Double trouble: on the value of twins-based estimation of the return to schooling

John Bound, Gary Solon *

Department of Economics, University of Michigan, Ann Arbor, MI 48109, USA

Received 4 August 1998; accepted 20 October 1998

Abstract

Several recent studies, including the Rouse (1999) and Behrman & Rosenzweig (1999) articles in this issue, use the schooling and wage variation between monozygotic twins to estimate the return to schooling. In this overview article, we summarize the results from this literature, and we examine the implications of endogenous determination of which twin goes to school longer and of measuring schooling with (possibly mean-reverting) error. Endogeneity of between-twins schooling variation is strongly suggested by the extensive (mostly non-economic) literature documenting that the between-twins difference in birth weight is correlated with the between-twins differences in both schooling and IQ. We conclude that twins-based estimation is vulnerable to the same sort of inconsistency that afflicts conventional cross-sectional estimation. We argue, however, that, if one starts with the presumption that endogenous schooling induces upward inconsistency in the estimated return to schooling, the new twins-based estimates may complement other approaches to tightening the upper bound on the return to schooling. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Twins data; Return to schooling

1. Introduction

Take two individuals who, with the same education, would receive the same wage rate. Then randomly assign one of them to go to school longer. How large a wage advantage should we expect the more educated worker to end up with compared to the less educated worker?

This question—how large is the wage return to schooling?—has preoccupied empirical researchers in economics and other disciplines for a long time. The most common research strategy has been to apply least squares to a regression of the log wage rate on years of schooling and some control variables with a sample from a general population survey such as the decennial census, the Current Population Survey, or the Panel Study of Income Dynamics. It has long been understood, though, that the resulting coefficient estimate for years of schooling does not consistently estimate the return to schooling.1 The problem is that the cross-sectional comparison of workers with different years of schooling does not replicate the above-described thought experiment even if the workers are identical with respect to the observed control variables. For example, if the more educated workers tend to be more intelligent, motivated, or blessed with advantageous family backgrounds, and if these advantages are not completely accounted for by the measured control variables, then the more educated workers typically would have received higher wages even without their additional schooling. It therefore is difficult to

---

1 We are referring to “the” return to schooling as if it were a single parameter, and we will maintain that fiction to simplify our analysis. In reality, however, returns to schooling are heterogeneous. Different individuals face different returns, and the same individual’s returns vary with the level and type of schooling being considered. See Card (1995) for a thorough discussion of heterogeneous returns.
ascertain how much of the empirical association between wages and schooling is due to the causal effect of schooling and how much is due to unobserved factors that influence both wages and schooling. This puzzle remains one of the most challenging identification problems in applied econometrics.

Over the years, researchers have adopted several approaches to tackling this problem. One popular approach has been to seek richer data sets that can be used to control more extensively for measures of intelligence, family background, and the like. The main problem with this approach, of course, is that the controls inevitably remain incomplete. A second approach, especially popular in recent years, is to treat the endogeneity of schooling with instrumental variables. The main problem here is to locate a valid instrument, i.e., a variable that is substantially correlated with years of schooling, but somehow is not correlated with the unobserved factors that influence wages.  

A third approach dates back to an Indiana University dissertation written almost 70 years ago by Donald Gorseline (1932). Gorseline took a sample of brothers and related the difference between the brothers’ incomes to the difference between their years of schooling. Insofar as brothers resemble each other in intelligence, motivation, family background, and so forth, Gorseline hoped that comparing brothers to each other would reduce the bias from failing to control for such factors. Indeed, if brothers were identical with respect to these factors and their schooling differences were randomly generated, Gorseline’s approach would correspond perfectly to the thought experiment above. Since Gorseline’s study, it has occurred to several researchers that, if comparing siblings is a good idea, comparing monozygotic twins might be even better. Monozygotic twins share the exact genetic endowment and usually experience even more similar environments than non-twin siblings or dizygotic twins do. Superficially, then, it seems that comparing monozygotic twins should control even more thoroughly for intelligence, family background, etc.

Over the last few years, several studies—including the Rouse (1999) and Behrman & Rosenzweig (1999) articles in this issue—have pursued this strategy of estimating the return to schooling by comparing monozygotic twins. In this paper, we ask what we have learned from these studies. We emphasize a point previously raised by Griliches (1979) and reiterated by Neumark’s (1999) article in this issue—that between-twins estimation does not eliminate the inconsistency of the conventional cross-sectional estimator and can even aggravate it. The intuition is that, even though differencing between twins does difference out much of the endogenous variation in schooling, it does not eliminate it, and it also filters out much of the exogenous variation. If endogenous variation comprises as large a proportion of the remaining between-twins variation as it does of the cross-sectional variation, then between-twins estimation is subject to as large an endogeneity inconsistency as the cross-sectional estimator. This important point is an old one, but we develop it in detail because it has been largely overlooked in the latest twins research. This research, however, has given more careful attention to another problem also stressed by Griliches (1979)—that differencing between twins exacerbates the errors-in-variables inconsistency due to measurement error in schooling.

Despite these problems, we recognize that between-twins comparisons still may generate evidence that usefully complements the information from other estimation approaches. The siblings-based literature originally was motivated by a suspicion that conventional cross-sectional comparisons tend to overestimate the return to schooling because schooling is positively correlated with unobserved factors that raise wages. If one starts with that presumption, twins-based estimates that correct for measurement error may by viewed as providing an upper bound on the return to schooling. If the twins-based estimates were to come out substantially smaller than those from conventional cross-sectional estimation and other approaches, then the twins-based estimates could be credited for bounding the return to schooling within a narrower range.

In the next section of this paper, we will provide a brief summary of the studies that have estimated the return to schooling by comparing monozygotic twins. In Section 3, we will examine how endogenous choice of schooling affects both conventional cross-sectional estimates and twins-based estimates of the return to schooling. In Section 4, we will consider the additional complexities induced by errors in measuring schooling. Section 5 will summarize our findings.
2. Summary of results


Table 1 lists the estimated coefficients of years of schooling in each of these studies. These estimates come from regressions of the wage measures listed in the table’s third column on years of schooling and, in some cases, a few control variables. In particular, except as indicated in the table’s footnotes, the ordinary least squares (OLS) and generalized least squares (GLS) estimates shown in the fourth and fifth columns come from regressions that control for age, gender (in those studies that include women as well as men), and race (in the US studies that include nonwhites). The sixth column, labeled \( \hat{\beta}_C \) for “covariance estimator”, shows the results from applying least squares to the regression of the between-twins difference in the wage measure on the between-twins difference in years of schooling. The age, gender, and race controls drop out of this differenced regression because monozygotic twins share the same values of these regressors. For this same differenced regression, the last column reports results from instrumental-variables estimation, which will be discussed later in Section 4.

