{S}(X)}[y|a,b,s] = [1-p_{a,b}(s)]f_{Y_0|\mathcal{S}(X)}[y|s] + p_{a,b}(s)f_{Y_1|\mathcal{S}(X)}[y|s]. \quad (3)$$

A similar argument applies if these probabilities only depend on a finite number of X's that can be non-parametrically controlled for in the analysis. The identification problem is similar to the one described in the previous section. If mixing weights can be recovered from (3), then one could identify the extent of misreporting in either measure of educational qualifications, and therefore the ATT, using the results in Battistin and Sianesi (2006).[9]

4 Estimation issues

4.1 Mixture representation of the problem

In this section we build on the mixture representation (3) to identify mixture weights (i.e. the $p_{a,b}(x)$'s) and mixture components. Clearly, non-parametric identification is not feasible, i.e. from raw mixture probabilities one cannot non-parametrically recover mixture weights and mixture components. For this reason, in what follows we shall consider mixtures in which all components come from the same *parametric* class. It is worth noting that the mixture we deal with is *finite* with a *known* number of components. More in general, if we were to consider the case of multiple 'treatments' (i.e. multiple

[9] However, identification of the $p_{a,b}(s)$'s is not easily achieved in (3), as conditional distributions are defined with respect to $\mathcal{S}(x)$ and not X. The problem is $\mathcal{S}(x)$ being a balancing score for (D_A, D_B) but *not* for D^*. Apparently, the proof of identification in Mahajan (2006) no longer holds for this case.

qualifications) the number of mixture components would correspond to the number of alternative qualifications.

We will assume throughout that within cells defined by $\mathcal{S}(x)$ the mixture components are *normally distributed* with cell-specific parameters. This correspond to assuming log-normality of wages conditional on the balancing score.[10] The reason for this choice is twofold. First, it can be shown that any finite mixture of univariate normal distributions is identifiable (see Everitt and Hand, 1981, and Titterington, Smith and Makov, 1985). Perhaps the most natural and intuitive way of addressing the identification problem for mixtures of parametric distributions is in Yakowitz and Spragins (1968), who show that a necessary and sufficient condition for the mixture to be identifiable is that the mixture components be a linearly independent set over the field of real numbers. This condition is met for the case of mixtures of normal distributions. Second, a variety of procedures have been developed to estimate mixture of normal distributions. Maximum likelihood estimators are well known to have desirable asymptotic properties (see Everitt and Hand, 1981), and it is natural to use the EM algorithm (Dempster, Laird and Rubin, 1977) for estimating the parameters of the mixture.[11]

A few technical considerations are worth mentioning. First, under the hypothesis of no 'treatment' effects (i.e of no returns) the two mixture components coincide and thus the mixture representation in invalid. More precisely, if we had

$$f_{Y_0|X}[y|x] = f_{Y_1|X}[y|x], \qquad \forall x$$

so that $\Delta^*(x)$ and the ATT are both zero, then the mixture representation in (1) would be invalid, which in turn would invalidate the mixture representation in (3). This problem is known in the statistical literature on mixtures as the problem of *homogeneity*. Note that testing homogeneity, that is testing no mixture against a mixture of two distributions, is a non-regular problem, in that the null hypothesis belongs to the boundary of the parameter space. However, using the results in Yakowitz and Spragins (1968), it follows that any non-degenerate finite mixture of normal distributions cannot itself be normal. It follows that in our application, testing for the presence of treatment effects under the maintained assumption of normal components and non-trivial $p_{a,b}$'s in (3) corresponds to testing normality of the observed mixture distributions.[12]

Second, the likelihood function for a two-component univariate normal distribution is unbounded, so that a global maximiser does not exist. Singularities occur at certain points on the boundary of

[10] Our estimation procedure can be extended to more general families of parametric distributions.

[11] Identification of returns through normality of the mixture components is in the spirit of the work by Heckman and Honore (1990), where it is shown that under normality it is possible to estimate the distribution of potential wages in the Roy model from a single cross-section of data. See also the discussion by Heckman (2001).

