The Evolution of the Returns to Observable and Unobservable Skills: Evidence from a Panel of Danish Twins

Paul Bingley
Danish National Centre for Social Research, Copenhagen

Kaare Christensen
The Danish Twin Registry, University of Southern Denmark

and

Ian Walker
Department of Economics, Lancaster University Management School
and Education Research Section, Princeton University

March 27, 2015

Keywords: returns to education, twins, ability bias
JEL Codes: I3

Abstract:

We estimate the returns to schooling using administrative data for a large and long panel of Danish twins. We use identical (MZ) twins that, arguably, enables us to estimate the causal effect of schooling purged of ability bias; and contrast these results with same-sex fraternal twins, which are prone to be affected by ability bias. The difference between these estimates provides an estimate of the extent of ability bias and we exploit this to decompose the evolution of returns over our 23 year panel. We deal with measurement error in the usual way – instrumenting with other measures. We also attempt to address the additional concern that even MZ twins are unobservably different in ways that affect earnings. Here we exploit the potential of the school environment of twins to generate exogenous within-twin differences. We find that the within twin schooling differences do differ by school size even though the twins share the same school because they may not share the same class. Nonetheless, we find that the resulting IV estimates are not significantly different from the conventional FE estimates for MZ’s but smaller for DZ’s.

* Our collaboration was supported by the Danish Social Science Research Council (FSE-24-04-0240), the Danish Council for Strategic Research (DSF-2139-08-0020), the British Academy who funded a visit to the Education Research Section at Princeton University, and the UK Economic and Social Research Council (RES-000-27-0200). We are grateful to Alan Krueger, Orley Ashenfelter, Cecilia Rouse, Costas Meghir and Erik Plug for helpful comments on this work. We are grateful to Vibeke Jensen for her help with the data. The usual disclaimer applies.
1. Introduction

Estimating the returns to education has been a major industry for the
economics research community and an important challenge has been to identify the
causal effect of education on wages. Moreover, the substantial rise in wage inequality
that has occurred since the 1970’s in the US, the UK and many other countries has
extended the motivation for this research agenda. For example Katz and Goldin
(2008) has emphasised the role of changes in the returns to skills in explaining the
evolution of wage inequality. An important issue has been the extent to which this
rising inequality is due to rising returns to observed skills, such as education, or rising
residual inequality, part of which may reflect changes in the return to 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 unobserved skills. The novelty of the paper is that we can throw light on all of these
issues by exploiting a large and long panel of twins of known zygocity. Identical
(monozygotic, MZ) twins are particularly valuable to researchers because they have
the same endowments (at the time of conception)\(^1\) so that differencing within MZ twin
pairs might eliminate any (or, at least, most) unobserved endowments. While MZ
twins share effectively all of their segregating genes, fraternal (dizygotic, DZ) twins
share, on average, 50\% of their segregating genes, so that differencing within DZ
pairs does not eliminate endowments – even for same sex DZs\(^2\). The basic idea of this
paper is that the extent of ability bias can, in principle, be inferred from comparing
schooling coefficient estimates in within-twin estimated wage difference equations
from same sex fraternal twin differences with estimates from identical twin
differences. The argument is that MZ twins reveal the causal effect purged of ability
bias while the DZ twins do not because the latter are contaminated by “ability’ bias.
So the difference between estimates for MZs and those for DZs reveals the effect of
unobserved ability.

---

\(^1\) However, other important differences may remain. For example, birthweight differences are thought
to have effects on education (see Berhman and Rosenzweig (2004)).


\(^3\) See Heckman and Urzua (2009) and Angist and Imbens (1994).


Griliches (1979) cautions that twins are “not a panacea”, and renewed reservations about the use of twins for estimating the returns to education have been expressed in Bound and Solon (1999) and in Neumark (1999). These papers draw attention to two distinct problems. The first problem is measurement error in differenced schooling, and the second is the potential endogeneity of within twin differences. Regarding measurement error, the tendency for estimates to be attenuated (i.e. biased towards zero) has typically been addressed using a second measure as an instrument for the first. An important innovation of Ashenfelter and Krueger (1994) was that they asked each twin about their own schooling and their co-twin’s schooling. The difference in twin cross-reported schooling is used as an instrument for difference in self-reported schooling. We use a similar idea here.

The second “double trouble” criticism has proved more troublesome and has so far been left unresolved in the literature: that within-twin pair schooling differences are potentially endogenous, 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. The presumption in the twins literature is that omitted ability is removed by differencing within identical-twin-pair, but 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. Very recently, Cesarini et al (2014) have empirically confirmed earlier fears that differencing, even for MZs, does not eliminate the ability bias problem. They use almost nine hundred Swedish MZ twin pairs and find that the twin difference in earnings is strongly correlated with the difference in IQ even controlling for schooling differences. The limited US evidence is inconclusive.

Directly addressing the consequences of unequal ability is difficult in the context of twins. To do so requires that we have an instrument for within twin differences in schooling. Of course, most instrumental variables that give rise to variation in schooling are usually driven by a mandatory policy reform, usually defined by cohort and location, and such a reform would invariably affect both twins equally because twins are in the same cohort and almost invariably schooled at the same location. For example, raising the mandatory minimum school leaving age has the same affect on the schooling of both twins who wanted to leave at the old minimum, and no effect on those who wanted to leave at the new minimum or later.
Neumark (1999) hints at the possibility that twins do not necessarily share exactly the same environment and gives the example of differential treatment in school – for example, one twin might experience a different quality of teacher than the other, and our data allows us to capture this possibility.

Thus, the contribution of this paper is threefold. First, we address both aspects of the “double trouble” critique. We provide MZ estimates that correct for both measurement error and for the violation of the equal ability assumption using instrumental variable estimation. Secondly, we exploit the distinction between MZ and DZ twins to uncover the return to unobserved ability – up to a constant. In a typical cross section dataset of twins this would not be interesting to do because we could not eliminate the unknown constant in this difference, but with our long panel of twins we can do this for each cross section which enables us to show the extent to which the return to unobserved differences, as well as the return to (observable) education, has been changing over time. That is, the panel allows us to difference out the unknown constant and so identify changes in unobservable returns. Thirdly, we use the largest sample of twins ever assembled. Indeed, the data used in the work here approximately doubles the size of the data in this literature as a whole.

