

Earnings, Education, Genetics, and Environment

Author(s): Paul Taubman

Source: *The Journal of Human Resources*, Vol. 11, No. 4 (Autumn, 1976), pp. 447-461

Published by: University of Wisconsin Press

Stable URL: <http://www.jstor.org/stable/145426>

Accessed: 20-02-2018 12:15 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

University of Wisconsin Press is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Human Resources*

EARNINGS, EDUCATION, GENETICS, AND ENVIRONMENT

PAUL TAUBMAN

ABSTRACT

A major and well-recognized difficulty in estimating the effects of education on earnings is that the more educated are likely to be more able, irrespective of education. If ability also determines earnings and is not controlled, ordinary least squares will yield biased estimates of the education coefficient. In this study, we use data on identical twins to control for differences in ability that arise from genetic endowments and family environment. Not controlling for genetics and family environment may cause a large bias, up to two-thirds of the noncontrolled coefficient.

I. INTRODUCTION

Much research in economics has been concerned with the sources of the inequality in earnings. While the human capital model has focused attention on the role of education, on-the-job training, and, more peripherally, innate and acquired cognitive skills, other economists have examined such diverse factors as risk aversion, healthiness, and compensating differences for nonpecuniary rewards.¹ While the development of better data, estimating techniques, and economic theories have all led to an improved understanding of the sources of inequality in earnings, our empirical knowledge is unsatisfactory on a number of important issues.

A widely recognized problem in prior studies of the impact of education on earnings is that people with more education are likely to be more "able," net of any additional ability produced by education. Unfortunately, since "ability" can encompass many particular skills including cognitive, affective, and psycho-

The author is Professor of Economics, University of Pennsylvania.

[Manuscript submitted September 1975; accepted March 1976.]

¹ For a summary and extension of the human capital model, see Mincer [8]. For a discussion of risk aversion and other factors, see Taubman [12].

logical, few if any of which are measured adequately, it is difficult to pin down the effects of education. This study uses data on twins to attack this ability problem from a different angle. When we regress differences in brothers' earnings on differences in schooling, we hold constant (by eliminating) those abilities that are common to the brothers. For fraternal twins, we eliminate skills produced by the common or family environment, while for identical twins we eliminate common environment and skills based on genetic endowments.

Data on twins can also be used to explore the extent to which the variance in earnings is attributable to the sum and the separate effects of genetic endowments and common (family) environment and the extent to which this variance is attributable to noncommon environment.² The verb *explore* was used in the preceding sentence both because this is the first study in economics to examine these issues and because there are major difficulties or unproven key assumptions associated with the model used in this paper. However, work undertaken subsequent to that reported here gives hope that many of these assumptions can be tested and, if valid, replaced with more tenable versions (see [1]).

In this paper I will use a version of the human capital model that is fairly general and not particularly rigorous. Our basic assumption will be that a person will be paid a real wage rate (the money wage rate divided by the price level, which will be set at 1) equal to his marginal product. A person's marginal productivity depends on a variety of skills and attributes. Unfortunately, about the only potential skill that has been widely examined is "cognitive" ability.³

Economists and sociologists have related earnings to education, years of experience, and other variables on the grounds that if schooling produces one or more skills that are reimbursed in the marketplace, then the more educated should receive higher earnings. We will generalize this model by assuming that skills are produced by a combination of genetic endowments and environment, where the latter is defined to encompass "everything else." If years of schooling happens to be the only input into skills that we measure, we can write:

$$(1) \quad \ln Y = aS + bG + cN + u$$

where $\ln Y$ is the log of earnings, S is years of schooling, G is genetic endowments, N is environment other than schooling, and u is a random error.⁴

2 The common environment will include neighborhood and other nonfamily effects and will exclude some family effects. See below for details.

3 It is difficult to get a clear picture of the effect of this skill because different studies use different concepts and measures of cognitive ability and study people with different years of labor market experience. See, for example, Taubman and Wales [14], Griliches and Mason [6], and Sewell and Hauser [9]. The latter study contains evidence that cognitive skills begin to have an effect only after five years of labor force experience. See Taubman [12] for an attempt to measure the effects of some noncognitive skills.

