DOUBLE TROUBLE: ON THE VALUE
OF TWINS-BASED ESTIMATION
OF THE RETURN TO SCHOOLING

John Bound
Gary Solon

Double Trouble: On the Value of Twins-
Based Estimation of the Return to Schooling
John Bound and Gary Solon
NBER Working Paper No. 6721
September 1998
JEL No. J24

ABSTRACT

Several recent studies use the schooling and wage variation between monozygotic

twins to estimate the return to schooling. In this paper, 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.

John Bound
Department of Economics
University of Michigan
Ann Arbor, MI 48109
and NBER
jbound@psc.lsa.umich.edu

Gary Solon
Department of Economics
University of Michigan
Ann Arbor, MI 48109
gsolon@umich.edu

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 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

---

[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.

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.[2]

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 if 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.[3] Monozygotic twins share

---

[2] Finding such an instrument is not easy. Bound, Jaeger, and Baker (1995) and Bound and Jaeger (1996), for example, have questioned the validity of Angrist and Krueger's (1991) quarter-of-birth instruments. The instrumental-variables (IV) and twins-based estimation strategies are not as distinct as they seem at first. Least squares estimation of the regression of the between-twins difference in log wages on the between-twins difference in schooling is exactly equivalent to IV estimation of the regression of the log wage on schooling with schooling instrumented by the difference between the individual's schooling and that of the individual's twin. The criticisms we will make of twins-based estimation can be thought of as arguments that the between-twins difference in schooling is an invalid instrument. The point that relying on between-twins variation can *worsen* the inconsistency due to endogenous schooling is just a special case of the point -- emphasized in Bound, Jaeger, and Baker (1995) -- that IV estimation with an invalid instrument can be more inconsistent than OLS estimation.

[3] In the final chapter of his dissertation, Gorseline himself recommended that, in future research, "The persons who should be studied preferably should be identical twins...."

the exact same 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 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 (forthcoming) -- 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-

3

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

In a series of studies published between 1976 and 1980, Jere Behrman, Paul Taubman, and various colleagues estimated the return to schooling with a sample of monozygotic twins from the National Academy of Sciences-National Research Council Twin Registry. Behrman and Taubman's estimates stood as the only ones of their kind until Ashenfelter and Krueger's (1994) study of participants in the 1991 Annual Twins Day Festival in Twinsburg, Ohio. Ashenfelter and Rouse's (1998) study incorporated data from the 1992 and 1993 Twinsburg festivals, and Rouse (forthcoming) added data from the 1995 festival. The Twinsburg studies have sparked interest in analyzing still other data on monozygotic twins. Behrman and Rosenzweig (forthcoming) used data from the Minnesota Twin Registry. In addition to these studies based on U.S. samples, Miller, Mulvey, and Martin (1995) used the Australian Twin Register, and Isacsson (1997) used the Swedish Twin Registry.

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

4

from regressions that control for age, gender (in those studies that include women as well as men), and race (in the U.S. studies that include nonwhites). The sixth column, labeled $\hat{\beta}_{cv}$ 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 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 U.S. twins studies range from 0.08 to 0.11 and are fairly typical of what is found in cross-sectional U.S. 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

5

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 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 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

$$(1) \qquad y_i = \alpha + \beta S_i + \varepsilon_i$$

where y_i is worker i's log wage rate, S_i is her years of schooling, and ε_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

(2) $\quad U_i = y_i - C_i(S)$

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.[4] 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

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

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

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

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 equation (4) in for the left side of equation (3) leads to an explicit solution for worker i's optimal years of schooling:

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

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 equation (1). Then the probability limit of the OLS estimator of the return to schooling β is

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

Thus, the conventional OLS estimator is consistent only if $Cov(w_i, \varepsilon_i) = 0$, which is to say, only if schooling is exogenous with respect to the wage equation. The common fear, however, has been that $Cov(w_i, \varepsilon_i)$ is negative. For example, suppose that individuals

---

[4] Rosen (1977) derives a special case of equation (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.

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 ε_i. Then

$Cov(w_i, \varepsilon_i)$ would be negative, and OLS would tend to overestimate β.

