THE CO-TWIN METHODOLOGY AND RETURNS TO SCHOOLING - TESTING A CRITICAL ASSUMPTION*

ÖRJAN SANDEWALL

DAVID CESARINI

MAGNUS JOHANNESSON

Abstract

Twins-based estimates of the return to schooling feature prominently in the labor literature. Their validity hinges critically on the assumption that within-pair variation in schooling is explained by factors which are unrelated to wage earning ability. This paper develops a framework for testing this assumption, and finds, using a unique dataset of monozygotic twins, strong evidence against it. Differences in adolescent IQ test scores predict differences in educational attainment and including IQ in the wage equation causes within-pair point estimates for the returns to schooling to decline significantly. Our results thus cast doubt on the validity of twins-based estimates.

Keywords: I21, J24.

JEL Codes: returns to schooling, twins, equal ability assumption.

---

*Örjan Sandewall is at NERA Economic Consulting and and the Research Institute of Industrial Economics (IFN). David Cesarini is at the Department of Economics, Massachusetts Institute of Technology and the Research Institute of Industrial Economics (IFN). Magnus Johannesson is at the Department of Economics, Stockholm School of Economics. We thank the Jan Wallander and Tom Hedelius Foundation, the Swedish Research Council, and the Swedish Council for Working Life and Social Research for financial support. Please direct all correspondence to Örjan Sandewall: orjan.sandewall@nera.com

1 INTRODUCTION

The returns to schooling is a much-studied parameter in labor economics. Knowing the causal effect of schooling on earnings and other economic outcomes has important implications for educational policy, for efforts to better understand the evolution of inequality and for studies examining the sources of economic growth (Card, 1995; Katz and Autor, 1999). Yet, it has long been known that efforts to obtain precise estimates of the causal effect of schooling on earnings are complicated by the endogeneity of schooling decisions. In particular, there is a widely shared view that estimates of the marginal returns to schooling will be biased unless proper account is taken of heterogeneities in latent ability. If the propensity to invest in further years of education is also directly related in a positive way to the ability to earn wages, then this will cause an upward bias in estimates of the effect of an additional year of schooling on wages (see for example Card, 1999).

A number of approaches to removing or mitigating this endogeneity problem have been proposed. One strand of work uses instrumental variable analysis to try to reduce the bias of the estimates (Angrist and Krueger, 1991; Card, 1995).[1] A second influential strand of the literature has exploited within-family variation in general, and variation within monozygotic (MZ) twin pairs in particular, to try to control for unobserved ability. Under the key identifying assumption that ability is common among siblings of the type at hand, this allows for consistent estimates, as long as problems of measurement errors in the schooling variable can be dealt with adequately. Especially with regards to MZ twins, the attraction of the assumption of equal ability is easily understood. MZ twins are the result of a fertilized egg splitting in two shortly after conception, resulting in two genetically identical individuals. Furthermore, MZ twins (or "identical" twins, as they are often referred to) are typically raised by the same parents, go to the same school, and are influenced by the same peer groups when growing up.

In labor economics, twins-based estimates of the return to schooling have featured prominently. For example, Card (1999) devotes a substantial section of his survey of the literature on the returns to schooling to a largely favorable review of the twins-based estimates. But he also cautions that the entire enterprise hinges crucially on the assumption that identical twins have identical abilities. Consequently, researchers with different priors about the validity of the assumption have arrived at very different conclusions about how twins-based estimates of the return to schooling ought to be interpreted. Our hope is that this paper will introduce more empirical rigor into the debate about the validity and usefulness of what we dub the equal ability assumption.

The idea that the latent wage earning ability of two individuals in a pair of identical twins would be virtually identical is not hard to accept, a priori. However, identical ability begs the question of what causes observed within-pair differences in schooling, as standard optimising models predict that two identically able individuals would choose the same level of schooling (Ashenfelter and Rouse, 1998; Becker, 1964; Ben-Porath, 1967). Any observed variation in schooling must then be explained by "optimizing errors", or differences in preferences for schooling which do not affect wage earning ability. Hence, it is assumed that differences in schooling across the population are caused by ability differences, but that this is not true within twin pairs. Such an assumption is difficult to reconcile with a vast behavior genetic literature which finds that though MZ are remarkably similar in a number of domains, there are also systematic and stable within-pair differences on variables such as intelligence and personality (Plomin *et al.*, 2001).[2]

A natural hypothesis is that within-pair variation in intelligence may explain within-pair variation in schooling, thereby violating the assumption of "optimization error". This potential problem with the co-twin methodology was first demonstrated by Griliches (1979); although twins may have very similar levels of ability, the observed similarities in years of schooling and income are also large. Therefore, even though within-pair differences are purged from most of the heterogeneities in ability, they also lack most of the useful variation in schooling and income. Griliches (1979) noted that when the degree of twin similarity is the same for ability and for schooling, first-differencing contributes nothing in terms of removing ability bias. This critique has been further developed both conceptually and empirically by Bound and Solon (1999), who also point out that *a priori* the relationship between the degrees of similarity in ability and schooling, respectively, is not clear. There is a literature outside economics which reports associations between birthweight and educational attainment within twin pairs, see the review in Bound and Solon (1999). Bound and Solon note that such findings are suggestive of within pair endogeneity in schooling decisions.

This paper develops a framework for testing the *equal ability assumption*. To test the assumption, we use a unique dataset of Swedish twins matched to conscription records, administrative income records kept by Statistics Sweden and survey data collected by the Swedish Twin Registry. The conscription data contains information on performance on a test of cognitive ability taken at the age of eighteen. The fact that the test is taken at the age of eighteen renders it less plausible that schooling differences are causing IQ differences, and not the other way round. The main findings of the paper are that within-pair differences in IQ are significantly associated with income even after accounting for differences in schooling, that within-pair differences in IQ have a statistically and economically significant effect on within-pair differences in schooling, and finally that inclusion of IQ reduces within-pair estimates of returns to schooling by about 15% across various specifications and variable definitions. These results cast doubts on the validity of the co-twin approach to estimating the returns to schooling, and provide some empirical evidence for the critique of within-family estimation advanced by Griliches (1979) and developed by Bound and Solon (1999) in the context of twins-based estimates. The evidence reported here suggests that the quasi-experiment of MZ twinning does not approximate the ideal experiment, namely random assignment of educational attainment holding ability and other background factors constant, particularly well. In fact, under plausible assumptions about the reliability ratio of the within-pair difference in IQ and educational attainment, the within-pair correlation between IQ and schooling is about 0.30.

An additional concern about twins-based estimates relates to measurement error in schooling. As was noted by one of the first authors to apply this methodology (Taubman, 1976), differencing within-pairs decreases the signal to noise ratio, and hence serves to exacerbate the problem of imperfectly observed schooling. Furthermore, even with valid instruments for number of years spent in an educational facility, this quantity may not perfectly reflect true education, a distinction pointed out at least as early as in Griliches (1977). Studies of such heterogeneities in the production function for human capital abound, see for example Sacerdote (2001), on peer group effects and Rivkin *et al.* (2005) on teacher quality. In this paper, we follow Isacsson (1999) and use administrative data on educational attainment as an instrument for self-reported educational attainment. As the data of this study present limited opportunity to examine the issue of mismeasured education, the twin methodology will be given the benefit of the doubt; the assumption of perfectly instrumented schooling will be maintained, and focus is instead directed towards the source of the alleged benefits from using twins data - the equal or virtually equal ability within twin pairs.

The paper proceeds as follows. In the next section, empirical findings from the co-twin literature on estimating the returns to schooling are briefly reviewed. Thereafter a framework for examining the equal ability assumption is outlined in Section 3, followed by a presentation of the data in Section 4. The results from the main analysis are presented in Section 5, and in Section 6 we provide the results of several robustness checks. The consistency of the data with two additional restrictions is considered and rejected in Section 7. A discussion of the main findings is provided in Section 8, after which Section 9 concludes.

2 Previous Literature

Behrman and Taubman (Behrman and Taubman, 1976; Taubman, 1976) pioneered the use of data on twins for studying the returns to schooling. Examining within-pair differences in annual earnings and schooling among male twin veterans in the NAS-NRC dataset, Taubman (1976) found evidence of substantial upward ability bias in traditional cross-sectional estimates of the returns to schooling. Taubman's (1976) estimates decreased from 8.8% to 4.8% when moving from regression on the cross-section to within-pair estimation, despite correcting for an assumed 10% measurement error in the schooling data. The results in Behrman and Taubman (1976) imply similarly that standard OLS estimates are considerably upward biased.

