

UCD GEARY INSTITUTE DISCUSSION PAPER SERIES

The Returns to Observable and Unobservable Skills over Time: Evidence from a Panel of the Population of Danish Twins*

Paul Bingley
Department of Economics, Aarhus School of Business, 8000 Aarhus, Denmark

Kaare Christensen
Department of Epidemiology, University of Southern Denmark, 5000 Odense, Denmark

and

Ian Walker
Department of Economics, University of Warwick, Coventry CV4 7AL, UK
Education Research Section, Princeton University, Princeton NJ 08540, USA

Tuesday, 12 June 2007, Version 2.04

Keywords: wage inequality, schooling, twins, education returns, ability bias
JEL Codes: I20, J31

Abstract:

This paper provides estimates of the private financial return to education based on large samples of monozygotic (MZ) and dizygotic (DZ) twins which we obtain from Danish population registers. Our estimation exploits the fact that our data is a long panel. We show that the rising inequality, which we observe in the raw data, is due to rising returns to observable skills. Indeed, our results suggest that the inequality associated with unobservable skills appears to have fallen since the late 1980's. The fact that we have both MZs and DZs allows us to separate the rising residual variance into changes in returns to unobservables and changes in the variance in unobservables across successive cohorts.

Measurement error has been a concern in the twins literature since the usual methodology is based on within-twin differences. We exploit two instruments that provide additional measures of the within twin schooling difference: differences in when the twins first join the labour force

on a full-time basis, which comes from a register that is independent of the education registers; and the strong assortative mating in the data which allows us to use twins spouse's education as an instrument. We also address a further concern in the literature: that differencing between twins fails to remove individual fixed effects as opposed to family fixed effects resulting in schooling differences being correlated with the residual. This would induce the within twin schooling difference coefficient to be biased. Here we exploit the Danish equivalent of Maimonides' rule which generates potential variation in education within twin pairs associated with being placed in different classes if they attended a small school in a larger than average cohort. This different experience across twin pairs is shown to generate differences in within twin schooling. Our baseline estimates suggests that correcting for self-selection in schooling, and measurement error, gives returns that are about two fifths higher than OLS for men and about one fifth higher for women.

* The construction of the dataset used in this research was funded of Her Majesty's Treasury's Evidence Based Policy Fund in conjunction with the UK Inland Revenue, Department for Education and Skills, the Department of Work and Pensions, and the Department of Culture, Media and Sport. Our collaboration was supported by the Aarhus University Research Fund (AU-2004-533-033), the Danish Social Science Research Council (24-04-0240), and the UK Economic and Social Research Council (R.ECAA.0071). The usual disclaimer applies.

Corresponding author: Paul Bingley Telephone +45 8948 6400 Email pbi@asb.dk

1. Introduction

The role that education plays in determining the distribution of wages has been the subject of extensive research. Estimating the returns to education has been a major industry for the research community and an important issue has been identifying the causal effect of education. The substantial rise in wage inequality that has occurred since the 1970's in the US, the UK and some (but not all) other countries has provided further motivation for this research agenda. The major issue in this branch of the literature has been the extent to which this rising inequality is due to rising returns to observable skills, like education, or rising residual inequality - which itself can be decomposed into an increase in the return to unobserved skills and an increase in the variance of unobserved skills.

This paper aims to provide estimates of the causal effect of education, how it has changed over time, and the extent to which there has been a change in the returns to unobservable skills. The novelty of the paper is that, by exploiting a large panel of twins, it can throw light on all of these issues that the existing literature has only addressed by making strong assumptions.

There are many studies of the private financial returns to education based on the standard human capital model of earnings determination (see the excellent review by Card (1999)). Bias may occur in ordinary least squares estimates because the error term in the earnings equation is likely to be correlated with schooling for a variety of reasons - most famously because of omitted "ability". Moreover, the large empirical literature of the effect of schooling on wages emphasizes that this parameter varies across individuals and that individuals sort themselves non-randomly across schooling levels (see, for example, Card (1999, 2001), and Carneiro, Heckman and Vytlacil, (2003)). As a result, considerable care is needed both in the estimation of such returns and in the interpretation of the estimates. In particular, instrumental variables estimators do not, in general, provide estimates of the average effect of education on wages.

Identical twins have been used, in around a dozen studies to date[1], to provide estimates of the causal effect of schooling on earnings that are, arguably, free from ability bias. "Ability" here is used to denote any unobserved attributes that are

[1] Card (1999) includes a review of twins research and Bonjour *et al* (2003) updates this.

specific to an individual, *fixed over time*, and associated with productivity in the labour market and hence wages. This covers a multitude of unmeasured endowments that are associated with a greater ability to make money, such as pre-school human capital investments and non-cognitive attributes like motivation and perseverance, as well as any purely genetic component of intellectual ability. Identical (monozygotic, MZ) twins are particularly valued by researchers because they have the same endowments at the time of conception[2]. It is the prospect that differencing within MZ twin pairs eliminates these unobserved endowments that makes such twins attractive for researchers. Thus, the extent of ability bias can, in principle, be inferred from comparing the schooling coefficient estimate using data on the fraternal (dizygotic, DZ) twins (or, indeed, any sample of unrelated individuals) with estimates based on MZ pairs of twins. While instrumental variable estimates identify local effects, a further advantage of twins is that, providing ability bias is eliminated by differencing, we would expect that a twins-based estimator would provide an estimate of the average effect of education on wages

Since schooling is an important determinant of wages it is likely to substantially affect inequality. Thus, changes in inequality might arise because of changing returns to observable skills like schooling, or because of rising residual inequality. Juhn, Murphy and Pierce (1993) showed that both observed (education) and unobserved dimensions of inequality had been growing over time. Until recently, there was a consensus in the literature that the large increase in wage inequality of the 1980s, in the US, UK and elsewhere, was due to rising returns to both dimensions of skill. It has be argued that skill biased technological change (SBTC) has been the main factor behind the rise in the relative demand for skills (e.g. Krueger (1993), and Berman, Bound and Griliches, (1994)) and this idea has been an important building block in recent research on growth and trade (see, for example, the review in Acemoglou (2002)). Some recent research (e.g. Card and diNardo (2002), Beaudry and Green (2005) and Lemieux (2006)) argue that the growth in wage inequality was concentrated in the 1980's. Further research by Mincer (1998) and Deschenes (2002) suggest the rise in inequality has been concentrated at the upper end of the wage

[2] Although other important differences may remain. For example, birthweight differences are thought to have effects on education – see Berhman and Rosenzweig (2004).

distribution. Both of these developments represent challenges to the simple SBTC thesis.

In contrast to the literature on wages and schooling, studies of wage inequality, such as Card and Lemieux (2001) and Katz and Murphy (1992), have largely ignored the problem of endogeneity induced by self-selection into schooling. It is far from obvious that the bias associated with schooling selection is constant over time so the observed changes in returns over time may not reflect changes in the causal effect of education. Taking an instrumental variables approach to this problem requires that the researcher has a valid instrument which identifies the same local average effect over time. Taber (2001) uses NSLY to estimate earnings functions using IV, Heckman selection methods, and a fully structural model of dynamic selection and, in each case, finds that there is no increase in education returns so that increasing inequality is associated entirely with an increase in residual variance..

More recently, an innovative paper by Deschênes (2003) has suggested that changes in the causal effect of education on wages can be recovered from changes in the degree of convexity in the relationship between wages and education. The argument here is that ability bias causes the relationship to be convex (or more convex than it would be without ability bias). This arises because, in a model with heterogeneity in returns, those with higher unobserved returns select themselves into higher levels of education so that, even though the underlying causal relationship may be linear, the observable relationship between log wages becomes convex. Contrary to Taber (2001), Deschênes finds a large rise in the degree of convexity, across all cohorts, and concludes that there has been a large rise in the causal effect of education and he calculates that this rise accounts for the bulk of the increase in the correlation between education and wages.