Our headline results, in Section 5 below, show returns of the order of 3 to 4% that is typical of the (small) Danish literature, using OLS on the MZ data in levels; i.e. simply stacking the MZ twins (and similar results can be obtained for singletons suggesting that twins have external validity) and for the DZ’s when they are stacked. Throughout we exploit the panel by pooling the data across years and correcting the standard errors accordingly to account for multiple observations. When we exploit the twins nature of the data and apply OLS we find large attenuation, and so the estimated returns are substantially smaller, for both male and female MZ twins. The results suggest very large measurement error in the differences in schooling. When we then instrument to deal with this measurement error in the administrative reports using information from self-reports we obtain somewhat larger estimates consistent with the idea that imputing a continuous schooling variable from the administrative qualifications data generates measurement error. The DZ results are somewhat higher than the MZ suggesting that differencing leaves some remaining ability bias in the DZ case. Finally, when we instrument to control for (any remaining) unequal ability using schooling background information then we find estimates for MZs that remain high
and smaller DZ results that are now approximately the same as the MZ results. The suggestion in these results is that FE estimation with MZs produce estimates that are indeed purged of ability bias, but the same is not true of DZs. This finding is important because, to the extent that it applies elsewhere, it reinstates the twin methodology despite the fears arising from the double trouble critique.

Nonetheless, we go on to consider DZs as well as MZs in an attempt to isolate in the returns to unobserved skills, that underlies ability bias, from the returns to observed skills by exploiting the panel nature of the data. In addition to considering variation in returns across calendar time we also consider variation in returns across cohorts. Overall, we feel justified in thinking that we have not only reinstated the twins methodology but we have also demonstrated the value of having a panel of DZ twins as well as MZs to address the issue of rising returns – a topic where the search for causal returns has not previously penetrated deeply.

2. Literature

There are many studies of the private financial returns to education based on the standard human capital model of earnings determination (see the 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 on 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 (2001), and Carneiro, Heckman and Vytlacil, (2003)). As a result, 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\(^3\). While instrumental variable estimates identify local effects, an advantage of twins is that, providing ability bias is eliminated by differencing, we would expect that a twins-based estimator \textit{would} provide an estimate of the \textit{average} effect of education on wages that could be directly compared with conventional least squares estimates.

\(^3\) See Heckman and Urzua (2009) and Angist and Imbens (1994).
Identical twins have been used, in more than a dozen studies to date\(^4\), 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 specific to an individual, fixed over time, and associated with productivity in the labour market and hence earnings. 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 ability.

Table 1 summarizes the most recent identical twins studies\(^5\) and extends the reviews of Behrman and Rosenzweig (1999) and Card (1999). Ashenfelter and Krueger use the original 1991 Twinsburg festival data which yielded just 147 identical twin pairs\(^6\), while the Ashenfelter and Rouse (1998) pooled the 1993 festival data to give 333 identical twin pairs, and Rouse (1998) used the 453 MZ pairs which included the 1995 festival\(^7\). The Twinsburg data has very few DZ twins and they have not been used in previous research. Behrman \textit{et al} (1994) used the NAS-NRC data on 141 identical twin pairs who were all white male World War II veterans, and Behrman and Rosenzweig used 720 MZ twin pairs from the Minnesota Twins Registry data. The Australian Twin Register for 1980/82 and 1988/89 yielded 602 MZ twin pairs and 568 DZ twin pairs\(^8\) who are analysed in Miller \textit{et al} (1995) and in Miller \textit{et al} (1997). A study by Bonjour \textit{et al} (2003) used UK data on 187 MZ female twin pairs obtained from the records of a large London hospital\(^9\).

Isaccson (1999, 2004) are the only studies that use a similarly large dataset to ourselves. His data is drawn from Swedish registers of the population and yields 2492 and 2609 identical twin pairs in each study respectively. In the each case the data


\(^5\) 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 \textit{et al} 1994) as an instrument for own education. Taubman (1976) is an early example of twins research that does not instrument.

\(^6\) This sample is re-examined in Flores-Lagunes and Light (2004) who were specifically concerned with the treatment of measurement error.

\(^7\) See also Arias \textit{et al.} (2001) who subsequently re-analysed this data.

\(^8\) The 1988/9 data appears to be drawn from the 1980/82 sample and so the data are not independent observations.

\(^9\) Amin (2011) reanalyses the UK data and finds the original results to be very sensitive to a single outlier.
were twins born between 1926 and 1958 and earnings were observed around 1990\textsuperscript{10}. The Cesarini \textit{et al} (2014) data also comes from merging administrative datasets with surveys of twins: education qualifications, income in 2005, and IQ at age 18 come from various registers while data on 890 MZ twin pairs come from the Swedish Twin Registry survey datasets collected in 1998-2002 and 2005-2006, restricted to male twins born from 1950 to 1975. The income data was censored to try to ensure that only full-time workers were included. The one Chinese study (Li \textit{et al} (2012)) uses a dataset of 976 MZ twins, pooled across men and women. They use monthly earnings and do not report whether earnings were adjusted for labour supply.

The first column of estimates in our table provides OLS using the stacked twins; while the second column provides the OLS on the within-twin differences that introduces the problem of attenuation because of measurement error; and the final column instruments to overcome this attenuation.

These studies are not strictly comparable because of the construction of both the dependent variable and the explanatory variable of interest. The Australian research 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. 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 Miller \textit{et al}. (1995 and 1997), Behrman \textit{et al}. imputes earnings from detailed occupation information. The Isacsson Swedish research uses annual earnings and drops very low earners but otherwise takes no account of labour supply differences. The more recent Swedish work by Cesarini \textit{et al} (2013) instruments self-reported schooling by administrative data on qualifications. The UK research in Bonjour \textit{et al} 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 US Twinsburg data uses the reported hourly wage and education is recorded in years. A distinctive aspect of the UK research is that it attempts to address

\textsuperscript{10} 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 that contained both qualifications and schooling duration. Rouse (1999) also averages wages across Twinsburg datasets for those observations that appear on more than one occasion.
the double trouble critique - that within-twin differencing fails to eliminate all of the unobserved ability. It attempts to do so in two ways: exploiting information on discordance in smoking and in an important test at age 11 that normally has important effects on schooling attainment since those that passed were tracked into an academic school. Unfortunately the extent of discordance was too small to provide the basis for convincing instruments.

