Measurement Error and Returns to Education: Evidence from the UK

Erich Battistin

Institute for Fiscal Studies - London

Research ideas based upon results in…

Battistin, E. and Chesher, A. (2004) "*The Effect of Measurement Error on Evaluation Methods Based on Strong Ignorability*"

Battistin, E. (2004) "*Misreported Schooling and Returns to Education: Evidence from the UK*"

The Idea in a Nutshell

Throughout my talk I will investigate the effect of measurement error on the identification of treatment effects when the assumption of selection on observables is maintained

Data are informative on the triple *(Y,D,X)*

- *D* is the participation status, with D=1 for participants and D=0 for non participants
- *X* is the set of observables controlled for to assume *ignorable* participation
- *Y* is the outcome observed for each individual, which can be expressed in terms of *potential outcomes* from participation and non participation

Identification of treatment effects builds on the comparison of Y for participants and Y for non participants

The Idea in a Nutshell (continued)

I will consider the case of data informative on

- (Y,D,Z), where Z is an error affected measure of X
- (Y,W,X), where W is an error affected measure of D

As the identification of treatment effects requires that (Y,D,X) is observable, in both cases we get biased results

The Idea in a Nutshell (continued)

I will consider the case of data informative on

- (Y,D,Z), where Z is an error affected measure of X

as participation is *ignorable* once X is controlled for, comparing participants to non participants similar with respect to Z accounts only partially for the selection problem (see Battistin and Chesher, 2004)

- (Y,W,X), where W is an error affected measure of D

As the identification of treatment effects requires that (Y,D,X) is observable, in either case we get biased results

The Idea in a Nutshell (continued)

I will consider the case of data informative on

- (Y,D,Z), where Z is an error affected measure of X
- (Y,W,X), where W is an error affected measure of D

participants and non participants are erroneously classified, and the bias depends on the misclassification probabilities (see Battistin, 2004)

As the identification of treatment effects requires that (Y,D,X) is observable, in either case we get biased results

The Idea in a Nutshell (continued)

I will consider the case of data informative on

- (Y,D,Z), where Z is an error affected measure of X
- (Y,W,X), where W is an error affected measure of D

participants and non participants are erroneously classified, and the bias depends on the misclassification probabilities (see Battistin, 2004)

as D is binary, the measurement error is not classical and attenuation effects do not hold in general

As the identification of treatment effects requires that (Y,D,X) is observable, in either case we get biased results

An Application to UK data

I will use uniquely rich data from the NCDS to assess the importance of measurement error in estimating returns to education for the UK (as in Blundell *et al.*, 2004)

D: I will consider a multiple treatments setup, with treatments defined by different qualification levels (“None”, “O Levels”, “A Levels” and “Higher Education” – academic qualifications only)

X: controls include information on parents’ education and background, ethnicity, type of school attended, regional dummies and a proxy for ability (defined as the sum of scores at tests taken by individuals at age 7 and age 11)

Y: the outcome of interest is individual wages at age 33, and the analysis is restricted to males

Selection on X will be assumed throughout

An Application to UK data (continued)

In the first part of my talk I will consider the case (Y,D,Z) by allowing for errors in the NCDS ability score and using results from Battistin and Chesher (2004)

An Application to UK data (continued)

In the first part of my talk I will consider the case (Y,D,Z) by allowing for errors in the NCDS ability score and using results from Battistin and Chesher (2004)

In the second part I will deal with the case of mismeasured qualifications, that is (Y,W,X)

- First, bounds on returns
- Then, point identification using self reported qualifications and school records (Kane et al., 1999, and Black et al., 2000, Lewbel, 2003)

<u>Some results</u>….

still work in progress, but measurement error seems to play a non-negligible role in the estimation of returns to low level qualifications

IFS

Identification of treatment effects

Start from comparing observed outcomes for participants and non participants *net of* compositional differences with respect to *observable* characteristics

$$\sum [\quad E(Y \mid D=1, x_i) - E(Y \mid D=0, x_i) \quad] P(x_i \mid D=1)$$

if selection takes place *only* with respect to X, the last expression is equal to the *average treatment effect*

$$ATT = E(Y_1 \mid D=1) - E(Y_0 \mid D=1)$$

where Y_1 and Y_0 are the *potential outcomes from participation and non participation, respectively*