In this paper we exploit a large and long panel of twins, where we can identify zygosity. The MZs allow us to estimate the causal effect of observable education and how it has changed over time. By comparing the MZ estimates with estimates from the DZs we are able to uncover how much of inequality variation over time can be attributed to the econometric residual that has usually associated with unobserved skills. The fact that our data is a long panel and consists of twins from a wide range of cohorts means that we are able to decompose the variation in residual inequality into variation in the returns to unobserved skills and the changes, across cohorts, of the

variance in unobservables. The advantages of using twins are that no exclusion restrictions are required in order to address ability bias and the estimates are average effects, rather than local ones[3]. However, Zvi Griliches (1979) cautions that twins are "not a panacea", for a number of good reasons and renewed reservations about the use of twins for estimating the returns to education have been expressed in John Bound and Gary Solon (1999). There are two important issues that make twins problematic for estimating the causal effects of education on wages.

Firstly, measurement error may be large and is an issue even if instruments are available because the error may not be classical in nature – something that would undermine the validity of IV estimation. Secondly, as Neumark (1999) and Bound and Solon (1999) (BS) argue, there may still be endogeneity that causes bias in the wage difference equation because the within differences in schooling may be correlated with the error term. The presumption in the twins literature is that the omitted ability is entirely made up of a genetic effect and a family effect both of which therefore disappear by differencing between family members with the same genes. There is, in general, no strong reasons for thinking that this is necessarily the case – for example, birthweight differs between twins (actually by considerably more than between non-twin siblings) and there is substantial evidence that birthweight has real effects[4]. Neumark and BS note that if differencing does not remove *all* of the omitted ability then the within-twin estimator may still be biased and, in fact, may even be more biased than least squares applied to individuals. Addressing this important criticism is difficult in the context of twins because any instrumental variable that gives rise to differences in schooling via some reform are quite likely to affect both twins equally.

The first important criticism relates to measurement error. Griliches (1979) notes that the use of estimates obtained from differencing in general, and differencing within twins in particular, is that the method exacerbates the extent of measurement

[3] We can also, in principle, use twins data to investigate the extent to which the relationship between log wages and ability is concave and, by comparing MZ and DZ, see how ability bias might affect this concavity. Previous twins studies have assumed that returns to education are constant across education levels – but this is an assumption that has been driven by the sparsity of observations with large education differences in twins data, and not by any requirement to impose some parametric identification restriction. Here, we have such a large sample that we can relax this assumption but we reserve this, and other extensions, for future work.

[4] Berhman and Rosenweig (2004) (BR) find significant effects of birthweight on college attendance. Currie and Thomas (2003) show, in their analysis of the long run effects of the HeadStart program, that birthweight affects subsequent outcomes.

error in schooling and so increases the tendency for estimates to be attenuated (i.e biased towards zero) because of this larger measurement error. The solution to a pure measurement error problem is to use a second measure of the variable that is measured with error. Provided that the measurement error is classical (i.e. that the errors are independent of the truth) and that the two measures are correlated, the second measure can be an instrument for the first. An important innovation, proposed by Ashenfelter and Krueger (1994) (AK), was to use instrumental variables to eliminate this measurement error bias in twin differenced data under the assumption that the errors are classical. When they collected their data they asked each twin about the other twin's schooling and this cross-reported schooling measure is used as the basis for an instrument for the within twin pair difference in education. This innovation has largely been responsible for the subsequent revival of the use of twins to estimate the returns to schooling[5].

Our administrative register data has the advantage over own reported survey data that education is the official record of the individual's activity. Education is recorded as the highest vocational qualification and highest academic qualification obtained to date. The Ministry of Education attributes a "typical" completion time to each qualification[6], and the information we have access to is the maximum of the completion times associated with these qualifications[7]. Consequently, there is no issue of recall or individual misrepresentation. There is still the possibility that administrative mistakes occur or that reporting practices differ between institutions. For measurement error to be classical they need to be uncorrelated with the true education. There are more than three thousand reporting educational institutions and it seems unlikely that there would be systematic variation in measurement between institutions within the same level[8]. However, administrative records may still contain

[5] Berhman (1994) uses, for male twins, the twin's child's report of his/her father's schooling.

[6] Negotiation between the relevant professional bodies, educational institutions and the Ministry determines "typical" completion times. These are used as an input into the financing formulae that determines the amount the institution receives for a completed year.

[7] Some studies use qualifications rather than duration of schooling. Flores-Lagunes and Light (2004) is concerned about the non-classical nature of the measurement error when education is constructed from qualifications information since IV in this context relies on the measurement error being classical. However, they find that their estimates are not sensitive to the non-classical nature of the errors.

[8] Isacsson investigates the nature of measurement error in Swedish education registers, and finds it to be non-classical in nature. However, there is a large difference in timing of the two education measures he compares: 1990 register-based data and the 1974 survey-based data. This 16 year difference casts doubt on the degree of comparability.

measurement error and even a small proportionate measurement error in schooling can lead to a large proportionate measurement error in schooling differences that will cause biased estimates in twin differenced modelling. Moreover, individual completion times are clearly grouped by this method.

Here we have two possible ways to purge the within twin education differences of measurement error. Firstly, we have co-twin's partners' education (we include cohabitees as well as spouses) which, although it limits us to only those twins who have partners (at some point during the panel we observe them), allows us to take advantage of the strong assortative mating in the data[9]. The identifying assumption here is that one twin's wage is independent of cross-twin's spouse's education conditional on own spouse's education. Although assortative mating was apparently not a feature of the US data used in several important previous twin studies, or of the Swedish or UK data, it has been noted in earlier US data and it is clearly present in the Danish data. Secondly, the pension contribution registers provide us with information on when individuals began their paid work, after completing schooling, since pension contributions are compulsory. This register is entirely independent of the education register and there are no administrative mechanisms that ensure that they are consisent. Thus, we also use the difference between the age of starting work as an IV for measured schooling differences.

The second possible source of endodgeneity of within-twin pair schooling differences arises if there is some omitted variable that induces the within-twin wage differences to be spuriously correlated with schooling differences even when they are not measured with error. This important problem has been highlighted in recent criticisms of twins research by Neumark (1999), as well as by Bound and Solon (1999). They argue that there may still be endogeneity that causes bias in the wage difference equation because the within differences in schooling may be correlated with the error term. The presumption in the twins literature is that the omitted ability is entirely made up of an environmental component which is shared by MZ and DZ twins and an endowment component which is only shared by MZ twins. The effects of shared environments are removed with all twins differences, but the effects of shared endowments are only removed for within MZ-twin-pair differences. However, this

[9] Using own partner's education would be invalid if, as seems likely, it affects own wages directly. The sample selection is unimportant since very few individuals record no partner over the panel.

might not be the case – for example, birth weight differs between twins and there is substantial evidence that birth weight has real effects[10]. These critics note that if differencing does not remove *all* of the omitted ability then the within-twin estimator may still be biased, and may even be more biased than least squares applied to individuals. Directly addressing this criticism is difficult in the context of twins because most instrumental variables that give rise to differences in schooling via, for example, a policy reform are quite likely to affect both twins equally. However, we are able to go some way towards a response to Neumark and Bound and Solon by virtue of DZ twins which can be thought of as an immune group in the sense that individual DZ endowments will not be removed by differencing. Thus, DZ differenced estimates ought to provide an upper bound on the bias, and MZ differences ought to be (well) below that level[11].

Here we exploit the variation in the number of classes in each entry cohort of the schools attended by the twins. If a pair of twins attended a school where there is only one class in each school year then they will necessarily belong to the same class. However, if there is a second class then this leaves open the *possibility* that the twins are taught in different classes. In Denmark it is usual for schoolchildren to have the same teacher for the whole of their duration of compulsory schooling and that teacher would be responsible for delivering all, or at least much of, the curriculum. Thus, having separate teachers could result in markedly different outcomes for the twins that affect their later lives. While some schools have a policy of keeping twins in the same class, some have a policy of separating them, and others leave the decision to the parents. Part of the variation in the number of classes in Denmark is due to the operation of an administrative rule that forces schools to split large cohorts to ensure that the class size does not exceed a specific level[12]. Thus, small variations in cohort

[10] Berhman and Rosenweig (2004) (BR) find significant effects of birthweight on college attendance. Currie and Thomas (2003) show, in their analysis of the long run effects of the HeadStart program, that birthweight affects several subsequent outcomes. Black, Devereux and Salvanes (2005) use a large sample of twins to estimate that a 10% increase in birthweight leads to a 1% increase in earnings.