Table 1  
Selected MZ Twins Estimates in the Literature

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Gender</th>
<th>Twin pairs</th>
<th>$\beta_{OLS}$</th>
<th>$B_{WT}$</th>
<th>$\beta_{WTIV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashenfelter and Krueger (1994)</td>
<td>US</td>
<td>Pooled</td>
<td>147</td>
<td>0.084 (0.014)</td>
<td>0.092 (0.024)</td>
<td>0.129 (0.030)</td>
</tr>
<tr>
<td>Ashenfelter and Rouse (1997)</td>
<td>US</td>
<td>Pooled</td>
<td>333</td>
<td>0.110 (0.009)</td>
<td>0.070 (0.019)</td>
<td>0.088 (0.025)</td>
</tr>
<tr>
<td>Ashenfelter and Rouse (1997)</td>
<td>US</td>
<td>Pooled</td>
<td>333</td>
<td>0.110 (0.009)</td>
<td>0.070 (0.019)</td>
<td>0.088 (0.025)</td>
</tr>
<tr>
<td>Berhman et al (1994)</td>
<td>US</td>
<td>Pooled</td>
<td>141</td>
<td>0.094a (0.011)</td>
<td>0.035 (0.004)</td>
<td>0.101 (0.012)</td>
</tr>
<tr>
<td>Berhman and Rosenzweig (1999)</td>
<td>US</td>
<td>Pooled</td>
<td>720</td>
<td>0.113a (0.005)</td>
<td>0.104 (0.017)</td>
<td>NA</td>
</tr>
<tr>
<td>Bonjour et al (2004)</td>
<td>UK</td>
<td>Female</td>
<td>187</td>
<td>0.077 (0.001)</td>
<td>0.039 (0.023)</td>
<td>0.077 (0.033)</td>
</tr>
<tr>
<td>Isacsson (1994)</td>
<td>Sweden</td>
<td>Pooled</td>
<td>2492</td>
<td>0.046 (0.001)</td>
<td>0.022 (0.002)</td>
<td>0.024b (0.008)</td>
</tr>
<tr>
<td>Isacsson (2004)</td>
<td>Sweden</td>
<td>Pooled</td>
<td>2609</td>
<td>0.066c (0.009)</td>
<td>0.028c (0.009)</td>
<td>0.052c (0.036)</td>
</tr>
<tr>
<td>Miller, Mulvey and Martin (1995)</td>
<td>Australia</td>
<td>Pooled</td>
<td>602</td>
<td>0.064 (0.002)</td>
<td>0.025 (0.005)</td>
<td>0.048 (0.010)</td>
</tr>
<tr>
<td>Miller, Mulvey and Martin (1997)</td>
<td>Australia</td>
<td>Male</td>
<td>282</td>
<td>0.071d (0.003)</td>
<td>0.023 (0.008)</td>
<td>0.033 (0.014)</td>
</tr>
<tr>
<td>Rouse (1998)</td>
<td>US</td>
<td>Female</td>
<td>320</td>
<td>0.057d (0.002)</td>
<td>0.028 (0.006)</td>
<td>0.058 (0.011)</td>
</tr>
<tr>
<td>Li, Liu and Zhang (2012)</td>
<td>China</td>
<td>Pooled</td>
<td>976</td>
<td>0.066 (0.004)</td>
<td>0.025 (0.015)</td>
<td>0.032 (0.019)</td>
</tr>
<tr>
<td>Cesarini, Sandewall, and Johannesson (2014)</td>
<td>Sweden</td>
<td>Male</td>
<td>890</td>
<td>NA</td>
<td>0.024 (0.008)</td>
<td>0.021 (0.008)</td>
</tr>
</tbody>
</table>

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.
Since schooling is an important determinant of wages it is likely to substantially affect inequality. Moreover, changes in the relationship between wages and schooling will affect the evolution of inequality. The seminal work by Juhn, Murphy, and Pierce (1993) suggested that both observed (education) and unobserved (ability) drivers of inequality had been growing over time in the US. Although some early work in this area has not directly addressed the role of ability bias (see, for example, Card and Lemieux (2001) and Katz and Murphy (1992)), a number of studies have attempted to address the issue using conventional sources of data. This is not an issue that has previously been addressed by exploiting data on twins.

Cawley et al (2001) reports that the proportion of wage variation accounted for by ability (AFQT) is very small when schooling is controlled for, although Neal and Johnson (1996) find an important role for ability even when schooling is controlled for. Blackburn and Neumark (1993) was an early example that considers how the ability effect may have changed over time. They use National Longitudinal Survey of Youth 1979 data that contains information about ability test scores, from an armed forces test, and their focus is on the coefficient on the interaction between ability, education and time. They report that the rise in return to education is concentrated amongst those with both high education and high ability, but they argue that there is no trend in the return to ability. Thus, they attribute to rising returns to the returns to observable skills. Two further studies use a similar methodology. Bishop (1991) uses NLSY who finds that the time trend of returns to ability is not statistically significant. Grogger and Eide (1995) use two longitudinal surveys from the late 1970s and mid 1980s which contain school test scores: National Longitudinal Study of the High School Class of 1972 and the High School and Beyond (HSB). They also find that the rise in the return to education is concentrated on the most able. However, the problem with such cohort studies is that it is impossible to distinguish between age and time effects. So an apparent rise in returns over time might be due to an interaction with age or experience rather than with time. Murnane, Willett and Levy (1995) identify both time and age effects. They use the NLSHSC72 and the HSB datasets to compare wages in 1978 with wages in 1986. They estimate the contribution of ability, measured by scores in a math test, to the rise in the return to education measured at the age of 24 in 1978 and 1986. They conclude that a sizeable minority of the rise in the return to education can be attributed to a rise in the return to
ability. However, the finding could reflect other explanations – for example a rise in the correlation between ability and schooling. Heckman and Vytlacil (2001), like the studies referred to above, rely on data that contains an ability measure. An important difficulty is that ability is a test score and such a measure may not capture the unobserved ability to make money very well. Moreover, there is a high correlation between test scores and completed education, especially for test scores collected after schooling has been completed, that make identification of the returns to observed schooling separately from ability difficult because of collinearity. The Heckman and Vytlacil (2001) results, in their most general non-parametric case, are inconclusive, and the parametric restrictions that might yield more precision are typically rejected by their data.