Twins studies of the return to schooling typically begin with OLS estimates as a way of replicating the conventional cross-sectional estimates. In fact, as shown in Greene (1997, pp. 618–620), the OLS estimator is a weighted average of the covariance estimator and the between-families estimator, which applies least squares to the regression of the average of the two twins’ wage measures on the average of their years of schooling. The weighting depends on how much of the sample variation in schooling occurs within twin pairs and how much

### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>Wage measure</th>
<th>( \hat{\beta}_{OLS} )</th>
<th>( \hat{\beta}_{GLS} )</th>
<th>( \hat{\beta}_C )</th>
<th>( \hat{\beta}_{IV} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taubman (1976)</td>
<td>NAS-NRC</td>
<td>Log annual earnings in 1973</td>
<td>0.079</td>
<td>0.027</td>
<td>(0.002)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Ashenfelter &amp; Krueger</td>
<td>Twinsburg</td>
<td>Log hourly wage in 1991</td>
<td>0.084</td>
<td>0.087</td>
<td>0.092</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>Ashenfelter &amp; Rouse</td>
<td>Twinsburg</td>
<td>Log real hourly wage in 1991, 1992, or 1993</td>
<td>0.110</td>
<td>0.102</td>
<td>0.070</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td></td>
<td>(0.010)</td>
<td>(0.019)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Rouse (1999)</td>
<td>Twinsburg</td>
<td>Log real hourly wage in 1991, 1992, 1993</td>
<td>0.105</td>
<td>0.101</td>
<td>0.075</td>
<td>0.110(^{b})</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
<td>(0.023)</td>
<td></td>
</tr>
<tr>
<td>Behrman &amp; Rosenzweig</td>
<td>Minnesota</td>
<td>Log annualized full-time earnings in 1993</td>
<td>0.118(^{c})</td>
<td></td>
<td></td>
<td>0.104(^{c})</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
<td>(0.017)</td>
</tr>
<tr>
<td>Miller, Mulvey &amp; Martin (1995)</td>
<td>Australia</td>
<td>Log of mean 1985 annual earnings in individual’s usual occupation</td>
<td>0.064(^{d})</td>
<td>0.025(^{d})</td>
<td>0.048(^{d})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Isacsson (1997)</td>
<td>Sweden</td>
<td>Three-year average of log annual earnings in 1987, 1990, and 1993</td>
<td>0.049(^{e})</td>
<td>0.023(^{e})</td>
<td>0.024(^{e})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)Numbers in parentheses are estimated standard errors. The standard error estimates in the OLS column are not corrected for the correlation between twins’ error terms.

\(^{b}\)We thank Cecilia Rouse for providing us this estimate, which does not appear in her paper.

\(^{c}\)Controls for tenure and experience (instead of age); does not control for race.

\(^{d}\)Controls for marital status.

\(^{e}\)Controls for marital and big-city status. \( \hat{\beta}_{IV} \) uses a between-twins schooling difference measure based on administrative data as the instrument for a measure based on self reports.

\(^{4}\)Behrman, Rosenzweig & Taubman (1994) also used the Minnesota twins data in an earlier study, but their estimates of the return to schooling used earnings imputations based on only nine occupational categories. We focus on Behrman and Rosenzweig’s more recent estimates because these rely on actual earnings reports from a re-survey of the twins.
between. Given that, in all these data sets, the correlation between monozygotic twins in their self-reported years of schooling is about 0.75, the OLS estimator gives only about 1/8 weight to the covariance estimator and about 7/8 weight to the between-families estimator. Consequently, OLS estimation in a twins sample really is driven mainly by the same between-families variation that drives the estimation in a conventional cross-sectional study. As shown in the fourth column of Table 1, the OLS estimates of the return to schooling in the US twins studies range from 0.08 to 0.11 and are fairly typical of what is found in cross-sectional US studies, while the estimates in the Australian and Swedish studies are somewhat smaller.

Some of the twins studies also report GLS estimates that take account of the positive correlation between the twins’ error terms. Again, as shown in Greene (1997, p. 625), the GLS estimator also is a weighted average of the covariance estimator and the between-families estimator, but it gives more weight to the covariance estimator than the OLS estimator does. Even so, as shown in the fifth column of the table, the GLS estimates still rely mostly on the between-families variation and therefore come out more similar to the OLS estimates than to the covariance estimates.

Of course, the main motivation for the twins-based literature is a concern that the OLS and GLS estimators are inconsistent because schooling is correlated with factors—such as intelligence, motivation, and family background—that contribute to the error term in the wage equation. If these factors differed only between families and not at all between monozygotic twins, and if monozygotic twins nevertheless differed in their schooling for reasons completely unrelated to the between-twins difference in error terms, then the covariance estimator, which applies least squares to the regression of the difference in error terms, then the covariance estimator, but it gives more weight to the covariance estimator than the OLS estimator does. Even so, as shown in the fifth column of the table, the GLS estimates still rely mostly on the between-families variation and therefore come out more similar to the OLS estimates than to the covariance estimates.

Of course, the main motivation for the twins-based literature is a concern that the OLS and GLS estimators are inconsistent because schooling is correlated with factors—such as intelligence, motivation, and family background—that contribute to the error term in the wage equation. If these factors differed only between families and not at all between monozygotic twins, and if monozygotic twins nevertheless differed in their schooling for reasons completely unrelated to the between-twins difference in error terms, then the covariance estimator, which applies least squares to the regression of the between-twins difference in wage measures on the between-twins difference in schooling, would be a perfect solution to the endogeneity problem of the conventional cross-sectional estimators.

The sixth column of the table shows each study’s covariance estimates of the return to schooling. In Taubman’s (1976) study, \( \hat{\beta}_{CV} = 0.027 \) came out much smaller than \( \hat{\beta}_{OLS} = 0.079 \). He therefore concluded that “it is very important to control for genetics and family environment when studying the effects of schooling on earnings” and that failing to do so “may cause a large bias, up to two-thirds of the noncontrolled coefficient”. During the nearly two decades that these results were the only ones available for a sample of monozygotic twins, it was presumed that the between-twins variation told a very different story than the between-families variation. It therefore came as quite a surprise when Ashenfelter and Krueger reported that their \( \hat{\beta}_{CV} = 0.092 \) exceeded their \( \hat{\beta}_{OLS} = 0.084 \). Based partly on this comparison, Ashenfelter and Krueger reached the provocative conclusions that “unobserved factors do not cause an upward bias in simple estimates of the economic returns to schooling” and that “the economic returns to schooling may have been underestimated in the past”. As the next two rows of the table show, however, Ashenfelter and Krueger’s original results were at least partly an artifact of an odd sample. When additional waves of the Twinsburg survey were added in the Ashenfelter-Rouse and Rouse studies, the old result that \( \hat{\beta}_{CV} < \hat{\beta}_{OLS} \) was restored. This result also was replicated in the Miller–Mulvey–Martin and Isacsson studies.

The bulk of the evidence, then, indicates that the covariance estimator, which relies on between-twins variation, produces smaller estimates of the return to schooling. If we were convinced that differencing between twins completely removes the endogenous component of schooling variation, so that the remaining schooling variation between twins is as exogenous as if it were determined by coin flips, we would be led to the conclusions that the OLS estimates have been too high and that the covariance estimates have identified the true, lower return to schooling. Instead, however, we suspect that the measured schooling variation between monozygotic twins, like that between families, is contaminated both by endogenous determination of which twin goes to school longer and by measurement error. The next two sections analyze the implications of these two sources of contamination.