The co-twin approach experienced a revival in the 1990s, following the innovation by Ashenfelter and Krueger (1994) to collect data on both own schooling and co-twin's schooling from each individual in the sample. Having two measures of schooling, they then use the first-difference of schooling reported by one member of a pair as an instrument for the first-difference reported by the other member. If measurement errors are uncorrelated, this allows for a correction of the problem of measurement error in the schooling variable. Under the equal ability assumption, their approach thus provides a consistent estimate of the returns from schooling.

Ashenfelter and Krueger's (1994) within-pair IV estimates were, surprisingly enough, considerably higher than standard least squares estimates on the cross-section. However, later studies strongly suggest that these initial results were due to sampling variation, as analyses of extensions of this sample produced within-pair IV estimates that were not higher than conventional cross-sectional estimates (Ashenfelter and Rouse, 1998; Rouse, 1999). These later findings are consistent with most other co-twin studies (Miller *et al.*, 1995; Isacsson, 1999; Behrman and Rosenzweig, 1999, Bonjour *et al.*, 2003), who likewise find only a small upwards ability bias.[3]

Two recent additions to the co-twin literature are Isacsson (2004) and Zhang *et al.* (2007). Isacsson (2004) has the benefit of working with a representative dataset comprising education and income data for a very large number of Swedish monozygotic twins born 1926-1958, 2609 pairs in total, and is therefore able to provide precise estimates of non-linearities in returns to schooling, and to allow for non-classical errors in the measurements of schooling. Zhang *et al.* (2007) analyse a dataset of 914 pairs of Chinese monozygotic twins and find that the returns to schooling during the Cultural Revolution (defined as 1966-1976 in their study) was roughly the same as that of later cohorts. In both these studies, the implied ability bias in cross-sectional estimates is positive.

3 EMPIRICAL FRAMEWORK

3.1 An Augmented Co-Twin Model

Consider the following simple model of wage determination, drawing on Card (1999):

$$y_{ij} = \alpha_y + \beta S_{ij} + \gamma A_{ij} + u_{ij}, \tag{1}$$

where y_{ij}, S_{ij} and A_{ij} are income in natural logarithms, years of schooling, and ability, respectively, for individual i of twin pair j, and where the ordering of the individuals in a twin pair is random. Returns to schooling, β, and the conditional return to ability, γ, are assumed to be equal across individuals. Let latent ability, A, be defined widely enough to allow S and u to be independent, and be measured in standard deviations about the population mean. Finally, α_y varies with a quadratic in the age of the individual, to capture experience and cohort-specific effects. Furthermore, assume the following causal model of schooling,

$$S_{ij} = \alpha_S + \delta A_{ij} + \epsilon_{ij}, \tag{2}$$

where ϵ is a summary measure of all determinants of schooling which are exogenous to the unobservables of the wage equation. Extend this exogeneity to apply across twins within a pair, so that $Corr(A_{ij}, \epsilon_{kj}) = 0$ and $Corr(u_{ij}, \epsilon_{kj}) = 0$, $\forall\ i, j$. Specify the sign of ability such that $\delta > 0$. Notice that this assumption is without loss of generality because A is not observed. Therefore, our approach does not make any assumptions about the direction of the ability bias.

To capture cohort-specific effects, the intercept again varies with a quadratic function of age. Let the ability of a twin be statistically related to the ability of his co-twin in the following manner:

$$A_{1j} = \phi A_{2j} + \alpha_{1j}. \tag{3}$$

Here, ϕ is the correlation between the abilities of each twin and his co-twin, and α_{1j} is uncorrelated with A_{2j} by construction. Equivalently, ϕ is the share of variance in ability explained by a variance factor common to both twins. Furthermore, assume that differences in ability within pairs are independent of all other errors (u, ϵ, and τ (below)). The main identifying assumption of the literature on estimating the returns to schooling using variation within twin pairs, is that twins have identical latent abilities such that $A_{1j} = A_{2j}$. In the above framework, this translates to assuming $\phi = 1$, which in turn implies $Var(\alpha) = 0$ due to the random ordering of twins. Under $\phi = 1$, consistent estimates of β can be obtained by estimating the model in first-differences,

$$\Delta y_j = \beta_{FD} \Delta S_j + \gamma_{FD} \Delta A_j + \Delta u_j, \tag{4}$$

where $\Delta y_j \equiv y_{1j} - y_{2j}$ and similarly for the explanatory variables. Since ΔA_j is a zero vector under the standard twin assumption, the within-pair difference in income can simply be regressed on the within-pair difference in schooling,

$$\Delta y_j = \beta^{-}_{FD} \Delta S_j + \Delta u^{-}_j. \tag{5}$$

This is the basic idea behind all within-pair estimators in the literature. The aim of this study is to determine whether $\phi = 1$. For this purpose, consider IQ measured at around the age of 18, and specify its relationship with ability as follows:

(6) $$T_{ij} = \pi A_{ij} + \tau_{ij},$$

where τ_{ij} is independent of A_{ij}. Let T_{ij} be measured in standard deviations about the population mean, and $\pi > 0$. Finally, let y_1 refer to own income, as opposed to y_2 for co-twin's income, and similarly for S, A, T, u, ϵ, and τ, so that $(y_1)_{ij} = (y_2)_{kj}; \forall i \neq k$. When not specified, as above, y refers to own income, y_1.

3.2 Two Tests of the Basic Twin Assumption

3.2.1 Auxiliary Assumptions A

Assume $Corr(u_1, \tau_1) = Corr(u_1, \tau_2) = 0$. Estimate the equation,

(7) $$\Delta y_j = \beta \Delta S_j + \lambda_1 \Delta T_j + \Delta u_j^*,$$

where the error term is,

(8) $$\Delta u_j^* = -\lambda_1 \Delta T_j + \gamma_{FD} \Delta A_j + \Delta u_j.$$

If $\phi = 1$, then $\Delta A_j = 0$ and $\Delta T_j = \Delta \tau_j$, and consequently $\lambda_1 = 0$. Furthermore, β and λ_1 are consistently estimated since $\lambda_1 \Delta T_j = 0$ and $\gamma_{FD} \Delta A_j = 0$, and hence independent of ΔS_j and of ΔT_j. The distribution of $\hat{\lambda}_1$ is different under the null and the alternative hypothesis. It follows that $\hat{\lambda}_1$ is a valid test statistic for the null hypothesis that $\phi = 1$. Measurement error in schooling can be dealt with using an alternative measure of schooling as an instrument, the approach championed in this literature by Ashenfelter and Krueger (1994), assuming, of course, that the exclusion restriction is satisfied.

3.2.2 Auxiliary Assumptions B

Alternatively, assume $Corr(\epsilon_1, \tau_1) = Corr(\epsilon_1, \tau_2) = 0$, and relax the above assumptions on $Corr(u_1, \tau_1)$ and $Corr(u_1, \tau_2)$. Let λ_2 be defined by the following estimating equation

$$\Delta S_j = \lambda_2 \Delta T_j + \Delta \epsilon_j^*, \tag{9}$$

where, analogously, the error term is,

$$\Delta \epsilon_j^* = -\lambda_2 \Delta T_j + \gamma_{FD} \Delta A_j + \Delta \epsilon_j. \tag{10}$$

If $\phi = 1$, then $\Delta A_j = 0$ and $\Delta T_j = \Delta \tau_j$, and consequently $\lambda_2 = 0$. Furthermore, λ_2 is consistently estimated since $\lambda_2 \Delta T_j = 0$ and $\gamma_{FD} \Delta A_j = 0$, and hence independent of ΔT_j. The distribution of $\hat{\lambda}_2$ is different under the null and the alternative hypothesis. It follows that $\hat{\lambda}_2$ is a valid test statistic for the null hypothesis that $\phi = 1$, under this alternative restriction on the error terms.

3.3 Remark

If Auxiliary Assumptions A or B (or both) hold, then it follows that the estimated return to education should change significantly when including IQ as a covariate in the fixed effects wage equation only if the equal ability assumption is invalid. Denote the coefficient on schooling in the fixed effects regression without IQ included by β_1 and denote the coefficient on schooling in the regression with IQ included by β_2. A simple bootstrap procedure to test the hypothesis that the difference in estimated coefficients is not purely due to sampling variation is as follows. First, draw 10000 pseudo-samples of twin pairs with replacement. For each bootstrap draw, estimate β_1 and β_2. An n-percent confidence interval for the quantity $\beta_1 - \beta_2$ can then be constructed by extracting the $\frac{n}{2}$th and $\left(100 - \frac{n}{2}\right)$th percentile of the empirical distribution of $\beta_1 - \beta_2$ obtained from the bootstrap draws.