Cuhna et al (2011), following Cuhna et al (2005), uses three datasets: the US National Logitudinal Study (NLS) of 1962, the NLSY79 and the Panel Study of Income Dynamics (PSID). They exploit the virtues of each to identify different parameters and this allows them to quantify the contribution of the different components of the rising college premium. They conclude that the rising college premium is largely due to the rise in the returns to college. Kaymak (2009) is novel because it exploits the differences in education across successive birth cohorts. He uses CPS data for 1964–2003 and the censuses from 1960 to 2000. Their argument is that if ability is similar across close cohorts, then differences in educational attainment lead to differences in earnings only if education is productive. The results show that the return schooling rose from 4.8 percent to 8.4 percent between 1964 and 2003, (ii) the ability bias rose from 1.8 percent to 4.7 percent during the same period, and (iii) the acceleration in the education premium after 1980 is explained almost entirely by the rise in the ability bias. The methodology relies on the covariance between ability and schooling varying smoothly across cohorts. While this is probably true across time it seems likely that discrete reforms in schooling policy might also play a role across even nearby cohorts.

A final branch of the literature takes a structural approach. Belzil (2007) provides an excellent survey of this literature. The most notable contribution in this particular context is Taber (2001) who constructs a dynamic discrete choice model based on the NLSY data. The results suggest that an increase in the demand for
unobserved ability makes an important contribution to the rise in the college premium. Finally, Deschenes (2001) notes that rising returns should imply increasing convexity to age earnings profiles and manages to back this out from his estimates.

As far as Danish research is concerned there are no studies that estimate the changes in inequality over time at all. Moreover, there are very few studies that even estimate the returns to schooling. Only Christensen and Westergård-Nielsen (2001) report estimates, based on applying least squares to data on wages and education drawn from administrative records, across time – rising slowly from a low base of approximately 3%.

3. **Methodology**

While it seems inappropriate to compare results across rows in Table 1, not least because they relate to 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

\[
\begin{align*}
\ln(w) &= \beta S + \alpha A + \epsilon \\
S &= \gamma A + \zeta
\end{align*}
\]

where $A$ is unobserved “ability”, $\epsilon$ is uncorrelated with $S$ and $A$, and the residual in the $S$ equation, $\zeta$, is uncorrelated with $\epsilon$. It is assumed that $\zeta$ and $w$ are correlated only through their joint dependence on $A$. We refer to $A$ as ability but captures any unobserved factors that affect both $w$ and $S$. Since $A$ is unobservable, OLS estimates of $\beta$ in $w = \beta S + u$ will be biased such that

\[
\text{plim}(\beta_{OLS}) = \beta + \alpha \frac{\sigma_{AS}}{\sigma_{S}^2}
\]

and if, as seems reasonable, $\gamma > 0$, $\alpha > 0$ and $\sigma_{AS} > 0$, then $\beta_{OLS} > \beta$. That is, OLS captures the effects of $S$, and unobservables correlated with $S$ such as $A$,\textsuperscript{11} on $w$.

\textsuperscript{11} The traditional solution to this problem is to instrument $S$ to purge the estimates of its ability bias. For example, education reforms have been a popular source of instruments for schooling. See Harmon and Walker (1995) for an early example. However, it is well known that IV methods do not, in general, identify the Average Treatment Effect (ATE) but, rather, the Local Average Treatment Effect (LATE) associated with the particular instrument used. While it may be possible to speculate about the type of individual that is affected by the variation in the instrument used, in general, the ATE will differ from the LATE and the difference is impossible to uncover. In contrast, the twin methodology estimates the ATE.
If a large sample of twins is available, and $A$ is the same within identical twin pairs, then differencing the wages within MZ twin pairs will result in the $A$ being differenced out of (1), and we are left with the within-twin pair equation

\[(3) \quad \Delta w = \beta \Delta S + \Delta \epsilon.\]

where $\Delta$ is the within twin pair difference operator. Applying OLS to this within twin pair equation yields $\beta_{WT} = \beta$ where $\beta_{WT}$ is sometimes referred to as the covariance estimator - because it is the covariance between $\Delta w$ and $\Delta S$.

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

\[(4) \quad \Delta w = \beta \Delta S - \beta \Delta \nu + \Delta \epsilon.\]

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

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

where $\rho$ is the within twin pair correlation between their reported schooling levels. Since this correlation is very likely to be large and positive the downward bias in $\beta_{WT}$ is likely to be substantially worse than in $\beta_{OLS}$. That is, differencing exacerbates the bias in OLS that is due to measurement error.

Ashenfelter and Krueger (1993) correct for the measurement error that biases $\beta_{WT}$ by instrumenting $\Delta S$ 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 there is no family effect in the measurement error. Providing $\Delta S'$ is a valid instrument for $\Delta S$ 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.\(^\text{12}\)

---

\(^\text{12}\) While instrumental variable estimation requires that the measurement error is classical Lagunes-Flores and Light (2005) address this using Generalized Method of Moments estimation to estimate the parameters of a range of measurement error models, including forms of both classical and mean-reverting error models, on twins data. They find that classical measurement error models fit the dataset about as well as more flexible models.
Of course, measurement error is an issue for this research also. We have two sources of information on $S$. The first comes from administrative data provided by the Ministry of Education. It is common to think of administrative data as being free of measurement error in which case we would not need a second measure. However, the administrative data records qualifications and to maintain consistency with the existing literature we convert the detailed qualifications, following the Ministry’s own procedure, to produce a “normed” continuous measure of schooling (in months). Our second source is from surveys of the twins provided by the Twins Registry. These contain self-reports of own highest qualification that we can transform into a continuous variable using the Ministry’s weights. The two sources are matched together for us using the unique Social Security (CPR) number. Despite the aggregation required to generate a continuous schooling variable, we think of the administrative data as being the more accurate, and so a more natural candidate for being used as the endogenous regressor in the wage equation. In which case the self-reported qualifications from the surveys form the instruments (or aggregated into a continuous single instrument to reduce the potential for instruments to be weak). So we would be treating the self-reported information as the equivalent of the Twinsburg cross-reports in our context. On the other hand, we have many more observations in the administrative records than we have in the surveys. Using the administrative normed data as the instrument we would be able to use