[11] Bonjour *et al* (2003) is the only paper to attempt to explore this issue using instrumental variables. Inspired by Evans and Montgomery (1994), who noted a strong correlation between smoking when young and education, they attempt to use youth smoking as an instrumental variable but found no correlation between within pair smoking and within pair education. A major factor in determining education was a test conducted at age 11. Passing this test would result in selection into an academic, as opposed to vocational, school with expectations that one would stay at school until 18 and most would then attend university. Unfortunately only three of their twin pairs had different test results.

[12] Angrist and Levy (1999) refer to such a rule in Israeli schools as Maimonides Rule. See Bingley, Jensen and Walker (2005) for further details in the Danish context.

size within a school district can cause large changes in class size and change the probability that twins will enter, and remain in, separate classes from 1st up to 9th grade.

The remainder of the paper is organised as follows. The next section reviews the literature and places our contribution within that. A data description is followed by estimation results, interpretation and discussion.

2. Literature

2.1 Twins

Table 1 summarizes the most recent identical twins studies[13] and extends the reviews in Behrman and Rosenzweig (1999) (BR) and in Card (1999). Table 2 lists the fraternal (DZ) results where available and updates Table 6 in Card (1999). AK use the original 1991 Twinsburg festival data which yielded just 147 MZ pairs[14], while the Ashenfelter and Rouse (1998) (AR) study used the pooled data of 333 MZ pairs by adding the 1993 festival, and Rouse (1998) used the 453 MZ pairs which added the 1995 festival[15]. The Twinsburg data has very few DZ twins and they have not been used in previous research. Behrman *et al* (1994) (BRT) used the NAS-NRC data on 141 MZ pairs who were all white male World War II veterans, and BR used the Minnesota Twins Registry data of 720 MZ pairs.

The Australian Twin Register for 1980/82 and 1988/89 yielded 602 MZ pairs and 568 DZ pairs[16] which are analysed in Miller *et al* (1995) (MMM1) and in Miller *et al* (1997) (MMM2). A study by Bonjour *et al* (2003) (BCHHS) used UK data on 187 MZ female twin pairs obtained from the records of a large London hospital. Although these were a self-selected group of women BCHHS show that they match population survey data in their observable characteristics. Finally, Isaccson (1999 and 2004) are the only studies that use a large dataset. His data is drawn from Swedish

[13] In each case the analysis is based just on those twin pairs who have complete information – in particular both twins had to be employed for a wage to be observed. These studies all use education cross reported by co-twin (or child in the case of Berhman *et* al 1994) as an instrument for own education. Taubman (1976) is an early example of twins research that does not instrument.

[14] This sample is re-examined in Flores-Lagunes and Light (2004) who were specifically concerned with the treatment of measurement error.

[15] See also Arias *et al.* (2001) who subsequently reanalysed this data.

[16] The 1988/9 data appears to be drawn from the 1980/82 sample and so the data are not independent observations.

registers of the population and yields 2492 and 2609 MZ, and 3368 and 3601 DZ, twin pairs in each study respectively. In the each case the data were twins born between 1926 and 1958 and earnings were observed around 1990[17].

These studies are not strictly comparable because of the construction of both the dependent variable and the explanatory variable of interest. The Australian research uses schooling imputed from grouped information and imputes annual earnings from detailed occupation information. It therefore estimates the effects of education differences on between-occupation wage differences and so underestimates the actual returns to the extent that education affects wages within an occupation. It seems likely that the grouping in the education data will give rise to greater measurement error than in the Twinsburg data. Moreover, there are also labour supply differences that drive occupational earnings differentials since different occupations have quite different distributions of annual hours of work. Like MMM1 and MMM2, BRT imputes earnings from detailed occupation information.

The Swedish research infers education duration from qualifications and uses annual earnings (averaged over up to three years) and drops very low earners but otherwise takes no account of labour supply differences. The UK research in BCHHS uses earnings adjusted for a time code (to convert to weekly) and then constructs an average hourly wage rate from weekly hours of work data, while the Twinsburg data uses the reported hourly wage and education is recorded in years. Therefore, BCHHS, and AK and AR probably come closest to overcoming concerns about within twin labour supply variation.

While it seems inappropriate to compare results across rows in Table 1, not least because they relate to four different countries, the methodology employed by each study has been very similar and this does facilitate comparisons across columns. In particular, the methodology has typically proceeded along the following lines. Log wages, w, and education, S, are assumed to be determined by

[17] In fact, Isacsson has a 3 wave panel, with each wave 3 years apart, but he collapses this to a cross section by averaging across waves. Duration of schooling was imputed from information on qualifications using an equation estimated from a 1991 sample survey which contained both qualifications and duration. This method is akin to complementary matching as used in Arellano and Meghir (1992). Rouse (1999) also averages wages across Twinsburg datasets for those observations that appear on more than one occasion.

$$(1) \qquad \begin{aligned} w &= \beta S + \alpha A + \varepsilon \\ S &= \gamma A + \zeta \end{aligned}$$

where A is "ability" , ε is uncorrelated with S and A, and ζ is uncorrelated with ε. That is, ζ and w are correlated only through their joint dependence on A. However, A is unobservable and so OLS estimates of β in $w = \beta S + \varepsilon$ will be biased such that

$$(2) \qquad \text{plim}\left(\beta_{OLS}\right) = \beta + \alpha \frac{\sigma_{AS}}{\sigma_S^2}$$

and if, as seems reasonable, γ>0 and α>0 then $\beta_{OLS} > \beta$. That is, OLS captures the effects of both S and of unobservables correlated with both S and w, such as A. But, if A is the same within MZ twin pair differencing the wages within pairs will result in the unobservable A being differenced out, and we are left with the within-twin pair equation

$$(3) \qquad \Delta w = \beta \Delta S + \Delta \varepsilon \, .$$

where Δ refers to the within twin pair difference. Applying OLS to this within twin pair equation yields $\beta_{WT} = \beta$ where β_{WT} is sometimes referred to as the covariance estimator - because it *is* the covariance between Δw and ΔS.

However, if S is measured with error such that $S = S^* + \nu$, where S^* is the true level of schooling, then (3) becomes

$$(4) \qquad \Delta w = \beta \Delta S - \beta \Delta \nu + \Delta \varepsilon \, .$$

Berhman *et al* (1994) show that the bias from applying OLS to this within twin regression is given by

$$(5) \qquad \text{plim}\left(\beta_{WT}\right) = \beta\left[1 - \sigma_{\Delta\nu}^2 / \sigma_{\Delta S}^2\right] = \beta\left[1 - \sigma_\nu^2 / \left(\sigma_S^2 \left(1-\rho\right)\right)\right]$$

where ρ is the within twin pair correlation between their reported schooling levels. Since this correlation seems likely to be positive the downward bias in β_{WT} is likely to be substantially worse than in β_{OLS} – differencing exacerbates the bias in OLS that is due to measurement error.

Table 1 Recent MZ Twins Estimates in the Literature

Study	Sample	Date	Country	Gender	# twin pairs	β_{OLS}	B_{WT}	β_{WTIV}
Ashenfelter and Krueger (1994)	Twinsburg	1991	US	Pooled	147	0.084 (0.014)	0.092 (0.024)	0.129 (0.030)
Berhman *et al* (1994)	NAS-NRC	1973	US	Pooled	141	0.094[a] (0.011)	0.035 (0.004)	0.101 (0.012)
Miller *et al* (1995)	Australian Twins Register	1985	Australia	Pooled	602	0.064 (0.002)	0.025 (0.005)	0.048 (0.010)
Ashenfelter and Rouse (1997)	Twinsburg	1991-93	US	Pooled	333	0.110 (0.009)	0.070 (0.019)	0.088 (0.025)
Berhman and Rosenzweig (1997)	Minnesota Twins Register	1993	US	Pooled	720	0.113[a] (0.005)	0.104 (0.017)	n.a.
Miller, Mulvey and Martin (1997)	Australian Twins Register	1985	Australia	Male	282	0.071[d] (0.003)	0.023 (0.008)	0.033 (0.014)
				Female	320	0.057[d] (0.002)	0.028 (0.006)	0.058 (0.011)
Rouse (1998)	Twinsburg	1991-93, 95	US	Pooled	453	0.105 (0.008)	0.075 (0.017)	0.110 (0.023)
Isacsson (1999)	Swedish Twin Registry	1990	Sweden	Pooled	2492	0.046 (0.001)	0.022 (0.002)	0.024[b] (0.008)
Isacsson (2004)	Swedish Twin Registry	1990	Sweden	Pooled	2609	0.066[c] (0.009)	0.028[c] (0.009)	0.052[c] (0.036)
Bonjour *et al* (2004)	St Thomas' Hospital twins register.	1999	UK	Female	187	0.077 (0.001)	0.039 (0.023)	0.077 (0.033)

Notes: Table 1 from Bound and Solon (1999) and Table 6 from Card (1999) updated. a – GLS estimate. b – not instrumented but evaluated at a reliability ratio of 0.88. c– evaluated at upper secondary level of schooling. d – pooled DZ and MZ.