[12] We could not reject this condition for the empirical exercise presented in Section 6 below.

the parameter space as the variances of the mixture components tend to zero.[13] Everitt and Hand (1981) state that "*empirically meaningful maximum likelihood solutions can be obtained if attention is restricted to the largest of finite local maxima of the likelihood function*", though singularities may present a problem when the sample size is small and the components are not well separated. Moreover, it has been shown in the literature that the limit of an EM sequence depends on the initial guess used, and that convergence to a spurious maximiser or singularity of the likelihood function may be possible.

4.2 Definition of the balancing score

Let G be a variable identifying the 2×2 groups obtained by tabulating D_A and D_B:

$$G = \begin{cases} 1 \Leftrightarrow (D_A = 1, D_B = 1), \\ 2 \Leftrightarrow (D_A = 1, D_B = 0), \\ 3 \Leftrightarrow (D_A = 0, D_B = 1), \\ 4 \Leftrightarrow (D_A = 0, D_B = 0). \end{cases}$$

To account for the curse of dimensionality arising from the X's, we define $\mathcal{S}(X)$ as a partition that maps all observations into *mutually exclusive* sets $s_1, \ldots, s_J$ according to the values taken by $e_1(x)$, $e_2(x)$ and $e_3(x)$, where

$$e_g(x) \equiv Pr(G = g|x), \qquad g = 1, \ldots, 4.$$

Accordingly, the $e_g(x)$'s are the *propensity scores* defined by the *multinomial* regression of G on the x's. As we have shown above, letting the partition depend on values of the propensity scores is convenient to reduce the dimensionality problem. Results in Imbens (2000) and Lechner (2001) can be directly applied to conclude that $\mathcal{S}(X)$ is actually a balancing score for (D_A, D_B).

Our estimation strategy to recover the ATT consists of the following steps. First, by assuming normality of the mixture components in (3) we maximize the likelihood function defined over the four groups given by $D_A \times D_B$ and the J elements of the partition, the latter being defined to ensure strata of reasonable sample size. The resulting likelihood depends on $J \times 8$ unknowns (note that we allow the $p_{a,b}$'s to be stratum-specific): $p_{0,0}(s_j)$, $p_{1,0}(s_j)$, $p_{0,1}(s_j)$, $p_{1,1}(s_j)$, μ_{0j}, σ_{0j}, μ_{1j}, and σ_{1j}, $j = 1, \ldots, J$, the μ's and the σ's being mean and standard deviation of a normal distribution for the Y_0 and the Y_1 components, respectively. The EM algorithm is particularly useful to obtain the maximum likelihood estimates of the mixing weights in this case (Dempster, Laird and Rubin, 1977).[14]

[13] Many examples of this can be made - see the discussion in Everitt and Hand (1981).

[14] Our estimates proved robust to different starting values for the EM algorithm, obtained by generating random draws from $[0, 0.10]$ for $p_{0,0}(s_j)$, from $[0, 0.50]$ for $p_{1,0}(s_j)$ and $p_{0,1}(s_j)$, and from $[0.90, 1.00]$ for $p_{1,1}(s_j)$ and by fixing the μ's and the σ's to mean and standard deviation of log-wages in cells defined by $(D_A = 0, D_B = 0)$ and $(D_A = 1, D_B = 1)$. The numbers 0.10 and 0.90 were chosen with our application in mind so to reflect results in Kane, Rouse and Staiger (1999). To avoid numerical problems, we set $J = 8$ and allowed stratum-specific $p_{a,b}$'s only in strata with at least 15 observations.