The remaining weakness in the twins 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, Bound and Solon note that differencing reduces the total variation and so the ratio of exogenous variation in within twin schooling differences may fall or rise. This would, even in the absence of measurement error, 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. The crucial assumption is that within-twin schooling differences are exogenous conditional on covariates and the family fixed effect. Neumark (1999) and Bound (1999) show that if differencing does not entirely remove ability differences then twins-based estimates of the return to education may be either more or less biased than OLS in cross-section data.
Unobserved differences may be associated with differential ability or with environmental differences - for example, if one twin experiences a different quality of teaching at school than the other. In both cases, equity considerations might motivate compensating actions by the parents. On the other hand efficiency considerations might possibly lead to actions that exacerbate the differences. In the first case, the OLS estimate of $\beta$ is biased downwards, while in the second case it would be upward biased. Although we might imagine that parents do compensate for adverse shocks, genetic or environmental, it would be surprising if it were perfect and the proponents of the twin 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 selected 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 differentiated observed variables does not imply that there are none with respect to unobservable differences. Moreover, measurement error may imply that these correlations in the observed characteristics in the data are themselves biased towards zero.

Cesarini et al (2014) uses twin data that includes IQ tested at age 18 as part of the military conscription process in Sweden. They use this to demonstrate that within-twin differences in IQ is a significant determinant of earnings differences and goes on to show that its inclusion induces a reduction in their, albeit biased, estimate in the schooling return by 18% - a large, and statistically significant, fall. They infer from this that the assumption of equal ability can be rejected. They note that their schooling return estimate in this case is biased because schooling differences are not exogenous. However, they are not able to use IQ difference as an instrument for schooling difference because IQ affects wages directly as well as through schooling. Thus, while Cesarini et al (2014) identify that there is an double trouble issue, at least in their Swedish data, they are not able to overcome it. To our knowledge Bonjour et al (2003) is the only twins paper to have previously attempted to overcome this issue.

---

13 Lundborg (2013) also notes that many of the factors that may vary within twin pairs, such as birth weight, early life health, parental treatment, and relation with parents, do not predict within-twin pair differences in schooling in the twins in the US MIDUS dataset.
empirically. They attempt to use discordance in in test scores as instruments – one at age 11 that determined the type of subsequent schooling, and a current reading score. There was almost no discordance in the first and the latter was found not to be correlated with education differences. They also considered smoking discordance and again found no significant effect on education differences.

4. Institutional Background and Data

Education is compulsory in Denmark for everyone between the ages of 6 or 7 to 16 for grades 0 to 9 (or, optionally, 10\textsuperscript{14}). Usually this takes place at a single junior school (Folkeskole) premises. Schooling is non-selective up to grade 9. While there has recently been an element of school competition the vast majority of children attend the school in their own catchment area and this competition does not affect the cohorts that we use here. The private sector educates only 15\% of children of this age and this is highly concentrated in the Copenhagen area. The private sector is effectively voucher financed because parents pay the difference between the fee and the costs of public schooling. The transition to High School, for grades 11 to 13, is at the recommendation of the junior school. Alternatively, junior school children may be recommended to move into vocational education at this stage. Admission to University from High School is determined solely by high school GPA. University education consists of a three year bachelor degree followed optionally, but usually until recently, by a two year Masters degree.

There are two education measures available to us. The first is a standard Statistics Denmark construction based upon the highest qualification recorded in the Education Ministry’s administrative records from which schooling duration is imputed by calculating the typical time required to take the most direct route to each possible highest qualification. We do this to maintain consistency with the existing twins literature that considers the effect of years of schooling rather than

\textsuperscript{14} In 1973/4 8\textsuperscript{th} grade was made compulsory and in 1974/5 9\textsuperscript{th} grade become compulsory but these reforms made no difference to the observed schooling because very few individuals left at such an early age. Denmark has an elective 10\textsuperscript{th} grade which is used as a second opportunity to enter high school for those that failed to do so directly from 9\textsuperscript{th} grade, and allows students destined for high school time to make a more informed choice of later specialization. The 10\textsuperscript{th} grade curriculum essentially broadens, but does not extend, the coverage of the 9\textsuperscript{th} grade one.
It is clear that there would be measurement error in such an imputation process - just as there is in self-reported schooling years in the Twinsburg data and many other datasets. The second measure is from self-reports of education from several large Twins Omnibus Surveys conducted by the Twins Registry. We can use the vector of qualifications available in the Elderly Twins Panel collected in 1995, 1999, 2001, and 2003, and from the Middle Aged Twins Survey of 1998. In addition, the most recent survey, the 2002 Twins Omnibus, contains self-reported years of schooling. The first Twins Omnibus Survey, of 29,344 twins was conducted in 1994, of twin birth cohorts from 1953 to 1983, which provides us with self-reported grade years of basic schooling (up to grade 9) and further qualifications including attending high school (from grades 11 to 13). The second Omnibus, in 2002, was sent to 46,418 twin individuals born between 1931 and 1983, and also provides details of highest qualification and years of basic schooling which we transform into a continuous measure of years of education. Two further twins surveys are available to us: the Middle Aged Twins (MAT) Survey of 1998 (for those born 1938-1952), and the Elderly Twins Panel (ETP) collected in 1993 for those aged 75 and above with subsequent waves in 1999, 2001 and 2003 which added those ages 70+. Thus there is a large overlap in the respondents to the 2002 Omnibus and the MAT Survey, and between the 2002 and 1994 Omnibuses, but the respondents to the ETP are not be to be found in any of the other surveys. None of the twin surveys contain cross-reports of the kind found in the Twinsburg data. Zygosity is ascertained from a separate questionnaire sent to all twins in Denmark by the Danish Twins Registry.

US research into education and earnings has often relied on the CPS data. In 1992 the old question on years of education, formed from grade years at school and years in college, has been replaced by highest attainment based on school grade and post-schooling qualifications. In 1997 the categorisation of qualifications was made more detailed. See Jaeger (2003) for analyses of the changes to CPS and its implications for estimated returns. In all CPS studies researchers have used the school grade plus years of college as a measure of the years of education. That is, they assume that education qualifications are completed in the shortest possible time. The Twinsburg research uses a survey modeled on the old CPS and so makes the same assumption.

There is considerable amount of grade repetition in Danish schools, especially for boys.