4 DATA

The dataset links information from the Swedish Twin Registry with administrative data from Statistics Sweden and Swedish enlistment records. The Swedish Twin Registry contains virtually all twins born in Sweden from 1926 and onwards, and is kept mainly for the purpose of performing epidemiological studies (see Lichtenstein *et al.*, 2006) for a description of the Swedish twin registry). The survey data used in this paper was collected in 1998-2002 (the "SALT" survey) from twins born 1950-1958, and in 2005-2006 (the "STAGE" survey) from twins born in 1959-1975. Response rates were 74% and 60%, respectively. Only data on monozygotic twins (about one quarter of the sample) is used, where zygosity has been determined by the Swedish Twin Registry using a battery of questions relating to physical similarity. The validity of this method of determining zygosity has been repeatedly estimated to be 95-98% (Lichtenstein *et al.*, 2002). The dataset is restricted to individuals born between 1950 and 1975.[4] The cohort studied is hence sufficiently old so that income is observed at a point in the lifecycle where research has shown that annual income is a good proxy for lifetime earnings (Böhlmark and Lindquist, 2006).

4.1 Education Data

The data contains two measures of educational achievement. One is a self-reported measure from the survey data collected by the Swedish Twin Registry. The other is based on administrative data from 2005. The self-reported data consists, for the SALT cohort, of an indicator of highest attained qualification, and for the STAGE cohort, of total years of schooling at the different levels of the education system. For the SALT cohort, years of schooling are assigned based on the standard years of schooling associated with the degree in question. The administrative data contains highest degree attained. Years of schooling based on the survey data are used as the explanatory education variable, with degree dummies based on administrative sources used as instruments.

4.2 Income Data

Data on income consists of yearly taxable earnings in 2005 as reported by employers to the tax authorities. The income measure used in this paper ("sammanräknad förvärvsinkomst") is defined as the sum of income earned from wage labor, income from own business, pension income and unemployment compensation. Capital income is not included in the measure. In the main specification, only pairs where both twins in a pair had an income exceeding SEK 70,000 (exchange rate 2005; $1=SEK 7.5) are included, in an attempt to capture only individuals working full-time so that income more or less corresponds to hourly earnings. The practice of either excluding data not corresponding to full-time work or using information on hourly wages is followed by practically all previous studies of the returns to schooling using twins back to at least Ashenfelter and Krueger (Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998; Behrman and Rosenzweig, 1999; Bonjour *et al.*, 2003; Isacsson, 1999; Isacsson, 2004; Miller *et al.*, 1995; Rouse, 1999; Zhang *et al.*, 2007).[5]

4.3 IQ Data

All Swedish men are required by law to participate in military conscription at or around the age of 18. Until 1999, exceptions were only granted to men with serious documented psychological or physical handicaps. The actual drafting procedure can span several days during which a number of tests are administered to the conscripts. These include; assessments of medical status, physical stamina, muscular strength, eyesight, cognitive ability and psychological aptitude. This paper uses the military data on cognitive ability. As the normal school starting age in Sweden is seven, the average individual in the main sample would have taken the test about one year prior to finishing high school.

The IQ test used by the Swedish military is a fairly standard test of general intelligence (Spearman, 1904). An early version of the test was developed during World War II, and it has subsequently been revised on seven occasions (Carlstedt, 2000). Its basic structure has, however, remained unchanged during the study period considered in this paper. Recruits take four subtests (logical, verbal, spatial and technical) which, for most of the study period, are graded on a scale from 0 to 40. Carlstedt (2000) discusses the history of psychometric testing in the Swedish military and provides evidence on the psychometric properties of the test. Test scores are normalized by year using all observations in the dataset for which there are test scores, and the sum across test scores is then used as the raw IQ measure[6]. This raw measure is then normalized against all observations in the dataset, to allow for an approximation of population standard deviations to be used as the metric for IQ.

Using IQ test scores which were gathered not in a school environment, but under the considerably different conditions of military conscription, reduces the risk that the test scores pick up factors related to, *i.e.*, a general affinity with school-like tests that yet do not translate into wage earning capacity. Using the terminology of the empirical framework outlined above, this renders it more plausible that $Corr(\epsilon_1, \tau_1)$ is zero.

4.4 Representativeness

The total sample size was determined as follows: Out of the 31824 respondents to the STAGE and SALT surveys in our cohorts, 3522 were male monozygotic twins of which 2753 had data on education from both administrative and survey data. Of these, 2353 had non-missing income, and 2288 had an income above 70000 SEK, the cut-off used to eliminate observations whose income unambiguously did not derive from full-time employment. Among these, 2129 individuals had valid IQ test scores from enlistment data. Finally, 1780 of these observations were from complete pairs of twins, *i.e.* where the co-twin was also in the sample.

Before turning to the main results, some comments on the representativeness of the sample are in order. In Table I the main sample is compared to the national average with regards to income, education, marital status and age. For IQ, the norm group is the approximately 12000 twins born between 1950 and 1975 who responded to the SALT or STAGE survey and for whom there is IQ data. For all other variables, the comparison is made to the population data from Statistics Sweden. Income in the sample is about 20% higher than in the general population. Both education and age are slightly higher in the sample than in the national average, but these differences are small. Oversampling of twins with better than average education and income was also reported by Ashenfelter and Kruger (1994) and Ashenfelter and Rouse (1998).

The main purpose of this study is not to generalize from a sample of twins to a population of non-twins, but rather from one sample of twins to other samples of twins. Therefore, it is also important to know how representative the dataset is of the datasets of twins used hitherto. Table II compares parameters from our dataset to parameters reported previously in the literature.

The first two parameters concern similarity between twins. In our data, measured years of schooling correlate 0.73 between a twin and his co-twin, a figure in line with what has been reported in the literature. Furthermore, results on IQ test scores correlate 0.82, which again is a standard degree of similarity (Bouchard and McGue, 1981). The next two parameters concern the structure of the measurement errors in reported years of schooling. In our sample, the reliability ratio[7] is 0.88, which is very similar to those reported in previous twin studies. The reliability ratio of the within-pair differences is 0.65, which is closer to the lower than to the higher estimates reported in Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998). The observed within-pair reliability ratio in the data is also close to that expected based on the cross-sectional reliability ratio and the twin correlation in schooling, as reported above. If all measurement errors are classical, the imputed within-pair reliability ratio would thus be 0.58.[8]. Note also that the cross-sectional reliability ratio of 0.88 implies, under classical errors, a within-pair correlation in schooling of 0.82 (0.73/0.88) when correcting for measurement errors. As shown by Griliches (1979), co-twin estimators are less biased than cross-sectional estimators if and only if ϕ is greater than the similarity in schooling, *i.e.* in this dataset 0.82.

The final four parameters concern impacts on wages (in logarithms), and as such we would expect them to vary depending on institutional factors in the countries where they are measured. The first parameter, β_{IV}, is a simple cross-sectional estimate of the returns to schooling in our sample. The second parameter is the within family estimate of the return to schooling in the sample of MZ twins. In both cases, to try to adjust for measurement error, a full set of dummy variables on educational attainment based on the administrative data are used as instruments for self-reported educational attainment.

The estimated returns to schooling from cross-sectional data are slightly lower than those found in studies from US and UK, but slightly higher than those of Isacsson (1999) using Swedish twins. However, Isacsson's (1999) sample includes both men and women, whereas our estimates are for men only. Our data yields larger differences between within-pair estimates and cross-sectional estimates than what is commonly found in twin studies. Notice that the result from Isacsson (1999) was constructed using an imputed within-pair measurement error, and as such is not strictly comparable to the other figures which apply instrumental variables techniques to correct for measurement error.

The final two parameters in Table II concern the relationship of IQ with labor market outcomes. The standardized regression coefficient in a regression of log income on the IQ test score is 0.16, *i.e.* an increase in IQ of one standard deviation is associated with an increase in income of about 16% in our sample of monozygotic twins. Bowles and Gintis (2002), based on a meta-study of 24 studies on US data, report an average coefficient of 0.27. This discrepancy corresponds reasonably to differences in income dispersion between US and Sweden, as reported in Gottschalk and Smeeding (1997).