Second, we derive the probabilities of exact classification from the mixture weights estimated in the previous step:

$$\begin{aligned}
\lambda_0^A(x_i) &= 1 - \sum_b p_{0,b}(s_i) P[D_B = b | D_A = 0, x_i], \\
\lambda_1^A(x_i) &= \sum_b p_{1,b}(s_i) P[D_B = b | D_A = 1, x_i], \\
\lambda_0^B(x_i) &= 1 - \sum_a p_{a,0}(s_i) P[D_A = a | D_B = 0, x_i], \\
\lambda_1^B(x_i) &= \sum_a p_{a,1}(s_i) P[D_A = a | D_B = 1, x_i].
\end{aligned}$$

These expressions depend on $P[D_B = b|D_A = a, x]$ and $P[D_A = a|D_B = b, x]$, which we estimate from parametric binary regressions on X.

At the end of this procedure, *two* estimates of the ATT can be recovered using either $(D_A, \lambda_0^A(x), \lambda_1^A(x))$ or $(D_B, \lambda B(x), \lambda_1^B(x))$. Estimates of the ATT can be obtained by using the characterisation result in Battistin and Sianesi (2006). Under the assumptions made in the paper, they are both consistent estimates of the same parameter. The two estimates are then combined so to define a new estimate with minimum variance.

5 Data and educational qualifications of interest

5.1 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.

As to educational attainment, of particular interest to our purposes is that cohort members were asked to report the qualifications they had obtained as of March 1981 (aged 23) not only in 1981 (aged 23), but also in 1991 (aged 33). We can thus construct two separates measures of qualifications

Table 1: Sample selection

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

obtained up to 1981, based either on responses in the 1981 or in the 1991 survey. Furthermore, in 1978 the schools cohort members attended when aged 16 provided information on the results of public academic examinations entered up to 1978 (i.e. by age 20).[15] For each individual we thus have *three* measurements of educational qualifications acquired up to age 20.

We focus on males, further restricting attention to those in work (and with wage information) in 1991 and for whom neither of the three educational measure is ever missing. These criteria, summarised in Table 1, leave us with a final sample of 2,713.

5.2 Educational qualifications of interest

Non-parametric identification of the misclassification probabilities requires access to at least two independent measures of educational attainment. In the NCDS data, such measures are offered by self-reported attainment and by the School Files, the latter however only recording academic qualifications and only those achieved by age 20, that is O- and A levels.[16]

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

First, academic qualifications are well defined and homogenous, with the central government traditionally determining their content and assessment. By contrast, the provision of vocational qualifications is much more varied and ill-defined, with a variety of private institutions shaping their content

[15] Similar details were collected from other institutions if pupils had taken such examinations elsewhere. Results were obtained for approximately 95% of those whose secondary school could be identified.

[16] In the British educational system, 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).

and assessment.[17] A second advantage of focusing on O and A levels is that they are almost universally taken through uninterrupted education, whereas vocational qualifications are often taken after having entered the labour market. It is thus more difficult to control for selection into post-school (vocational) qualifications, since one would ideally want to control also for the labour market history preceding the acquisition of the qualification.

A highly policy relevant parameter and the one we focus on in our application is the return from attaining any academic qualification (that is, from acquiring at least O levels) compared to leaving school at the minimum age of 16 without any formal qualification.[18] Special interest in O levels arises from the finding that in the UK, reforms raising the minimum school leaving age have impacted on individuals achieving low academic qualifications, in particular O levels. In particular, Chevalier *et al.* (2003) show that the main effect of the reform was to induce individuals to take O levels. Del Bono and Galindo-Rueda (2004) similarly show that changes in features of compulsory schooling have been biased towards the path of academic attainment; the main effect of the policy was not to increase the length of schooling, but rather to induce individuals to leave school with an academic certification. In such a context it is thus of great policy interest to estimate the returns to finishing school with O levels compared to leaving with no qualifications. Indeed, Blundell, Dearden and Sianesi (2005) found a non-negligible return of 18% for those who did leave with O levels and of 13% for those who dropped out at 16 without any qualifications. Furthermore, the return to acquiring at least O levels compared to nothing captures all the channels in which the attainment of O levels at 16 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.

It is important to highlight that since we compare O level attainment recorded by the schools by the time the individuals were aged 20 to O level attainment self-reported by individuals by the time they were aged 23, we need to further assume that O level qualifications are completed by age 20. We can however safely consider this assumption to be met, at the very least for our NCDS cohort. From the school files, we could verify that only a negligible fraction of O levels achieved by age 20 had in fact been achieved after the typical age of 16.