4 If measures of other skill producers are available, they can be included with a suitable redefinition of G and N .

This paper will attempt to shed light on some of the influences of genetics, family, and other environment for earnings for white men aged about 50 in 1973. We will use both regression and variance analyses to study several different but related issues.

II. GENETICS AND ENVIRONMENT

By genetic endowments (*G*), we mean the innate capabilities that are based on a person's genes, half of which are contained in the egg and the other half in the sperm.⁵ Environment (*N*) includes all other systematic and nonsystematic determinants of skills, including prenatal development. While environment is "everything else," some particular aspects that are usually thought to be important include family, peer group, on-the-job training, schooling, and military. As this list indicates, an individual's environment includes elements over which he has both little and much choice.

Twin Types

Males and females normally have 23 pairs of chromosomes. The genes are contained in the chromosomes. Each gene contains two halves that may be the same or different, for example, AA or AB. We will assume that each skill or trait is influenced by many genes. Only a randomly determined half of each gene and of each chromosome is contained in the egg or sperm, each of which is a gamete of one parent. But once the egg is fertilized, that is, the two gametes combine, the two halves of each gene are merged to form the individual's genes.

There are two types of twins—monozygotic (*MZ*) and dizygotic (*DZ*). The *MZ*s, often known as "identical," are the result of the splitting of an already fertilized egg, while the *DZ*s, or "fraternal twins," are the result of two different eggs fertilized by two different sperm. Thus, *DZ* pairs do not have the same genetic composition, although they will be more alike than randomly drawn individuals. The *MZ* pairs, however, have the same genetic makeup because each piece of the split fertilized egg contains all and only the genetic information of the original fertilized egg (barring mutations).⁶

III. THE NAS-NRC TWIN SAMPLE

In this study, we will use the NAS-NRC twin sample, which is described in more detail in the Appendix. Briefly, however, for this study we have a maximum of

5 We are ignoring mutations which occur very rarely, i.e., about once in 100,000 or less.

For discussion of the biological and statistical aspects of genes, see [4].

6 For a more complete discussion of the biological aspects, see [4].

2,478 *pairs*, where each brother answered a questionnaire mailed in 1974.⁷ Most of these pairs also answered several earlier questionnaires to which we have access. To be included in the mailing, the twins had to be born between 1917 and 1927, to be white, to have served at some point in the military, and to be alive in 1974. These last restrictions suggest an underrepresentation of people with low intelligence or education or with poor mental and physical health, as compared to the corresponding white male cohort. Even compared to a population of veterans, the respondents to our questionnaire have more earnings and education.^{8,9} Regression estimates from stratified samples with nonpopulation weights still yield unbiased coefficients over the sample space.¹⁰

In some of our analysis, we assume that MZ and DZ pairs in our sample are random drawings from the same population.¹¹ The DZ pairs may have different distributions of genes and/or environment because DZ pairs as a percentage of births occur more frequently among older women. The corresponding percentage for MZ pairs is independent of mother's age and SES class.¹²

The means and variances of earnings, schooling, and several other variables by twin types are given in Table 1. It is evident that in our sample, the DZ twins come from families in which the parents have a bit less education and occupational status, and in which the number of siblings and older siblings are somewhat greater, although the differences are not statistically significant. The religious distributions are also very similar for MZs and DZs, which is a bit surprising to me since I would have expected Catholics to be a larger portion of

-
- 7 We eliminate pairs if either does not have earnings whose derivation is described in Taubman [11]. About 100 pairs whose zygosity is unknown are not used in this analysis. If a person's education is unknown, we set it equal to his brother's since even for DZ pairs, the brothers' correlation for education is almost .5. If neither brother responded to the education question, we used the mean of 13. Less than 20 cases were adjusted.
- 8 The average 1973 earnings and education in our sample are \$18,000 and 13 years. In the population as a whole, the corresponding figures for white veterans of the same cohort are about \$12,000 and 12 years. About one quarter of the differential can be eliminated if we reweight by parental education and region of birth so as to produce the average of white males born during the period 1917–27 on these variables.
- 9 While our sample is not representative of the population, it seems likely that we have over- and underrepresented various population groups rather than excluded all their members. For example, in Stauffer and others [10], it is indicated that there were huge differences in disqualification for mental problems by induction camp, ranging from .5 to 50 percent.
- 10 However, since less than 5 percent of the sample had less than a ninth grade education, our results may not be appropriate for those with low education.
- 11 In our statistical analysis, it is necessary to distinguish between MZ and DZ twins. For the most part, the twins' zygosity is determined by their answers to: "As children, were you and your twin alike as 'two peas in a pod' or of only ordinary resemblance?" This simple question assigns pairs accurately almost 95 percent of the time. See [11], App. A, for details.
- 12 See [11], App. A, for discussion and references.