Finally, the correlation between self-reported schooling and measured IQ is 0.51, a figure roughly in line with the average of 0.55 reported by Neisser *et al.* (1996) in an authoritative report on the state of intelligence research. It should be noted that the latter figure is based on IQ test scores from early years, mainly primary school. The fact that the correlation with schooling is lower in our data suggests that simultaneity in test scores, whereby differences in schooling cause differences in IQ, is not a major concern.

5 RESULTS

We now turn to our main results. In Figures 1-3, we report the three most important bivariate relationships in the data. Figure 1 is a scatter diagram of the intrapair difference in income against the intrapair difference in schooling. It is clear from the figure that a large number of identical twins do indeed have identical levels of educational attainment, and that within pair variation in educational attainment is associated with within pair variation in earnings. Figure 2 is a scatter diagram of the intrapair difference in IQ against the intrapair difference in schooling and the relationship is positive. Figure 3 shows that there is also positive a relationship between IQ and income within pairs.

In examining these figures, it is useful to recall that the signal-to-noise ratio is lower within family than it is in the cross-section. Assuming classical measurement error and a cross-sectional reliability ratio for IQ of 0.9 implies a measurement-error corrected within-pair correlation of educational attainment and IQ of 0.30.[9] This number is in and of itself a quite damning indictment of co-twin methodology. And yet, it does not take into account concerns about the validity of IQ as a measure of actual ability and is in this sense a lower bound on the extent to which these estimates are plagued by endogeneity problems. To develop this point more formally, in columns 1 through 5 of Table III, we report, for the sample of identical twins in our sample, the full set of regressions of income on educational attainment and IQ test scores. All regressions have family fixed effects, so the only source of variation is the within-family differences. Standard errors are clustered at the family level.

5.1 The Partial Effect of IQ in the Within-Pair Wage Equation (Auxiliary Assumptions A)

Columns 1 and 2 show the results from the regression of income on schooling, with and without the set of administrative dummies used as instruments. As expected, the attempted measurement error correction raises the estimated return to schooling. Columns 3 and 4, show the results from an augmented model with IQ included as a control. The coefficient on IQ demonstrates that within-pair differences in IQ have a direct relationship with income differences, and that this relationship is statistically significant and strong. The magnitude of the coefficient implies that a twin with an IQ one population standard deviation higher than his co-twin, has an income which is on average 7.4% higher than his co-twin, when controlling for schooling. The coefficient on schooling drops from 3.4% to 2.9%, or by about 15%.[10] Under the assumptions underpinning this test ($Corr(u_1, \tau_1) = Corr(u_1, \tau_2) = 0$), the hypothesis that $\phi = 1$ is hence rejected. Using the previously described bootstrapping procedure, the null hypothesis that the schooling coefficients are the same in the specifications with and without IQ included can be rejected at the one percent level ($p<0.01$).

5.1.1 Interpretation

If $\phi = 1$, the data is only consistent with the model if differences within twin pairs in the errors in test scores, $\Delta\tau$, are related both to $\Delta\epsilon$ and to Δu. In other words, under the standard twin assumption of equal ability, differences in test scores, ΔT, must be correlated both with $\Delta\epsilon$ (unobservable preferences for schooling which are not directly related to wage earning capacity), and with Δu (unobservable capacity to earn wages which is not related to schooling), yet be uncorrelated with ΔA (actual unobservable ability).

5.2 The Impact of IQ in the Within-Pair Schooling Equation (Auxiliary Assumptions B)

Column 5 of Table IV contains the results from the second simple test of $\phi = 1$, based instead on the assumption that $Corr\left(\epsilon_1, \tau_1\right) = Corr\left(\epsilon_1, \tau_2\right) = 0$. The estimated within-pair relationship with IQ is statistically significant and large and a difference of one population standard deviation is associated with a difference of 0.52 years of schooling[11]. Under these alternative test assumptions, $\phi = 1$ is rejected as well.

6 Robustness

There are a number of legitimate concerns which may be raised with regards to the above findings. In this section, five such issues are presented and addressed. This is followed by a summary of the results of these robustness checks.

6.1 Misclassification

Some of the twins in the sample may have been misclassified as monozygotic twins despite being in fact dizygotic twins. If ability differences are for some reason relatively less familial (*i.e.*, compared to the family share of variance of the exogenous determinant of schooling) in dizygotic twins, this will cause the above findings to be overstated. To examine this issue, the 5% of pairs which were the most dissimilar with respect to IQ were dropped and the main equations were re-estimated. This is a conservative test in that no more than 2-5% of monozygotic twins are normally misclassified as dizygotic using the type of classification algorithm employed by the Swedish Twin Registry (Lichtenstein *et al.* (2002)). It should also be noted that all twin studies referred to in the literature review above have employed similar classification algorithms as does the Swedish Twin Registry.

6.2 IQ Construction

To examine the sensitivity of our findings to variations in the construction of the aggregate test score, a so called factor "g", *i.e.* the first principal component, was calculated from the four subtests of the IQ test. This measure was standardized by year against all twins for whom there was data on IQ, and used as an alternative measure of IQ.

6.3 Choice of Instrument

As a further robustness check, the roles of instrument and regressor were reversed for the two sources of schooling data. As the administrative data, which were used as instruments in the main analysis, consist of dummy variables for highest degree attained, they were converted into years of schooling using population averages estimated by Isacsson (2004)[12].

6.4 Full-Time Threshold

Finally, the sensitivity of the main results to variations in the threshold on yearly earnings was examined, by applying alternative thresholds of 50,000 and 180,000 Swedish krona (about $6,700 and $24,000, respectively). Regarding the lower threshold, it should be noted that it corresponds to a full-time hourly wage of about $3.4, *i.e.* implausibly low in the context of Sweden. Furthermore, because of the logarithmic conversion of wages, the 24 observations below the lower threshold are between 4 and 10 standard deviations away from the mean (in a sample of around 2000). The lower threshold is indeed very low for the purposes of approximating a full-time proxy.

6.5 Endogeneity of IQ Test Scores

The IQ measure is taken at about the age of 18, after individuals have completed compulsory schooling but before they enter college. The fact that the IQ tests are taken at a relatively early age renders it less likely that the differences in test scores are endogenous to differences in educational attainment. Yet, there is evidence suggesting that differences in education can drive differences in test scores (Cascio and Lewis, 2006). This argument is particularly compelling for twin pairs where at least one twin has less than 12 years of schooling and hence either failed to complete high school or only completed a two-year high school curriculum. As a crude robustness check, we therefore restrict the sample to individuals whose education was still ongoing when they took the test, and rerun the analyses.[13]

6.6 Results

In Tables IV to VII we present results from the two simple tests of the *equal ability assumption*, under the alternative samples and variable specifications described in the previous section. In all cases, the coefficient on T, the variable for IQ, is statistically and economically significant in both the wage equation and the schooling equation. The estimated return to schooling also declines when IQ is included as a covariate.

The results obtained when excluding the most dissimilar pairs, and when using the alternative measure of IQ, are reported in Table IV. Exclusion of the dissimilar pairs with respect to IQ yields a higher estimated effect of IQ on income (0.100), and a lower estimated effect of IQ on schooling (0.322). However, neither of these results is far from the other robustness outcomes or the main results.

Excluding the 5% most dissimilar pairs also has as its effect that the coefficient on schooling in the wage equation, with IQ included, is less precisely estimated, and is in fact just shy of significance at the five percent level. However, the point estimate does not change much. The results for the alternative measure of IQ are highly similar to the results from the main specification in Table III.

In Table V, we report the results from regression models with the instruments interchanged and in Table VI, we report results with different income thresholds than SEK 70 000. The results are quite insensitive to these alternative specifications. Finally, Table VII shows the results omitting twin pairs where at least one sibling failed to complete three years of high school. The coefficient of schooling in the wage equation is estimated with less precision, but the point estimate is very similar to the other specifications. The coefficient on IQ remains highly significant, with a slightly *higher* point estimate than in the other samples.

Finally, for all alternative samples, the null hypothesis that the schooling coefficients are the same in the specifications with and without IQ included can be rejected at the five percent level.[14]

7 Evaluating Additional Model Restrictions

This section expands on the previous analysis by considering whether there are additional restrictions on the variance-covariance matrix of the errors which are consistent with the data but which have not been imposed thus far.