3. Endogenous schooling

We will frame our discussion of endogenous schooling in terms of a simple model in the tradition of Becker (1975), Rosen (1977), and Card (1995). Then we will extend that model to the case of twins data. Our model is extremely simple, ignoring, for example, that returns to schooling vary across the population. The model is sufficient, however, to capture the notion that originally motivated the siblings-based estimation approach—that the empirical association between wages and schooling confounds the causal effect of schooling with other factors that influence both wages and schooling.

Suppose that wages are determined by the familiar semi-logarithmic wage function

\[
y_i = \alpha + \beta S_i + \epsilon_i
\]  

(1)

where \( y_i \) is worker \( i \)'s log wage rate, \( S_i \) is her years of schooling, and \( \epsilon_i \) is an error term reflecting the combined effects of everything else that influences her wage. Assume that worker \( i \) chooses schooling to maximize the objective function

\[
U_i = y_i - C_i(S)
\]  

(2)

where \( C_i(S) \) is a cost-of-schooling function that includes...
forgone earnings, tuition, and non-pecuniary costs such as the effort of studying. We will assume that the marginal cost of schooling is positive and increasing: $C_i'(S) > 0$ and $C_i''(S) > 0$. Then worker $i$’s optimal schooling level $S_i$ must satisfy the first-order condition

$$C_i'(S_i) = \beta$$  (3)

Now let us specify the marginal cost of schooling as the simple linear function

$$C_i'(S) = \delta S + w_i$$  (4)

where $\delta > 0$ represents the increasing marginal cost of schooling and $w_i$ reflects the combined effects of everything besides the level of schooling that influences individual $i$’s marginal cost of schooling. Then substituting Eq. (4) in for the left side of Eq. (3) leads to an explicit solution for worker $i$’s optimal years of schooling:

$$S_i = (\beta - w_i)/\delta$$  (5)

Not surprisingly, optimal schooling increases with the return to schooling and decreases with its marginal cost.

Now suppose we follow conventional cross-sectional studies of the return to schooling by applying OLS to Eq. (1). Then the probability limit of the OLS estimator of the return to schooling $\beta$ is

$$\text{plim} \hat{\beta}_{\text{OLS}} = \beta + \text{Cov}(S_i, \epsilon_i)/\text{Var}(S_i) = \beta - \delta \text{Cov}(w_i, \epsilon_i)/\text{Var}(w_i)$$  (6)

Thus, the conventional OLS estimator is consistent only if $\text{Cov}(w_i, \epsilon_i) = 0$, which is to say, only if schooling is exogenous with respect to the wage equation. The common fear, however, has been that $\text{Cov}(w_i, \epsilon_i)$ is negative. For example, suppose that individuals with tastes or circumstances such that their marginal cost of schooling is relatively low also tend to have the ability or inclination to work more productively, engage in more on-the-job human capital investment, or do the other things that lead to a positive $\epsilon_i$. Then $\text{Cov}(w_i, \epsilon_i)$ would be negative, and OLS would tend to overestimate $\beta$.

This, of course, is just the sort of situation that motivated Gorsline and his successors to use between-siblings comparisons in hopes of reducing the inconsistency from endogenous schooling. How realistic were their hopes? To obtain a formal answer, let’s extend the above model to the case of monozygotic twins. First, replace Eq. (1) with

$$y_{ij} = \alpha + \beta S_{ij} + \epsilon_{ij}$$  (7)

where $y_{ij}$ denotes the log wage of the $j$th twin in the $i$th family and so forth. Assume that $\epsilon_{ij}$ follows the error-components model

$$\epsilon_{ij} = f_i + u_{ij}$$  (8)

where, $\text{Var}(f_i) = \sigma_f^2$, $\text{Var}(u_{ij}) = \sigma_u^2$, and $\text{Cov}(f_i, u_{ij}) = \text{Cov}(u_{ij}, u_{ij}) = 0$. This statistical characterization of the twin covariance in $\epsilon_{ij}$ defines $f_i$ as the portion of $\epsilon_{ij}$ that is common to both twins and $u_{ij}$ as the portion purely idiosyncratic to twin $j$. The implied restriction is that the twin correlation in $\epsilon_{ij}$ is positive. This accords with the supposition that $\epsilon_{ij}$ partly reflects factors like genetic endowment and family environment that are shared by monozygotic twins. In addition, if monozygotic twins exert upon each other a (probably stronger) version of the peer effects that have been much discussed with respect to neighborhood and school influences, the positive twin correlation in $\epsilon_{ij}$ is amplified. On the other hand, the efforts that monozygotic twins sometimes make to differentiate themselves from each other cut in the opposite direction. In the context of our model, such forces toward differentiation between twins reduce $\text{Corr}(\epsilon_{1i}, \epsilon_{2i})$ in Eq. (9) by reducing the share of $\sigma_f^2$ in $\sigma_u^2 + \sigma_f^2$. As an empirical matter, however, it is clear that the factors pushing toward a positive correlation dominate.

Similarly, replace Eqs. (2) and (4) with

$$U_{ij} = y_{ij} - C_i(S)$$  (10)

and

$$C_i'(S) = \delta S + w_{ij}$$  (11)

and let $w_{ij}$ also follow an error-components model

$$w_{ij} = g_i + z_{ij}$$  (12)

with $\text{Var}(g_i) = \sigma_g^2$, $\text{Var}(z_{ij}) = \sigma_z^2$, and $\text{Cov}(g_i, z_{ij}) = \text{Cov}(z_{ij}, u_{ij}) = \text{Cov}(z_{ij}, u_{ij}) = 0$. Then the optimal schooling for the $j$th twin in family $i$ is

$$S_{ij} = (\beta - g_i - z_{ij})/\delta$$  (13)

and the twin correlation in schooling is

$$\text{Corr}(S_{1i}, S_{2i}) = \sigma_f^2(\sigma_f^2 + \sigma_z^2)$$  (14)

The implication that the twin correlation in schooling is positive accords with the recurrent finding that the corre-

---

5 Rosen (1977) derives a special case of Eq. (2) by assuming that individuals maximize the present value of their real lifetime earnings with a constant real interest rate, constant hours of work, infinite lifetimes, and forgone earnings as the only cost of schooling.