This, of course, is just the sort of situation that motivated Gorseline 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 equation (1)

with

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

where y_{ij} denotes the log wage of the jth twin in the ith family and so forth. Assume

that ε_{ij} follows the error-components model

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

where $Var(f_i) = \sigma_f^2$, $Var(u_{ij}) = \sigma_u^2$, and $Cov(f_i, u_{ij}) = Cov(u_{i1}, u_{i2}) = 0$. The implied

twin correlation in ε_{ij},

(9) $\quad Corr(\varepsilon_{i1}, \varepsilon_{i2}) = \sigma_f^2 / (\sigma_f^2 + \sigma_u^2)$,

is positive. This accords with the supposition that ε_{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,[5] the

positive twin correlation in ε_{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, but, as an empirical matter, it is clear that the factors pushing toward a

positive correlation dominate.[6]

---

[5] See Solon (forthcoming).

[6] For example, if it were not for the positive twin correlation in ε_{ij}, the GLS estimates in Table 1 would not lie in between the OLS and covariance estimates.

Similarly, replace equations (2) and (4) with

(10) $\quad U_{ij} = y_{ij} - C_{ij}(S)$

and

(11) $\quad C_{ij}{}'(S) = \delta S + w_{ij}$,

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

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

with $Var(g_i) = \sigma_g^2$, $Var(z_{ij}) = \sigma_z^2$, and $Cov(g_i, z_{ij}) = Cov(z_{i1}, z_{i2}) = Cov(z_{i1}, u_{i2}) =$

$Cov(z_{i2}, u_{i1}) = 0$. Then the optimal schooling for the jth twin in family i is

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

and the twin correlation in schooling is

(14) $\quad Corr(S_{i1}, S_{i2}) = \sigma_g^2 / (\sigma_g^2 + \sigma_z^2)$.

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

finding that the correlation between monozygotic twins in their self-reported schooling is

about 0.75.

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:

(15) $\quad y_{i1} - y_{i2} = \beta(S_{i1} - S_{i2}) + u_{i1} - u_{i2}$.

The probability limit of this estimator is

(16) $\quad plim \, \hat{\beta}_{CV} = \beta + Cov(S_{i1} - S_{i2}, u_{i1} - u_{i2}) / Var(S_{i1} - S_{i2}) = \beta - \delta Cov(z_{ij}, u_{ij}) / \sigma_z^2$.

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 β 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, twin 1 also may have a

higher wage error component u_{ij} because he is more inclined to study hard, work hard on

the job, invest in on-the-job training, and so forth. If the between-twins variation in schooling is generated by this sort of scenario, then $Cov(z_{ij}, u_{ij})$ is negative, and the covariance estimator tends to overestimate β.

The covariance estimator therefore is vulnerable to much the same sort of inconsistency from endogeneity as the conventional cross-sectional estimator is. Indeed, the inconsistency of the covariance estimator is *more* severe if the regression of the twin-specific wage error component u_{ij} on the twin-specific schooling error component z_{ij} has a coefficient of greater magnitude than does the regression of the family wage error component f_i on the family schooling error component g_i. It is hardly obvious whether or not that condition is satisfied.

To gain better insight into that condition, Griliches (1979) has modeled the schooling error term w_{ij} as

(17) $w_{ij} = h_{ij} + k_{ij}$

and the wage error term ε_{ij} as

(18) $\varepsilon_{ij} = \theta h_{ij} + m_{ij}$

where h_{ij}, k_{ij}, and m_{ij} are mutually uncorrelated. This formulation partitions the schooling error term into the endogenous component h_{ij}, which also appears in the wage error term, and the exogenous component k_{ij}. The probability limit of $\hat{\beta}_{OLS}$ shown in equation (6) can then be reexpressed as

(19) $plim \; \hat{\beta}_{OLS} = \beta - \delta\theta Var(h_{ij}) / [Var(h_{ij}) + Var(k_{ij})]$.

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}:

(20) $plim \; \hat{\beta}_{CV} = \beta - \delta\theta Var(h_{i1} - h_{i2}) / [Var(h_{i1} - h_{i2}) + Var(k_{i1} - k_{i2})]$.

11

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?[7] 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 parent 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

---

[7] Of course, without such differences, the regression of the between-twins wage difference on the between-twins schooling difference would not be identified. Ashenfelter and 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, and Taubman (1994) report corresponding figures of 52 percent for the NAS-NRC twins and 56 percent for the Minnesota twins.

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."[8] Less extreme differences are quite common. Behrman, Rosenzweig, and Taubman (1994) report that, in their sample of monozygotic twins from Minnesota, 69 percent of the twin pairs had birth weights differing by at least four ounces, and 48 percent differed by at least eight ounces. They also report that these differences in birth weight are significantly correlated with differences in subsequent educational attainment as well as occupational position. They find, for example, that a four-ounce 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.[9]

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 doesn't), 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.