7.1 Is IQ a (nearly) Perfect Measure of Ability?

The precision with which the test score, T, proxies for ability, A, can be crudely evaluated by considering whether the variance of τ is zero:

$$\sigma_\tau^2 = 0. \tag{11}$$

A simple way of testing this is to consider the best linear within-pair predictor of schooling using test scores, as presented in Table III:

$$\Delta S_j = \lambda_3 \Delta T_j + \Delta \epsilon_j^*, \tag{12}$$

where, by construction, $E(\Delta T_j, \Delta \epsilon_j^*) = 0$. If $Var(\tau) = 0$, then $\lambda_3 = \delta$, and an efficient estimator of λ_3 is given by the best cross-sectional predictor of schooling:

$$S_{ij} = \alpha_S + \lambda_4 T_{ij} + \epsilon_j^*. \tag{13}$$

Hence, if $Var(\tau) = 0$, then $\lambda_3 = \lambda_4$, where λ_4 is efficient. Under the alternative hypothesis that $Var(\tau) \neq 0$, only λ_3 is consistent. Applying random effects GLS, $\hat{\lambda}_4 = 1.24$ with a standard error of 0.06 , to be compared to $\hat{\lambda}_3$ which is 0.52, with a standard error of 0.11[15].The null hypothesis of $Var(\tau) = 0$ is rejected at the 0.1% level using a standard Hausman (1978) test.

7.2 Is Schooling as Familial as Ability?

As pointed out earlier, Griliches (1979) established that if the proportion of the variance in ability which is explained by a common family component is identical to the degree of within-pair similarity in the non-ability determinants of schooling (ϵ), then the within-pair estimator of returns to schooling is equally biased as the cross-sectional estimator. In the above presented framework, this condition amounts to the restriction that:

$$\phi = Corr(\epsilon_1, \epsilon_2). \tag{14}$$

In analogy with the previous section, consider the best linear within-pair predictor of income using schooling:

$$\Delta y_j = \lambda_5 \Delta S_j + \Delta u_j^*, \tag{15}$$

where, by construction, $E(\Delta S_j, \Delta u_j^*)$. If $\phi = Corr\,(\epsilon_1, \epsilon_2)$, then $\lambda_5 = (\beta + \gamma r)$, where r is the coefficient on schooling when regressing ability on schooling and a quadratic in age. In other words, as an indicator of β, λ_5 is equally biased as the standard cross-sectional OLS estimator of income on schooling in the cross-section. Consequently, an efficient estimator of λ_5 is given by the best cross-sectional predictor of income:

$$y_{ij} = \alpha_y + \lambda_6 S_{ij} + u_{ij}^* \tag{16}$$

Hence, if $\phi = Corr\,(\epsilon_1, \epsilon_2)$, then $\lambda_5 = \lambda_6$, where λ_6 is efficient. Under the alternative hypothesis that $\phi \neq Corr\,(\epsilon_1, \epsilon_2)$, only λ_5 is consistent. Using a random effects IV estimator to account for measurement error in schooling yields $\hat{\lambda}_6 = 0.068$ (0.005)[16], to be compared with the fixed effects estimator $\hat{\lambda}_5$, which as reported in Table III is 0.034 (0.012). The null hypothesis of $\phi = Corr\,(\epsilon_1, \epsilon_2)$ can be rejected at the 0.2% level using a standard Hausman (1978) test.

8 Discussion

The main finding in the previous sections is that the assumption of equal ability within pairs of monozygotic twins is violated in our sample. Within-pair variation in IQ test scores predicts within-pair variation in schooling and including within-pair variation in IQ in the fixed effects regressions lowers the estimated return to schooling by approximately fifteen percent. These results are robust across different alternative samples and likely understate the extent of the bias, for two reasons. First, even if the schooling instruments are valid, differencing of the IQ test scores obviously exacerbates the problem of errors in variables (Griliches, 1979). Additionally, the estimated decline in the schooling coefficient does not take into account concerns about the validity of IQ tests as measures of actual ability and in this sense likely understates the extent to which the twins-based estimates are plagued by endogeneity problems.

There are indications that the *equal ability assumption* is violated even in the previous co-twin literature on returns to schooling. For example, Ashenfelter and Rouse (1998) report that for pairs where two twins had obtained different levels of schooling, 11% stated a reason which could be directly interpreted as indicating an ability difference (such as "one twin was better at books"). Bonjour *et al.* (2003) provide evidence that ability differences do matter, as out of the 38% of twins who went to different classes, half indicated ability differences as the reason. In both of those studies however, these findings are interpreted as providing support for the idea that ability differences are relatively unimportant.[17]

The only directly comparable finding that we are aware of is in Griliches (1979), who reports a regression coefficient of just 0.13 for the within-pair effect of one standard deviation in IQ on years of schooling, based on a small sample of just 76 pairs of male monozygotic twins. In our much larger and more representative sample, this figure is 0.52, suggesting the problem is quite serious. Though we have reported evidence suggesting that the co-twin approach to estimating the returns to schooling produces biased estimates, it does not necessarily follow that the entire enterprise should be abandoned. For example, in their otherwise quite critical assessment of the co-twin method, Bound and Solon (1999) suggest that although we do not know whether ability is more familial than is schooling, within-pair estimates can still be used as an upper bound on the returns to schooling, under the assumption that ability bias is positive as is commonly thought (Bound and Solon (1999))[18].

Given that within-pair IV estimates are generally lower than the cross-sectional OLS estimates, co-twin estimates then contain information allowing us to tighten the bounds on the possible values of the returns to schooling. However, the central premises of this type of bounds argument, that ability bias can be taken to be positive *a priori* and that the suitability of an identification method therefore can be determined on the basis of the results it provides - if lower than OLS, then accept as an improvement - is dubious from a methodological perspective. Furthermore, as Bound and Solon (1999) note, such reasoning naturally rests on the assumption that the instruments for schooling are valid. Such an assumption is far from innocuous, and the potential reduction in bias must be weighed carefully against the plausibility of this assumption. At the very least, the results reported here suggest that it was probably premature of Greene to claim that the co-twin approach "ameliorates" (Greene, 2003, p. 381) the problems with standard cross-sectional regressions.

A crucial question is to what extent these findings generalize to other countries. Returns to schooling vary across countries (Harmon *et al.* (2003)), and in principle so could the importance of ability bias. However, the argument in this paper rests not mainly on the magnitudes of the estimates from the wage equation, but on (i) the direct within-pair association of IQ with wages holding schooling constant, and (ii) on the significant within-pair associations between IQ and schooling. As long as within-pair dynamics are not substantially different in Sweden compared to other countries, these empirical findings constitute a general methodological argument against using within-twin differences for the estimation of returns to schooling, and in favor of placing greater weight on estimates based on alternative identification approaches.

9 Conclusion

Monozygotic twins' schooling decisions have been used in a number of prominent papers to estimate the returns to schooling. The key identifying assumption in these studies is that within-pair variation in schooling is explained by factors which are unrelated to wage earning ability. Using a unique dataset of 890 pairs of male monozygotic twins' schooling, income and adolescent IQ test scores, this paper finds strong evidence against the *equal ability assumption*. Within-pair differences in IQ test scores, obtained around the age of eighteen, are found to be a significant predictor of both income and schooling differences. Controlling for within-pair IQ differences reduces the estimated returns to schooling by about 15%. These results hence challenge the usefulness of the empirical results derived from the co-twin literature.

10 REFERENCES

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106(4): 979–1014.

Ashenfelter, Orley, and Alan B. Krueger. 1994. "Estimating the Returns to Schooling Using a New Sample of Twins." *American Economic Review*, 84(5): 1157-1173.

Ashenfelter, Orley, and Cecilia Rouse. 1998. "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins." *Quarterly Journal of Economics*, 113(1): 253-84.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. New York: Columbia University Press for the National Bureau of Economic Research.

Behrman, Jere R., and Mark Rosenzweig. 1999. "'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates." *Economics of Education Review*, 18(2): 159-67.

Behrman, Jere R., and Paul Taubman. 1976. "Intergenerational Transmission of Income and Wealth." *American Economic Review*, 66(2): 436-40.

Ben-Porath, Yoram,. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4): 352-65.

Böhlmark, Anders , and Matthew J. Lindquist. 2006. "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics*, 24(4): 879-900.

Bonjour, Dorothe, Lynn F. Cherkas, Jonathan E. Haskel, Denise D. Hawkes, and Tim D. Spector. 2003. "Returns to Education: Evidence From UK Twins." *American Economic Review*, 93(5): 1799-1812.