Table 2 Recent DZ Twins Estimates in the Literature

Study	Sample	Date	Country	Gender	# twin pairs	β_{OLS}	B_{WT}	β_{WTIV}
Berhman *et al* (1994)	NAS-NRC + Minnesota	1973	US	Pooled		0.073 (0.003)	0.057 (0.005)	n.a.
Miller *et al* (1995)	Australian Twins Register	1985	Australia	Pooled	568	0.066 (0.002)	0.045 (0.005)	0.074 (0.008)
Berhman and Rosenzweig (1997)	NAS-NRC + Minnesota	1993	US	Pooled				
Miller, Mulvey and Martin (1997)	Australian Twins Register	1985	Australia	Male	164	0.071[d] (0.003)	0.029 (0.011)	0.051 (0.019)
				Female	158	0.057[d] (0.002)	0.049 (0.007)	0.071 (0.011)
Isacsson (1999)	Swedish Twin Registry	1990	Sweden	Pooled	3368	0.047 (0.001)	0.039 (0.002)	0.053[b] (0.006)
Isacsson (2004)	Swedish Twin Registry	1990	Sweden	Pooled	3601	0.066[c] (0.008)	0.047[c] (0.009)	0.056[c] (0.003)

Notes: Table 6 from Card (1999) updated. a – GLS estimate. b – not instrumented but evaluated at a reliability ratio of 0.88. c– evaluated at upper secondary level of schooling. d – pooled DZ and MZ.

Ashenfelter and Krueger (1993) correct for this measurement error that biases β_{WT} by instrumenting ΔS with the difference in the cross-reported level of S, $\Delta S'$, assuming that the measurement error is classical, i.e. $\Delta S' = \Delta S^* + \Delta \nu$, where S^* is the true education level, and that there is no family effect in the measurement errors. Providing $\Delta S'$ is a valid instrument and the measurement error is well behaved then the resulting estimate $\beta_{WTIV} = \beta$. However, if the measurement error is mean reverting then the classical properties will, in general, fail to hold and IV will not produce consistent estimates.

The presumption in the twins literature is that differencing eliminates bias due to unobserved ability but exacerbates measurement error, and that instrumenting the differenced schooling eliminates the resulting attenuation towards zero. AK further proposed a solution to the problem of correlated measurement errors which would otherwise lead β_{WTIV} to be biased. They suggest replacing ΔS by the schooling difference reported by one twin and instrumenting this with the schooling difference reported by the other, which effectively eliminates any measurement error that is common within twins. Subsequent studies have followed this lead.

The remaining weakness in the method is that differencing may not remove all of the ability bias if there is some *individual* component to MZ ability that is not removed by differencing. Indeed, since the bias is determined by the ratio of exogenous variation to total variation, BS note that differencing reduces the total variation and the ratio of exogenous variation in within twin schooling differences may fall or rise. Even in the absence of measurement error, this would imply that within twin estimates would suffer from ability bias which may be smaller or larger than the ability bias experienced in regular cross-section data. Neumark (1999) and Bound (1999) show that if differencing does not entirely remove ability then twins-based estimates of the return to education may be either more or less biased than OLS in cross-section data. Thus, proponents of the twins method have attempted to show that schooling differences between twins are uncorrelated with other observed differences. The papers based on the Twinsburg, UK and Swedish datasets all show no significant correlations between differences in education and differences in other observables. However, this is not an entirely convincing response to the criticism. An inability to find, in the limited data available, significant correlations between the within twin education differences and other within twin differenced variables does not

imply that there are none with respect to unobservable differences. Moreover, measurement error may imply that these correlations in the data are biased downwards.

Bound (1999) argues that since, in the ability bias framework, least squares provides an upper bound twins data is useful because it can tighten that bound.. However, typical IV estimates in non-twin studies largely rely on policy reforms that generate natural experiments that have, almost invariably, generated estimates returns that are *higher* than least squares. It seems likely that typically such estimates do not reflect the average return and may, therefore, be larger or smaller than the least squares estimate. Thus, twins provide us with an estimate of the average effect purged of endogeneity in a way that IV, in general, does not.

2.2 Rising returns

Much of the research on education returns in the US suggest that the strength of the correlation between wages and education has risen, perhaps by as much as 100%, since the 1980's. The residual wage variance is also thought to have risen, something that is often attributed to rising returns to unobserved skills such as cognitive ability (see Richard J. Murnane, John B. Willett, and Frank Levy (1995)). One interpretation of these facts is that the rise in the correlation between log wages and education is an increase in the causal effect of education, while the rise in residual variance reflects a corresponding change in the returns to unobserved skills. This interpretation favours the SBTC view of increased wage inequality. An alternative interpretation is that the correlation between log wages and schooling is partly spurious and the rising correlation is a reflection of an increase in the degree to which the correlation is spurious through increasing ability-education sorting across cohorts.

Resolving this requires estimates of how the causal effect of education has changed over time. Christopher Taber (2001) addresses this directly and finds that, although the estimated effect of education on log wages obtained from applying OLS rose from 5% to 10% over the 1980's, the estimated effect of a year of schooling using IV (and a selectivity model) remained close to 10% throughout the period. His further estimates of a dynamic selection model also suggested returns to observable education was essentially static. The implication of the causal effect of education

being broadly static is that the rising inequality over the period is largely due to rising returns to unobserved skills.

Olivier Deschênes (2003) addresses the problem indirectly by exploiting the fact that, in a model with heterogeneous returns, those with larger returns will choose more education so that the relationship between log wage and schooling ought to be convex. If the selectivity into schooling increases, so that the bias in OLS rises, then this would be reflected in a rising degree of convexity in the log wage schooling relationship. He estimates a quadratic relationship between log wages and schooling and shows that, controlling for cohort, the linear term has fallen over time and the quadratic term has risen. He then uses these estimates to infer that there has been a rise in the causal effect over the 1980's and 90's of around 30% as opposed to the 60% rise in the OLS coefficient on schooling in a log-linear specification. In contrast, the return to unobserved ability is estimated to have risen by about 10% over the 1980's and 1990's and so this can explain only explain about one fifth of the rise in wage inequality. Deschênes concludes that there is no support for the idea that ability bias in OLS estimates has risen over time. These inferences are supported by quantile regressions in Lemieux (2006a) using US CPS data which shows that returns to post-secondary education in the upper quartiles of the wage distribution has risen faster than the returns in lower quartiles.

2.3 Contribution

Our contribution to these two literatures is to exploit the size of of our twins dataset, and its long panel nature, to provide estimates of the causal effect of education, broken down by calendar time, to cast light on whether returns are rising. Indeed, the fact that our data is a panel allows us to estimate the returns to education at a point in time jointly with how that return changes over time[18]. Moreover, because we have data for both MZ and DZ pairs we can provide corresponding estimates from MZs that are, arguably, immune from selection issues, and so give an estimate of the causal effect; and from DZs, which will be contaminated by unobservable differences[19]. Thus, the difference between these two sets of estimates is informative

[18] The fact that it is a panel also frees us from any concern about changes in the composition of the workforce over time – something that had been shown to account for a high proportion of wage growth in the UK during the 1980's and 90's (see Richard Blundell, Howard Reed, and Thomas Stoker (2003)).