Table 2 presents cross tabulations between our three educational measurements.[19] Before briefly

[17]In fact, there is a wide assortment of options ranging from job-specific, competence-based qualifications to more generic work-related qualifications, providing a blend of capabilities and competencies in the most disparate fields.

[18]Although the British system is quite distinct from the one in the US, one could regard the no-qualifications group as akin to the group of high-school drop-outs. The 'None' category also includes very low-level qualifications at NVQ level 1 or less, in particular the academic CSE grade 2 to 5 qualifications. Students at 16 could take the lower-level Certificates of Secondary Education (CSE) or the more academically demanding O levels. The top grade (grade 1) achieved on a CSE was considered equivalent to an O level grade C. Most CSE students tended to leave school at 16.

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

Table 2: Academic qualifications: conditional probabilities by qualifications in the School files

age 20	age 23, at 23			age 23, at 33			
	None	**O**	**A**	**None**	**O**	**A**	**HE**
None	80.6	18.3	1.1	76.1	22.5	1.1	0.3
O	4.8	82.5	12.7	10.5	74.8	10.1	4.6
A+	0.2	1.2	98.6	3.7	1.2	48.2	46.9

discussing the table, it is reassuring to note that the patterns that emerge from them are the same irrespective of the samples selected on the basis of non-missing educational information ever or non-missing wage information in 1991 (the latter obviously also restricting attention to those employed in 1991).

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

Overall, we have the needed variation in the two measures. If we take the school files at face value, we find much more over- than under-reporting, although some under-reporting still remains. Always trusting the school files, we find most reporting errors to concern adjacent levels, though some non-negligible errors are still present even if move two levels. Finally, reporting errors get worse when individuals are asked to recall their qualifications.

Moving now to our more aggregate educational measure of any (academic) qualifications versus none, Table 3 highlights significant differences between the information contained in the two reports, with 10% of student-school matches disagreeing. Of particular interest for later comparison with our results, the incidence of academic qualifications in the population is 58.8% according to transcript

[20]This discussion on over-reporting crucially relies on the assumption that O and A level be completed by age 20. Although there is indirect support for this from the data and anecdotal evidence, in future research we will explore the sensitivity of our estimates to violations of this assumption.

Table 3: Sample size: academic qualifications (cell frequencies in brackets)

	Self-reported at age 23		
Admin	None	Any	
None	901 (33.2%)	218 (8.0%)	1,119 (42.2%)
Any	49 (1.8%)	1,548 (57.0%)	1,597 (58.8%)
	950 (35.0%)	1,766 (65.0%)	2,716 (100.0%)

information, whilst according to self-reports it is considerably higher, 65%.

6 Results

This section presents estimation results for our parameter of interest, namely the return to acquiring any academic qualification. We first combine administrative data from the School Files and information on academic qualifications self-reported at age 23 to recover the misclassification probabilities λ's, thus providing the first estimates for the UK of both exam transcript errors and individual misreporting probabilities.

6.1 Misclassification probabilities

Figures 1 and 2 present our estimates of the probabilities of correctly reporting attainment of academic qualifications. In particular, Figure 1 presents the percentage of individuals who report correctly to have any academic qualification (λ_1) and who report correctly not to have any academic qualification (λ_0). In Figure 2 we present the same quantities calculated from administrative data. The probabilities of exact classification have been calculated for all individuals in the sample using the methodology described in Section 4.2.[21] Table 4 reports their modes and corresponding bootstrap standard errors.