TABLE 1
SOME SUMMARY STATISTICS FOR INDIVIDUALS IN THE NAS-NRC SAMPLE
(calculated separately for MZs and DZs)

	MZs		DZs	
	Mean	Variance	Mean	Variance
1973 annual earnings	18.4 ^a	150 ^b	18.1 ^a	166 ^b
ln 1973 annual earnings	9.67 ^a	.28	9.64 ^a	.32
1967 or 1972 occupational score ^d	50.4	472	49.8	445
Years of schooling	13.5	9.1	13.3	9.8
Initial full-time civilian occupation ^{c,d}	36.7	610	35.0	590
Age	51.0	8.4	51.2	8.8
Mother's education years	10.0	9.6	9.7	11.9
Father's education years	9.3	12.6	9.1	14.8
Father's occupational status ^d	29.6	532	28.6	503
% Catholic	26	19	23	18
% Jewish	4	4	5	5
% other non-Protestant	2	2	3	3
Number of siblings alive 1940	2.6	4.9	3.0	5.6
Number of older siblings alive 1940	1.6	3.3	2.1	3.7
Number of pairs	1019		907	

Note: Calculations are for those for whom earnings are nonzero for both brothers. For other variables, if one brother answered and the other did not, nonrespondent is set equal to his brother. If both did not answer, both are set at mean or put in "other category." In the whole tape, no responses were .5, 10, and 18 percent for education, initial occupation, and 1967 occupation, respectively. For mother and father data, if brothers' answers differ, mean of responses is used.

^aThousands of dollars.

^bMillions of dollars.

^cAs recalled in 1974.

^dCensus occupational status score.

the DZ pairs. The means of schooling, initial and later occupational status, and earnings are nearly the same across twin type, although the variances differ by up to 10 percent.¹³

The conclusion that MZs and DZs are random drawings from the same population is further strengthened by a comparison of the simple correlations given in Table 2.¹⁴ The left-hand portion of that table treats both brothers as

13 For samples of this size, the 5 percent level of significance in an F test is about 1.2.

14 See also the comparison below of our regression results with those based on Census data.

TABLE 2
INDIVIDUAL AND CROSS-SIB CORRELATIONS

	Individuals			Cross-Sib		
	<i>S</i>	<i>OC</i> ₆₇	<i>lnY</i>	<i>S</i>	<i>OC</i> ₆₇	<i>lnY</i>
MZs						
<i>S</i>	1	.54	.44	.76	.44	.40
<i>OC</i> ₆₇		1	.35		.43	.27
<i>lnY</i>			1			.54
DZs						
<i>S</i>	1	.51	.44	.54	.29	.29
<i>OC</i> ₆₇		1	.35		.20	.19
<i>lnY</i>			1			.30

individuals, and the results are close for MZs and DZs. The right-hand portion consists of cross-sib correlations, defined, for example, as $\sigma_{SY'}/\sigma_S\sigma_{Y'}$ where $\sigma_{SY'}$ is the covariance of one brother's years of schooling and his sib's log of earnings. The cross-sib correlations are uniformly lower than the comparable ones for individuals. The DZ cross-sib correlations are uniformly lower than the comparable MZ ones.

IV. REGRESSION ANALYSIS

In previous studies of the effects of education on earnings, it has not been possible to control completely for the other determinants of earnings that might be correlated with schooling. With twins, however, it is possible to eliminate genetic differences for the identical twins and common background for both types by studying the within-pair differences in earnings.

