Twin-based Estimates of the Returns to Education: Evidence from the Population of Danish Twins*

Paul Bingley

Department of Economics, Aarhus Business School, 8000 Århus, Denmark

Kaare Christensen

Department of Epidemiology, University of Southern Denmark, 8000 Odense, Denmark

and

Ian Walker

Department of Economics, University of Warwick, Coventry CV4 7AL, UK
Institute for Fiscal Studies, 7 Ridgmount St., London WC1E 7AE, UK

Wednesday, 14 September 2005, Version 2.01

Keywords: schooling, twins, measurement error
JEL Codes: I3

Abstract:

This paper provides estimates of private financial returns to education based on large panels of monozygotic and dizygotic twins which we obtain from Danish population registers. Measurement error has been a concern in the literature since the usual methodology is based on within-twin differences. Our data is from administrative registers and so any measurement error is likely to be classical in nature which allows us to use instrumental variable estimation to eliminate it. We exploit the strong assortative mating in the data to provide us with an instrument. Our baseline estimates suggests that correcting for self-selection and measurement error gives estimated returns that are about two fifths higher than OLS for men and about one fifth higher for women. Further estimation shows that the rising returns which we observe in the raw data are due to strongly rising returns to observable skill, and that the returns to unobservable skills appear to be falling. We also find strong nonlinearities with no effect of schooling *per se* but large returns to college. We allow for complementarity of parental education and account for differences in returns over the ability distribution.

* 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) and the Danish Social Science Research Council (24-04-0240). The usual disclaimer applies.

Corresponding author: Paul Bingley, Department of Economics, Aarhus Business School, 8000 Århus, Denmark
Tel. +45 8948 6400, Email pbi@asb.dk

1. Introduction

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)). An important issue has been endogeneity of schooling that causes bias 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".

Identical twins have been used to control for ability bias in estimating the causal effect of schooling on earnings in around a dozen studies to date. "Ability" here is used to denote any unobserved attributes that are specific to an individual, fixed over time and associated with productivity in the labour market. This covers a multitude of unmeasured earnings endowments, 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. The extent of ability bias can, under certain assumptions, be inferred from comparing the schooling coefficient estimate using data on the twins as individuals (or, indeed, any sample of unrelated individuals) with estimates based on *within* identical (monozygotic, MZ) pairs of twins. MZ twins are particularly valued by researchers because they have the same DNA and are therefore genetically identical[1]. It is the prospect of estimation based on differences between MZ twin pairs eliminating the influence of unobserved endowments (or at least for genetic endowments at the time of conception) that makes twins attractive for researchers. On the other hand Griliches (1979) cautions that twins are "not a panacea".

There are three main issues that make twins problematic[2] for estimating the causal effects of education on wages: small samples limit what can reasonably be estimated so that restrictive functional forms have to be imposed on the data; measurement error may be large and is an issue even if strong valid instruments are available because the error may not be classical in nature which undermines the validity of IV estimation; and the within-twin pair differences in education may be correlated with the error term in the within-twin pair wage difference equation – that

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

[2] See Bound and Solon (1999).

is, the wage difference equation also suffers from some endogeneity because differencing has not removed all of the ability bias. This paper uses particularly informative data to address the first two issues to provide less problematic twins-based estimates. We reserve the last issue for further research.

The first issue is that almost all previous studies are based on small samples and this has limited what could reasonably be estimated. In particular most estimates allow for little or no heterogeneity in returns – the return to a year of education is assumed to be one number (or one per gender in some cases). Only three studies have allowed for nonlinear returns. And only one study has allowed for heterogeneity in returns with respect to other observable characteristics. This paper provides estimates for a sample of MZ twin pairs which provides us with a dataset that is larger even than the sum of all datasets used in all other twins research on the topic to date. Furthermore, we provide estimates for a corresponding DZ sample, which can be thought of as an "immune group" to the MZ experiment. This is in the sense that DZ pairs have 50% of their DNA in common, and differencing within-DZ-pair should remove some, but not all, ability differences. Our large dataset allows us to provide estimates that allow for observable differences not only by gender but also by parental background. Moreover, even though the differences in schooling length within twin pairs are quite small, we can, by exploiting the fact that the twin average of education differs considerably between pairs, allow for nonlinearities in returns. Finally, our dataset is an unbalanced panel of up to 23 years duration. We exploit this by way of a mixed model to allow for unobserved heterogeneity in returns which we can decompose into individual heterogeneity, family heterogeneity, and a residual variance which can thought of as luck or risk. We also exploit the time-span of the data to allow for changes in returns over time which, by comparing MZ and DZ twins, we can use to show the extent to which returns to observable skills have changed over time compared to the returns to unobservable skills.