Our results suggest that individuals are appreciably *more* accurate than transcripts when it looks as if they don't have any academic qualification. The difference between the mode of the distributions of λ_0 from self-reported and administrative data is 5.2 percentage points and is significant at the conventional level. In other words, individuals are over 5 percentage points less likely than schools to under-report (forget) their qualifications. On the other hand, by comparing the modes of the distributions of λ_1 it seems that individuals are slightly *less* accurate than transcripts when it looks as if they do have academic qualifications (the difference between the modes of 1.8 percentage points is again significant at the conventional level). Individuals are thus almost 2 percentage points more likely than schools to over-report their attainment. Thus, and in line with the little evidence available

[21] After estimating cell-specific $p_{a,b}$'s depending on values of the balancing score, we were not able to reject the hypothesis of constant mixture weights across cells. We therefore estimated the mixture in (3) by imposing this restriction. The quantities $P[D_B = b|D_A, x]$ and $P[D_A = a|D_B, x]$ were predicted from parametric regressions considering only the coefficients of significant regressors.

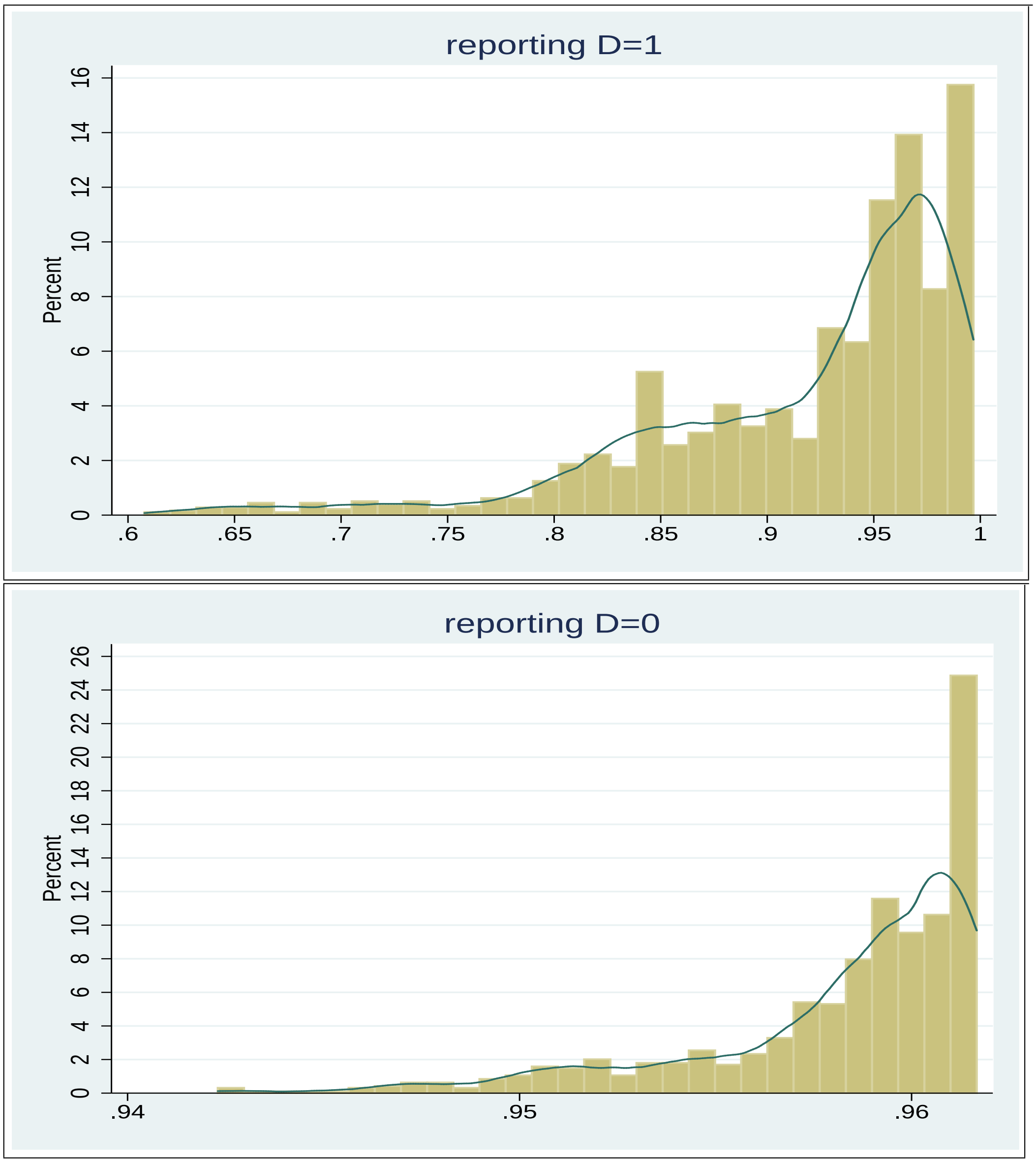

Figure 1: Extent of misclassification in *self-reported* information: percentage of individuals with any academic qualification amongst those reporting so (λ_1, top panel) and percentage of individuals with *no* academic qualification amongst those reporting so (λ_0, bottom panel)