To understand what can and what cannot be done with twins, it is necessary to compare the estimates obtained when using individuals and within-pair differences. As an aid in making this comparison, let us order individuals within each pair randomly, for example, alphabetically, and denote the within-pair difference by Δ . We can rewrite equation (1) as:

$$(1a) \quad \Delta \ln Y = a\Delta S + b\Delta G + c\Delta N + \Delta u$$

We can estimate both (1) and (1a) using OLS and compare the estimates. Denote the estimate from equation (1) as \hat{a}_1 and that from (1a) as \hat{a}_2 . Using standard methods, it can be shown that:

$$(2) \quad \text{plim}(\hat{a}_1) = a + [\text{plim cov}(S, bG + cN + u)] / \text{plim var}(S)$$

$$(3) \quad \text{plim}(\hat{a}_2) = a + [\text{plim cov}(\Delta S, b\Delta G + c\Delta N + \Delta u)] / \text{plim var}(\Delta S)$$

TABLE 3
 INDIVIDUAL AND WITHIN-PAIR REGRESSIONS FOR lnY
 (t-statistics in parentheses)

Based on individuals:				
1. lnY =	9.019 +	.0789S -	.0084Age	$\bar{R}^2 = .201$
	(65.1)	(32.2)	(3.2)	
2. lnY =	8.58 +	.0795S		$\bar{R}^2 = .199$
	(262.5)	(32.5)		
MZ within-pair				
3. ΔlnY =	-.0064 +	.0270ΔS		$\bar{R}^2 = .012$
	(.2)	(3.6)		
DZ within-pair				
4. ΔlnY =	-.0090 +	.0594ΔS		$\bar{R}^2 = .069$
	(.3)	(8.2)		

As is well known, (2) yields biased estimates if $\text{plim cov}(S, bG + cN + u)$ is nonzero, which is generally thought to be the case for the earnings-schooling model.

For MZ twins ΔG is zero. Making the usual assumption that Δu is uncorrelated with ΔS , $\text{plim } \hat{a}_2$ will be unbiased provided either c is zero or ΔN is uncorrelated with ΔS . The first condition means that the differences in MZ brothers' environments have no direct effect on earnings, though they may affect schooling. The second condition means that the differences in environment that determine earnings are not correlated with schooling. If the bias in \hat{a}_1 arises only because of genetics or common environment, \hat{a}_2 will be a consistent estimate.

For DZ pairs, ΔG is not zero and \hat{a}_2 will not be consistent if $\text{plim cov}(\Delta S, b\Delta G)$ is not zero. Let us assume that the sibling environmental correlations are the same across twin types.¹⁵ Then we can determine if genetics should be controlled for by comparing MZ and DZ within equations. It is interesting to note that the bias in \hat{a}_2 based on DZ data can be larger than that for \hat{a}_1 because both the numerator and denominator in the plim expression change when we go from levels to within-pair differences.

Much recent empirical work on earnings has employed a semilog functional form such as:

$$(4) \quad \ln Y = d + aS + e(\text{Age} - S) = d + (a - e)S + e \text{Age}$$

We choose to use the same form partly to maintain comparability and partly because a variety of tests suggested that the semilog form was statistically better than double log or linear forms.

15 The same test can be used if the following somewhat weaker conditions prevail. Let N be divided into common and specific components as $N_{ij} = m_i + h_{ij}$, where i indicates family and j sib. Then $\Delta N = \Delta h_{ij}$. As long as Δh is uncorrelated with ΔS , the test is valid.

Table 3 contains four regressions for 1973 earnings. The first two are estimated from observations on individuals; the third and fourth are estimated from within-pair differences. Since the men in the sample range in age from 45 to 55 in 1973, it is not surprising that age has a small and even negative coefficient and that the education coefficient is barely changed when age is included. In both equations (1) and (2), in this table, the coefficient on education is about .08 and highly significant. In studies using Census data and a semilog form, Mincer [8, p. 92] also estimates the coefficient on education to be about .08.