The second important criticism relates to measurement error. Griliches (1979) (and elsewhere) notes that the use of estimates obtained from differencing in general, and differencing between MZ twins in particular, is that the method exacerbates the extent of measurement 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 classic solution to a pure measurement error problem is to use a second measure of the variable that is measured with error. Providing the measurement error in the two measures is uncorrelated the second measure can be an instrument for the first. Therefore, it was an important innovation proposed by Ashenfelter and Krueger (1994) (AK) who used instrumental variables to eliminate this measurement error bias in twin differenced data. 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[3]. Our administrative register data has the advantage over own reported survey data that education is the official record of the individual's activity. However, administrative records do not contain the cross-reported differences in schooling that has been used as an instrument in much recent work. Rather, we instrument education with co-twin's partners' education (we include cohabitees as well as spouses) which allows us to take advantage of the strong assortative mating in the data[4]. Although assortative mating was not a feature of the US data used in several important previous studies, or of the Swedish or UK data, it has been noted in earlier US data and is clearly present in the Danish data.

In the Danish case 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, and the data we have access to is the maximum of the completion times associated with these qualifications. Consequently, there is no issue of recall or misrepresentation and rounding errors are likely to be very small. However, there is still the possibility that administrative mistakes occur, or of systematic differences in reporting practices between institutions, and this cannot be ignored. There are two advantages to knowing that measurement error comes just from institutional mis-recording or mis-reporting (and a small amount of rounding): this error is likely to be classical in nature and not mean-reverting; and the variance in the measurement error should be the same for MZs as DZs. The first property allows us to use instrumental variables to correct for measurement error. This second property is useful since it allows us to assume that the extent of bias induced by measurement error is the same for MZs as DZs.

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

[4] Using own partner's education would be invalid if, as seems likely, it affects own wages directly.

A final major criticism has recently been levelled at twin-based methods by Neumark (1999), as well as by Bound and Solon (1999) (BS). 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 a genetic effect and a family effect which therefore disappears with 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 more than between non-twin siblings) and there is substantial evidence that birthweight has real effects[5]. 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 may even be more biased than least squares applied to individuals[6]. Addressing this important criticism is difficult in the context of twins because any instrumental variable that gives rise to differences in schooling via, for example, some reform are quite likely to affect both twins equally[7].

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. Finally we conclude with an agenda for further research.

2. Literature

Table 1 summarizes the most recent identical twins studies[8] and extends the reviews in Behrman and Rosenzweig (1999) (BR) and in Card (1999). Table 2 lists the fraternal (DZ) results where available and extends Table 6 in Card (1999). AK use

[5] 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.

[6] 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 the virtue of IV in this context relies on the measurement error being classical. However, they find that their estimates are not sensitive to the non-classsical nature of the errors.

[7] In future work we hope to exploit the random nature of military conscription that is a feature of Denmark to do this.

[8] 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.

the original 1991 Twinsburg festival data which yielded just 147 MZ pairs[9], 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[10]. 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 and xxx DZ pairs who were all white male World War II veterans, and BR used the Minnesota Twins Registry data of 720 MZ pairs and xxx DZ pairs.

The Australian Twin Register for 1980/82 and 1988/89 yielded 602 MZ pairs and 568 DZ pairs[11] 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 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 substantial dataset. His data is drawn from Swedish 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 observed having earnings around 1990[12].

These studies are not strictly comparable because of the construction of both the dependent variable and the explanatory variable. 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 of cross 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

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

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

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

[12] 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.

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 - Something about where our earnings data comes from.

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 period 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 . So BCHHS, and AK and AR probably come closest to overcoming concerns over 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 Table 1.

In particular, the methodology has typically proceeded along the following lines. Log wages, w, and education, S, are assumed to be determined by

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

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, $\gamma>0$ and $\alpha>0$ 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 $\beta_{WT} = \beta$. 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

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

where ρ is the within twin pair correlation between their reported schooling levels. Since this correlation seems likely to be strongly 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.

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 $\beta_{WTIV} = \beta$. 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 a 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 if least squares does provide an upper bound then 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. The implication is that existing twins estimates have failed to tighten the bound that least squares provides because least squares does not bound standard IV estimates.

Our contribution to this literature is twofold. Firstly, since we use administrative data our measurement error is likely to be less prone to depart from classical properties, and we provide an alterative instrument that seems likely to be less problematic than one based on cross reported schooling. There is no reason to expect the measurement error in our chosen measure of education to exhibit mean reversion. Moreover, our 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[13]. Indeed, we would argue that our instrument corrects for both measurement error and any bias arising from the self-selection of level of education and so addresses the criticism that has been raised relating to differences in education due to self-selection.