Estimation of treatment effects

✓ Different estimators are discussed in the literature depending on how we estimate the quantities in the expressions above (*para-semipara-nonpara*-metric stuff)

✓ *Propensity score matching* is a fancy and popular choice to make

✓ Since all methods use the *same* idea (selection on observables), they are all consistent for the *same* parameter (obvious, but it is worth pointing this out)

Mismeasured regressors

Assume that selection on X holds, but Z in place of X is unwittingly observed in the data

$$\sum [\quad E(Y \mid D=1, z_i) - E(Y \mid D=0, z_i) \quad] P(z_i \mid D=1)$$

It can be shown that, even if selection on X holds, the last expression does not identify the average treatment effect!

Propensity score matching does *not* work: even if participants and non participants are balanced with respect to Z, they are not necessarily balanced with respect to X

Mismeasured regressors (continued)

Characterising the bias that arises from using Z in place of X needs some work, but can be done (on a case by case basis)

An approximation to the bias can be derived when measurement error of classical form affects only one *continuous* variable in the X's (ability, in what follows)

In the latter case, if σ^2 is the variance of the error

$$\text{Bias} = \sigma^2 B(Z) + o(\sigma^2)$$

and B(Z) is identified from observed data (details in Battistin and Chesher, 2004)

Mismeasured regressors (continued)

$$\text{Bias} = \sigma^2 B(Z) + o(\sigma^2)$$

A sensitivity analysis can be conducted at conjectured values of the measurement error variance σ^2

Instrumental variables can solve for the problem, but only in a *linear* setting (non parametric identification is dealt with in the paper)

Attenuation bias does not hold in general

An application to NCDS data (continued)

Incremental returns

	O Level	A Level	HE
Ols			
Matching			
Weighting			
Stratification			

Bias (given the noise-to-signal ratio)

	O Level	A Level	HE
10%			
20%			
30%			

An application to NCDS data (continued)

Incremental returns

	O Level	A Level	HE
Ols	0.2092	0.0719	0.1619
Matching	0.2002	0.0813	0.1730
Weighting	0.1960	0.0830	0.1809
Stratification	0.2010	0.0830	0.1980

Bias (given the noise-to-signal ratio)

	O Level	A Level	HE
10%			
20%			
30%			

An application to NCDS data (continued)

Incremental returns

	O Level	A Level	HE
Ols	0.2092	0.0719	0.1619
Matching	0.2002	0.0813	0.1730
Weighting	0.1960	0.0830	0.1809
Stratification	0.2010	0.0830	0.1980

Bias (given the noise-to-signal ratio)

	O Level	A Level	HE
10%	0.0059	0.0012	0.0010
20%	0.0118	0.0024	0.0018
30%	0.0177	0.0037	0.0028

Back to NCDS data

The treatment variable in my example refers to different qualification types (HE,A Level, O Level or None)

Misclassification may arise because of misreporting of the qualification level. Respondents may either lie, not know if the schooling they've had counts as a qualification or simply not remember

According to evidence from other studies (Kane et al., 1999)

- misreporting is more likely to happen for low levels of qualification
- over reporting is more likely than under reporting

Mismeasured treatment status

Assume that selection on X holds, and that W in place of D is unwittingly observed in the data

$$\sum [\ E(Y \mid W=1, x_i) - E(Y \mid W=0, x_i)\ ]\, P(x_i \mid W=1)$$

Sadly enough, it can be shown that the last expression is not equal to the average treatment effect (see Battistin, 2004, for details)

The intuition for this is that individuals for whom we observe W=1 are a mixture of participants (D=1) and non participants (D=0), with mixing weights given by misclassification probabilities

Mismeasured treatment status (continued)