[19] See Lemieux (2006b) for evidence suggesting that the allegedly rising returns to unobservables is due to rising measurement error in CPS data.

about the effect of unobservables and it is possible to decompose the unexplained variance into changing returns to unobservable and changes in the variance of unobservables.

First, consider a model without measurement error in schooling. We can write the MZ levels equations as

$$w_{it} = \beta_t S_i + \alpha_t A_i + u_{it} \tag{6}$$

and so the corresponding within MZ twin difference at time *t* is $\Delta^{ij} w_t = \beta_t \Delta S_t + \Delta^{ij} u_t$ where the notation is $\Delta^{ij} w_t = w_{it} - w_{jt}$ and, by assumption, $\Delta^{ij} A_t = 0$. The corresponding equation for within DZ twin differences at time *t* is

$$\Delta^{ij} w_t = \beta_t \Delta^{ij} S_t + \left(\alpha_t \Delta^{ij} A_t + \Delta^{ij} u_t \right). \tag{7}$$

Note that the variance of the residual in the DZ differences differs from the variance of the residual in the MZ differences case by $\alpha_t^2 \operatorname{var}\left(\Delta^{ij} A_t\right) = 2\alpha_t^2 \left(\operatorname{var} A_{it} - \operatorname{cov} A_{it} A_{jt}\right)$. Note, however, that if ability is a fixed effect then it should be the case that $\left(\operatorname{var} A_{it} - \operatorname{cov} A_{it} A_{jt}\right)$ is a constant across *t* for any twin pair. Thus, estimates of the variances from the MZ and DZ differences equations across time allow us to identify how α has changed across time.

Moreover, with panel data we can time difference the within twin differenced equations to obtain an expression for the growth in wage differences for DZ's as

$$\Delta^t \Delta^{ij} w_t = \Delta^t \beta_t \Delta^{ij} S_t + \left(\Delta^t \alpha_t \Delta^{ij} A_t + \Delta^t \Delta^{ij} u_t \right) \tag{8}$$

and a corresponding expression for the MZ wage difference growth $\Delta^t \Delta^{ij} w_t = \Delta^t \beta_t \Delta^{ij} S_t + \Delta^t \Delta^{ij} u_t$. Estimation of (8) allows us to recover $\Delta^t \beta_t$ and, indeed, estimation of this wage difference growth equation jointly with the wage difference level provides cross equation restrictions that should improve the precision of our estimates. Thus, by exploiting both the panel nature of our data and the fact that we can identify zygosity we can recover both α and β for each year of our panel. Estimates of α and β, together with least squares estimates, allow us to use (2) to recover σ_{AS} / σ_S^2 and since we can observe the variance of *S* we can compute the covariance between schooling and ability. Finally, our panel is not a single cohort of

twins. Rather the twins vary in age. Thus, we can parameterise the model to allow us to recover differential changes in α and β by cohort.

Of course, measurement error is also an issue for us. Since we use administrative data this is likely to be less prone to depart from classical properties. Bronars and Grogger (2006) investigated the validity of cross-reported schooling amongst multiple sibships in NLSY79. Multiple sibling reports of education for the same individual were used to test, and reject, the overidentifying restrictions on the validity of sibling-reported schooling as an instrumental variable. Isacsson (2004) use a large sample of Swedish MZ twins found that estimates of the average return to years of schooling that relied on a classical measurement error assumption were biased upwards by 30%. However, Lagunes-Flores and Light (2005) show, in a re-analysis of the Twinsburg data, that the measurement error is close to being classical so that IV, at least in that data, is an appropriate estimation method.

We provide two alterative instruments that seem likely to be less problematic than one based on cross reported schooling. Our first proposed instrument is the difference in the education of the twins' spouses and we use this to demonstrate the fragility of the view that there are no unobservable differences between twins[20]. Our second instrument is the within difference in the age when full-time work started.

Finally, we deal with the possibility that differencing in the MZ's does not entirely eliminate the unobserved *A* using an IV approach. Our IV to deal with this issue is based on differences in the education experienced by the twins who were taught in separate classes at school.

3. Data Description

Attendance at primary and lower secondary school (grades 1-9, corresponding roughly to ages 7-15) is compulsory in Denmark. Schooling is a requirement from 1 August in the year that the child turns seven years old until 31 July in the year which regular instruction has been received for 9 years. Grade retention is rarely practiced. During the period 1981-1990 analysed in this paper, 89% of children attended public (i.e. state funded) schools. These 1826 (in 1990) schools are run by 275 municipalities, and are attended by an average of 309 students. Municipalities have a

[20] AR found no such correlation between education and spouse's education in the Twinsburg data – albeit for a small sample of just 91 twin pairs.

mean population of 36,094 residents, but this ranges from 2,512 to 466,723 (Copenhagen), and the number of schools per municipality ranges from 1 to 76 accordingly. Public school expenditure is financed through municipal income tax, together with a complex between-municipality redistribution scheme, which subsidises expenditures in low income municipalities. Average total expenditure per student per year was DKK 31,360 in 1990 (corresponding to €4,248 in 2005 prices), having risen steadily from DKK 18,447 in 1981 (€3,713 in 2005 prices). The total number of students fell consistently throughout the period, from 728,900 in 1981 to 559,600 in 1990 due to smaller birth cohorts. The net effect was a reduction in expenditure on public schools between 1981 and 1990 from €2.629 billion to €2.365 billion (2005 prices).[21]

The dataset is derived from merging data from several administrative databases containing individual information for *all* residents of Denmark via the Central Person Register. The Central Person Register (CPR) is a national administrative database, started in 1968, that contains social security (i.e. CPR) numbers that are allocated at birth. The census in 1970 enables links between all children and their legal (and biological) mother and legal father to be established[22] and so allows us to match siblings, and the date of birth allows us to identify twins (indeed, all multiple births). This enables us to identify relationships for 99% of births from 1956 onwards. We call this our child-centric database. The zygosity information is contained in responses to a special questionnaire, sent to all twins in Denmark, to four questions (including "two peas in a pod")[23]. The CPR enables us to match in any available administrative database – called registers - to our child-centric database and to the twins subset in particular[24].

[21] See Danish Ministry of Education and Research (1993) for further details.

[22] We use only observations for individuals who were born in Denmark and are resident in Denmark at the time of observation.

[23] Christiansen *et al* (2003) checked self-reported zygosity from these questions against DNA tests for 873 Danish twin pairs and in 96% of twin pair cases the self-reported zygosity is confirmed. We drop triplets and above because our data does not classify their zygosity.

[24] Our database contains not only all twins, but also a 5% random sample of all children from each year cohort matched to their parents and siblings, plus a randomly chosen child from the same cohort who attended the same school at 8th grade with their parents and siblings, plus the geographically nearest child from the same cohort based on distance between residences when the child was at 8th grade with their parents and siblings. The registers that we have matched in to date contain information on incomes, hours and education. Here we use the data on twins alone.

The Danish Twins Registry has played a key role in the development and maintenance of the twins data (see Harvald *et al* (2004) for background to the register and see, Christiansen (2003) for details of zygocity). We select all MZ twins and same gender DZ twins in the age range 25-55 inclusive, to avoid sample selection due to education and retirement decisions. These need to be observed at some point 1980-2002. This is the longest period over which consistent labour income and hours of work information is presently available to us.

Throughout we use annual real log gross income from work because our data on hours of work information is grouped[25]. The labour earnings data itself is taken from tax returns and tax filing is compulsory. We have restricted our sample to twin pairs where BOTH are observed to be full-time full-year workers (annual hours at least 60% of full year hours) to reduce the impact of labour supply variation on our estimates[26]. The education measure available to us is a standard Statistics Denmark construction based upon highest qualification from which length is imputed by "typical completion times" compiled by the Education Ministry. We also restrict attention to those twin pairs who were *both* observed to be in married or cohabiting partnerships at some point in 1980-2002, and for whom we observe education of the partner, because we are going to use co-twin's partner's education as our instrument for own education, and within-pair differences in partner's education as our instrument for education difference[27].

Most of our attention will be directed towards our MZ twins but we will also consider same sex DZ twins. After allowing for all of the selection criteria above, we have approximately 107 thousand such twin-pair*year observations over an unbalanced 23 year sample where we have complete information. Approximately 40% of these are MZ, and approximately 40% are female (because of their lower labour

[25] Hours of work are derived from mandatory pension contributions which are a step function of hours worked (on a weekly basis the steps are 10-19, 20-29, 30+). The hours information that we have access to is a function of the sum of these contributions over the calendar year.