Next let us examine the within-pair regressions. In the DZ within-pair version, our estimate of a is .059, while in the MZ within-pair version it is .027. Mincer [8] and others have shown that the difference in earnings by education level to be greatest around age 50. Thus, the MZ within-pair results suggest that years of schooling per se have only a small, though statistically significant, effect on earnings.

We have used the analysis of covariance (Chow test) to test the null hypothesis that the two within-pair equations are the same. With a calculated F of about 4.8 and with 2 and 1932 degrees of freedom, we can reject this null hypothesis at the 5 percent level. Thus, in this sample it appears necessary to control for genetic endowments when estimating the returns to schooling. Since the DZ within-pair estimates are less than the individual estimates, it is also necessary to control for family environment.

The combined bias from not controlling for genetics and common environment can be estimated as .051/.079, or about 65 percent. (Below we consider the possible importance of measurement error on this estimate.) Prior studies have tried to estimate the bias by including measures of IQ and/or family background. To the best of my knowledge, no one has found such a large bias. Of course, as Griliches [5] has pointed out, the bias depends on the effect of the omitted variable on earnings and the slope coefficient of the omitted variable on schooling. The latter coefficient may vary by cohort.¹⁶ But the NBER-TH sample, which has yielded some of the larger estimates of the bias from omitting IQ and family background, is drawn from about the same cohort.¹⁷ Yet the bias in that sample does not approach two-thirds. Thus, it appears that regressions that control for ability by using IQ, parental education and occupation, and other often available measures will not adequately control for ability.

While we do not currently have measures of IQ available, we do have a fairly wide list of measures that often have been used as proxies for environment and/or genetic endowments.¹⁸ We have reestimated equation (2) including many

16 Taubman and Wales [13] demonstrate that the IQ education slope has changed by cohort for those with at least a high school diploma.

17 See, for example, [14].

18 The information available will not let us determine if a variable is a proxy for environment and/or genetic endowments.

of these variables. The resulting equation is:

$$\begin{aligned}
 (5) \quad \ln Y = & 8.98 + .069S - .0074Age + .0052(S\ Mother) + .0059(S\ Father) \\
 & (63.9)(25.8) \quad (2.9) \quad (1.8) \quad (2.2) \\
 & + .00066(OC\ Father) - .0078Sibs + .051Cath + .321Jew \\
 & (2.0) \quad (2.3) \quad (2.8) \quad (8.8) \\
 & + .0068(Born\ South) - .0231(Raised\ rural) \quad R^2 = .224 \\
 & (.3) \quad (1.4)
 \end{aligned}$$

The education of the mother and father and his occupation (socioeconomic) status are generally included in such equations to account for differences in family income, child-rearing practices, and preferences inculcated in the offspring.¹⁹ In our equation the variables for the father are positive and significant and that for mother’s education positive and nearly significant. These findings are in rough accord with those of other researchers such as Blau and Duncan [3].²⁰ Number of siblings (alive in 1940, other than one’s twin) was included primarily to represent the reduction in time and resources available for each child as family size increases. As would be expected from this argument, the coefficient is negative and significant. The religious variables are included primarily to represent different family environments and stress on achieving that may be associated with various religious upbringings and family income levels in the 1920s and 1930s. We find that Jews and Catholics earn 33 and 5 percent more than Protestants and others. Both findings are in approximate quantitative accord with those in the NBER-TH sample, which contains men born between 1917 and 1924 (see [11]). Dummies for having been raised in a rural area and having been born in the South are included to capture opportunity and value differences. These variables may also capture school-quality differences. Both dummies are insignificant here, though they do have significant effects on years of schooling.

Our primary interest in including these proxy variables for environment and genetics is to determine their effect on the education coefficient. Once we include these proxies, this coefficient drops only to .07, or about 12 percent. Thus, as suggested earlier, these variables do not completely represent the variables for which they are intended to be proxies.