Two types of misclassification are to be considered

proportion of participants amongst those with W=0

$$P(D=1 \mid W=0) = 1 - P(D=0 \mid W=0)$$

proportion of non participants amongst those with W=1

$$P(D=0 \mid W=1) = 1 - P(D=1 \mid W=1)$$

✓ if both are zero, then we get standard identification of treatment effects by taking X into account

✓ they may depend on X (this makes things slightly more complicated)

✓ in the absence of further information, bounds on ATT can be derived exploiting priors and/or results from other studies

Mismeasured treatment status (continued)

Let λ_1=P(D=1|W=1) and λ_0=P(D=0|W=0)

Two types of restrictions on the misclassification probabilities are often imposed

✓ observations of W are more accurate than pure guesses

$$\lambda_0 > 0.5 \qquad \lambda_1 > 0.5$$

can be weakened by assuming that the sum of these probabilities is greater than one, that is $\lambda_0+\lambda_1>1$

Mismeasured treatment status (continued)

Let λ_1=P(D=1|W=1) and λ_0=P(D=0|W=0)

Two types of restrictions on the misclassification probabilities are often imposed

- ✓ observations of W are more accurate than pure guesses

$$\lambda_0 > 0.5 \qquad \lambda_1 > 0.5$$

- ✓ over reporting is more likely than under reporting

$$\lambda_1 < \lambda_0$$

These restrictions can hold within groups defined by X (for example, groups defined by ability). Since

$$ATT=ATT(\lambda_0,\lambda_1,Y,W,X)$$

bounds can be derived by looking at the max and the min value of the last expression with respect to (λ_0,λ_1)

Bounds

λ_1 (columns); λ_0 (rows)

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1									
0.2									
0.3									
0.4									
0.5									
0.6									
0.7									
0.8									
0.9									