Secondly, the size of the dataset we use, and the fact that we have both MZ and DZs, allows us to trade precision for greater flexibility in several ways. Bias due to self-selection in education clouds several important issues that our analysis can help clarify. Previous attempts have had to grapple with eliminating the self-selection into work - which twins data allows us to address less problematically. Moreover, the panel nature of our data allows us to provide estimates broken down by calendar time, to cast light on whether returns are rising because of changes in the returns to observed skills or to unobserved skills by comparing estimates for MZ and DZ pairs over time. We also interact parental education levels with the difference in education within twins to investigate the extent to which schooling can compensate for lower family endowments. Further, we interact within pair education differences with the average level of education for the pair to investigate how returns vary across the distribution of ability. Finally, we estimate an explicit random coefficients model and decompose the variance in returns into family heterogeneity, individual heterogeneity, and risk or luck.

[13] 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.

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.

3. Data Description

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 19xx, that contains social security (i.e. CPR) numbers that are allocated at birth. The census in 1956 enables links between all children and their legal (biological??) mother and legal father to be established[14] and so allows us to match siblings, and the date of birth allows us to identify twins (indeed, all multiple births). 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")[15]. 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[16].

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 sample 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 highly grouped[17]. The labour earnings data itself is taken from tax returns and tax filing is compulsory even for those without earnings.

---

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

[15] 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.

[16] 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.

[17] 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.

We have restricted our sample to twin pairs where BOTH are observed to be full-time workers (annual hours at least 60% of full year hours) to reduce the impact of labour supply variation on our estimates[18]. We have several measures of education but we choose to use the measure of education that Statistics Denmark construct from reported information on highest qualification[19]. We also restrict attention to those twin pairs who BOTH were 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[20].

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[21]. Approximately 40% of these are MZ, and approximately 40% are female (because of their lower labour 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 very similar except for

[18] 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.

[19] Alternatively, we could use education duration as defined by the month and year when the highest qualification was recorded by the institution granting the qualification. This information is available for all those competing education after 1972 when this register was first computerised (from which we could subtract the minimum compulsory schooling, which conveniently changed in 1957 from 15 to 16, and divide by 12). However, to the extent that there are systematic differences in reporting practices across institutions we choose not to use this since this would give rise to potential correlation in measurement error within twins because they are likely to have attended the same institutions.

[20] Alternatively, we could use education duration as defined by the month and year when the highest qualification was recorded by the institution granting the qualification. This information is available for all those competing education after 1972 when this registration of students was first computerised (from which we could subtract age at starting school). However, to the extent that there are systematic differences in reporting practices between institutions we choose not to use this since this would give rise to potential correlation in measurement error within twin pairs because they are likely to have attended the same institutions.

[21] Adding different-sex DZ pairs would add approximately another 40 thousand pair*year observations. Adding the non-participants (i.e. less than full-time) pair*years would also increase our sample sizes by almost 50% for men and 90% for women.

age which is accounted for by the recent growth in the number of multiple births associated with fertility treatment which are inevitably DZ. Table 4 shows descriptive statistics for our samples of twin pairs. There are clearly smaller absolute differences in education and earnings, on average between MZ twins compared to DZ twins. Table 5 shows the frequency of education differences: a much higher proportion of MZs report exactly the same education length.

Table 3 Means (standard deviations) for various samples of individuals

Sample	Same sex twin pairs	Same sex DZ twin pairs	MZ twin pairs
MEN			
Education years	12.49 (2.99)	12.41 (2.93)	12.62 (2.79)
Ln Earnings	12.681 (0.341)	12.679 (0.344)	12.684 (0.335)
Age	38.97 (8.24)	39.34 (8.18)	38.37 (8.29)
N * Years	63299	38929	24370
N	11438	7068	4370
WOMEN			
Education years	12.30 (2.72)	12.16 (2.75)	12.50 (2.65)
Ln Earnings	12.375 (0.334)	12.362 (0.328)	12.394 (0.337)
Age	38.38 (8.20)	38.85 (8.16)	37.69 (8.22)
N * years	43425	25817	17608
N	9618	5618	4000

Note: N refers to number of individuals. Age, education, and wage rate is averaged over all years observed. Earnings are January 2005 Danish kroner.

Table 4 Means (standard deviations) of within twin differences

Sample	Same sex twin pairs	Save sex DZ twin pairs	MZ twin pairs
MEN			
Education years	1.780 (2.230)	2.037 (2.321)	1.370 (2.010)
Ln Earnings	0.266 (0.284)	0.291 (0.299)	0.226 (0.251)
N * years			
N	5719	3534	2185
WOMEN			
Education years	1.709 (1.946)	1.969 (2.012)	1.327 (1.775)
Ln Earnings	0.275 (0.279)	0.291 (0.283)	0.250 (0.271)
N * years			
N	4809	2809	2000

Note: N refers to numbers of pairs. Age, education, and wage rate is averaged over all years observed. Earnings are January 2005 Danish kroner.