Bouchard, Thomas J. Jr., and Matt McGue. 1981. "Familial Studies of Intelligence: A Review." *Science*, 212(4498): 1055-1059.

Bound, John, and Gary Solon. 1999. "Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling." *Economics of Education Review*, 18(2): 169-82.

Bound, John and David A. Jaeger, "On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist and Krueger's 'Does Compulsory School Attendance Affect Schooling and Earnings?'" NBER Working Paper No. 5835, 1995.

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 Association*, 90(430), 443–450.

Bowles, Samuel and Herbert Gintis,. 2002. "The Inheritance of Inequality." *Journal of Economic Perspectives*, 16(3): 3-30.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of labor economics, Vol. 3A*, ed. Orley Ashenfelter and David Card. Amsterdam, NL: North Holland.

Card, David, "Estimating the Return to Schooling: Progress on Some Persistent Problems." NBER Working Paper No. 7769, 1995.

Carlstedt, Berit, 2000, *Cognitive Abilities: Aspects of Structure, Process and Measurement* (Göteborg, Sweden: PhD dissertation, University of Gothenburg, 2000).

Cascio, Elizabeth U. and Ethan G. Lewis. 2006. "Schooling and the Armed Forces Qualifying Test." *Journal of Human Resources*, 41(2): 294-318.

Gottschalk, Peter, and Timothy M. Smeeding. 1997. "Cross-National Comparisons of Earnings and Income Inequality." *Journal of Economic Literature*, 35(2): 633-687.

Greene, William H. 2003. *Econometric Analysis: Fifth Edition*. Upper Saddle River, New Jersey: Prentice-Hall Inc.

Griliches, Zvi. 1997. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica*, 45(1): 1-22.

Griliches, Zvi. 1979. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy*, 87(5): S37-S65.

Harmon, Colm P., Hessel Oosterbeek, and Ian Walker. 2003. "The Returns to Education: Microeconomics." *Journal of Economic Surveys*, 17(2): 115–156.

Hausman, Jerry. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6): 1251-1271.

Isacsson, Gunnar. 1999. "Estimates of the Return to Schooling in Sweden from a Large Sample of Twins." *Labour Economics*, 6(4): 471-89.

Isacsson , Gunnar. 2004. "Estimating the Economic Return to Educational Levels Using Data on Twins." *Journal of Applied Econometrics*, 19(1): 99-119.

Katz, Lawrence F., and David H. Autor. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of labor economics, Vol. 3A*, ed. Orley Ashenfelter and David Card. Amsterdam, NL: North Holland.

Lichtenstein, Paul, Ulf de Faire, Birgitta Floderus, Magnus Svartengren, Pia Svedberg, and Nancy L. Pedersen. 2002. "The Swedish Twin Registry: A Unique Resource for Clinical, Epidemiological and Genetic Studies." *Journal of Internal Medicine*, 252(3): 184-205.

Lichtenstein, Paul, Patrick F. Sullivan, Sven Cnattingius, Margaret Gatz, Sofie Johansson, Eva Carlström, Camilla Björk, Magnus Svartengren, Alicja Volk, Lars Klareskog, Ulf de Faire, Martin Schalling, Juni Palmgren, and Nancy L. Pedersen. 2006. "The Swedish Twin Registry in the Third Millennium : An Update." *Twin Research and Human Genetics*, 9(6): 875-882.

Miller, Paul, Charles Mulvey, and Nick Martin. 1995. "What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings." *American Economic Review*, 85(3): 586-99.

Molenaar, Peter C. M., Dorret I. Boomsma, and Conor V. Dolan. 1993. "A third source of developmental differences." *Behavior Genetics*, 23(6): 519–524.

Neisser, Ulric, Gwyneth Boodoo, Thomas J Bouchard. Jr., A. Wade Boykin, Nathan Brody, Stephen J. Ceci, Diane F. Halpern, John C. Loehlin, Robert Perloff, Robert J. Sternberg, and Susana Urbina, 1996. "Intelligence: Knowns and Unknowns." *American Psychologist*, 51(2): 77-101.

Plomin, Robert D., John C. DeFries, Gerald E. McClearn, and Peter McGuffin, *Behavioral Genetics* 4th ed. (New York, NY: Freeman, 2001).

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417-458.

Rouse, Cecilia,. 1999. "Further Estimates of the Economic Return to Schooling from a New Sample of Twins." *Economics of Education Review*, 18(2): 149-57.

Sacerdote, Bruce. 1976. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116(2): 681-704.

Spearman, Charles. 1904. "General Intelligence, Objectively Determined and Measured." *The American Journal of Psychology*, 15: 201-292.

Taubman, Paul. 1976. "Earnings, Education, Genetics, and Environment." *Journal of Human Resources*, 11(4): 447–461.

Turkheimer, Erik. 2004. "Spinach and Ice Cream: Why Social Science is So Difficult." In Behavior Genetics Principles: Perspectives in Development, Personality, and Psychopathology. ed. Lisabeth F. DiLalla. Washington, D.C.: American Psychological Association.

Zhang, Junsen, Pak-Wai Liu, and Linda Yung. 2007. "The Cultural Revolution and Returns to Schooling in China: Estimates Based on Twins." *Journal of Development Economics*, 84(2): 631-639.

Notes

[1]For critiques of the instrumental variable approach, see Bound, Jaeger and Baker (1995) and Bound and Jaeger (1996).

[2]The evidence of such differences is quite conclusive, as studies have repeatedly found that the reliability of various IQ tests significantly exceed the correlation in the test scores of two identical twins. These difference between identical twins may stem from neurodevelopmental noise (Molenaar *et al.*, 1999) or other, etiologically relevant, stochastic developmental processes that affect one twin and not the other (Plomin *et al.*, 2001; Turkheimer, 2004).

[3]For two older summaries of the literature on returns to schooling using twins data, see Card (1999) and Behrman and Rosenzweig (1999).

[4]1950 is the first year for which we have data on IQ test scores.

[5]It can also be noted that due to the logarithmic transformation, some outliers in the full dataset are more than 10 standard deviations lower than the average.

[6]Assigning equal weights to each sub-test is in accordance with the standard practice of the Swedish Armed Forces.

[7]With classical measurement errors, the reliability ratio is the square root of the R^2 from a regression of measured years of schooling on its instruments. If there is only one instrument, this is equivalent to the correlation, as stated in the table.

[8]The imputed within-pair reliability ratio is $(r - Corr(S_1, S_2))/(1 - Corr(S_1, S_2))$, where r is the reliability ratio based on cross-sectional measures.

[9]The imputed within-pair correlation is derived as $\frac{Corr(\Delta S^1, \Delta T)}{\sqrt{\rho_{\Delta S^1}}\sqrt{\rho_{\Delta T}}}$, where $\rho_{\Delta S^1}$ and $\rho_{\Delta T}$ are the reliability ratios of the two respective first-differenced variables, as derived in footnote 8.

[10]It should be noted that since T is an imperfect measure of ability, the estimated returns to schooling are biased and inconsistent when $\phi \neq 1$, i.e. when the equal ability assumption is violated.

[11]This can be compared with the sample standard deviation of schooling of 2.6 years.

[12]Isacsson (2004) examined a representative sample with high quality data on years of schooling and regressed this on the same type of administrative data that are used in this paper.

[13]In principle, it is also possible that that differences in IQ test scores are caused by differences in program choices in high school. Unfortunately, we do not have any data that would allow us to examine this possibility.

[14]The difference is significant at the one percent level in four out six cases, the two exceptions being the high threshold sample and the sample without high school dropouts. For those two samples, the hypothesis is only rejected at the five percent level.

[15]This is numerically identical to $\hat{\lambda}_2$ as reported in Table III although the interpretations are different.

[16]Note that the 7.2% reported in Table II is not based on a random effects estimator.

[17]Ashenfelter and Rouse (1998) also report that the within-pair variation in schooling is uncorrelated with both order of birth and spousal education, and take this lack of correlation as indirect evidence that determinants of MZ in educational attainment are random.

[18]We emphasize that Bound and Solon (1999) do not strictly advocate the co-twin approach, but merely point out this logical implication.