---

[8] See Munsinger (1977), Kamin (1978), Marsh (1979, 1980), and James (1982).

[9] See Babson et al. (1964), Churchill (1965), Willerman and Churchill (1967), Kaelber and Pugh (1969), Babson and Phillips (1973), and Fujikura and 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).

Another plausible source of variation between some monozygotic twins is their psychological need to differentiate themselves from each other.[10] 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.[11] 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 comprises a smaller share of the between-twins schooling variation than it does of the between-families variation, which means it is uncertain whether the covariance estimator based on between-twins variation is subject to less inconsistency than the conventional OLS estimator.

So what, if anything, can be learned from twins-based estimates of the return to schooling? One optimistic answer would be that, if endogeneity of schooling were the only problem with estimating the wage-schooling regression, and if we were confident that

---

[10] 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."

[11] 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.

the cost-of-schooling error term and the wage error term are negatively correlated both in the cross-section and within twin pairs, then both the OLS estimator and the covariance estimator would be upward-inconsistent. In that case, since the covariance estimates in Table 1 are typically smaller than the OLS estimates, we could credit the twins-based literature for having tightened 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 inconsistency in estimation of the return to schooling. These include the instrumental-variables strategy, Gorseline's approach of comparing non-twin siblings, comparing dizygotic twins, and embellishing cross-sectional regressions with observable controls for academic performance (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.[12]

Unfortunately, though, several researchers -- such as Griliches (1977, 1979) and Ashenfelter and 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 educational attainment and its relationship with wages (Willis, 1986), the sign of the endogeneity inconsistency is theoretically indeterminate. To make matters worse, endogeneity of true schooling is not the only source of inconsistency 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.

---

[12] Just as it is unclear whether the monozygotic-twins approach is more or less subject to endogeneity inconsistency than cross-sectional OLS estimation, it also is unclear how it compares to the other approaches. This is just a generalization of Chamberlain's (1977) point that monozygotic twins may not provide any more identification leverage than dizygotic twins or non-twin siblings do.

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 equation (1) with the true S_i proxied by worker i 's report leads to a modest attenuation inconsistency 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 underestimation of the return to schooling. If, however, we suspect that the endogeneity of schooling causes an upward inconsistency, 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 inconsistency.

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.

16

Classical Measurement Error

Suppose that self-reported schooling S_{ij}^* is generated by the classical errors-in-variables model

(21) $S_{ij}^* = S_{ij} + v_{ij}$

where the measurement error v_{ij} is uncorrelated with both the family components and *each* twin's individual-specific components of both schooling and the wage equation's error term. That is, $Cov(v_{ij}, f_i) = Cov(v_{ij}, u_{i1}) = Cov(v_{ij}, u_{i2}) = Cov(v_{ij}, g_i) = Cov(v_{ij}, z_{i1}) = Cov(v_{ij}, z_{i2}) = 0$. Then applying least squares to the cross-sectional regression of y_{ij} on the error-ridden S_{ij}^* yields an estimator of the return to schooling with probability limit

(22) $plim \ \hat{\beta}_{OLS} = \beta + Cov(S_{ij}, \varepsilon_{ij}) / [Var(S_{ij}) + Var(v_{ij})]$

$\qquad\qquad\qquad - \beta Var(v_{ij}) / [Var(S_{ij}) + Var(v_{ij})]$.

Now, with measurement error incorporated into the analysis, the inconsistency of $\hat{\beta}_{OLS}$ consists of two terms. The first term, caused by the endogeneity of true schooling, has the same numerator as the inconsistency term in equation (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, $\hat{\beta}_{OLS}$ cannot be downward-inconsistent by very much. If the endogeneity inconsistency is positive and large, $\hat{\beta}_{OLS}$ 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 $\hat{\beta}_{OLS}$ 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

$$(23) \quad plim \; \hat{\beta}_{CV} = \beta + Cov(S_{i1} - S_{i2}, u_{i1} - u_{i2}) / [Var(S_{i1} - S_{i2}) + Var(v_{i1} - v_{i2})]$$

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

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 equation (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, 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. [13] The second term, the errors-in-variables attenuation inconsistency, also is 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.