Table 5 *Distribution of Education Differences (%)*

	All twin pairs	DZ twin pairs	MZ twin pairs
MEN same education	34.22	28.18	43.87
0-1 year difference	13.22	13.18	13.30
1-2 years difference	17.86	18.68	16.54
2-3 years difference	7.60	8.96	5.43
3-4 years difference	6.91	7.36	6.20
4+ years difference	20.18	23.64	14.66
N * years	63299	38929	24370
WOMEN same education	38.45	31.69	48.36
0-1 year difference	4.18	4.15	4.22
1-2 years difference	17.08	17.37	16.65
2-3 years difference	9.97	11.47	7.76
3-4 years difference	13.60	15.66	10.57
4+ years difference	16.73	19.66	12.44
N * years	43425	25817	17608

Table 6 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 7 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.

Figures 1 show the scatter of raw wage within twin pair differences by education difference for MZ and DZ men and women (trimmed??). The education data is coded to the nearest month so it is possible to see a scatter across the whole range of data, not just at integer differences. The solid line is a local linear regression which smooths the relationship. The MZ twins lines are close to being flat above a one year difference, but the wage gain for a one year difference is approximately 8%

for females and 4% for males??? In contrast, the DZ samples show some positive return that is close to being linear across the range of the data above one year.

We can illustrate the value of our spouses' education difference instrument in Figure 2 which shows the scatter of the within-twin pair difference in education and their partner's educations differences. There is a clear association between the two for both MZs and DZs, and it is stronger for females than males. Finally Figure 3 shows the wage difference against the partners' education differences.

Table 6 Correlation with education

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
Co-twins education	0.620	0.444	0.651	0.476
Own partner's education	0.394	0.390	0.427	0.360
Co-twin partner's education	0.348	0.312	0.350	0.269

Table 7 Within twin correlation with co-twin's partner's education

	MEN		WOMEN	
	MZ	DZ	MZ	DZ
Difference in partner's education	0.094	0.123	0.150	0.143

Figure 1 Differences in wages and education:

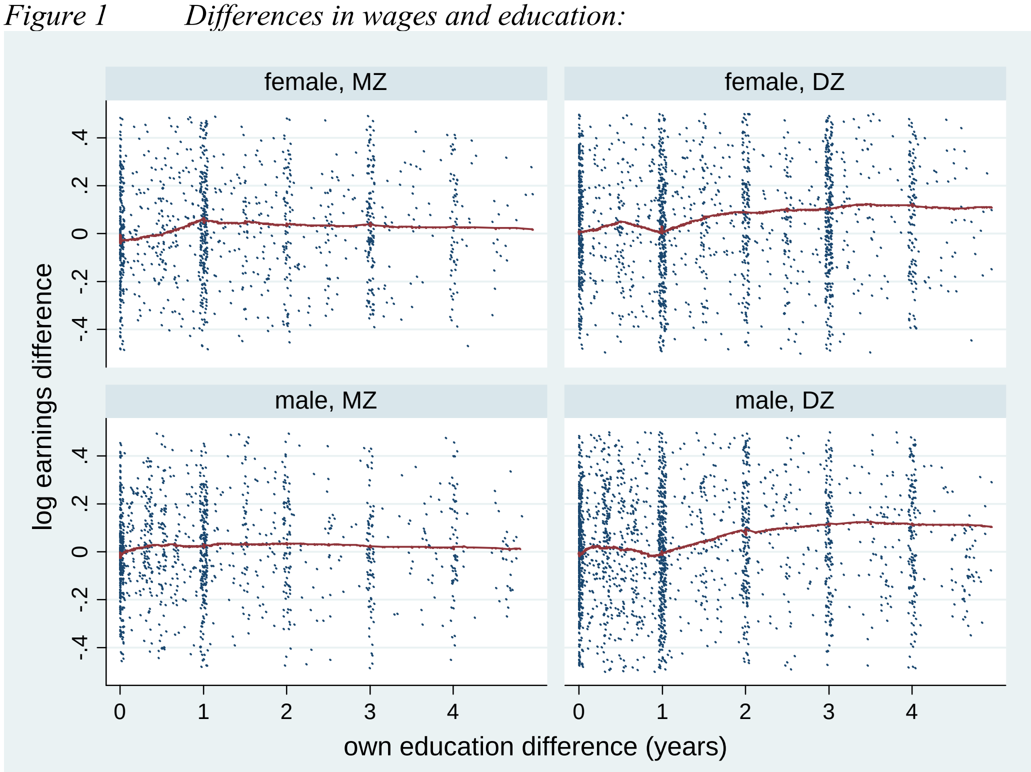


Figure 2 *Differences in education and partner's educations:*

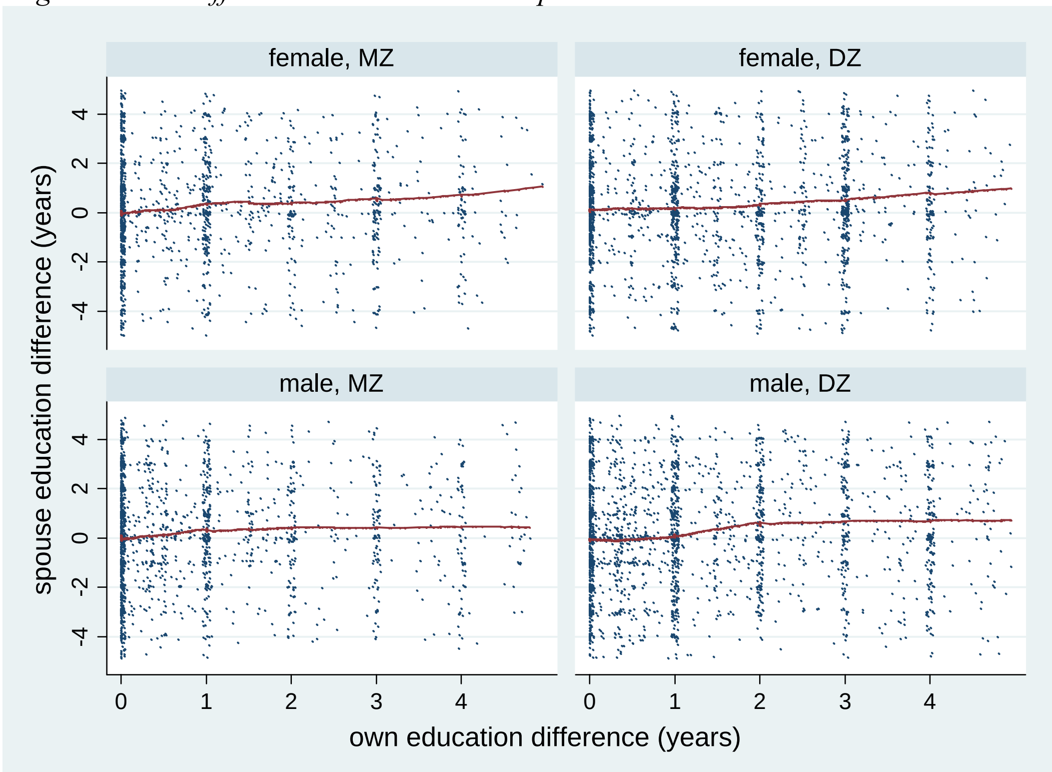
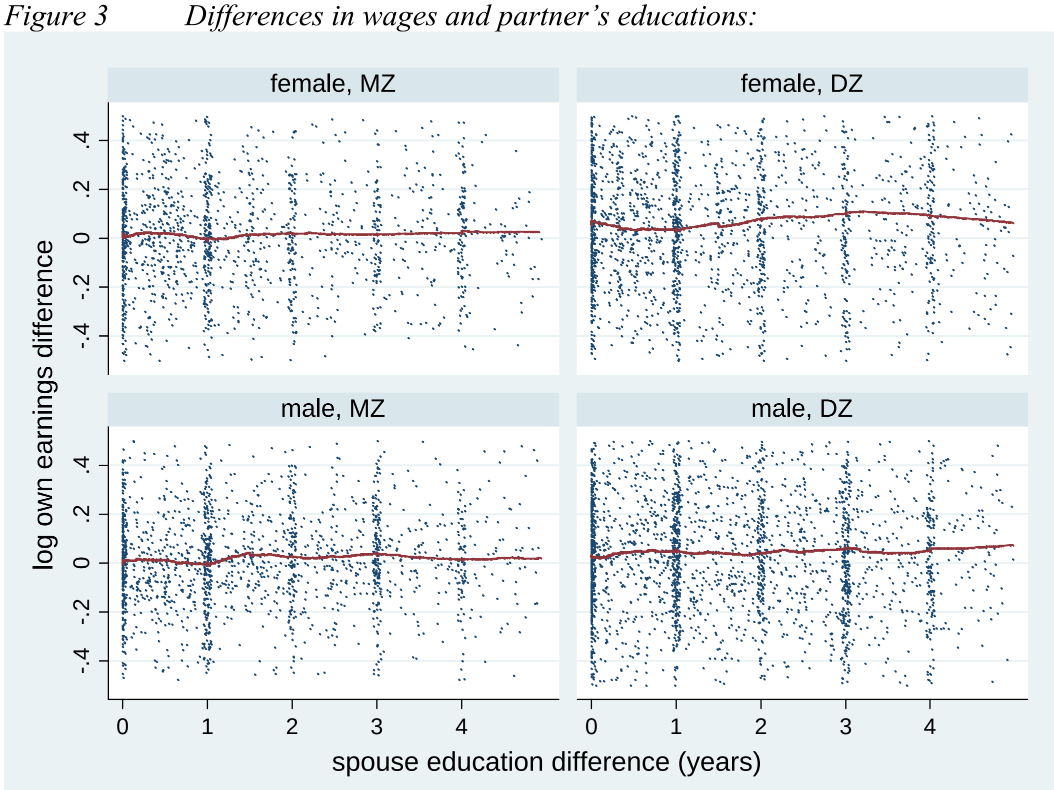