[26] The joint full-time labour force participation rates (both members of the pair working full-time in the same year) for MZ (DZ) pair*years are 58.1% (54.7%) for females and 72.6% (67.9%) for males. Headline estimates using samples that impute earnings, from a regression of wages on gender interacted with a quadratic in age and year dummies, are contained in the Appendix and are very close to the ones reported below. The pattern of the more detailed results for these larger samples is very similar and are available from the authors on request.

[27] 82% of female pairs and 76% of male pairs meet this selection criteria.

market participation rate). There are 2185 (3534) MZ (DZ) male pairs, and 2000 (2809) female MZ (DZ) pairs – approximately the same number of MZ pairs as in all previous existing datasets put together.

Table 3 shows the basic descriptive statistics for individuals. MZ and same sex DZ individuals are quite similar except for age which is accounted for by the recent growth in the number of multiple births associated with fertility treatment which are inevitably DZ. There are clearly smaller absolute differences in education and earnings, on average between MZ twins compared to DZ twins. Table 4 shows the frequency of education differences: a much higher proportion of MZs report exactly the same education length. Table 5 reports the correlation between: each twins education level and the co-twins education, the own partner's education level, and the co-twins partner's education level. The previous literature has been concerned to show that differences in schooling are random. If it were the case that within MZ twin differences in schooling were correlated with other choice variables then this would undermine the case for thinking that differencing removes the endogeneity bias. Both AK and AR show that the Twinsburg data appears to exhibit no correlation between the within MZ twins schooling difference and the difference in their partner's schooling and other variables. Similar results are reported in the Swedish and British datasets. However, failure to find evidence of correlations in observables is not the same as success in finding lack of correlation with unobservables.

Table 3 *Descriptive statistics in levels and differences for the twin samples Means (standard deviations)*

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
Levels				
Years of education	12.62 (2.79)	12.49 (2.99)	12.50 (2.65)	12.16 (2.75)
Log real annual earnings	12.68 (0.33)	12.68 (0.34)	12.39 (0.34)	12.36 (0.33)
Age	38.37 (8.29)	39.34 (8.18)	37.69 (8.22)	38.85 (8.16)
Twin-pair differences				
Years of education	1.37 (2.01)	2.04 (2.32)	1.32 (1.77)	1.97 (2.01)
Log real annual earnings	0.22 (0.25)	0.29 (0.30)	0.25 (0.27)	0.29 (0.28)
# person-year obs.	24,370	38,929	17,608	25,817
# twin pairs	2,185	3,534	2,000	2,809

Table 6 shows the correlation between the within twin differences in education and the within twin differences in partners education and in the within twin differences in other variables. It is this correlation that we exploit to instrument for measurement error in our within-twin estimation.

Table 4 Distribution of education differences
% of sample in category

	MEN		WOMEN	
Absolute difference	MZ	DZ	MZ	DZ
Same duration	43.87	28.18	48.36	31.69
1-12 months	13.30	13.18	4.22	4.15
13-24 months	16.54	18.68	16.65	17.37
25-36 months	5.43	8.96	7.76	11.47
37-48 months	6.20	7.36	10.57	15.66
Above 49 months	14.66	23.64	12.44	19.66

Table 5 Education correlations within twin pairs and spouses

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
T_A, T_B	0.620	0.444	0.651	0.476
T_A, S_A	0.394	0.390	0.427	0.360
T_A, S_B	0.348	0.312	0.350	0.269
S_A, S_B	0.359	0.283	0.247	0.238
(T_A-T_B), (S_A-S_B)	0.094	0.123	0.150	0.143

Note: T_A refers to twin A, T_B refers to twin B, S_A refers to this spouse of twin A, S_B refers to the spouse of twin B. Table entries are pairwise education correlations.

4. Estimation

4.1 Least Squares and Instrumental Variable Estimates

For comparison with other studies that are not twin-based, this section estimates the returns to education using the twins as individuals, by gender, using a minimal specification that includes only education, age, age squared and region of residence controls. The coefficients on education years from this exercise are reported below in Table 7. To maintain comparability with other research we stack this data for individuals whenever earnings are observed and we correct the standard errors for repeated observations accordingly. These OLS results reflect earlier Danish OLS results in Christensen and Westergaard-Nielsen (2001) which suggest estimated returns of the order of 4%. The OLS DZ and MZ results are identical for women, and significantly different (by about 10%) for men.

We also provide instrumental variable estimates. The existing IV literature that uses cross-section data, instruments to control for bias induced either through measurement error or through self-selection. The use of cross-reported twins education in the twins literature is explicitly aimed at eliminating measurement error and is unlikely to be able to help explain why one twin selects more schooling than the other. Here we exploit the strong assortative mating that Table 5 suggested is in the data and so we use co-twin partner's education as our instrument. Our instrument could plausibly be thought of as dealing with both sources of bias so we also report IV estimates using the individual data.

Despite the attention given to ability bias and the implication that OLS is biased upwards much, if not most, of the existing IV literature suggests that OLS is biased downwards rather than upwards (see Card (1999)). Common assumptions about the size of the reliability ratio for schooling data would suggest that OLS would be biased downwards by something of the order of 10% and yet IV typically exceeds OLS by much more than this, which denies the conventional story about the direction of ability bias. Card (1999) has presented a more sophisticated argument than the traditional ability bias story – that ability is multidimensional and individuals make their choice of education on the basis of comparative advantage. In this case, the endogeneity arising from self-selection into education can imply that OLS is biased

Table 7 *OLS and IV education returns treating twins as individuals*
Model estimates (standard errors)

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
β_{OLS}	0.0298	0.0331	0.0373	0.0370
	(0.0005)	(0.0004)	(0.0006)	(0.0005)
	Pooled MZ & DZ		Pooled MZ & DZ	
β_{OLS}	0.0317		0.0373	
	(0.0003)		(0.0003)	
β_{IV}	0.0652		0.0542	
	(0.0011)		(0.0014)	
Partial R^2	0.1061		0.0910	
F-test	107.49		83.61	
Hausman t-test	18.75		16.15	
# person-year obs.	?		?	

Note: Specifications also include a constant, age, age squared and regional fixed effects. Instrument is the education level of the co-twins's partner. Standard errors in parentheses are corrected for clustering by year and family. Results using imputed annualised earnings for those years where one twin is working less than full-time are in the Appendix..

either up or down. In the light of this, correcting for endogeneity could easily imply larger IV estimates than OLS. Indeed, in Table 7 this is precisely what we find[28].

The IV literature has been subject to important criticism by Bound, Jaeger and Baker (1995) and Staiger and Stock (1997) for using instruments that are only weakly correlated with the endogenous variable of interest – something that they show leads to IV being more biased than OLS in finite samples. In Table 8 we therefore present Hausman t-tests of the null that OLS is consistent. For both males and females we strongly reject the null of exogenous education. These authors recommend that an F-test and partial R-squares from including instruments in the first stage be computed and that an F below 10 would be cause for concern. These are also reported in Table 8 and we find F statistics that far outside the danger zone and partial R^2 that seem large in comparison with their analyses. The IV results exceed OLS results – around double for males and around 45% higher for females. These increases are much larger than measurement error alone would imply – especially because it seems likely that our data would be more reliable than conventional self-reported data. Thus, it is tempting to conclude that the estimated return to education is substantially higher than OLS suggests – in line with other findings in the literature. However, OLS estimates the average treatment effect while IV estimates a LATE (see Angrist and Imbens (1994)) and it is unclear why these would be the same in the context of this instrument.

Table 8 Fixed Effect Estimated Education Returns on Samples of Twin Pairs

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
β_{OLS}	0.0049	0.0183	0.0093	0.0250
	(0.0009)	(0.0007)	(0.0012)	(0.0009)
β_{IV}	0.0451	0.0946	0.0441	0.0535
	(0.0099)	(0.0063)	(0.0083)	(0.0061)
Partial R^2	0.0036	0.0176	0.0065	0.0130
F-test	4.86	13.45	7.84	12.37
Hausman t-test	4.24	4.10	4.31	4.81
# person-year obs.	24,370	38,929	17,608	25,817

Note: Specifications also include a constant, age, age squared and regional fixed effects. Results using imputed earnings for those years where one twin is working less than full-time are in the Appendix.