7 For example, if it were not for the positive twin correlation in $\epsilon_{ij}$, the GLS estimates in Table 1 would not lie in between the OLS and covariance estimates.
lation between monozygotic twins in their self-reported schooling is about 0.75.\footnote{The assumption that Cov(z_{ij},u_{ij}) = Cov(z_{i2},u_{i2}) = 0 simplifies the notation below, but does not seriously restrict the analysis. Without this assumption, Cov(z_{ij},u_{ij})/\sigma^2 in the last term of Eq. (16) would be replaced by Cov(z_{i1} - z_{i2}, u_{i1} - u_{i2})/Var(z_{i1} - z_{i2}).}

The covariance estimator of the return to schooling applies OLS to the regression of the between-twins difference in log wages on the between-twins difference in schooling:

\[ y_{i1} - y_{i2} = \beta(S_{i1} - S_{i2}) + u_{i1} - u_{i2} \]  

The probability limit of this estimator is

\[ \text{plim} \ \hat{\beta}_{CV} = \beta + \text{Cov}(S_{i1} - S_{i2}, u_{i1} - u_{i2})/\text{Var}(S_{i1} - S_{i2}) = \beta - \delta \text{Cov}(z_{ij},u_{ij})/\sigma^2 \]  

Thus, the covariance estimator is consistent only if the twin-specific error component in the schooling equation is uncorrelated with the twin-specific error component in the wage equation. If instead Cov(z_{ij},u_{ij}) < 0, then the between-twins difference in schooling is endogenous, and the covariance estimator of \( \beta \) is upward-inconsistent. For example, suppose that twin 1 in family \( i \) has a stronger "work ethic" than twin 2 and feels less burdened by the effort of school work. Because twin 1 faces a lower non-pecuniary cost of schooling, he may go to school longer. By the same token, why do monozygotic twins with the same schooling show any difference in wages? A crucial

\[ \text{plim} \ \hat{\beta}_{OLS} = \beta - \delta \text{Var}(h_{ij})/\text{Var}(h_{ij}) + \text{Var}(k_{ij}) \]  

In this formulation, the endogeneity inconsistency of the cross-sectional OLS estimator is seen to be proportional to the fraction of cross-sectional schooling variation due to the endogenous component \( h_{ij} \). A parallel analysis of the covariance estimator shows that the inconsistency of that estimator is similarly proportional to the fraction of the between-twins schooling variation due to the between-twins difference in the endogenous component \( h_{ij} \):

\[ \text{plim} \ \hat{\beta}_{CV} = \beta - \delta \text{Var}(h_{i1} - h_{i2})/\text{Var}(h_{i1} - h_{i2}) + \text{Var}(k_{i1} - k_{i2}) \]  

Therefore, the inconsistency of the between-twins estimator is less than that of the conventional cross-sectional estimator only if endogenous variation comprises a smaller share of the between-twins variation in schooling than it does of the between-families variation. We are not aware of any compelling a priori reason to be confident that this is the case.

Nonetheless, the twins-based literature has been motivated by the hope that, although conventional cross-sectional estimation of the return to schooling is rendered inconsistent by the endogeneity of schooling, the covariance estimator is not. That hope rests on the very strong assumption that any within-family variation in the endogenous schooling factor \( h_{ij} \) is purely genetic, so that any schooling differences between monozygotic twins are as exogenous as if they were determined by coin flips. Why would anyone be convinced that this is so? Maybe wishful thinking has played a role, but perhaps it has seemed plausible to some that, since monozygotic twins are identical, any variation in their schooling must be purely random.

But, if monozygotic twins are perfectly identical, why do they ever display any schooling difference at all?\footnote{Of course, without such differences, the regression of the between-twins wage difference on the between-twins schooling difference would not be identified. Ashenfelter & Krueger (1994) report that 51 percent of the pairs of monozygotic twins in their Twinsburg sample differ in their self-reported years of schooling. Behrman, Rosenzweig & Taubman (1994) report corresponding figures of 52 percent for the NAS-NRC twins and 56 percent for the Minnesota twins.} By the same token, why do monozygotic twins with the same schooling show any difference in wages? A crucial
part of the answer is that monozygotic twins, notwithstanding their remarkable similarity and identical genetic endowment, are not exactly identical. Any parents of monozygotic twins will tell you that their kids do differ in temperament and abilities. Often these differences are subtle, but presumably it is these differences, rather than coin flips, that account for the twins' divergent choices about schooling. And, if these same differences in temperament and abilities also exert other influences on wages, the empirical association of the between-twins wage difference with the between-twins schooling difference reflects more than just the causal effect of the latter on the former.  

Why is it that twins with identical genetic endowments turn out to be different? A complete answer would require a deeper understanding than anyone now possesses of how a person's genetic heritage and experiences interact to influence the individual she becomes. It is quite clear, though, that monozygotic twins are shaped by somewhat different experiences. Some of these occur even before the twins emerge from the womb. In some extreme cases of the so-called twin transfusion syndrome, monozygotic twins are dramatically different at birth because the twins have competed for nourishment in the uterus, and the "winner" has flourished at the great expense of the "loser". Less extreme differences are quite common. Behrman, Rosenzweig & Taubman (1994) reported that, in their sample of monozygotic twins from Minnesota, 69 percent of the twin pairs had birth weights differing by at least 4 oz, and 48 percent differed by at least 8 oz. They also reported that these differences in birth weight are significantly correlated with differences in subsequent educational attainment as well as occupational position. They found, for example, that a 4-oz difference in birth weight is associated with about a half-year difference in years of schooling. Numerous earlier studies similarly documented an association between differences in birth weight and differences in IQ.  

Subsequent to birth, monozygotic twins usually experience very similar environments, but they continue to undergo some divergent experiences as well. Some of these occur more or less at random (e.g., one twin happens to break his arm, and the other does not), and some are more systematic. For example, it is common to separate twins when they enter school and put them in different classrooms, where they proceed to be influenced by different teachers and classmates. Another plausible source of variation between some monozygotic twins is their psychological need to differentiate themselves from each other. In some cases, this conscious or unconscious motivation to part ways may stem from a desire to avoid what might otherwise be unbearably fierce competition. Clearly, this force toward between-twins variation might amplify the variation due to different experiences, with small differences between twins in their non-genetic endowments influencing subsequent specialization.

The key point is that even monozygotic twins are a little different, and their (often small) differences in abilities and temperament may contribute to their (often small) differences in schooling. What this implies for the inconsistency of twins-based estimation of the return to schooling depends on the extent to which the between-twins differences that generate their schooling differences also contribute in other ways to their wage differences. In terms of our simple model, the question is the extent to which their schooling differences are differences in \( h_{ij} \) or differences in \( k_{ij} \). As far as we can tell, there is no a priori basis for answering that question. It therefore is unclear whether endogenous variation com-

---

10 In principle, just as numerous researchers have tried to use instrumental-variables estimation to treat the endogeneity of cross-sectional schooling variation, one might also use instrumental variables to treat the endogeneity of between-twins schooling variation. Indeed, Behrman, Rosenzweig & Taubman (1994) perform IV estimation of Eq. (15) with the between-twins difference in birth weight as the instrument for the between-twins difference in schooling. As discussed below, however, a large literature has documented that the between-twins difference in birth weight is associated with the between-twins difference in IQ. It therefore is hard to believe Behrman, Rosenzweig, and Taubman’s assumption that the between-twins difference in birth weight is uncorrelated with the error term in Eq. (15).