Figure 3 *Differences in wages and partner's educations:*



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 8. 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 Westergaard-Nielsen (2001) which suggest estimated returns of the order of 4%. The OLS DZ and MZ results are identical for women, and approximately (and significantly) different by 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 6 suggested is in the data and 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 up or down. In the light of this, correcting for endogeneity could easily imply larger IV estimates than OLS. Indeed, in Table 8 this is precisely what we find[22].

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

Table 8 Estimated Education Returns on Samples of Individuals

Sample	DZ only	MZ only	Pooled DZ and MZ	Pooled DZ and MZ
Method	OLS	OLS	OLS	IV
MEN				
β	0.0331 (0.0004)	0.0298 (0.0005)	0.0317 (0.0003)	0.0652 (0.0011)
R^2				-
N*years	77,858	48,740	126,598	126,598
Years	11.0	11.2	11.1	11.1
N	7068	4370	11,438	11,438
Partial R^2				0.1061
F-test				107.49
Hausman t				18.75
WOMEN				
β	0.0370 (0.0005)	0.0373 (0.0006)	0.0373 (0.0003)	0.0542 (0.0014)
R^2				-
N*years	51634	35216	80,850	80,850
Years	9.2	8.8	9.0	9.0
N	5618	4000	9618	9618
Partial R^2				0.0910
F-test				83.61
Hausman t				16.15

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. See Appendix for results with imputed earnings for those years where one twin is working less than full-time.

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 below10 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.

4.2 Fixed Effects

Table 9 takes the data used in Table 8 above 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.

Table 9 Fixed Effect Estimated Education Returns on Samples of Twin Pairs

Sample	DZ twins		MZ twins	
Method	OLS	IV	OLS	IV
MEN				
β	0.0183 (0.0007)	0.0946 (0.0063)	0.0049 (0.0009)	0.0451 (0.0099)
R^2		-		-
Partial R^2		0.0176		0.0036
F-test		13.45		4.86
Hausman t		4.10		4.24
N	3534		2185	
Years	11.0		11.2	
N*years	38929		24370	
WOMEN				
β	0.0250 (0.0009)	0.0535 (0.0061)	0.0093 (0.0012)	0.0441 (0.0083)
R^2		-		-
Partial R^2		0.0130		0.0065
F-test		12.37		7.84
Hausman t		4.81		4.31
N	2809		2000	
Years	9.2		8.8	
N*years	25817		17608	

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

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 Extensions to the Simple Fixed Effect Models

In this subsection we extend the benchmark FEIV model to allow for: variation in returns across calendar time and its separation into returns to observable and unobservable skills; nonlinear effects of education on wages to allow for the possibility that education complements (or substitutes for) unobserved ability; and interactions by parental education to allow for the possibility that education complements (or substitutes for) the home production of labour market productivity.

4.2.1 Rising returns

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.

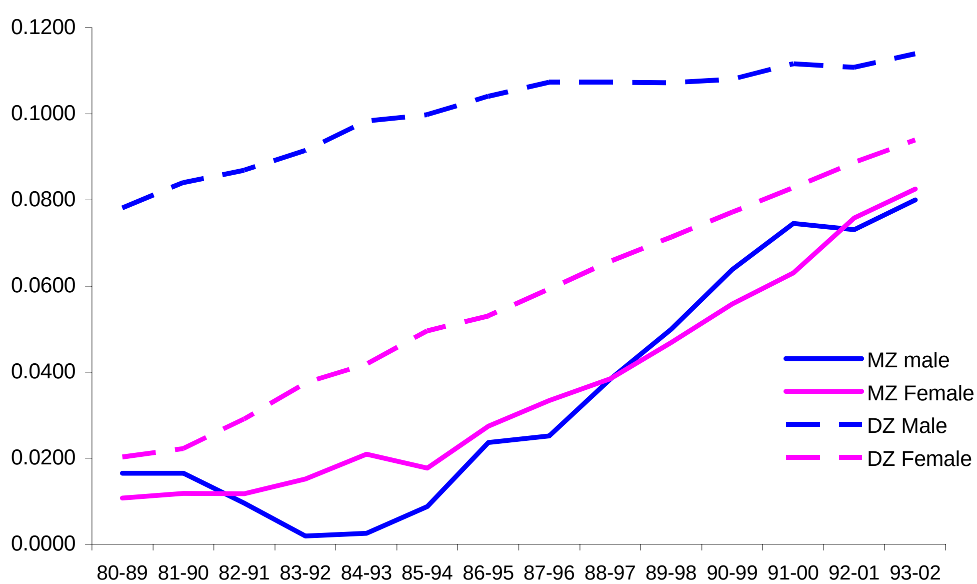
The virtue of MZ twin estimates is that they are purged of unobservables and therefore reflect only the return to observable skills. In contrast DZ estimates reflect returns to observable and unobservable skills. Therefore the difference between MZ and DZ estimates are informative about the returns to unobservables. To address this issue we constructed ten year averages and rotated this through our 23 year 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