11 Tables and Figures

TABLE I.
SUMMARY STATISTICS AND SAMPLE REPRESENTATIVENESS

	Monozygotic Twins	Population
Income (in SEK 1000)	360	298
S.D.	(228)	-
Schooling (in years)	12.9	12.2
S.D.	(2.6)	-
IQ	0.12	0.00
S.D.	(0.92)	(1.00)
Age (in years)	42.9	42.3
S.D.	(7.6)	-
1 if Married	0.51	0.45
S.D.	0.50	-
# Observations	1780	-

Note: Income is in thousand SEK and is defined as the sum of income earned from wage labor, income from own business, pension income and unemployment compensation. The schooling data for the sample is based on self-reported education. IQ is measured in standard deviations. IQ is standardised against the group of all twins, including opposite sex twins, in the dataset for whom there is conscription IQ data (12366 observations in total). All other population variables are constructed using data from Statistics Sweden for the universe of men in Sweden aged 30-55 years in 2005.

TABLE II.
COMPARABILITY WITH PREVIOUS LITERATURE

	Sample	Literature	Country/ies	Source
CO-TWIN SIMILARITY				
Corr(S^1_1,S^1_2)	0.73	0.66	US	AK1994
		0.75	US	AR1998
		0.70	Australia	M1995
Corr(T_1,T_2)	0.82	0.86	Various	BM1981
EDUCATION INSTRUMENTS				
Cross-Sectional Reliability Ratio	0.88	~0.90	US	AK1994
		0.92	US	AR1998
		0.88	Australia	M1995
		0.88	Sweden	I1999
Within Family Reliability Ratio	0.65	0.57-0.83	US	AK1994
		0.62-0.76	US	R1999
LABOUR MARKET				
β_{IV}	7.2%	~8%	US, UK.	C1999, B2003
		6.4%	Australia	M1995
		5.2%	Sweden	I1999
$\beta_{FE,IV}$	3.4%	~7%	US, UK	C1999, B2003
		4.5%	Australia	M1995
		4.2%	Sweden	I1999
$\delta y/\delta T$	0.16	0.27	US	BG2002
$Corr(S^1,T)$	0.51	0.55	US	N1996

Notes: S^1 is self reported schooling, T is measured IQ, y is log income. Subscripts refer to a twin's order in a pair. The sample "correlation" of schooling and instrument for schooling was derived as the square root of the R^2 when regressing self-reported years of schooling on the set of administrative schooling dummies used as instruments. β_{IV} is the regression coefficient from the cross-sectional regression of log income on schooling (S^1), using a set of dummies on educational attainment categories from administrative records as instruments. $\beta_{FE,IV}$ is the within-family estimate of the return to schooling, using the within pair difference in the set of dummy variables as instruments. $\delta y/\delta T$ is the standardized regression coefficient in the regression of log income on measured IQ.

Abbreviations of sources: AK1994 - Ashenfelter and Kruger (1994); AR1998 - Ashenfelter and Rouse (1998); M1995 - Miller et al. (1995); BM1981 - Bouchard and McGue (1981); I1999 - Isacsson (1999); R1999 - Rouse (1999); C1999 - Card (1999); B2003 - Bonjour et al. (2003); BG2002 - Bowles and Gintis (2002); N1996 - Neisser et al. (1996). Card (1999), Bowles and Gintis (2002) and Neisser et al. (1996) are all surveys or meta-analyses.

TABLE III.
RESULTS OF THE TWO TESTS OF THE EQUAL ABILITY ASSUMPTION

Dependent Variable	(1) FE, IV Income	(2) FE Income	(4) FE/IV Income	(5) FE Income	(6) FE Schooling
Schooling	0.034***	0.024***	0.029**	0.021**	-
s.e.	(0.013)	(0.008)	(0.013)	(0.008)	-
IQ	-		0.074***	0.078***	0.517***
s.e.	-		(0.026)	(0.026)	(0.135)
Schooling Instrumented?	Yes	No	Yes	No	-
Family Fixed Effects?	Yes	Yes	Yes	Yes	Yes
R^2	0.009	0.011	0.019	0.020	0.022
# Observations	1780	1780	1780	1780	1780
Groups	890	890	890	890	890

Notes: Standard error within parentheses, clustered at the family level. Adminstrative dummies for highest degree attained are used as instruments for years of schooling. All regressions are run with family fixed effects. Three stars (***) denote statistical significance at the one percent level, two stars (**) denote statistical significance at the five percent level and one star (*) denotes statistical significance at the ten percent level.

TABLE IV.
ROBUSTNESS CHECKS OF THE FIXED EFFECTS REGRESSIONS

Dependent Variable		(1) Income	(2) Income	(3) Schooling	Pairs	# Obs
Exclude 5%	Schooling	0.030**	0.025*	-	846	1692
	s.e.	(0.015)	(0.014)	-		
	IQ	-	0.100***	0.322**		
	s.e.	-	(0.034)	(0.130)		
	R^2	0.002	0.014	0.007		
Alternative IQ	Schooling	0.034**	0.028**	-	890	1780
	s.e.	(0.013)	(0.013)	-		
	IQ	-	0.076***	0.522***		
	s.e.	-	(0.026)	(0.135)		
	R^2	0.009	0.019	0.023		

Notes: "Exclude 5%" are the results from the baseline specification with the 5% of twin pairs most dissimilar on IQ omitted. "Alternative IQ" are results from the baseline specification with IQ defined as the principal component of the four subtests. Standard error within parentheses, clustered at the family level. Adminstrative dummies for highest degree attained are used as instruments for years of schooling. All regressions are run with family fixed effects. Three stars (***) denote statistical significance at the one percent level, two stars (**) denote statistical significance at the five percent level and one star (*) denotes statistical significance at the ten percent level.

TABLE V.
ROBUSTNESS CHECKS OF THE FIXED EFFECTS REGRESSIONS:
INSTRUMENTS REVERSED

Dependent Variable		(1) Income	(2) Income	(3) Schooling	Pairs	# Obs
Regr⇆Instr	Schooling	0.035***	0.031**	-	890	1780
	s.e.	(0.012)	(0.012)	-		
	IQ	-	0.071*	0.600***		
	s.e.	-	(0.027)	(0.140)		
	R^2	0.007	0.017	0.026		

Notes: This table reports results from the baseline specification but with the instruments interchanged. The administrative data which were used as instruments in the main analysis consist of dummy variables for highest degree attained. Thy were converted into years of schooling using population averages estimated by Isacsson (2004). Standard error within parentheses, clustered at the family level. All regressions are run with family fixed effects. Three stars (***) denote statistical significance at the one percent level, two stars (**) denote statistical significance at the five percent level and one star (*) denotes statistical significance at the ten percent level.

TABLE VI.
ROBUSTNESS CHECKS OF THE FIXED EFFECTS REGRESSIONS:
DIFFERENT TRESHOLDS

Dependent Variable		(1) Income	(2) Income	(3) Schooling	Pairs	# Obs
Low Threshold	Schooling	0.034**	0.029**	-	895	1790
	s.e.	(0.013)	(0.014)	-		
	IQ	-	0.071***	0.510**		
	s.e.	-	(0.027)	(0.135)		
	R^2	0.011	0.019	0.022		
High Threshold	Schooling	0.026**	0.023**	-	791	1582
	s.e.	(0.011)	(0.012)	-		
	IQ	-	0.059***	0.430**		
	s.e.	-	(0.023)	(0.122)		
	R^2	0.005	0.015	0.015		

Notes: This table reports from our baseline specification with thresholds set at 50,000 ("Low Threshold") and 180,000 ("High Threshold") instead of 70,000. Standard error within parentheses, clustered at the family level. Adminstrative dummies for highest degree attained are used as instruments for years of schooling. All regressions are run with family fixed effects. Three stars (***) denote statistical significance at the one percent level, two stars (**) denote statistical significance at the five percent level and one star (*) denotes statistical significance at the ten percent level.

TABLE VII.
ROBUSTNESS CHECKS OF THE FIXED EFFECTS REGRESSIONS:
EXCLUDING HIGH-SCHOOL DROPOUTS

Dependent Variable		(1) Income	(2) Income	(3) Schooling	Pairs	# Obs
	Schooling	0.041*	0.034	-	453	906
	s.e.	(0.022)	(0.022)	-		
	IQ	-	0.129***	0.368**		
	s.e.	-	(0.042)	(0.169)		
	R^2	0.007	0.024	0.011		

Notes: This table reports results from our baseline specifcation excluding twin pairs where at least one member completed less than 12 years of schooling. Standard error within parentheses, clustered at the family level. Adminstrative dummies for highest degree attained are used as instruments for years of schooling. All regressions are run with family fixed effects. Three stars (***) denote statistical significance at the one percent level, two stars (**) denote statistical significance at the five percent level and one star (*) denotes statistical significance at the ten percent level.