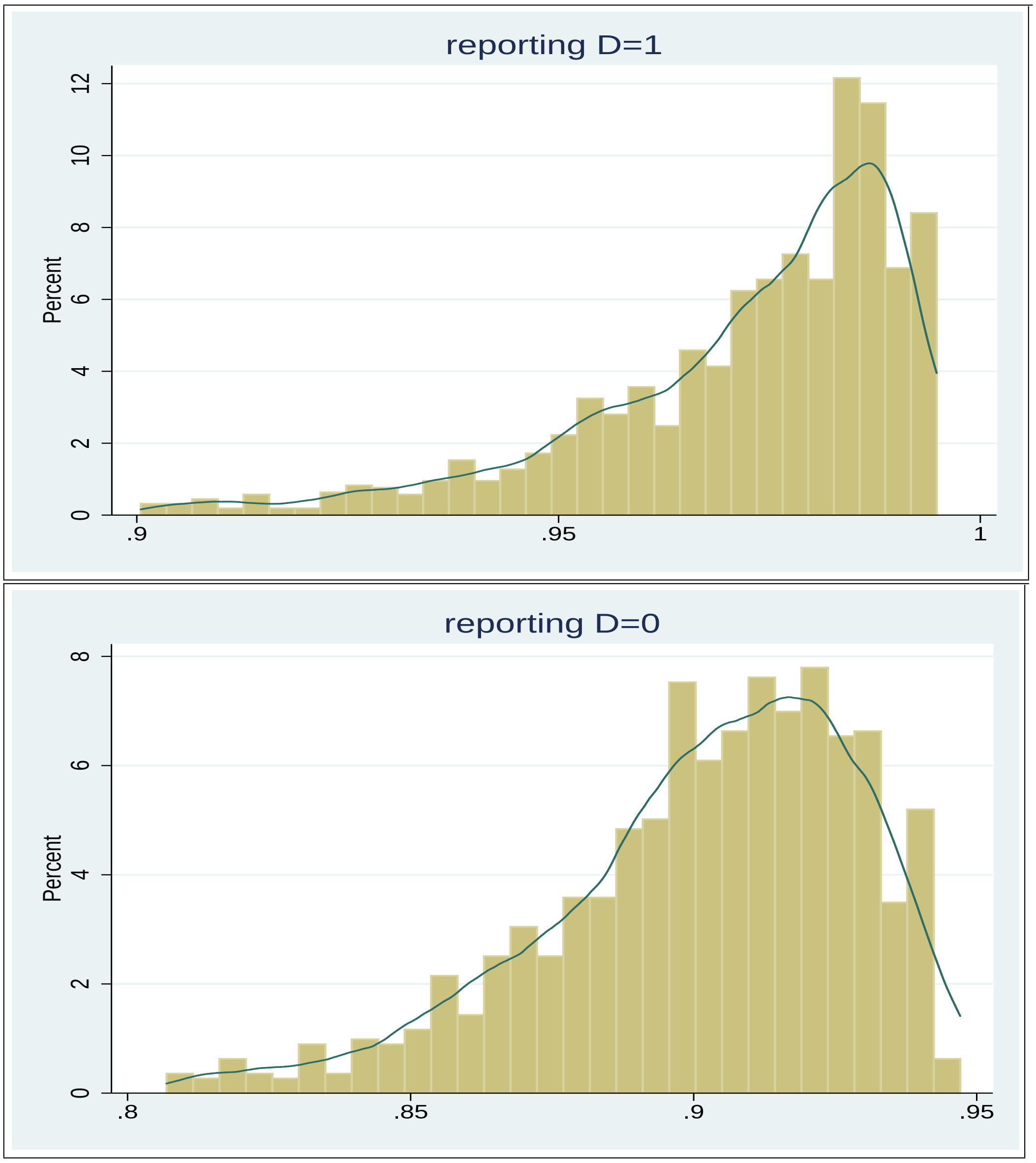

Figure 2: Extent of misclassification in *administrative* information: percentage of individuals with any academic qualification amongst those reporting so (λ_1, top panel) and percentage of individuals with *no* academic qualification amongst those reporting so (λ_0, bottom panel)

Table 4: Extent of misclassification

	Self-reported		**Administrative**	
	Mode	Std. Err.	Mode	Std. Err.
λ_1	0.970	0.119	0.988	0.120
λ_0	0.961	0.043	0.909	0.046

from the US, no source appears to be uniformly better.

Using the misclassification probabilities $\lambda(x)$'s we have recovered and the observed propensity score $e(x)$, from the law of iterated expectations (see Section 2) we obtain an estimate of the true incidence of academic qualifications in the population, namely $P(D^* = 1|x)$, of 55.6%. Interestingly, this is considerably lower than the incidence according to either educational measure; at 65%, however, individuals tend to over-report to a much larger extent than schools (58.8%).

6.2 Returns to any academic qualification

With the misclassification probabilities in hand, we can then proceed to estimate the returns to achieving any academic qualification as outlined in Section 4.2 above. In the current version of the paper, all the ATT's have been computed using constant probabilities λ_0 and λ_1 set to the mode of their distributions, and using the characterisation result in Battistin and Sianesi (2006). In future work, we plan to use the results in our companion paper and to allow for individual specific λ's. Our estimation results are reported in Table 5.[22]

We start by reporting the return as estimated from the raw data (Δ_{LFS}). Such an estimate relies on the self-reported educational measure and is obtained by controlling *only* for the Labour Force Survey-style variables of gender, age, ethnicity and region (gender and age via sample choice). Ignoring both omitted-ability bias and potential misclassification in self-reported attainment we thus find a return to academic qualifications of 32.9%.