Bounds

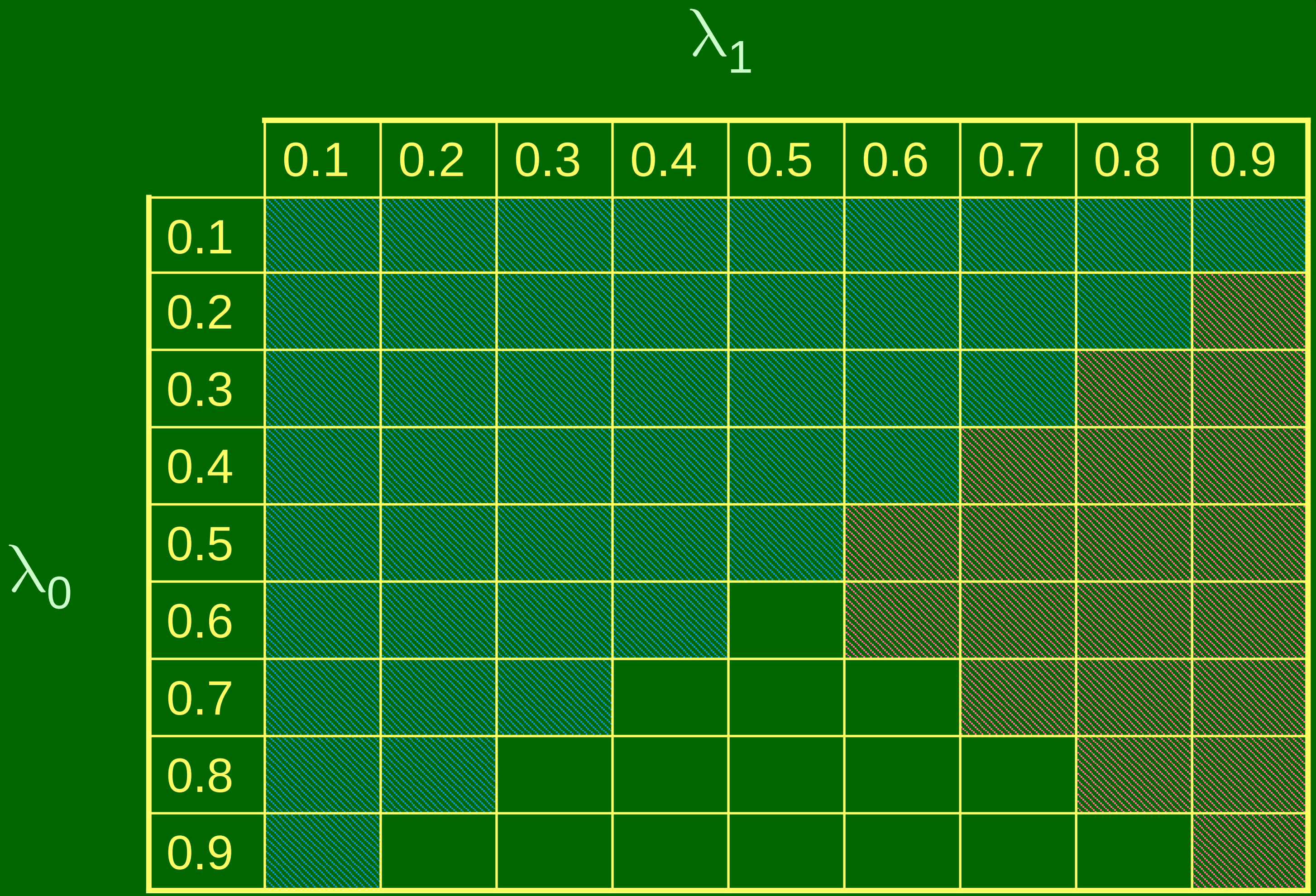
λ_1

λ_0

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1									
0.2									
0.3									
0.4									
0.5									
0.6									
0.7									
0.8									
0.9									