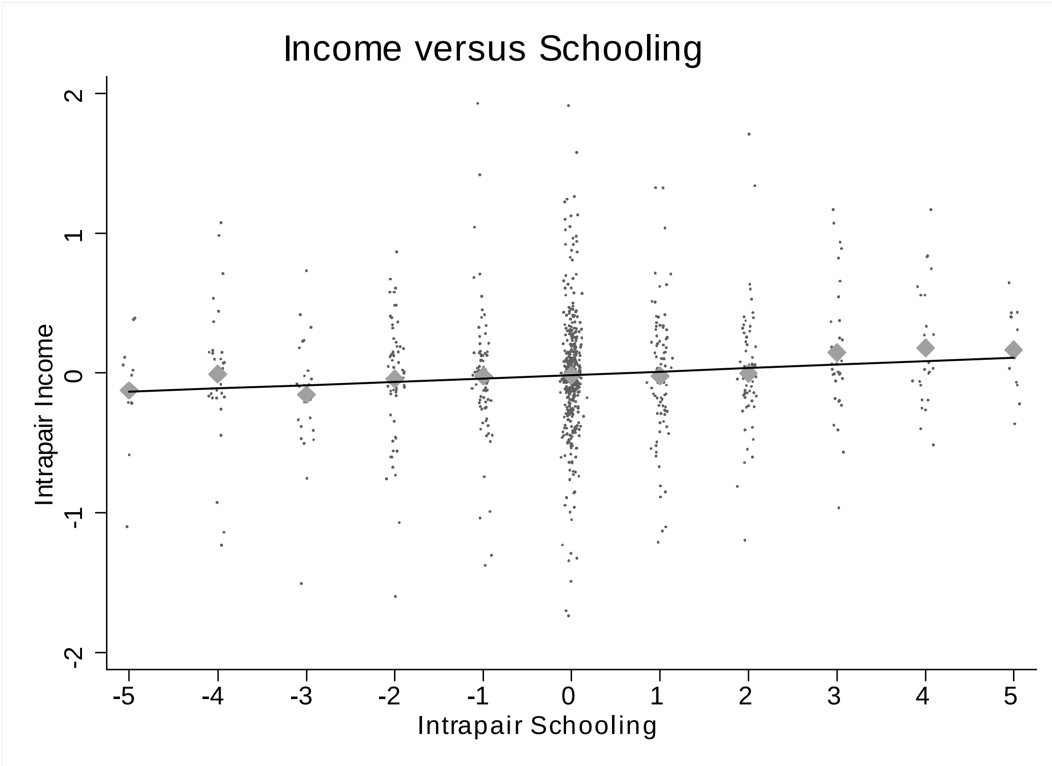

Figure I. intrapair difference in schooling plotted against intrapair difference in education. diamonds denote the mean ln income difference by category. for expositional clarity, the education variable has been jittered.

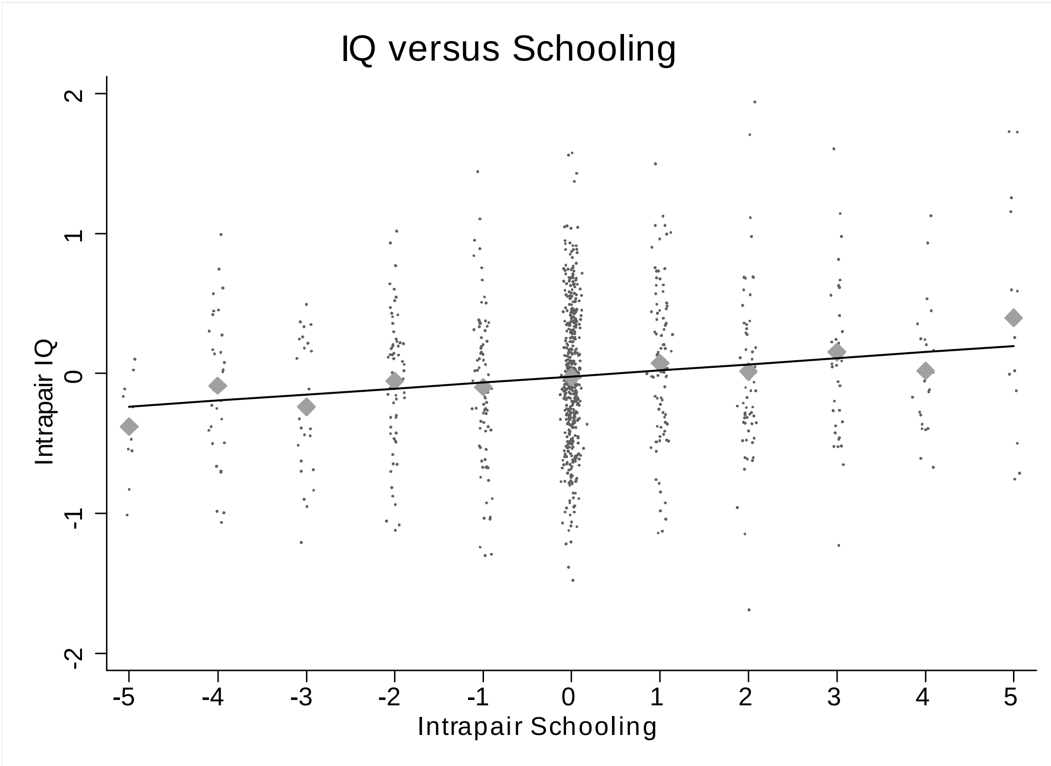

Figure II. intrapair difference in schooling plotted against intrapair difference in iq. diamonds denote the mean ln income difference by category. for expositional clarity, the education variable has been jittered.

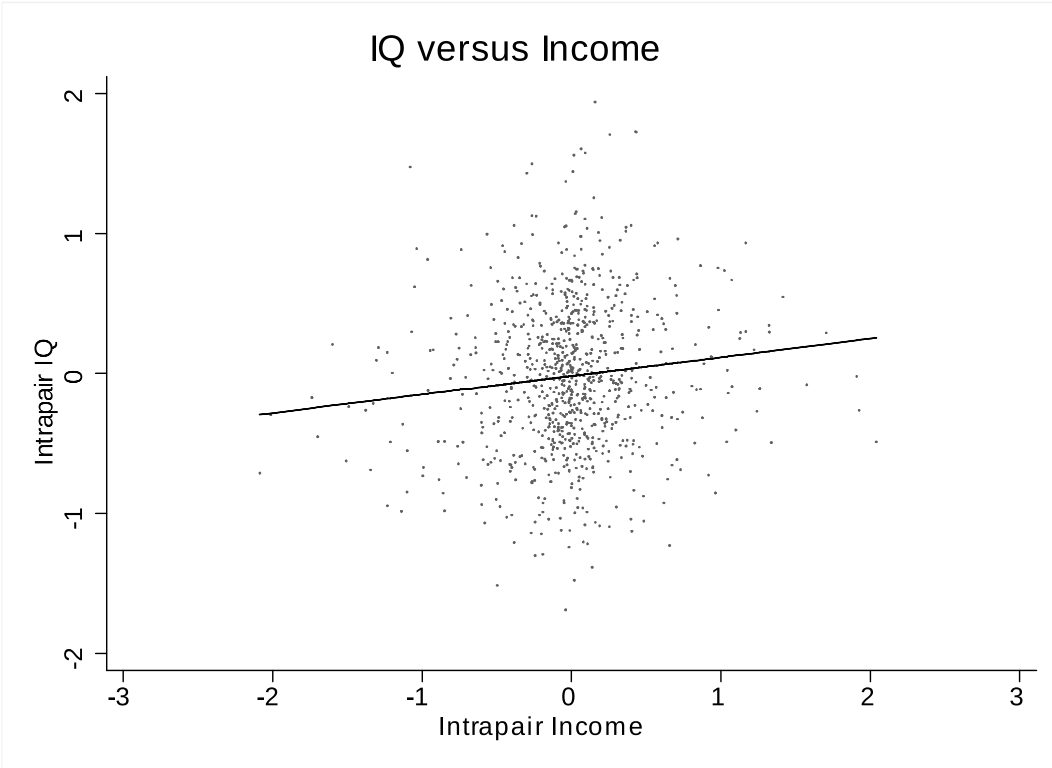

FIGURE III. INTRAPAIR DIFFERENCE IN LN INCOME PLOTTED AGAINST INTRAPAIR DIFFERENCE IN IQ.