The Danish Twins Registry is the oldest survey based twin registry in the world. Christiansen et al (2006) 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. Zygosity was inferred from the answers, provided independently by the twins, to four questions. To classify twins as MZ required all four answers to be the same. DZ was assumed if none of the answers were the same. The small proportion who answered two or three questions identically were coded as XZ’s and we experimented with how to classify these. In the results reported here we dropped these XZ’s but the results were not sensitive to how they were classified. Details are available on request.
The survey data can be matched to the extensive administrative registers that Denmark has maintained since 1970.\textsuperscript{18} We have the choice of using the survey data to instrument the imputed schooling durations from the administrative register or, if we impute duration from the self-reported qualifications, \textit{vice versa}. Since, we regard the administrative data to be more accurate than the self-reports there would be an efficiency gain from doing the former. However, there is some advantage of adopting the latter approach. This would enable a split sample approach that uses the estimates of the first stage, based on the Omnibus Surveys, to predict the schooling differences of all twin pairs in the administrative data. Since the administrative data contains many more twin pairs than the twin surveys there would be a gain in precision from doing this.\textsuperscript{19}

The earnings data was only available from the administrative registers that contained tax returns for the period 1980 to 2011. This is the longest period over which consistent labour income information is available in Denmark from the tax return data. Tax filing is compulsory at the individual level. This information was extracted for twins aged 25 to 55 to avoid heavily censoring the data by late completed education and early retirement.\textsuperscript{20} We dropped the small proportion of observations where income was top-coded. We cannot compute an hourly (or even weekly) wage rate because our data on hours of work information is grouped\textsuperscript{21} and so, throughout, we use \textit{annual} real log gross income from work. We have restricted our sample to twin pairs where both are observed to be full-time full-year workers (annual

\textsuperscript{18} When the administrative data was first created it was populated from the 1970 census, the last ever conducted in Denmark. So the data for older individuals in the data is effectively a self-report. For those who had not completed their education by the 1970 census the data was populated from administrative records and the self-reported data was overwritten. The Census question was only coded for those with highest qualification dated 1920 and later. Since the most recent data comes from institutional sources we expect measurement error for observations beyond the mid 1970’s to be lower.

\textsuperscript{19} In addition, the duration of schooling can also be inferred directly from Ministry registers, independently of qualifications, using the age at which individuals first joined the labour market - but only for relatively young twins in our data. Here we would rely on information on the age at which individuals made their first payments into pension schemes, known collectively as ATP – something which is compulsory.

\textsuperscript{20} The official retirement age over this period was 67 and our restriction to 55 resulted in the loss of 15\% of the sample. The censoring at age 25 resulted in the loss of a further 6\% of observations.

\textsuperscript{21} Hours of work data 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.
hours at least 60% of full-time full-year hours) to reduce the impact of labour supply on earnings\textsuperscript{22} variation on our estimates.

Finally, we need to be conscious of the possibility of the availability of further instruments. In particular, although our data provides a rich variety of instruments that might be regarded as alternative measures of schooling duration the double trouble critique prompts us to consider the possibility of instrumenting within-twin differences in schooling.

We select all MZ twins and same sex DZ twins with: at least one (and as many as 25) years of earnings data for both twins from the administrative tax records who were born between 1966 and 2001 and aged 25 to 55 within that period; the self reported years of schooling from the 1994 Twins Omnibus which provides data on cohorts born from 1953 to 1982. After allowing for all of the selection criteria above, we have over one hundred thousand twin-pair*year observations where we have complete information. Approximately 40% of the total are MZ twins, and approximately 40% are female (because of their lower full-time employment rate).

Table 2 shows the basic descriptive statistics for individuals. MZ and same-sex DZ individuals are very 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 twins. Table 3 shows descriptive statistics for our samples of twin pairs. The important difference is in the education year differences: which average 1.37 years (1.33) for MZ and 2.04 (1.97) for DZ men (women).

Figure 1 shows the frequency of imputed education differences from the administrative data. There are a much higher proportion of MZ twins that have exactly the same imputed education length relative to DZs. Nonetheless, there are significant proportions of observations, for both MZs and DZs which have large differences in education. Figure 2 shows a scatter of the relationship between log earnings differences and the imputed education differences (with jitter). The relationship is clearly close to flat (with a slope coefficient of just 0.01), and this will be reflected later in low estimated returns using OLS in differences.

\textsuperscript{22} The joint full-time labour force participation rates (both members of the pair working full-time in the same year) for MZ twins (DZ twins) pair*years are 58.1% (54.7%) for females and 72.6% (67.9%) for males. Headline estimates using samples that up-rate earnings to full-time full-year are contained are very close to the ones reported below and are available on request.
Table 2 Means (standard deviations) for various samples of individuals

<table>
<thead>
<tr>
<th>Sample</th>
<th>Same sex twin pairs</th>
<th>Same sex DZ twin pairs</th>
<th>MZ twin pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEN Education years</td>
<td>12.49 (2.99)</td>
<td>12.41 (2.93)</td>
<td>12.62 (2.79)</td>
</tr>
<tr>
<td>Ln Earnings</td>
<td>12.681 (0.341)</td>
<td>12.679 (0.344)</td>
<td>12.684 (0.335)</td>
</tr>
<tr>
<td>Age</td>
<td>38.97 (8.24)</td>
<td>39.34 (8.18)</td>
<td>38.37 (8.29)</td>
</tr>
<tr>
<td>N * Years</td>
<td>63299</td>
<td>38929</td>
<td>24370</td>
</tr>
<tr>
<td>N</td>
<td>11438</td>
<td>7068</td>
<td>4370</td>
</tr>
</tbody>
</table>

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 2015 Danish kroner.

Table 3 Means (standard deviations) of within twin differences

<table>
<thead>
<tr>
<th>Sample</th>
<th>Same sex twin pairs</th>
<th>Same sex DZ twin pairs</th>
<th>MZ twin pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEN Education years</td>
<td>1.780 (2.230)</td>
<td>2.037 (2.321)</td>
<td>1.370 (2.010)</td>
</tr>
<tr>
<td>Ln Earnings</td>
<td>0.266 (0.284)</td>
<td>0.291 (0.299)</td>
<td>0.226 (0.251)</td>
</tr>
<tr>
<td>N</td>
<td>5719</td>
<td>3534</td>
<td>2185</td>
</tr>
</tbody>
</table>

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               | 4809                | 2809                   | 2000         |

Note: N refers to numbers of pairs. Earnings are January 2015 Danish kroner.