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.

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



first third of the period and then rise quickly over the late 1980's and through the 1990's. 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.

4.2.2 Nonlinear returns

AR found no significant variation in returns across the own schooling distribution. Table 11 examines this issue in the Danish case. We partition the data by the average value of the level of education of the twin pairs into overlapping broad groups defined as lower secondary basic education, higher secondary, and tertiary. Here the MZ results suggest that the modest overall returns mask no returns at school, and large returns at tertiary level. Moreover, in Table 12 we present estimates for four year schooling ranges and show that the returns are only significantly different from zero at the higher levels of schooling.

Table 11 Fixed Effect IV Estimated Returns across the Schooling Distribution

Sample:	MZ				DZ			
	Male		Female		Male		Female	
Average schooling:	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)
Lower secondary	0.0159	*0.0103*	0.0071	*0.0121*	0.0719	*0.0099*	0.0020	*0.0082*
Upper secondary	0.0057	*0.0248*	0.0225	*0.0164*	0.1134	*0.0119*	0.1285	*0.0233*
Tertiary	0.1157	*0.0202*	0.1001	*0.0132*	0.0905	*0.0093*	0.0853	*0.0084*
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.

Table 12 Fixed Effect IV Estimated Returns across the Schooling Distribution

Sample:	MZ				DZ			
	Male		Female		Male		Female	
Schooling range:	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)
7-10	0.0159	*0.0103*	0.0071	*0.0121*	0.0719	*0.0099*	0.0020	*0.0082*
8-11	-0.0150	*0.0075*	0.0218	*0.0077*	0.0429	*0.0052*	0.0320	*0.0068*
9-12	-0.0115	*0.0085*	0.0262	*0.0080*	0.0464	*0.0052*	0.0236	*0.0069*
10-13	-0.0014	*0.0109*	0.0316	*0.0131*	0.0973	*0.0087*	0.0433	*0.0083*
11-14	0.0627	*0.0177*	0.0041	*0.0136*	0.1267	*0.0107*	0.0823	*0.0107*
12-15	0.1228	*0.0239*	-0.0005	*0.0158*	0.1750	*0.0139*	0.0659	*0.0104*
13-16	0.1188	*0.0235*	0.0277	*0.0155*	0.1679	*0.0125*	0.0962	*0.0103*
14-17	0.1140	*0.0199*	0.0917	*0.0127*	0.0905	*0.0093*	0.0853	*0.0084*
15-18	0.0686	*0.0352*	0.1738	*0.0161*	0.0484	*0.0103*	0.1499	*0.0177*
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.

4.2.3 Parental Background

Table 13 examines the extent to which the effect of additional schooling varies across the parental schooling distribution. AR also investigated this issue and found large but insignificant effects of parental background as measured by education. We are only able to link parents to children born after 1955 (the date of the last ever census in Denmark), and completed education is only recorded for observations who are born after 1920. Thus, the samples that we use here are now somewhat smaller. We estimate a parsimonious specification with own education interacted with paternal education, and with maternal education. Maternal education has a statistically significant positive effect on sons' returns and a significantly negative effect on daughters' returns.

Table 13 Fixed Effect IV Estimated Education Returns and Parental Schooling

Sample:	MZ				DZ			
	Male		Female		Male		Female	
	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)
Own education	-0.0314	0.1057	0.4145	0.1732	-0.0021	0.005	0.0030	0.0402
Paternal interaction	-0.0207	0.0113	-0.0072	0.0100	0.0012	0.0004	0.0038	0.0038
Maternal interaction	0.0318	0.0134	-0.0216	0.0076	0.0016	0.0005	0.0068	0.0037

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

To facilitate interpretation, we use the estimates in Table 13 to compute, in Table 14, estimated returns by parental education levels for MZ twins across a range of parental education from 7 years to 14 years which captures the bulk of the joint distribution. For males, increasing maternal education increases returns strongly, while increasing maternal education decreases returns - maternal education has a bigger effect at low levels of paternal. While for females we find the opposite – maternal education reduces returns especially at low paternal education. We use the standard errors estimated in Table 13 to construct confidence intervals around the predictions in Table 14 and indicate predictions that are significantly different from zero with bold type.