John Bishop and several other economists have pointed out to me that the previous statements on bias must be qualified because of measurement error. As is well known, if our true variable is s , but we measure S which has a measurement error of v , then the bias from the measurement error (assuming v is uncorrelated with Y and with s) depends on σ_v^2/σ_s^2 . Similarly, in the within-pair equations, the bias depends on $\sigma_{\Delta v}^2/\sigma_{\Delta s}^2$. We would expect $\sigma_{\Delta v}^2/\sigma_{\Delta s}^2$ to be

19 See [11, ch. 1] for a further discussion. The occupational status measure is Duncan’s SES score.

20 See, however, Sewell and Hauser [9] for a different picture.

TABLE 4
ESTIMATES OF THE TRUE EFFECT OF SCHOOLING ON $\ln Y$,
FOR VARIOUS VALUES OF MEASUREMENT ERROR VARIANCE

	From MZ Within-Pair Equation	From Individual Equation	True Bias
If:			
$\sigma_v^2 = .05\sigma_s^2$.032	.084	62%
$\sigma_v^2 = .10\sigma_s^2$.048	.088	45%
$\sigma_v^2 = .15\sigma_s^2$.070	.091	23%
$\sigma_v^2 = .20\sigma_s^2$.121	.096	-26%

greater than σ_v^2/σ_s^2 since the brothers' true schooling will be correlated, while the brother's measurement error either will not be correlated if we are dealing with wrongly reported numbers or will be less highly correlated if measurement error is expanded to include quality (denominated in units of earnings potential) as Welch has suggested.²¹

However, we can calculate the effect of any measurement error variance on the estimates from within and between estimates. The results (under the assumption that each estimate would be unbiased if there were no measurement error) given in Table 4 assume that the measurement error across brothers is independent.²² As is evident in the table, if independent measurement error for years of schooling is no greater than 10 percent, which seems large based on CPS-Census comparisons, the MZ within-pair equations still imply a large bias.²³

Thus far we have shown that it is necessary to control for genetics and family environment when we estimate the effects of education on earnings. In the next section we will attempt to estimate the contribution of genetics and environment to earnings.

V. ANALYSIS OF VARIANCE

Suppose we assume that schooling also is a function of genetic endowments and environment (though not necessarily the same G and N as in the earnings equation (1)). Then we can write a reduced-form equation for earnings in terms

21 $\sigma_{\Delta S}^2 = 2\sigma_s^2 - 2\sigma_{ss'}$, where prime indicates a sib's brother.

22 If measurement error arises because of differences in quality of schooling, which presumably are correlated across brothers, the MZ within-pair biases will be smaller than those shown in Table 4.

23 See Bishop [2]. The Census-CPS match may overstate measurement error since in the Census in some cases wives provide data for their husbands. This source of error is not found in our study.

of a single genetic index, which we will continue to label G , and a single environmental index, which will include the random errors and which will still be labeled N . Since both G and N are unobserved, we can dimension them such that their coefficients are each 1 in our earnings equation. Let Y stand for the log of earnings. Then our reduced-form equation is:²⁴

$$(6) \quad Y = G + N$$

Using (6), we can write the variance of Y as:

$$(7) \quad \sigma_Y^2 = \sigma_G^2 + 2\sigma_{GN} + \sigma_N^2$$

Let a brother be denoted by a prime and MZ pairs by an asterisk. Then we can calculate the cross-sib covariance as:

$$(8) \quad \sigma_{YY'}^* = \sigma_{GG'}^* + 2\sigma_{GN'}^* + \sigma_{NN'}^*$$

For MZ twins, $\sigma_{GG'}^* = \sigma_G^2$ and $\sigma_{GN'}^* = \sigma_{GN}^*$ because G is the same for the sibs. Thus, for MZ twins

$$(9) \quad \sigma_Y^2 - \sigma_{YY'}^* = \sigma_N^2 - \sigma_{NN'}^*$$

The term $\sigma_{NN'}^*$, which is the covariance of the brothers' environment, can be thought of as their common environment. This common environment is broader than family environment to the extent that neighborhood and peer group effects occur, but less than family environment to the extent that parents treat the twins differently.²⁵

For many purposes we would wish to include the neighborhood and peer group effects with the family effect. Thus, we can treat the variance in (9) as an upper bound to the contribution of nonfamily environment to the variance of the log of Y . We can estimate $(\sigma_Y^2 - \sigma_{YY'}^*)/\sigma_Y^2$ from Table 2 by taking the complement of the cross-sib correlation for an $\ln Y$. Using this approach, we find that noncommon environment accounts for at most 45 percent of the total variance of $\ln Y$. For comparison, noncommon environment at most accounts for 60 percent of occupational status and 25 percent of years of schooling.²⁶