Figure 2 Distribution of Education Differences (Years)
A scatter of the relationship between the imputed years of schooling from the register data against imputed schooling duration from the self-reported qualifications in the survey datasets is shown in the Figure 3. Figure 4 shows the same relationship in differences. The size of the marker represents the sample weights and the solid line is the simple regression line. The coefficient on the self-report is close to one in Figure 3 suggesting that our instrument works quite well. The coefficient on the difference in schooling in Figure 4 is still close to one half.

*Figure 3*  \hspace{1cm}  *Differences in Log Earnings and Education (Years)*

![Figure 3](image1)

*Figure 4*  \hspace{1cm}  *Scatter of Education and Self-reported Education*

![Figure 4](image2)
The instrument that we exploit in an attempt to control for bias associated with remaining ability differences comes from the possibility for the twins to receive different schooling experience because their school contains more than one class per cohort. The number of classes in Danish schools is driven by cohort size and a Maimonides rule – an idea that was pioneered in Angrist and Lavy (1999) in their analysis of Israeli schools which used a rule (of 40) based on the teachings of Maimonides. The rule that Danish kommunes follow is seldom explicit but there is a federal rule that kommunes are bound by. While we do not observe the number of classes in each school attended by all individuals in our sample, we do know their school cohort size and impute the number of classes based on assuming that kommunes follow a rule of a maximum class size of 26 to ensure that they do not breach the federal rule (of 27). The distribution of eighth grade cohort size is shown in Figure 5 below. The corresponding proportion of twins in small 1-class schools is 22%. Many schools have policies on how to best educate twins: some teach twins together in the same class, some separate them, others let parents decide with advice form the school. We do not know what school policies are or how the twins are taught. We simply know that if the twins attend a one class cohort in which case they cannot be separated. Table 4 reports a breakdown of the education years differences by 1 and 2+ class schools as well as by zygosity. MZ twins have lower education
differences than DZ’s, more so for girls than boys. It is clear that attending a 1-class school is associated with a smaller difference in completed years of education compared to a 2+ class school. The differences remain even when we control for school size and rely only on the discontinuity associated with the rule. Note that it is usual for children in Danish schools to remain in the same class with the same teacher right up to eighth grade and for the class teacher to teach all subjects. Thus, experiencing different classes in the Danish schooling system could make for dramatic differences.

Figure 5    Distribution of 8th grade enrolment size

Table 4    Means (standard deviations) of within twin differences in education years

<table>
<thead>
<tr>
<th>Sample</th>
<th>All same sex twin pairs</th>
<th>Same sex DZ twin pairs</th>
<th>MZ twin pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MEN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.78 (2.23)</td>
<td>2.04 (2.32)</td>
<td>1.37 (2.01)</td>
</tr>
<tr>
<td>1 class school</td>
<td>1.66 (1.00)</td>
<td>1.83 (1.11)</td>
<td>1.201 (0.92)</td>
</tr>
<tr>
<td>2+ class school</td>
<td>2.02 (1.46)</td>
<td>2.30 (1.44)</td>
<td>1.51 (1.31)</td>
</tr>
<tr>
<td>N</td>
<td>5719</td>
<td>3534</td>
<td>2185</td>
</tr>
<tr>
<td></td>
<td>WOMEN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.71 (1.95)</td>
<td>1.97 (2.01)</td>
<td>1.33 (1.78)</td>
</tr>
<tr>
<td>1 class school</td>
<td>1.60 (0.88)</td>
<td>1.77 (0.89)</td>
<td>1.22 (0.89)</td>
</tr>
<tr>
<td>2+ class school</td>
<td>1.88 (0.96)</td>
<td>2.12 (0.78)</td>
<td>1.45 (1.01)</td>
</tr>
<tr>
<td>N</td>
<td>4809</td>
<td>2809</td>
<td>2000</td>
</tr>
</tbody>
</table>

Note: N refers to numbers of pairs.
5. Estimates

5.1 Headline Estimates

To facilitate comparison with other studies that are not twin-based, we begin by estimating the returns to schooling using the MZ twins as individuals, by gender, using a minimal specification that includes only schooling, age, age squared, and crude region of residence controls. We then instrument for measurement error in the administrative measure of schooling using own self-reported education from the survey data. Our data does not provide an instrument for the level of education years – there was a reform to the minimum school leaving age but by the time it was introduced it was irrelevant. And, while there was a school building program that will have changed proximities the rollout of this occurred too early to affect the education of the twins in our data. Thus, we are not able to produce estimates of the causal effect of education using individual data.

The results are presented in Table 5. 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 in the first row reflect earlier Danish OLS results in Christensen and Westergaard-Nielsen (2001) that suggest estimated returns of the order of 4%. We find OLS estimates of 3% (3.1%) for MZ (DZ) men and 3.7% (3.7%) for MZ (DZ) women. Moreover we have a large sample of non-twins drawn from the same administrative tax records and get identical, but even more precise results. We conclude that twins are the same as non-twins, and MZs are the same as DZs, as far as labour market outcomes are concerned. However, the results in the second row suggest considerable measurement error in the self-reported education measure that we use. Here we instrument the years of education in the self-reported education measure with the administrative data. Not surprisingly, the first stage estimates are very strong and the precision of the second stage estimates remains very high. However, the contrast between the OLS and IV results suggest low reliability in the levels of education attributed to individuals in the self-reported schooling data - around 55% for men and 60% for women.

We then move on to consider twins based FE estimation. We begin with OLS using twin differences. Not surprisingly, the OLS FE estimates in the first row of Table 6 are heavily attenuated although they are still significantly different from zero
The second row uses the differences in our alternative schooling year measure based on the administrative data. This corresponds to the typical twins methodology used in the literature. The MZ columns would be those that would be highlighted in the literature on the grounds that they instrument with an alternate education difference to control for measurement error bias. Our prior is that the DZ results would still be contaminated by some ability bias and the coefficients would be higher than the corresponding MZ estimates. This is true for men, but not significantly for women.