---

12 See Babson, Kangas, Young & Brahhall (1964), Churchill (1965), Willerman & Churchill (1967), Kaelber & Pugh (1969), Babson & Phillips (1973), and Fujikura & Froehlich (1974), as well as the previous footnote’s references on the role of transfusion syndrome. Some researchers also have found an association between twins’ birth order and intelligence, but here the evidence is much weaker (Pencavel, 1976; Scarr, 1982).

13 This phenomenon may be part of the explanation for why monozygotic twins reared apart show nearly as much resemblance in many dimensions as monozygotic twins reared together. Although the reared-apart twins grow up in less similar environments, they also face less need to differentiate from each other. In Farber’s (1981) words, “the differences also speak eloquently to the need of each individual to be an individual—unique and clearly bounded. And, since twins evidently make themselves different as well as being made different by external forces, the pattern suggests that an approach that views the individual as an active participant in the creation of his reality is essential to understand the data”.

14 See Ainslie (1985). One manifestation noted by Ainslie is the tendency in some twin pairs for one twin to favor the mother while the other prefers the father.
years of schooling also may play an important role. In inconsistence in the estimation of the return to schooling. 

endogeneity of true schooling is not the only source of (Willis, 1986), the sign of the endogeneity inconsistence cational attainment and its relationship with wages 

recognizes the multi-faceted nature of human capital and earned higher wages in any case. Especially once one go to school longer tend to be the ones who would have 

Griliches (1977, 1979) and Ashenfelter & Krueger 

These include the instrumental-variables strategy, Gorse-line’s approach of comparing non-twin siblings, comparing dizygotic twins, and embellishing cross-sectional regressions with observable controls for academic per-formance (such as standardized test scores) and family background. As long as one is willing to presume that cross-sectional OLS estimation exaggerates the causal effect of education on earnings, most of these strategies have the potential to tighten the upper bound on the return to schooling. 

From this point of view, comparing monozygotic twins serves the same purpose as other methods researchers have tried for reducing the endogeneity inconsistence in estimation of the return to schooling. These include the instrumental-variables strategy, Gorse-line’s approach of comparing non-twin siblings, comparing dizygotic twins, and embellishing cross-sectional regressions with observable controls for academic per-formance (such as standardized test scores) and family background. As long as one is willing to presume that cross-sectional OLS estimation exaggerates the causal effect of education on earnings, most of these strategies have the potential to tighten the upper bound on the return to schooling. 

Unfortunately, though, several researchers—such as Griliches (1977, 1979) and Ashenfelter & Krueger (1994)—have questioned the assumption that those who go to school longer tend to be the ones who would have earned higher wages in any case. Especially once one recognizes the multi-faceted nature of human capital and the role of comparative advantage in determining edu-cational attainment and its relationship with wages (Willis, 1986), the sign of the endogeneity inconsistence is theoretically indeterminate. To make matters worse, endogeneity of true schooling is not the only source of inconsistence in the estimation of the return to schooling. As many researchers have stressed, errors in measuring years of schooling also may play an important role. In the next section, we extend our analysis to encompass this issue of measurement error.

4. Errors in measurement of schooling

Results from the twins-based studies, as well as additional results they cite from previous research on measurement of schooling, suggest that between 5 and 20 percent of the cross-sectional variance in self-reported years of schooling is due to measurement error. As a result, the standard econometrics textbook treatment of errors in variables implies that applying OLS to Eq. (1) with the true $S_t$ proxied by worker $t$’s report leads to a modest attenuation inconsistence in the estimated return to schooling. If not for the additional endogeneity problem discussed in the preceding section, we therefore would expect the conventional cross-sectional approach to produce understimation of the return to schooling. If, however, we suspect that the endogeneity of schooling causes an upward inconsistence, we would expect the net effect to be only a slight underestimation or an overestimation of unknown magnitude.

Although the impact of measurement error probably is small for the cross-sectional estimator, Taubman (1976), Griliches (1979), and many subsequent researchers have explained that its impact may be much greater for the covariance estimator. Because monozygotic twins are highly correlated in their years of schooling, between-twins differencing filters out most of the “signal” component of schooling variation without a commensurate reduction in the “noise” from measurement error. As a result, the covariance estimator probably is subject to a much more severe errors-in-variables inconsistence.

In the first part of this section, we formalize these points by specifying a mostly classical model of measurement error and using it to modify the previous section’s results on the probability limits of the OLS and covariance estimators. We also discuss Ashenfelter and Krueger’s (1994) instrumental-variables approach to correcting for measurement error. In the second part, we extend the analysis further to encompass the (probably more realistic) possibility that the measurement error in schooling is mean-reverting rather than classical.

4.1. Classical measurement error

Suppose that self-reported schooling $S'_t$ is generated by the classical errors-in-variables model

$$S'_t = S_t + v_t$$

(21)

where the measurement error $v_t$ is uncorrelated with both the family components and each twin’s individual-spe-cific components of both schooling and the wage equation’s error term. That is, $\text{Cov}(v_t, \tilde{f}_t) = \text{Cov}(v_t, u_{1t}) = \ldots$
Cov(v_i, u_{12}) = Cov(v_{i|z_i}, u_{12}) = Cov(v_{i|z_i}, v_{i|z_i}) = 0. Then applying least squares to the cross-sectional regression of y_{i|z_i} on the error-ridden S_{i | y} yields an estimator of the return to schooling with probability limit

\[ \text{plim} \hat{\beta}_{\text{OLS}} = \beta + \text{Cov}(S_{i|y}, e_y)/[\text{Var}(S_{i|y}) + \text{Var}(v_{i|y})] \]  

(22)

- \beta \text{Var}(v_{i|y})/[\text{Var}(S_{i|y}) + \text{Var}(v_{i|y})]

Now, with measurement error incorporated into the analysis, the inconsistency of \( \text{plim} \hat{\beta}_{\text{OLS}} \) consists of two terms. The first term, caused by the endogeneity of true schooling, has the same numerator as the inconsistency term in Eq. (6), but the denominator is slightly inflated by the additional variance from measurement error. The second term is the textbook errors-in-variables attenuation inconsistency, which is proportional to the fraction of the cross-sectional variance in self-reported schooling that arises from measurement error. Most of the evidence indicates that fraction is only about 0.1, so the errors-in-variables inconsistency of the conventional cross-sectional estimator is modest. Consequently, if the endogeneity inconsistency is non-negative, \( \hat{\beta}_{\text{OLS}} \) cannot be downward-inconsistent by very much. If the endogeneity inconsistency is positive and large, \( \hat{\beta}_{\text{OLS}} \) may be substantially upward-inconsistent. This is the reasoning that has led to the conventional view that cross-sectional OLS estimation of the return to schooling is either nearly consistent or upward-inconsistent. To put the point a little differently, if we believe that the endogeneity inconsistency is upward, and if we correct for the downward errors-in-variables inconsistency by multiplying \( \hat{\beta}_{\text{OLS}} \) by 1.1 or so, we can presume that the resulting adjusted estimator is upward-inconsistent and, with large enough samples, provides an upper bound on the return to schooling.