We then include the *full* set of rich controls of Blundell, Dearden and Sianesi (2005): in addition to gender, age, ethnicity and region, we now control for math and reading ability test scores at 7 and 11, various family background measures (mother's and father's education, mother's and father's age, father's social class, mother's employment status and number of siblings when the child was 16) and school type. As one would expect in a situation of positive selection into education, the corresponding return, Δ_{FULL}, falls to 25.5%.

[22]Numbers in the table should be read as follows. FULL: Misclassification probabilities and returns estimated using the full set of controls, as in Blundell, Dearden and Sianesi (2005): gender and age (implicitly), math and reading ability test scores at 7 and 11, family background (mother's and father's education, mother's and father's age, father's social class when child was 16, mother's employment status when child was 16 and number of siblings when child was 16), school type, ethnicity and region. LFS = Misclassification probabilities and returns estimated using Labour Force Survey-style controls: gender and age (implicitly), ethnicity and region.

We subsequently implement our empirical strategy to correct for potential error in reported qualifications by exploiting the second, independent measure provided by the School files. In a first step we pretend to be using a Labour Force Survey-type of data and use the restricted (LFS) set of controls. Note that we use such limited set of variables both to estimate the misclassification probabilities and to then estimate the return. The corresponding estimate of Δ^*_{LFS}, at 36.1%, has increased by over 3 percentage points once accounting for measurement error.