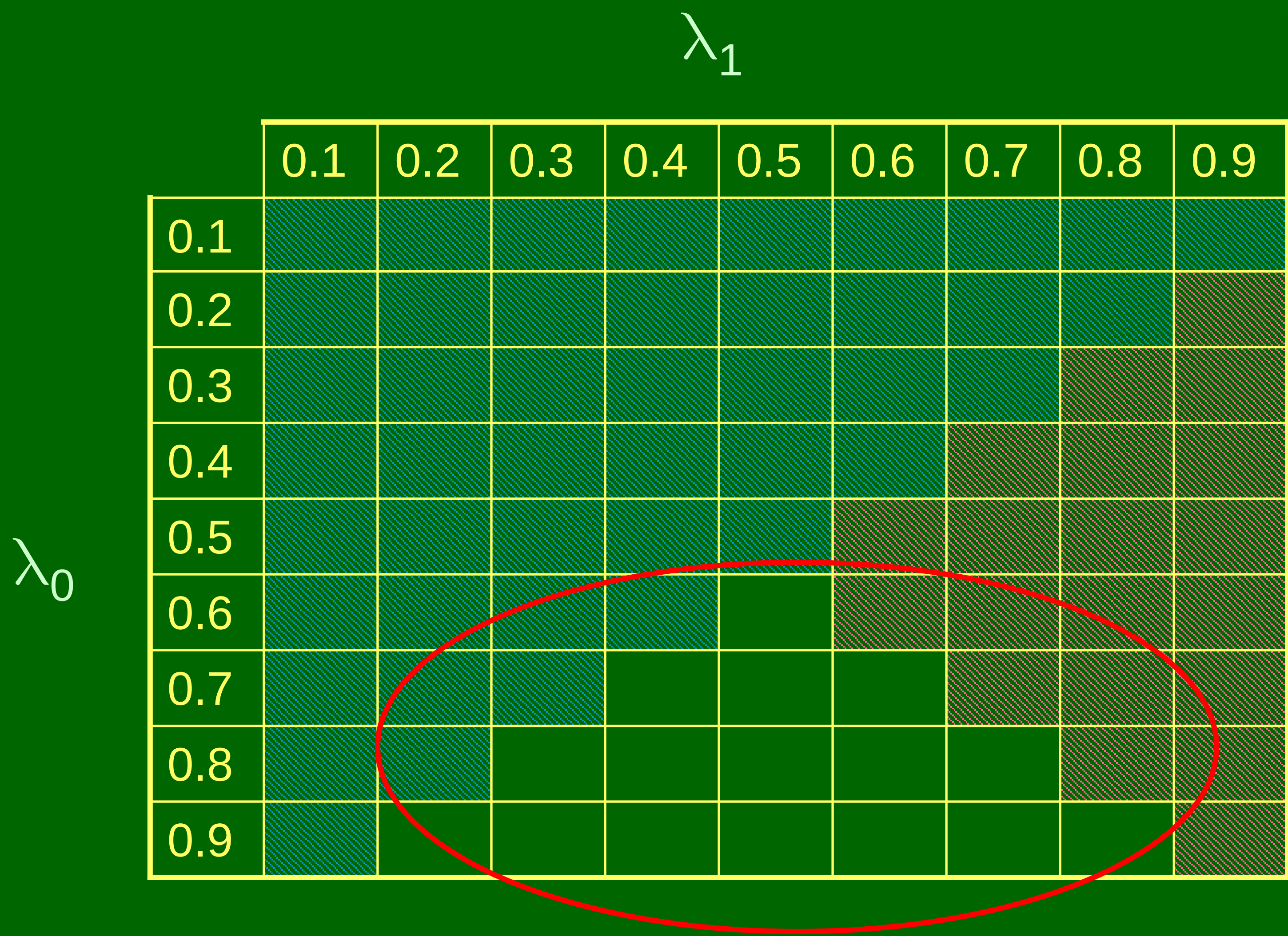
$$\lambda_0 + \lambda_1 > 1$$

Bounds



$\lambda_0 + \lambda_1 > 1$ $\lambda_1 < \lambda_0$

Bounds



$\lambda_0 + \lambda_1 > 1$ $\lambda_1 < \lambda_0$

An application to NCDS data (continued)

Returns to any qualification and to HE

	$\lambda_0 + \lambda_1 > 1.4$		$\lambda_0 + \lambda_1 > 1.5$		$\lambda_0 + \lambda_1 > 1.6$	
Any						
HE						
	$\lambda_0 + \lambda_1 > 1.7$		$\lambda_0 + \lambda_1 > 1.8$		$\lambda_0 + \lambda_1 > 1.9$	
Any						
HE						

An application to NCDS data (continued)

Returns to any qualification and to HE

	$\lambda_0 + \lambda_1 > 1.4$		$\lambda_0 + \lambda_1 > 1.5$		$\lambda_0 + \lambda_1 > 1.6$	
	lower	**upper**	**lower**	**upper**	**lower**	**upper**
Any						
HE						
	$\lambda_0 + \lambda_1 > 1.7$		$\lambda_0 + \lambda_1 > 1.8$		$\lambda_0 + \lambda_1 > 1.9$	
	Lower	**upper**	**lower**	**upper**	**lower**	**upper**
Any						
HE						

An application to NCDS data (continued)

Returns to any qualification and to HE

	$\lambda_0 + \lambda_1 > 1.4$		$\lambda_0 + \lambda_1 > 1.5$		$\lambda_0 + \lambda_1 > 1.6$	
	lower	**upper**	**lower**	**upper**	**lower**	**upper**
Any	0.289	0.723	0.289	0.525	0.289	0.482
HE	0.253	0.677	0.253	0.488	0.253	0.446
	$\lambda_0 + \lambda_1 > 1.7$		$\lambda_0 + \lambda_1 > 1.8$		$\lambda_0 + \lambda_1 > 1.9$	
	Lower	**upper**	**lower**	**upper**	**Lower**	**upper**
Any	0.289	0.409	0.289	0.361	0.289	0.304
HE	0.253	0.374	0.253	0.328	0.253	0.270

NCDS qualifications

Three measurements of qualification available at age 23

Self-reported qualifications by age 23 (Wave 4, 1981)

Self-reported qualifications by age 23 (Wave 5, 1991)

- First, ask about qualifications obtained after 1981
- Then, general question about qualifications obtained in life

NCDS qualifications

Three measurements of qualification available at age 23

Self-reported qualifications by age 23 (Wave 4, 1981)