[28] Bingley and Rasmussen (2005) report IV estimates based on proximity of high school of 4.7%.

4.2 Fixed Effects

Table 9 uses the same data and takes sibling differences to estimate, using least squares, the within differenced equation. Again we stack the longitudinal data and correct standard errors for repeated observations. The OLS column makes no attempt to deal with measurement error and the estimates are considerably lower than the OLS levels estimates in Table 8 - a finding that is consistent with there being large measurement error bias in the differenced education. The IV column instruments the reported schooling difference by the difference in the twins partners' education levels. The Hausman test of exogeneity again rejects, albeit not as strongly as when the data is used as a cross-section of individuals. According to the F and partial R^2 the instrument remains valid even in its differenced form.

The conventional wisdom asserts that the DZ data fails to fully control for ability differences and, assuming that ability bias is positive, are therefore biased upwards. The MZ IV results, which we assume is such that differencing together with instrumenting has removed all ability bias and any measurement error bias, are now significantly lower than the IV results in Table 8. The MZ IV results are now 40% higher for males and 20% for females than the OLS estimates. OLS DZ returns are higher than the OLS for MZ because differencing does not remove all ability bias differences in DZs. However, they are still higher than the OLS results in levels.

4.2 Rising Returns

In this section we extend the benchmark FEIV model to allow for: variation in returns across calendar time and its separation into returns to observable and unobservable skill.

Juhn *et* al (1993) raised the question of how returns to observable and unobserved skills have been changing over time. Card and Lemieux (2001) use the US CPS 1974 to 1996 and microdata from Great Britain and Canada, and argue that the increase in the US college premium that occurred over the 1980's largely reflects a rising return to unobserved skill rather than a rise in the return to observed education. Deschenes (2003) used cohorts formed from successive CPS cross sections and found little evidence that the rise in the conventional measure of the return to education is due to variation in the extent of unobserved ability bias over time. He

argued that there was some rise in the causal effect. Taber (2001) uses IV and finds no change in the causal effect.

The virtue of MZ twin estimates is that they are purged of unobservables and therefore reflect only the return to observable skills. In contrast DZ estimates reflect returns to both observable and unobservable skills. Therefore the *difference* between MZ and DZ estimates is informative about the returns to unobservables. To address this issue we constructed a ten year window and rotated this through our 23 year observation period. In Table 10 our DZ estimates show returns that rise strongly over the 1980's and the 1990's in Denmark, faster for women than men. In contrast the MZ data, which we would argue are purged of any unobservable ability, show returns that are static in the first third of the period and then rise quickly over the late 1980's and through the 1990's.

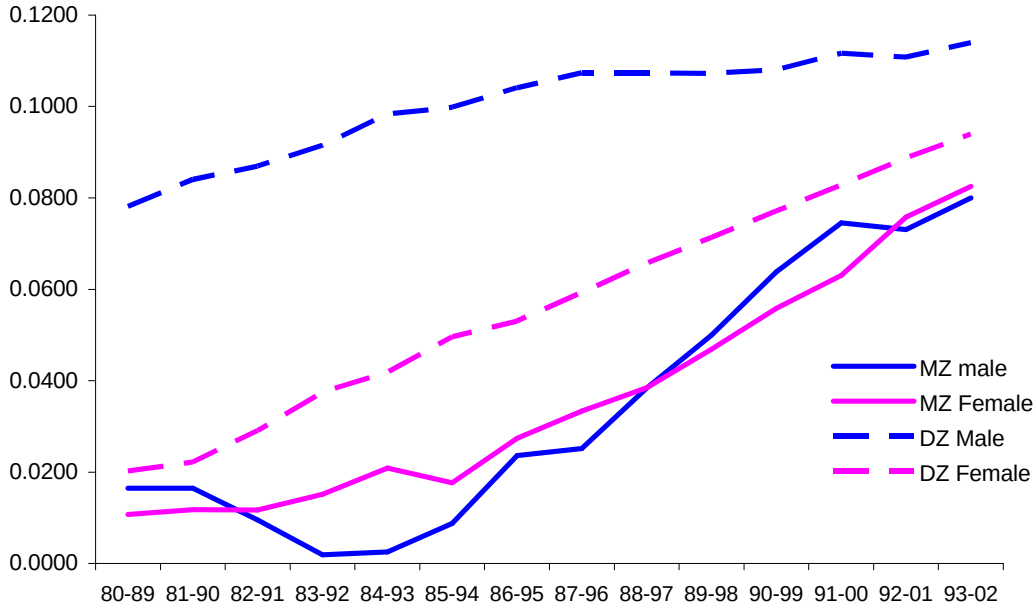
Table 10 *Fixed Effect IV Estimated Education Returns over Time*

Sample:	MZ				DZ			
	Male		Female		Male		Female	
Year range:	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)
80-89	0.0165	*0.0129*	0.0107	*0.0107*	0.0781	*0.0080*	0.0203	*0.0078*
81-90	0.0165	*0.0135*	0.0118	*0.0103*	0.0840	*0.0085*	0.0222	*0.0075*
82-91	0.0096	*0.0138*	0.0117	*0.0101*	0.0869	*0.0088*	0.0291	*0.0074*
83-92	0.0019	*0.0146*	0.0152	*0.0094*	0.0915	*0.0091*	0.0374	*0.0074*
84-93	0.0025	*0.0148*	0.0209	*0.0098*	0.0983	*0.0096*	0.0419	*0.0079*
85-94	0.0087	*0.0147*	0.0177	*0.0102*	0.0998	*0.0099*	0.0496	*0.0082*
86-95	0.0236	*0.0158*	0.0274	*0.0110*	0.1041	*0.0102*	0.0530	*0.0086*
87-96	0.0252	*0.0159*	0.0334	*0.0110*	0.1074	*0.0102*	0.0592	*0.0090*
88-97	0.0384	*0.0158*	0.0384	*0.0114*	0.1073	*0.0101*	0.0657	*0.0091*
89-98	0.0499	*0.0162*	0.0468	*0.0118*	0.1072	*0.0101*	0.0713	*0.0093*
90-99	0.0638	*0.0164*	0.0558	*0.0123*	0.1080	*0.0101*	0.0772	*0.0094*
91-00	0.0745	*0.0162*	0.0630	*0.0128*	0.1116	*0.0104*	0.0828	*0.0098*
92-01	0.0731	*0.0158*	0.0758	*0.0137*	0.1108	*0.0102*	0.0886	*0.0104*
93-02	0.0800	*0.0167*	0.0825	*0.0151*	0.1140	*0.0109*	0.0939	*0.0114*
All years	0.0451	*0.0099*	0.0441	*0.0083*	0.0946	*0.0063*	0.0535	*0.0061*

Note: Specifications also include a constant, age, age squared and regional fixed effects.

The results, graphed in Figure 4, strongly suggest that there were indeed slowly increasing returns to *unobservable* skills (the gap between the dashed and solid lines) in the 1980's, but that by the late 1980's the returns to unobservable skills seems to be falling for men and are static for women.

Figure 4 MZ and DZ returns over calendar time: 10 year averages



5. Conclusions and Further Research

In this study we present estimates of the returns to schooling based on a large sample of Danish twins. The data is drawn from population administrative registers and form a 23 year long unbalanced panel of more than eight thousand MZ twins and more than twelve thousand DZ twins for whom we have the requisite information, about 40% of whom are women. We present baseline estimates that suggest that OLS on cross section data is biased downwards, for both MZ and DZ, because of measurement error. Predictably, we find that the simple fixed effects estimators are biased downwards to a much larger degree. When we instrument the FE estimators to counter measurement error we find estimates that are fully 50% higher than OLS in the MZ twins. In the DZ twins case we find that the FEIV estimates are higher than OLS for DZs - in the case of males almost treble, and close to 50% in the case of females. This is consistent with there being some remaining ability bias in the DZ case. Thus, our modelling resembles the previous earlier US research by AK that found FEIV MZ estimates that were larger than the corresponding cross section OLS.