It is also possible to calculate the cross-sib covariance for DZ brothers. The expression is, of course:

$$(10) \quad \sigma_{YY'} = \sigma_{GG'} + 2\sigma_{GN'} + \sigma_{NN'}$$

To be able to identify any additional parameters, it is necessary to make a

24 In (6) we have adapted a linear and additive form for $\ln Y$. Taubman [11] has shown that this form is acceptable for \ln earnings but not for earnings, though the results that follow are the same for $\ln Y$ and Y .

25 The twin literature on other subjects emphasizes the different treatment of twins.

26 When we redid the occupational status calculations after excluding the no-responses, the results were hardly changed.

number of assumptions which may or may not be valid. One set of assumptions proposed in Taubman [11] is that all genetic effects are additive, there is random mating, one brother's specific environment is not correlated with his sib's genes, and $\sigma_{NN'} = \sigma_{NN'}^*$. Even with these strong assumptions, the model is underidentified. Restricting the analysis to the cases where all the estimated variances and covariances are nonnegative and less than σ_Y^2 , Taubman [11] indicates that genetics and common environment would account for from 5 to 50 percent and from 18 to 5 percent of the total variance in Y , respectively.²⁷ Other assumptions made to identify the model could give different results, but our assumptions seem at least as plausible as any others.²⁸

In assessing the results in this section or more definitive ones based on more complex techniques, it is important to keep in mind certain limitations and caveats. First, the results may not be robust to changes in model specification. Second, the twin data yield estimates that are applicable to the population of which the twins are representative, which presumably is white males born between 1917 and 1927 who served in the military.²⁹ In World War II, men were rejected for service because of mental and physical defects that arose from a variety of genetic and environmental causes. It is not clear if our results generalize to all white males about 50 years old in 1973, because it is not known if rejections from the service were more severe with respect to genetic endowments or environment.

There are several other caveats that can best be understood if we rewrite equation (6) with the coefficients on G and N treated as prices and not standardized to equal 1. That is:

$$(6b) \quad Y_t = P_{Gt}G_t + P_{Nt}N_t$$

In order for our results to apply to white males at all times in the U.S., it is necessary that for all t , $P_{Gt} = P_G$, $P_{Nt} = P_N$, $\sigma_{Gt}^2 = \sigma_G^2$, $\sigma_{Nt}^2 = \sigma_N^2$, and $\sigma_{GtNt} =$

27 There is also a covariance between G and N .

28 Behrman, Taubman, and Wales [1] have recently shown that embedding a twin model in a latent variable framework may allow us to distinguish between various models and to identify the separate contributions of genetics and family environment.

29 This assumes that there are no biases arising from differential response rates. Average earnings in our sample exceed 1971 earnings of veterans of the same age by \$7,000, which is more than can be accounted for by inflation. The education level of our twins also exceeds that of veterans. Moreover, in this sample, the average earnings of those with an eighth grade education or less exceeds that of high school graduates for both MZ and DZ pairs. In [11] it is shown that for our responders, the allocation of variance is approximately the same for occupational status and earnings. But we have information on occupational status for many of our nonresponders. The results for occupational status are comparable for responders and nonresponders, although genetics is accorded more emphasis in the responder analysis.

σ_{GN} . Certainly prices can change if supply or demand for any relevant skills shifts. If P_{Gt} and P_{Nt} do not change proportionately, the contributions of G and N to σ_Y^2 will alter. Even with prices fixed, the distribution of environment can change; for example, the distributions of schooling, of family size, and of size of city of upbringing have altered during this century. Similarly, the distribution of genetic endowments can still be changing because of “recent” successful mutations, changes in environment that alter the advantage of particular gene combinations, or migration. These three reasons suggest that it would be desirable to study twins from other cohorts.