Self-reported qualifications by age 23 (Wave 5, 1991)

Admin information by age 21 (School records, 1978)

- schools which cohort members had attended at age 16 were asked to supply results for O Level and A Level examinations
- information was collected from other institutions if pupils had taken such examinations elsewhere

NCDS qualifications

Three measurements of qualification available at age 23

Self-reported qualifications by age 23 (Wave 4, 1981)

Self-reported qualifications by age 23 (Wave 5, 1991)

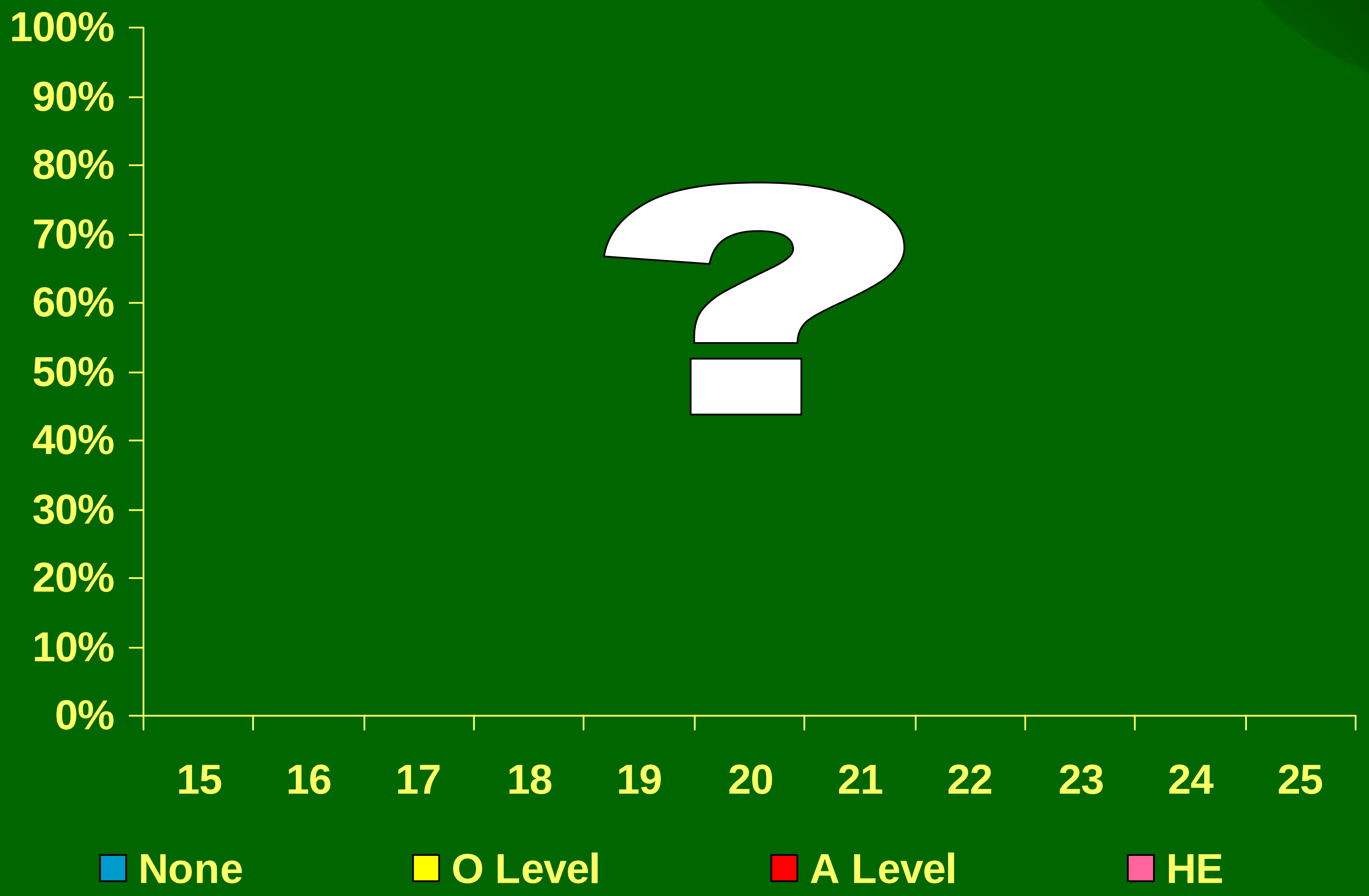
Admin information by age 21 (School records, 1978)