With measurement error, the covariance estimator applies least squares to the regression of the between-twins difference in y_{ij} on the between-twins difference in S_{ij}. The probability limit of the resulting estimator of the return to schooling is

\[ \text{plim} \hat{\beta}_{\text{CV}} = \beta + \text{Cov}(S_{i1} - S_{i2}, u_{i1} - u_{i2}) \]

/ \left[ \text{Var}(S_{i1} - S_{i2}) + \text{Var}(v_{i1} - v_{i2}) \right] \]  

(23)

- \beta \text{Var}(v_{i1} - v_{i2})/[\text{Var}(S_{i1} - S_{i2}) + \text{Var}(v_{i1} - v_{i2})]

In this instance, too, the inconsistency consists of two terms. The first term, caused by the endogeneity of the between-twins variation in true schooling, is the same as the inconsistency term in Eq. (16) except that the denominator is now inflated by the variance of the between-twins difference in measurement error. In this case, unlike the cross-sectional case, the inflation of the denominator is more than slight. For example, suppose that the share of measurement error in the cross-sectional variance of reported schooling is indeed 0.1, the correlation between twins’ reporting errors is 0, and the correlation between twins’ true schooling is 5/6 (so that the correlation between their reported schooling is 0.75). Then, in the cross-section, the measurement error adds on only 1/9 of the variance from true schooling variation. In contrast, the measurement error accounts for 40 percent of the variance in the reported between-twins schooling differences and therefore inflates the denominator of the endogeneity inconsistency term by 2/3. 16

The second term, the errors-in-variables attenuation inconsistency, is also strongly affected by differencing between twins. With the numbers assumed above, the proportional attenuation goes from 10 percent for the cross-sectional OLS estimator to 40 percent for the covariance estimator. As Griliches (1979) and many others have stressed, an errors-in-variables problem that was mild in the cross-section becomes egregious in the between-twins differences.

Accordingly, consideration of measurement error complicates the interpretation of the covariance estimates in Table 1. Before allowing for measurement error, we argued that, if one believes that those who go to school longer typically would have earned higher wages even without the extra schooling, then both the OLS estimator and the covariance estimator would be upward-inconsistent. In that case, the covariance estimates, which usually come out smaller than the OLS estimates, could be viewed as providing a tighter upper bound on the return to schooling. But now, with measurement error mixed into the stew, we see that the covariance estimator is subject to a severe downward errors-in-variables inconsistency at the same time that the measurement error also reduces the magnitude of the endogeneity inconsistency. The sign of the covariance estimator’s inconsistency therefore becomes unclear because it depends on which inconsistency dominates, the downward errors-in-variables inconsistency or the upward endogeneity inconsistency. The bounding result is lost, and it seems questionable whether twins-based estimation has taught us anything.

This is where the Princeton researchers who conducted the Twinsburg studies entered with their greatest potential contribution to this literature. In a shrewd effort to account for measurement error, they collected two measures of each twin’s schooling. In addition to eliciting the self report S_{ij}, they also measured twin j’s schooling as reported by j’s twin:

\[ \bar{S}_{ij} = S_{ij} + \bar{v}_{ij} \]  

(24)

Ashenfelter & Krueger (1994) then devised a clever estimation procedure that uses the multiple measures of schooling to correct for the inconsistency from measure-

16 Under the given assumptions, Var(S_{i1} - S_{i2}) = Var(S_{ij})/3 and Var(v_{i1} - v_{i2}) = 2 Var(v_{ij}) = 2 Var(S_{ij})/9 = (2/3) Var(S_{i1} - S_{i2}).
ment error. If the correction is successful, the only remaining inconsistency stems from the endogeneity of schooling, and the possibility of signing the inconsistency of between-twins estimation is restored.

Ashenfelter and Krueger’s approach rests on a mostly classical model of measurement error. In particular, it assumes that the measurement error \( v_i \) in the twin report obeys the same assumptions listed for \( v_j \) after Eq. (21)—that it is uncorrelated with \( f_i, g_i \), and both twins’ values of \( u \) and \( z \). It loosens the classical assumptions a bit, though, in that it allows for a structured variety of correlation among the twins’ reporting errors. In particular, Ashenfelter and Krueger assume that

\[
\begin{align*}
v_{i1} &= a_1 + b_{i1} + c_{i1} \\
v_{i2} &= a_1 + b_{i2} + c_{i2} \\
\tilde{v}_{i1} &= a_1 + b_{i1} + \tilde{c}_{i1} \\
\tilde{v}_{i2} &= a_1 + b_{i1} + \tilde{c}_{i2}
\end{align*}
\]

where the \( a, b, \) and \( c \) terms are mutually uncorrelated and are uncorrelated with \( f_i, g_i \), and both twins’ \( u \) and \( z \). This model implies

\[
\text{Cov}(v_{i1}, v_{i2}) = \text{Cov}(\tilde{v}_{i1}, \tilde{v}_{i2}) = \sigma^2_u
\]

so that a common family effect in reporting error induces a positive covariance among the errors. In addition, it implies

\[
\text{Cov}(v_{i1}, v_{i2}) = \text{Cov}(\tilde{v}_{i1}, \tilde{v}_{i2}) = \sigma^2_u + \sigma^2_b
\]

so that an individual-specific propensity for reporting error induces an even more positive covariance. The idea here is that, if twin \( j \) overreports her own schooling, she is likely to overreport her twin’s schooling, too.\(^\text{17}\)

Under these assumptions, Ashenfelter and Krueger eliminate the downward errors-in-variables inconsistency by using one twin’s report of the between-twins difference in schooling as an instrument for the other twin’s report of the difference. That is, they perform instrumental-variables (IV) estimation of the regression of \( y_{12} \) on \( S_{i1}^* - S_{i2}^* \) with \( S_{i1} = S_{i2}^* \) as the instrument for \( S_{i1}^* - S_{i2} \). The probability limit of this estimator is

\[
\text{plim} \, \hat{\beta}_{IV} = \beta + \text{Cov}(S_{i1}^* - S_{i2}^*, u_{i1} - u_{i2}) / \text{Var}(S_{i1}^* - S_{i2}^*)
\]

Comparing this expression to Eq. (16), we see that Ashenfelter and Krueger’s IV estimator has the same probability limit that the covariance estimator would have if there were no measurement error in schooling. It therefore resurrects the case for between-twins estimation as a means of obtaining an upper bound on the return to schooling.

Another useful reference point is Eq. (23), the probability limit of the covariance estimator when there is measurement error. A comparison of Eqs. (23) and (28) shows that Ashenfelter and Krueger’s correction for measurement error in the between-twins regression has two effects. First, their correction succeeds in eliminating the covariance estimator’s errors-in-variables inconsistency. Second, as emphasized by Neumark’s article in this issue, their correction also increases the magnitude of the endogeneity inconsistency term by undoing the inflation of its denominator by measurement error. If the endogeneity inconsistency is upward, both of these effects induce a tendency for \( \hat{\beta}_{IV} \) to exceed \( \hat{\beta}_{CV} \).