We finally make full use of our rich NCDS data and implement our strategy to account for reporting errors. The estimates of the misclassification probabilities λ based on the full set of controls are those we discussed in the previous subsection. The resulting return is Δ^*_{FULL} (which in the rest of the paper we had simply termed Δ^*). Our estimate of the ATT from acquiring any academic qualification is 26.4%.

So what can we say about the relative importance of omitted ability and measurement error biases, and about the possibilities that the two cancel out? By comparing the true return Δ^*_{FULL} to the one ignoring both types of potential biases Δ_{LFS}, we do *not* find any evidence of balancing biases; quite to the contrary, ignoring both biases leads to a sizeable upward bias in estimated returns of 19%. The resulting calibration rule to get the LFS-style estimate of the average return to academic qualifications for males close to the true return suggests to multiply the 'raw' estimate by 0.81.

As to the relative importance of ability and measurement error biases, we find that what is crucial is to control for ability bias. In particular, estimates that correct for measurement error but not for omitted ability incur a 27% upward bias (compare Δ^*_{FULL} to Δ^*_{LFS}), whilst controlling for ability but ignoring misclassification error leads to a 4% downward bias, which is not statistically significant at conventional levels (compare Δ^*_{FULL} to Δ_{FULL}).

In this application, we thus find that the policymaker or analyst can not simply rely on measurement error to cancel out the ability bias. It is important to note that such a result pertains to measurement error in educational reports that were obtained relatively close to the completion of the qualification of interest. We also intend to provide estimates of individual misreporting probabilities at different elapsed times after completion of the qualification, which will allow us to assess whether the relative balance of the two biases is shifted once we turn to recall information, which is likely to suffer from a larger extent of measurement error.

7 Conclusions

In this paper we provide reliable estimates of the returns to educational qualifications in the UK that allow for the possibility of misreported attainment. We additionally identify the extent of misreport-

Table 5: Estimates of returns

Parameter	**Control for**	**Estimate**	**Std. Err.**
Δ_{LFS}	nothing	0.329	0.015
Δ_{FULL}	ability bias	0.255	0.026
Δ^*_{LFS}	meas. err. bias	0.361	0.025
Δ^*_{FULL}	both	0.264	0.047

ing in different types of commonly used data sources on educational qualifications and thus provide estimates of their relative reliability. In particular, we provide estimates of exam transcript errors, as well as of individual misreporting probabilities at different elapsed times after completion of the qualification of interest.

We also provide evidence on the relative importance of ability and measurement error biases, and produce some simple calibration rules as to how to correct returns estimated on data that rely on recall about individual qualifications and contain limited or no information on individual ability and family background characteristics (such as the Labour Force Survey).

On the methodological front, we propose a semi-parametric estimation based on balancing scores and mixture models.

References

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

[2] Battistin, E. and Sianesi, B. (2006), , unpubliblished manuscript, Institute for Foscal Studies, London.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] Dempster, A.P. Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete data via the EM Algorithm, Journal of the Royal Statistical Society Series B, 39, 1-38.

[15] Everitt, B.S. and Hand, D.J. (1981), Finite Mixture Distributions, Chapman and Hall: London

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

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

[18] Heckman, J.J. (2001), *Micro data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture*, Journal of Political Economy, Vol. 109, No. 4, 673-748.

[19] Heckman, J.J. and Honore, B.E. (1990), *The Empirical Content of the Roy Model*, Econometrica, Vol. 58, No. 5, 1121-1149.

[20] 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.

[21] Lechner, M. (2001), *Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption*, in M. Lechner and F. Pfeiffer (Eds.), Econometric Evaluation of Labour Market Policies, Heidelberg: Physica, 43-58.

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

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

[24] 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.

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

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

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

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

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

[30] 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.

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

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

[33] 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.

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

[35] 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.

[36] Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985), Statistical Analysis of Finite Mixture Distributions, John Wiley and Sons: Chichester.

[37] Yakowitz, S.J. and Spragins, J.D. (1968), On the Identifiability of Finite Mixtures, Annals of Mathematical Statistics, 39, 209-214.