**Table 5  Estimated coefficients on education levels (pooled DZ and MZ twins)**

<table>
<thead>
<tr>
<th></th>
<th>MEN</th>
<th></th>
<th>WOMEN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td>0.030</td>
<td>0.031</td>
<td>0.037</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>IV for measurement error</td>
<td>0.065</td>
<td>0.066</td>
<td>0.071</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0014)</td>
<td>(0.0012)</td>
<td>(0.0017)</td>
</tr>
</tbody>
</table>

Note: Standard error in parentheses clustered at family level. Controls for age, age squared, and region of residence included.

**Table 6  FE estimated coefficients on education levels (DZ and MZ twins)**

<table>
<thead>
<tr>
<th></th>
<th>MEN</th>
<th></th>
<th>WOMEN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td>0.005</td>
<td>0.018</td>
<td>0.009</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>IV for measurement error</td>
<td>0.055</td>
<td>0.095</td>
<td>0.061</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>IV for endogenous ΔS</td>
<td>0.051</td>
<td>0.059</td>
<td>0.060</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.023)</td>
</tr>
</tbody>
</table>

Note: Standard error in parentheses clustered at family level. First stage estimates for the final row include school cohort size (when in eight grade) and a dummy variable for a 1-class school. The F statistic from this is over 20 for men and women, and for MZ and DZ.

The third row is motivated by the different ability critique and attempts to instrument for any ability bias that remains after within-twin differencing using proximity; as well as the measurement error instrument. Our prior is that instrumenting would reduce the estimated coefficients, more so for DZs than MZs. The results in the third row does suggest lower estimates than the second row for the MZ men, but not significantly so; while the women exhibit an insignificant fall. However, the DZ results are much higher than the second row. The final MZ results are not significantly different from the conventional FE estimates that are highlighted in previous literature – although this is not the case for the DZ results. We conclude
that, although we have a strong instrument for $\Delta S$ instrumenting makes no difference to the FE MZ results - at least in the Danish case the second of the double trouble criticisms turn out to be empirically unimportant.

5.2 The effect of time and cohort

There has been a great deal of interest in the evolution of inequality over time and the role that variation in the returns to education plays in this. However, this literature has never considered the traditional argument that the virtue of MZ twin estimates is that they are purged of unobservables and therefore reflect only the return to observed skills. In contrast same sex DZ twin estimates (or OLS estimates in general) reflect returns to both observed and unobserved skills. Therefore the difference between MZ and DZ twins estimates (or between MZ twin and OLS estimates in general) is informative about the returns to unobserved skills multiplied by the ability bias term. This ability bias term is the ratio of a covariance between observed and unobserved skills to the variance in observed skills. The former is, of course, unobservable. However, this term reflects the selectivity of the education system and so is a constant (at least for a given cohort of individuals who all faced the same system). The within cohort variation across time tells us how the returns to unobservable skill has changed since the constant differences out.

Figure 6  DZ and MZ FE estimates over time

Note: Averages of moving 10 year window on the data.
6. Conclusions

In this study we present estimates of the returns to schooling based on a large sample of Danish twins. Earnings and education data are drawn from population administrative registers over a 23 year period and we merge this with information from two surveys that provide self-reports on education and allows us to distinguish between identical and same-sex fraternal twins. We present baseline estimates that suggest that applying OLS to the pooled cross section data results in downwards biased estimates because of measurement error. Predictably, we then find that the simple fixed effects estimators are biased downwards to a much larger degree. When we instrument the within-twin-pair estimates to counter the measurement error, our results are dramatically higher for male and female MZs (and for female DZs but not for male DZs). Thus, our modelling resembles the earlier US research by Ashenfelter and Krueger that found fixed effect instrumental variables estimates for MZ twins that were larger than the corresponding cross section OLS.

The first contribution of the paper is to show that there is substance to the ability difference critique. The existing literature is largely uninformative about this important issue but several papers have suggested this could be an issue and recent research provides a convincing test that rejects the equal ability assumption. This recent work then obliges researchers to address the extent of the resulting bias. We do this using IV estimation where our instruments are based on a reform whose implementation was delayed and so rolled out across time. It mandated additional provision for post-compulsory education and so allowed for a different impact within twins that gave rise to discordant schooling. When we apply IV to our conventional FE twins estimates which instruments for measurement error we find statistically significantly smaller male MZ returns – suggesting that differencing alone does not remove all ability bias (at least for men). This finding suggests that existing MZ twins studies do not reveal the causal effect and that the effects of unequal ability could be important.

Our second contribution is to exploit the panel element of the data to produce estimates by cohort and calendar time. Our aim is to contribute to the debate about the impact of the evolution of returns on overall inequality. The MZ FE results here suggest rising returns, for both men and women, since the early 80’s. The DZ results suggest higher returns rising at a lower rate. The difference between the two falls
slowly for men but remains approximately constant for women. A test that the returns to unobserved skills remain constant over the period cannot be rejected by the estimates.
References


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


## Appendix

**Table A1**  
Tests for Table 7: IV Estimates in Levels

<table>
<thead>
<tr>
<th></th>
<th>MEN</th>
<th>WOMEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td>OLS</td>
<td>0.0312</td>
<td>0.0341</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>IV for measurement error</td>
<td>0.0616</td>
<td>0.0661</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>IV for measurement error and ability bias</td>
<td>0.0604</td>
<td>0.0671</td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td>(0.0018)</td>
</tr>
</tbody>
</table>

Note: Standard error in parentheses. Controls for age, age squared, and region of residence included. The second row uses self-reported schooling as an IV, and the final row adds the own proximity controls.

**Table A2**  
Tests for Table 8: IV Estimates in Differences

<table>
<thead>
<tr>
<th></th>
<th>MEN</th>
<th>WOMEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td>OLS</td>
<td>0.0066</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>IV for measurement error</td>
<td>0.0779</td>
<td>0.1040</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0077)</td>
</tr>
<tr>
<td>IV for measurement error and ability bias</td>
<td>0.0572</td>
<td>0.0901</td>
</tr>
<tr>
<td></td>
<td>(0.0097)</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>IV for measurement error and ability bias</td>
<td>0.0334</td>
<td>0.0406</td>
</tr>
<tr>
<td></td>
<td>(0.0219)</td>
<td>(0.0125)</td>
</tr>
</tbody>
</table>

Note: Standard error in parentheses. Controls for age, age squared, and region of residence included. The second row uses self-reported schooling as an IV, the third row adds the own proximity controls, and the final row adds the own and spouse’s proximity controls.

First stages?