Returning to Table 1, Ashenfelter and Krueger’s \( \hat{\beta}_{CV} \) does indeed exceed their \( \hat{\beta}_{IV} = 0.092 \). Similarly, \( \hat{\beta}_{IV} \) exceeds \( \hat{\beta}_{CV} \) in every other study that reports both estimates, and the difference is substantial in every study except Isacsson’s.\(^\text{18}\)

Finally, let’s compare Eq. (28) to Eq. (22), the probability limit of the conventional cross-sectional OLS estimator. If we inflate the OLS estimator by \( \text{Var}(S_{i1}^*) + \text{Var}(v_{i1}) / \text{Var}(S_{i1}) \approx 1.1 \), both the adjusted OLS estimator and the Ashenfelter–Krueger estimator eliminate the errors-in-variables inconsistency. Then the probability limits of the two estimators differ only in their endogeneity inconsistencies. As discussed in Section 3, which endogeneity inconsistency is larger is not clear a priori. It therefore may seem at first that the Ashenfelter–Krueger approach makes no progress toward identifying the return to schooling. In our view, though, the Ashenfelter–Krueger methodology does still have some potential for providing new information. Although it is theoretically unclear which estimator is subject to the smaller endogeneity inconsistency, that question can be viewed as an empirical matter if one is willing to presume that

\(^\text{17}\) Although Ashenfelter and Krueger’s model obviously is restrictive, it still is a considerable advance, and it could not have been achieved without their innovation of eliciting each twin’s report of the other’s schooling. Before their study, it was possible to guess at the impact of measurement error with back-of-the-envelope calculations such as the one we performed above. These calculations, however, required arbitrary assumptions about the correlation between the twins’ reporting errors (e.g., ours used a dubious assumption of zero correlation). They also required estimates of the share of measurement error in the variance of self-reported schooling. These estimates came from studies of reinterview data, which made the dubious assumption that reporting errors by the same person at different times are uncorrelated with each other.

\(^\text{18}\) Some of the twins studies also report alternative IV estimates that appropriately correct for measurement error only under the assumption that \( \sigma^2_v = 0 \). That is, they assume that, if twin \( j \) overreports his own schooling, this implies no tendency for him to overreport his twin’s schooling. Both Ashenfelter & Krueger (1994) and Rouse (1999) report strong evidence against this assumption.
schooling and the wage equation’s error term are positively correlated both in the cross-section and within twin pairs. In that case, both estimators are upward-inconsistent, and the one that delivers smaller empirical estimates can be viewed as providing a tighter upper bound on the return to schooling.

From that perspective, Ashenfelter and Krueger’s initial results, from the first wave of the Twinsburg survey, do not appear very useful. As shown in Table 1, their \( \hat{\beta}_{\text{IV}} = 0.129 \) exceeds their \( \hat{\beta}_{\text{OLS}} = 0.084 \) by more than 50 percent, so that, even after the OLS estimator is inflated to correct for measurement error, the OLS estimator seems to provide a much tighter upper bound on the return to schooling than the between-twins IV estimator does. This is the result that leads Neumark’s (1999) study reported in this issue to conjecture that Ashenfelter and Krueger’s estimator may be more upward-inconsistent than the conventional cross-sectional estimator.

But subsequent results, from additional waves of the Twinsburg survey and from other twins surveys, look quite different. Inspection of Table 1 shows that, in four of the five studies that have followed in Ashenfelter and Krueger’s footsteps, \( \hat{\beta}_{\text{IV}} \) comes out smaller than \( \hat{\beta}_{\text{IV}} \) even before the OLS estimate is inflated to compensate for measurement error. In the one exception, Rouse’s (1999) study reported in this issue, \( \hat{\beta}_{\text{IV}} \) exceeds \( \hat{\beta}_{\text{OLS}} \) so slightly that it also provides a smaller estimate once \( \hat{\beta}_{\text{OLS}} \) is adjusted for measurement error.

Here, then, is our interpretation of the current twins-based evidence. The studies that apply Ashenfelter and Krueger’s methodology to US twins data estimate the return to schooling at somewhere around 0.10. Because we believe that between-twins differences in schooling are not randomly assigned, but are endogenously chosen, we suspect that these estimates are upward-inconsistent. They therefore fortify our suspicion, already held on the basis of conventional cross-sectional estimates, that the return to schooling in the United States is really less than 0.10. How much less remains unknown.

---

4.2. Mean-reverting measurement error

So far, our analysis has rested on the classical assumption that the measurement error in reported schooling is uncorrelated with true schooling. The literature, however, has not delved seriously into the question of why schooling is sometimes mismeasured, so it has never been clear whether the classical assumption should be trusted. Two recent studies—by Kane, Rouse & Staiger (1997) and Berger, Black & Scott (1997)—have argued both theoretically and empirically that, contrary to the classical assumption, the measurement error in schooling is mean-reverting. The basic idea is that individuals with very low true schooling cannot underreport their schooling by much, but can overreport by a lot; individuals with very high schooling cannot overreport by much, but can underreport by a lot. In this subsection, we develop a very simple model to represent the resulting negative correlation between reporting error and true schooling, and we use it to derive new results on how such mean-reverting measurement error affects twins-based estimation of the return to schooling.

Our model of measurement error is just the same as in the previous subsection except that now we replace Eqs. (21) and (24) with

\[
S'_{ij} = \gamma S_{ij} + \nu_i
\]

and

\[
\bar{S}_{ij} = \gamma S_{ij} + \bar{\nu}_i
\]

where \( 0 < \gamma < 1 \). The condition that \( \gamma < 1 \) signifies mean-reversion. The condition that \( \gamma > 0 \) assures that, despite the measurement error, the schooling reports are still positively correlated with true schooling. This is an exceedingly simple model—it assumes that the same mean-reversion applies regardless of whether the schooling variation stems from the family or twin-specific component, and, like the classical model and variations such as Ashenfelter and Krueger’s, it ignores the discreteness of years of schooling.

Nevertheless, it is an appealingly tractable way to begin exploring the implications of mean-reverting measurement error.

How does the introduction of mean-reversion modify the previous subsection’s results concerning the probability limits of the OLS, covariance, and between-twins IV estimators? Starting with the OLS estimator, the prob-

---

19 Behrman & Rosenzweig (1999) do not report an OLS estimate, but the OLS estimates in the table regularly exceed the GLS estimates, and Behrman and Rosenzweig’s GLS estimate does exceed their IV estimate. It therefore seems safe to assume that their OLS estimate also would exceed their IV estimate.

20 Ironically, if we agreed with Ashenfelter and Krueger’s conjecture that “unobserved ability may be negatively related to schooling level”, we would conclude that their methodology has not proved to be useful at all. In that case, both the between-twins IV estimator and the OLS estimator would be downward-inconsistent. Since the between-twins IV estimates usually come out smaller than the cross-sectional OLS estimates, we would then judge that the Ashenfelter–Krueger methodology has failed to tighten the lower bound provided by the OLS estimates.