---

[13] Under the given assumptions, $Var(S_{i1} - S_{i2}) = Var(S_{ij}) / 3$, and $Var(v_{i1} - v_{i2}) = 2Var(v_{ij}) = 2Var(S_{ij}) / 9 = (2/3)Var(S_{i1} - S_{i2})$.

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:

$$(24) \quad \widetilde{S}_{ij} = S_{ij} + \widetilde{v}_{ij}.$$

Ashenfelter and Krueger (1994) then devised a clever estimation procedure that uses the multiple measures of schooling to correct for the inconsistency from measurement 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 $\widetilde{v}_{ij}$ in the twin

report obeys the same assumptions listed for v_{ij} after equation (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

$$
\begin{aligned}
v_{i1} &= a_i + b_{i1} + c_{i1} \\
v_{i2} &= a_i + b_{i2} + c_{i2} \\
\tilde{v}_{i1} &= a_i + b_{i2} + \tilde{c}_{i1} \\
\tilde{v}_{i2} &= a_i + b_{i1} + \tilde{c}_{i2}
\end{aligned}
$$

(25)

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

(26) $\quad Cov(v_{i1}, v_{i2}) = Cov(v_{i1}, \tilde{v}_{i1}) = Cov(v_{i2}, \tilde{v}_{i2}) = Cov(\tilde{v}_{i1}, \tilde{v}_{i2}) = \sigma_a^2$

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

(27) $\quad Cov(v_{i1}, \tilde{v}_{i2}) = Cov(\tilde{v}_{i1}, v_{i2}) = \sigma_a^2 + \sigma_b^2$

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.[14]

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 $y_{i1} - y_{i2}$ on $S_{i1}^* - \tilde{S}_{i2}$ with $\tilde{S}_{i1} - S_{i2}^*$ as the instrument for $S_{i1}^* - \tilde{S}_{i2}$. The probability limit of this estimator is

(28) $\quad plim\ \hat{\beta}_{IV} = \beta + Cov(S_{i1} - S_{i2}, u_{i1} - u_{i2}) / Var(S_{i1} - S_{i2})$.

---

[14] 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.

20

Comparing this expression to equation (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 equation (23), the probability limit of the covariance estimator when there *is* measurement error. A comparison of equations (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 (forthcoming), 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 $\hat{\beta}_{IV} = 0.129$ does indeed exceed their $\hat{\beta}_{CV} = 0.092$. 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.[15]

Finally, let's compare equation (28) to equation (22), the probability limit of the conventional cross-sectional OLS estimator. If we inflate the OLS estimator by $[Var(S_{ij}) + Var(v_{ij})] / Var(S_{ij}) \approx 1.1$, 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

---

[15] Some of the twins studies also report alternative IV estimates that appropriately correct for measurement error only under the assumption that $\sigma_b^2 = 0$. 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 and Krueger (1994) and Rouse (forthcoming) report strong evidence against this assumption.

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 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 $\hat{\beta}_{IV} = 0.129$ exceeds their $\hat{\beta}_{OLS} = 0.084$ 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 (forthcoming) 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, $\hat{\beta}_{IV}$ comes out smaller than $\hat{\beta}_{OLS}$ even before the OLS estimate is inflated to compensate for measurement error.[16] In the one exception, Rouse (forthcoming), $\hat{\beta}_{IV}$ exceeds $\hat{\beta}_{OLS}$ so slightly that it also provides a smaller estimate once $\hat{\beta}_{OLS}$ 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 U.S. 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

---

[16] Behrman and Rosenzweig 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.

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.[17]

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. But two recent studies -- by Kane, Rouse, and Staiger (1997) and Berger, Black, and Scott (1997) -- have argued both theoretically and empirically that the measurement error in schooling is instead 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 equations (21) and (24) with

$$(29) \quad S_{ij}^{*} = \gamma S_{ij} + v_{ij}$$

and

$$(30) \quad \widetilde{S}_{ij} = \gamma S_{ij} + \widetilde{v}_{ij}$$

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

---

[17] 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.

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.[18] 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 probability limit previously shown in equation (22) now needs to be multiplied through by a correction factor

(31) $\quad \lambda_{OLS} = \gamma[Var(S_{ij}) + Var(v_{ij})] / [\gamma^2 Var(S_{ij}) + Var(v_{ij})]$.

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

(32) $\quad Var(S_{ij}) > Var(v_{ij}) / \gamma$.