We find very important differences by level of education of the twins. Indeed, our estimates suggest that, on average across the period, there is no return to school level education, only to college. We exploit the panel nature of the data to show that, over time, there is some evidence that returns to tertiary and higher secondary education may have risen.

Finally, by contrasting MZ and DZ results we conclude that, since the mid 1980's, although the return to observed skills have been rising, the returns to unobserved skills appear appear to have been falling.

References

Acemoglu, Darren (2002), "Technical Change, Inequality and the Labor Market", *Journal of Economic Literature*, 40, 7-72.

Angrist, Joshua D. and Guido Imbens (1994) "Identification and estimation of local average treatment effects*", Econometrica*, 62, 467-75.

Angrist, Joshua D. and Victor Lavy (1999), "Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement", *Quarterly Journal of Economics*, 114, 533-575

Arellano, Manuel and Costas Meghir (1992), "Female labour supply and on-the-job search: an empirical model estimated using complementary data sets'", *Review of Economic Studies*, 59, 537-559.

Arias, Omar, Kevin Hallock, and Walter Sosa-Escudero (2001), "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data", *Empirical Economics*, 26, 7-40.

Ashenfelter, Orley and Alan Krueger (1994), "Estimates of the economic return to schooling from a new sample of twins", *American Economic Review*, 84, 1157–1174.

Ashenfelter, Orley and Cecilia Rouse (1998), "Income, schooling and ability: Evidence from a new sample of identical twins", *Quarterly Journal of Economics*, 113, 253–284.

Behrman, Jere and Mark Rosenzweig (2004) "The returns to birthweight", *Review of Economics and Statistics*, 86, 586-601.

Behrman, Jere, Mark Rosenzweig and Paul Taubman (1994), "Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment", *Journal of Political Economy* 102, 1131–1174

Black, Sandy, Paul Devereux and Kjell Salvanes (2005), "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes", IZA Discussion Paper 1864.

Bingley, Paul and Annedde Rasmussen (2005), "Intergenerational transmission of human capital: Reform based evidence from Denmark", mimeo, Arhus University.

Bingley, Paul, Vibeke Myrup Jensen, and Ian Walker (2005), "The effects of school resources on participation in post-compulsory education in Denmark", mimeo, Arhus University.

Blundell, Richard, Howard Reed, and Thomas Stoker (2003), "Interpreting Aggregate Wage Growth: The Role of Labor Market Participation," *American Economic Review*, 93, 1114-1131.

Bonjour, Dorothe, Lynn Cherkas, Jonathon Haskel, Denise Hawkes and Tim Spector (2003), "Returns to education: Evidence from UK twins", *American Economic Review*, 93, 1799-1812.

Bound, John, David A. Jaeger, and Regina M. Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak", *Journal of the American Statistical Associatioon*, 90, 443-450.

Bound, John and Gary Solon (1999), "Double trouble: on the value of twin-based estimation of the return to schooling", *Economics of Education Review*, 18, 169-182.

Bronars, Steven G. and Gary S. Oettinger (2006), "Estimates of the return to schooling and ability: evidence from sibling data", *Labour Economics*, 13, 19-34

Card, David (1999), "The causal effect of education on wages", in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Volume 3, Elsevier, Amsterdam.

Card, David (2001), "Estimating the return to schooling: Progress on some persistent econometric problems", *Econometrica*, 69, 1127-1160.

Card, David (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.

Card, David and Thomas Lemieux (2001), "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis." *Quarterly Journal of Economics*, 116, 705-746.

Carneiro, Pedro. and Sokbae Lee (2005), "Ability, Sorting and Wage Inequality", mimeo, UCL.

Carneiro, Pedro, James J. Heckman and Edward Vytlacil, (2003), "Estimating the Return to Schooling When It Varies Among Individuals," mimeo.

Christensen, J. and Niels Westergard-Nielsen (2000), "Denmark", in Colm Harmon, Ian Walker, and Niels Westergard-Nielsen (eds.) *Education and Earnings in Europe: A Cross Country Analysis of Returns to Education*, Edward Elgar, London.

Christiansen, L, H. Fredericksen, K. Schousboue, A. Skytthe, N. von Wurmb-Schwark, Kaare Christensen, and K. Kyvik (2003), "Age and sex-based differences of questionnaire-based zygosity in twins", *Twin Research*, 6, 275-278.

Currie, Janet and Duncan Thomas (1995), "Does Head Start Make a Difference?", *American Economic Review*, 83, 341-364.

Danish Ministry of Education and Research (1993), *Facts and Figures* (in Danish), Danish Ministry of Education and Research: Copenhagen.

Danish Ministry of Education (2002), *Education key facts and figures for 2002*, Danish Ministry of Education: Copenhagen.

Deschênes, Olivier (2003), "Unobserved ability, comparative advantage, and the rising return to education in the United States, 1979-2002", *mimeo*, University of California, Santa Barbara.

Flores-Lagunes, Alfonso and Audrey Light (2004), "Measurement error in schooling: Evidence from samples of siblings and identical twins", mimeo, University of Arizona. http://uaeller.eller.arizona.edu/~alfonso/MESchool.pdf .

Griliches, Zvi (1979), "Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87, S37–S64.

Harvald B, G., K. Hauge, K. Kyvik, K. Christensen, A. Skytthe, and N. Holm (2004), "The Danish Twin Registry: Past and present", *Twin Research*, 7, 318-35.

Heckman, James J. (1974), "Lifecycle consumption and labor supply: An explanation of the relationship between income and consumption over the lifecycle, *American Economic Review*, 64, 188-94

Isacsson, Gunnar (1999), "Estimates of the return to schooling in Sweden from a large sample of twins", *Labour Economics*, 6, 471-489.

Isacsson, Gunnar (2004), "Estimating the economic returns to educational levels using data on twins", *Journal of Applied Econometrics*, 19, 99-119.

Juhn, Chinhui., Kevin M. Murphy and Brooks Pierce (1993), "Wage inequality and the rising returns to skill", *Journal of Political Economy*, 101, 410-442.

Katz, Lawrence and Kevin M. Murphy, (1992), "Changes in relative wages, 1963-1987: Supply and demand factors", *Quarterly Journal of Economics*, 107, 35-78.

Lemieux, Thomas (2006a), "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?", forthcoming, *American Economic Review*.

Lemieux, Thomas (2006b), "Post-secondary education and increasing wage inequality", mimeo, UBC.

Miller, Paul, Charles Mulvey and Nick Martin (1995), "What do twins studies tell us about the economic returns to education? A comparison of US and Australian findings", *American Economic Review*, 85, 586–599

Miller, Paul, Charles Mulvey and Nick Martin (1997), "Family characteristics and the returns to schooling: Evidence on gender differences from a sample of Australian twins", *Economica*, 64, 137-154.

Murnane, Richard J., John B. Willett, Frank Levy (1995), "The Growing Importance of Cognitive Skills in Wage Determination", *The Review of Economics and Statistics*, 77, 251-66.

Neumark, David (1999), "Biases in twin estimates of the return to schooling", *Economics of Education Review*, 18, 143-148.

Rouse, Cecilia (1998), "Further estimates of the return to schooling from a new sample of twins", *Economics of Education Review*, 18, 149-157.

Staiger, Douglas and James H. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557-86.

Taber, Christopher R. (2001), "The rising college premium in the eighties: Return to college or return to unobserved ability?", *Review of Economic Studies*, 68, 665-691.

Appendix

Table A1 Benchmark estimates using sample with wage imputation

Sample	Female		Male		Female		Male	
Estimator	MZ		MZ		DZ		DZ	
OLS	0.0363	*0.0005*	0.0320	*0.0004*	0.0376	*0.0004*	0.0365	*0.0003*
IV	0.0675	*0.0019*	0.0765	*0.0016*	0.0561	*0.0016*	0.0620	*0.0013*
FE OLS	0.0075	*0.0010*	0.0078	*0.0008*	0.0258	*0.0007*	0.0215	*0.0006*
FE IV	0.0251	*0.0078*	0.0359	*0.0075*	0.0649	*0.0047*	0.1095	*0.0055*
N*years	60588		67164		94332		114602	

Note: Standard errors in parentheses adjusted for clustering by family and year.