21 In addition, a more complete analysis of mean-reverting measurement error would model the error in measuring wages as well as schooling. The results in Table 3 of Bound & Krueger (1991), however, indicate that measurement error in earnings has little effect on the estimation of the return to schooling. This is to be expected if, as Pischke (1995) suggests, the mean-reversion in earnings measurement error pertains to the transitory component of earnings and not the permanent component.
ability limit previously shown in Eq. (22) now needs to be multiplied through by a correction factor

\[ \lambda_{\text{OLS}} = \frac{\gamma[\text{Var}(S_{ij}) + \text{Var}(v_{ij})]}{\gamma^2 \text{Var}(S_{ij}) + \text{Var}(v_{ij})} \]  

(31)

The new probability limit exceeds the old one if and only if

\[ \text{Var}(S_{ij}) > \text{Var}(v_{ij})/\gamma \]  

(32)

Given the evidence that, in the cross-section, the “signal” variance \( \text{Var}(S_{ij}) \) far exceeds the “noise” variance \( \text{Var}(v_{ij}) \), this condition is met unless \( \gamma \) is implausibly small. Therefore, if we were right in the preceding subsection when we conjectured that the endogeneity of schooling causes \( \hat{\beta}_{\text{OLS}} \) to be upward-inconsistent, that presumption is only strengthened by consideration of mean-reverting measurement error.

Moving on to the covariance estimator, the probability limit previously shown in Eq. (23) now needs to be multiplied through by a correction factor

\[ \lambda_{\text{CV}} = \frac{\gamma[\text{Var}(S_{i1} - S_{i2}) + \text{Var}(v_{i1} - v_{i2})]}{\gamma^2 \text{Var}(S_{i1} - S_{i2}) + \text{Var}(v_{i1} - v_{i2})} \]  

(33)

The new probability limit exceeds the old one if and only if

\[ \text{Var}(S_{i1} - S_{i2}) > \text{Var}(v_{i1} - v_{i2})/\gamma \]  

(34)

Given the evidence that differencing between twins filters out much more “signal” than “noise,” it is not clear at all whether this condition is met. In the previous subsection, we said we could not sign the inconsistency of the covariance estimator because we do not know which dominates, the downward errors-in-variables inconsistency or the upward endogeneity inconsistency. Now the uncertainty about whether inequality (34) is satisfied only adds to the ambiguity.

The neatest result of all pertains to Ashenfelter and Krueger’s between-twins IV estimator. The probability limit previously shown in Eq. (28) now needs only to be divided through by \( \gamma \). One implication is that one of Neumark’s main points—that Ashenfelter and Krueger’s correction for measurement error exacerbates the endogeneity inconsistency of the covariance estimator—applies all the more strongly when the error in measuring schooling is mean-reverting instead of classical. Another implication is that, even if \( \text{Cov}(S_{i1} - S_{i2}, u_{i1} - u_{i2}) = 0 \), so that there is no endogeneity inconsistency in the between-twins estimation, the probability limit of Ashenfelter and Krueger’s estimator is \( \beta/\gamma \neq \beta \). Thus, their IV strategy, which is designed to correct for a variant of classical measurement error, overcorrects when the measurement error really is mean-reverting. This result is reminiscent of similar cross-sectional results reported in Kane, Rouse & Staiger (1997), and Berger, Black & Scott (1997). In any case, if we were right in the previous subsection when we argued that Ashenfelter and Krueger’s approach may be useful for providing an upper bound on the return to schooling, that argument is only strengthened by consideration of mean-reverting measurement error.

5. Conclusion

Ashenfelter and Krueger’s innovative study of participants in the Twinsburg festival has rekindled interest in the twins-based approach to estimating returns to schooling that Behrman, Taubman, and their colleagues pioneered in the 1970s. That interest is vividly illustrated by the articles that several leading researchers have contributed to this issue of the Economics of Education Review. The new twins-based literature already has had considerable influence on the profession’s thinking. For example, the concluding section of Card’s (forthcoming) Handbook of labor economics chapter on the return to schooling treats the new twins-based estimates as the “gold standard” according to which other estimates ought to be evaluated.

In this review article, we have stressed a point made by Griliches (1979) almost two decades ago and reiterated in this issue by Neumark (1999)—that between-twins differences in schooling are not randomly assigned, but are chosen endogenously. As a result, between-twins estimates of the return to schooling are vulnerable to an endogeneity inconsistency similar to the one that afflicts conventional cross-sectional estimates. We therefore disagree with the proposition that the new twins-based studies have solved the return-to-schooling literature’s chronic identification problem.

Nevertheless, we have argued that, even though the new between-twins estimates are inconsistent, they still may be useful. If one believes that the correlation between schooling and the wage equation’s error term is positive both in the cross-section and within twin pairs, then, once measurement error has been treated, both conventional cross-sectional estimation and between-twins estimation tend to overestimate the return to schooling. We have shown that this is especially true if the error in measuring schooling is mean-reverting but the correction for measurement error does not take this into account. While it is theoretically unclear which estimator’s upward inconsistency is less severe, most of the empirical evidence so far suggests that the between-twins estimates tend to be at least a little smaller and therefore provide a tighter upper bound on the return to schooling. In this way, the monozygotic-twins literature has complemented other approaches used to treat endogeneity inconsistency: the instrumental-variables strategy, comparisons of non-twin siblings and of dizygotic twins, and cross-sectional analyses that control for observable
measures of family background and academic performance.

We should note that this view of the contribution of the monozygotic-twins literature is a relatively benign one. If, contrary to the usual presumption, one suspects that schooling is negatively correlated with the wage equation’s error term, then both the cross-sectional and between-twins estimates are downward-inconsistent, and the usually lower between-twins estimates have failed to tighten the lower bound already provided by the cross-sectional estimates. And, if one is agnostic about the sign of the correlation between schooling and the wage equation’s error term, we still do not have a clue. Similarly, the bounding results can be undermined by adopting a more flexible model of measurement error than the models we have reviewed here.22

In his much-cited review article on uses of siblings data, Griliches (1979) cautioned that “one has to keep in mind that they are not a panacea and that simple within (between brothers or between twins) estimates are not necessarily closer to the ‘truth’”. After noting that between-siblings estimates of the return to schooling usually do not differ dramatically from the corresponding cross-sectional estimates, Griliches inferred, “Either the role of the left-out variable… is not particularly large or its structure is not much more ‘familial’ than that of the other variables in the model”. In other words, either the cross-sectional estimates and the between-siblings estimates both are approximately consistent, or else both are inconsistent in much the same way. We hope that the contributors to the new twins literature will begin to take this indeterminacy more seriously.

Acknowledgements

The authors thank Charlie Brown, Jere Behrman, Mark Rosenzweig, and an anonymous referee for commenting on an earlier draft, and Aidan and Charlie Geronimus (John Bound’s twin sons) for supplying loads of anecdotal evidence.

References


---

22 For example, Ashenfelter and Krueger’s (1994) IV estimator could be either upward- or downward-inconsistent if errors in measuring schooling were correlated with the error term in the wage equation or if the a, b, and c measurement error components in Eq. (25) were accompanied by an additional component specific to the twin being reported on, rather than the twin doing the reporting.