Given the evidence that, in the cross-section, the "signal" variance $Var(S_{ij})$ far exceeds the "noise" variance $Var(v_{ij})$, this condition is met unless γ is implausibly small.

Therefore, if we were right in the preceding subsection when we conjectured that the endogeneity of schooling causes $\hat{\beta}_{OLS}$ 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 equation (23) now needs to be multiplied through by a correction factor

(33) $\quad \lambda_{CV} = \gamma[Var(S_{i1} - S_{i2}) + Var(v_{i1} - v_{i2})] / [\gamma^2 Var(S_{i1} - S_{i2}) + Var(v_{i1} - v_{i2})]$.

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

(34) $\quad Var(S_{i1} - S_{i2}) > Var(v_{i1} - v_{i2}) / \gamma$.

---

[18] 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 and 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.

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 equation (28) now needs only to be divided through by γ. 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 $Cov(S_{i1} - S_{i2}, u_{i1} - u_{i2}) = 0$, so that there is no endogeneity inconsistency in the between-twins estimation, the probability limit of Ashenfelter and Krueger's estimator is $\beta / \gamma > \beta$. Thus, their IV strategy, which is designed to correct for a variant of classical measurement error, *over*corrects when the measurement error really is mean-reverting. This result is reminiscent of similar cross-sectional results reported in Kane, Rouse, and Staiger (1997) and Berger, Black, and 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 1970's. 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 recently reiterated by Neumark (forthcoming) -- 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

26

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 don't 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.[19]

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.

---

[19] For example, Ashenfelter and Krueger's 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 equations (25) were accompanied by an additional component specific to the twin being reported on, rather than the twin doing the reporting.

References

Ainslie, R. C. (1985) *The Psychology of Twinship.* University of Nebraska Press, Lincoln.

Angrist, J. D. and Krueger, A. B. (1991) Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106, 979-1014.

Ashenfelter, O. and Krueger, A. (1994) Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* 84, 1157-1173.

Ashenfelter, O. and Rouse, C. (1998) Income, schooling, and ability: evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113, 253-284.

Babson, S. G., Kangas, J., Young, N., and Brahhall, J. L. (1964) Growth and development of twins of dissimilar size at birth. *Pediatrics* 33, 327-333.

Babson, S. G. and Phillips, D. S. (1973) Growth and development of twins dissimilar in size at birth. *New England Journal of Medicine* 289, 937-940.

Becker, G. S. (1975) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* 2nd ed. University of Chicago Press, Chicago.

Behrman, J. R. and Rosenzweig, M. R. (forthcoming) "Ability" biases in schooling returns and twins: a test and new estimates. *Economics of Education Review.*

Behrman, J. R., Rosenzweig, M. R., and Taubman, P. (1994) Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of Political Economy* 102, 1131-1174.

Berger, M. C., Black, D. A., and Scott, F. A. (1997) Bounding parameter estimates with non-classical measurement error. Unpublished manuscript, Department of Economics, University of Kentucky.

Bound, J. and Jaeger, D. A. (1996) 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?" Working paper no. 5835, National Bureau of Economic Research.

Bound, J., Jaeger, D. A., and Baker, R. M. (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, 443-450.

Bound, J. and Krueger, A. B. (1991) The extent of measurement error in longitudinal earnings data: do two wrongs make a right? *Journal of Labor Economics* 9, 1-24.

Card, D. (1995) Earnings, schooling, and ability revisited. *Research in Labor Economics* 14, 23-48.

Card, D. (forthcoming) The causal effect of education on earnings. In *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card. North-Holland, Amsterdam.

Chamberlain, G. (1977) Are brothers as good as twins? In *Kinometrics: Determinants of Socioeconomic Success within and between Families*, ed. P. Taubman. North-Holland, Amsterdam.

Churchill, J. A. (1965) The relationship between intelligence and birthweight in twins. *Neurology* 15, 341-347.

Farber, S. L. (1981) *Identical Twins Reared Apart: A Reanalysis*. Basic Books, New York.

Fujikura, T. and Froehlich, L. A. (1974) Mental and motor development in monozygotic co-twins with dissimilar birth weights. *Pediatrics* 53, 884-889.

Gorseline, D. E. (1932) *The Effect of Schooling upon Income*. Indiana University Press, Bloomington.

Greene, W. H. (1997) *Econometric Analysis*, 3rd ed. Prentice Hall, Upper Saddle River.

Griliches, Z. (1977) Estimating the returns to schooling: some econometric problems. *Econometrica* 45, 1-22.