All measures are likely to be a reasonably good indicator for all qualification levels but not for higher education

In what follows, I will assume that O levels and A levels qualifications are attained by age 21 (sounds plausible, as I consider only academic qualifications)

- O Levels generally obtained by age 16 if undertaken at school
- A Levels generally obtained at the end of secondary school

Age when obtained highest qualification



NCDS qualifications (continued)

Self-reported 91

Self-reported 81

	None	O Level	A Level
None			
O Level			
A Level			

NCDS qualifications (continued)

Self-reported 91

Self-reported 81

	None	O Level	A Level
None	63.27	31.97	4.76
O Level	8.40	71.76	19.85
A Level	2.95	16.58	80.47

NCDS qualifications (continued)

School Records

Self-reported 81

	None	O Level	A Level
None	95.32	4.09	0.58
O Level	40.65	59.15	0.20
A Level	19.04	40.74	40.23

NCDS qualifications (continued)

School Records

Self-reported 91

	None	O Level	A Level
None	91.19	6.99	1.82
O Level	41.55	58.03	0.42
A Level	15.20	39.59	45.21

Point identification

Multiple reports of D can solve for misclassification, provided that errors are independent across reports (see Kane et al., 1999, and Black et al., 2000, Lewbel, 2003)

To fix ideas, let W_1 be the qualification that results from the school files and let W_2 be self reported qualification

$$W_1 = D + e_1 \qquad W_2 = D + e_2$$

Identification of returns is possible when $W_1 \perp W_2 \mid D$

➢ this appears to be the case for qualifications as they result from the school files and from either 81 or 91 reports

Point identification

Multiple reports of D can solve for misclassification, provided that errors are independent across reports (see Kane et al., 1999, and Black et al., 2000, Lewbel, 2003)

To fix ideas, let W_1 be the qualification that results from the school files and let W_2 be self reported qualification

$$W_1 = D + e_1 \qquad W_2 = D + e_2$$

Hopefully partial identification if $Cov(W_1, W_2 \mid D) > 0$

- one can assume that errors in 91 reports are positively correlated with errors in 81 reports, since they come from the same person

Point identification

Multiple reports of D can solve for misclassification, provided that errors are independent across reports (see Kane et al., 1999, and Black et al., 2000, Lewbel, 2003)

To fix ideas, let W_1 be the qualification that results from the school files and let W_2 be self reported qualification

$$W_1 = D + e_1 \qquad W_2 = D + e_2$$

Not quite as IV: actually, it can be shown that instrumenting one report with the other produces upward biased estimates of treatment effects

Point identification (continued)

GMM methods can be used to estimate

- the misclassification probabilities for W_1 and W_2 (conditional on X)
- the returns to qualifications corrected for misreporting

four equations result from the mean of Y in cells defined by the 2X2 cross tabulation of W_1 and W_2

$$E(Y|W_1 = w_1, W_2 = w_2, x)$$

three equations result from the sample proportions

$$P(W_1 = w_1, W_2 = w_2, x)$$

it can be shown that the seven equations above define seven unknowns, so that point identification is achieved

if λ_1 and λ_0 do not depend on X (or are constant within groups defined by X), the seven unknowns are over identified

An application to NCDS data (continued)

returns to any qualification

	81 reports	91 reports
λ_1		
λ_0		
Effect		
Raw data		

IFS

An application to NCDS data (continued)

returns to any qualification

	81 reports	**91 reports**
λ_1	0.9804	
λ_0	1.0000	
Effect	0.2895	
Raw data	0.3026	

Summary

Nice idea, isn't it?

YES NO