Table 14 Estimated Education Returns across Parental Schooling

		Paternal education							
	Male	9	10	11	12	13	14	15	16
Maternal education	9	0.0684	0.0477	0.0270	0.0062	-0.0145	-0.0352	-0.0560	-0.0767
	10	**0.1003**	**0.0795**	0.0588	0.0381	0.0173	-0.0034	-0.0241	-0.0449
	11	**0.1321**	**0.1113**	**0.0906**	**0.0699**	0.0491	0.0284	0.0077	-0.0131
	12	**0.1639**	**0.1432**	**0.1224**	**0.1017**	**0.0810**	0.0602	0.0395	0.0188
	13	**0.1957**	**0.1750**	**0.1543**	**0.1335**	**0.1128**	**0.0921**	0.0713	0.0506
	14	**0.2275**	**0.2068**	**0.1861**	**0.1653**	**0.1446**	**0.1239**	**0.1032**	0.0824
	15	**0.2594**	**0.2386**	**0.2179**	**0.1972**	**0.1764**	**0.1557**	**0.1350**	**0.1142**
	16	**0.2912**	**0.2705**	**0.2497**	**0.2290**	**0.2083**	**0.1875**	**0.1668**	**0.1461**
	Female								
	9	**0.1554**	**0.1481**	**0.1409**	**0.1337**	**0.1265**	**0.1193**	**0.1121**	**0.1049**
	10	**0.1338**	**0.1266**	**0.1194**	**0.1121**	**0.1049**	**0.0977**	**0.0905**	**0.0833**
	11	**0.1122**	**0.1050**	**0.0978**	**0.0906**	**0.0833**	**0.0761**	0.0689	0.0617
	12	0.0906	0.0834	0.0762	**0.0690**	0.0618	0.0545	0.0473	0.0401
	13	0.0690	0.0618	0.0546	0.0474	0.0402	0.0330	0.0257	0.0185
	14	0.0474	0.0402	0.0330	0.0258	0.0186	0.0114	0.0042	-0.0030
	15	0.0259	0.0187	0.0114	0.0042	-0.0030	-0.0102	-0.0174	-0.0246
	16	0.0043	-0.0029	-0.0101	-0.0174	-0.0246	-0.0318	-0.0390	-0.0462

Note: Standard errors in parentheses.

4.2.4 Unobservable Differences in Returns

Finally, because the data is a panel we can employ multilevel modelling to estimate a mixed model with both fixed and random effects. This essentially estimates a random coefficient (on education) model and decomposes the variance around the mean return into family heterogeneity, individual heterogeneity, and luck or risk. Table 14 presents estimates where the standard errors are bootstrapped to correct for the use of a generated regressor as an explanatory variable.

Table 14 Fixed Effect IV Estimated Returns: Random Coefficients

Sample:	MZ				DZ			
	Male		Female		Male		Female	
	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)	β	s.e.(β)
Mean	0.0273	0.0013	0.0326	0.0015	0.0351	0.0010	0.0388	0.0011
	σ	s.e.(σ)	σ	s.e.(σ)	σ	s.e.(σ)	σ	s.e.(σ)
Std. dev.								
Family	0.0094	0.0008	0.0085	??	0.0054	??	0.0050	??
Individual	0.0084	0.0012	0.0062		0.0959		0.0077	
Luck	0.1915	0.0006	0.1958		0.1904		0.1882	

5. Conclusions and Further Research

In this study we present estimates of the returns to schooling based on a very large sample of Danish twins. The data is drawn from population administrative registers and form a 23 year long unbalanced panel of more than 8000 MZ twin pairs and 12000 DZ 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. We find that the simple fixed effects estimators are biased downwards to a much larger degree. When we instrument the FE estimators we find estimates that are fully 50% higher that 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.

Unlike the AR results we do find significant interactions with parental education, and we find very important differences by level of education of the twins. Indeed, our estimates suggest that there is no return to school level education, only to college. We exploit the panel nature of the data to show that estimated returns have been rising and, by contrasting MZ and DZ results we conclude that, since the mid 1980's returns to unobserved skills appear to have been falling while the return to observed skills have been rising. We also exploit the panel data to show that there are large and significant variances to the estimates returns and we decompose these into family, individual and luck. We find that individual variance in returns is smaller for MZs than for DZs.

References

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

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

Bingley, Paul and xxx xxxxx (2005), "xxxxxxxxxxxx" mimeo.

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.

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 (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,* xx, 705-746.

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.

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.

Deschenes, 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, 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.

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.

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

Westergard-Nielsen, Niels (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.

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.