Griliches, Z. (1979) Sibling models and data in economics: beginnings of a survey. *Journal of Political Economy* 87, S37-S64.

Isacsson, G. (1997) *Estimates of the Return to Schooling in Sweden from a Large Sample of Twins*. Swedish Institute for Social Research, Stockholm.

James, W. H. (1982) The IQ advantage of the heavier twin. *British Journal of Psychology* 73, 513-517.

Kaelber, C. T. and Pugh, T. F. (1969) Influence of intrauterine relations on the intelligence of twins. *New England Journal of Medicine* 280, 1030-1034.

Kamin, L. J. (1978) Transfusion syndrome and the heritability of IQ. *Annals of Human Genetics* 42, 161-171.

Kane, T. J., Rouse, C. E., and Staiger, D. (1997) Estimating returns to schooling when schooling is misreported. Unpublished manuscript, Kennedy School of Government, Harvard University.

Marsh, R. W. (1979) The identical-twin transfusion syndrome and intelligence: a second look. *British Journal of Psychology* 70, 413-415.

Marsh, R. W. (1980) The significance for intelligence of differences in birth weight and health within monozygotic twin pairs. *British Journal of Psychology* 71, 63-67.

Miller, P., Mulvey, C., and Martin, N. (1995) What do twins studies reveal about the economic returns to education? A comparison of Australian and U.S. findings. *American Economic Review* 85, 586-599.

Munsinger, H. (1977) The identical-twin transfusion syndrome: a source of error in estimating IQ resemblance and heritability. *Annals of Human Genetics* 40, 307-321.

Neumark, D. (forthcoming) Biases in twin estimates of the return to schooling. *Economics of Education Review*.

Pencavel, J. H. (1976) A note on the IQ of monozygotic twins raised apart and the order of their birth. *Behavior Genetics* 6, 455-460.

Pischke, J.-S. (1995) Measurement error and earnings dynamics: some estimates from the PSID validation study. *Journal of Business and Economic Statistics* 13, 305-314.

Rosen, S. (1977) Human capital: a survey of empirical research. *Research in Labor Economics* 1, 3-39.

Rouse, C. E. (forthcoming) Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review*.

Scarr, S. (1982) Effects of birth weight on later intelligence. *Social Biology* 29, 230-237.

Solon, G. (forthcoming) Intergenerational mobility in the labor market. In *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card. North-Holland, Amsterdam.

Taubman, P. (1976) Earnings, education, genetics, and environment. *Journal of Human Resources* 11, 447-461.

Willerman, L. and Churchill, J. A. (1967) Intelligence and birthweight in identical twins. *Child Development* 38, 623-629.

Willis, R. J. (1986) Wage determinants: a survey and interpretation of human capital earnings functions. In *Handbook of Labor Economics*, vol. 1, eds. O. Ashenfelter and R. Layard. North-Holland, Amsterdam.

Table 1

Estimated Returns to Schooling Based on Samples of Monozygotic Twins

Study	Sample	Wage Measure	$\hat{\beta}_{OLS}$	$\hat{\beta}_{GLS}$	$\hat{\beta}_{CV}$	$\hat{\beta}_{IV}$
Taubman (1976)	NAS-NRC	Log annual earnings in 1973	.079 (.002[a])		.027 (.008)	
Ashenfelter and Krueger (1994)	Twinsburg	Log hourly wage in 1991	.084 (.014)	.087 (.015)	.092 (.024)	.129 (.030)
Ashenfelter and Rouse (1998)	Twinsburg	Log real hourly wage in 1991, 1992, or 1993	.110 (.009)	.102 (.010)	.070 (.019)	.088 (.025)
Rouse (forthcoming)	Twinsburg	Log real hourly wage in 1991, 1992, 1993, or 1995	.105 (.008)	.101 (.009)	.075 (.017)	.110[b] (.023)
Behrman and Rosenzweig (forthcoming)	Minnesota	Log annualized full-time earnings in 1993		.118[c] (.005)		.104[c] (.017)
Miller, Mulvey, and Martin (1995)	Australia	Log of mean 1985 annual earnings in individual's usual occupation	.064[d] (.002)		.025[d] (.005)	.048[d] (.010)
Isacsson (1997)	Sweden	Three-year average of log annual earnings in 1987, 1990, and 1993	.049[e] (.002)		.023[e] (.004)	.024[e] (.008)

[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.