VI. CONCLUSIONS

In this paper we have used a new and very rich data source, the NAS-NRC twin sample, to examine the relationship of schooling and some of the effects of genetic endowments and family environment to earnings.

In our regression analysis, we found that our sample, though it is not a random drawing of the population, yields results quite similar to those obtained with Census data when we do not control for genetics and family environment. However, once we control for G and N , we find that the coefficient of schooling declines by two-thirds. It seems likely that only a modest portion of this decrease is due to measurement error. Not controlling for genetics appears to account for some of the decline, but part is probably also attributable to not controlling for family environment.

In other studies people have tried to control for genetic endowments and family environment by including a variety of proxy variables. When we use such proxies as parental education, father’s occupation, number of sibs, religion, region of birth, and rural upbringing, we find many of the variables significant, but that the education coefficient declines by only 12 percent. Thus, this list of proxies is not complete enough to represent fully genetic endowments and family environment.

Of course, G and N could influence earnings even if neither caused a bias on the education coefficient. We calculate that noncommon environment, which is the complement of common environment and genetic endowments, accounts for about half of the variance in log Y around age 50. We have tried to separate the other half into its genetic and family environment components, but have had only limited success.

In summary, we can conclude that it is very important to control for genetics and family environment when studying the effects of schooling on earnings, that the types of proxies for G and N generally available in Census studies are far from adequate, and that a large proportion of the variance in earnings at age 50 is accounted for by a combination of family environment and genetic endowments.

APPENDIX: THE SAMPLE

One of the major and recognized problems in earlier medical and psychological research based on twins was that the sample of twins studied was not a random drawing of the population of twins.³⁰ Partly for this reason and partly to have available a large enough group to study relatively rare diseases, a group of geneticists and medical researchers decided to assemble a “random” sample of white male twins. The sample construction and techniques are described in Jablon and others [7], from which the following quotation is taken:

In 1955 experiments were initiated to explore methods of identifying twins who served in the Armed Forces during World War II. The method settled on was to obtain from the various state and city vital statistics offices in the U.S. copies of the birth records of all white male twins born in the years 1917–1927 and to match the names thus obtained against the VA Master Index (VAMI) to determine which twins survived with both entering military service. About 99% of all World War II veterans are represented in VAMI.

Cooperation of 42 vital statistics offices was obtained (all of the continental U.S. except Arizona, Connecticut, Georgia, Maine, Missouri, New Orleans, Utah, and Vermont). Over 54,000 eligible pairs were found by the participating offices, and, of these, 16,000 pairs were identified by the VA as both having served in the armed services. For 15,000 pairs, one member only was identified as a veteran, and for 23,000 pairs neither was identified. Thus, 108,000 names were searched against the VAMI, of which 47,000, or 43.5% were matched.

It is not possible to tell just why the proportion of matches was so low. For a white male cohort born in 1920, about 86% survived in 1942. About 80 percent of the survivors served in the military forces in World War II, so that we might have expected to match about 69% rather than 43.5%. Possible reasons for the discrepancy include higher mortality of twins than in singletons born in the same year, higher rates of rejection for physical disability, and failures to match correctly at VAMI because of changes in name or inaccurate birth dates shown on the VAMI index card. It must be realized that the VAMI file is ordinarily searched only when a military identifying number is known, thus assuring correct identification. The file clerks, when searching for twins, had to rely on name and date of birth, and there probably were many failures to match men for whom cards were in fact on file.³¹

We mailed our survey on April 15, 1974, to 12,500 pairs for whom we had recent addresses and who had cooperated with recent studies. As of August, we

30 For example, if the twins were selected because one had a particular illness, doctors are likely to be more thorough in examining the other brother for the illness.

31 I have been told that inaccuracies in the VAMI index are no longer considered a major reason for the low match rate. Infant mortality was much higher for twins than for single births in the relevant time period. See Woodworth [15].

had 6,600 replies out of a possible 12,000 people with up-to-date addresses, even though on the third mailing we did not try to contact the nearly 6,500 people where neither brother had previously responded. The 6,600 replies contain 2,468 matched pairs and 1,600 unmatched individuals.

REFERENCES

1. J. Behrman, Paul Taubman, and Terence Wales. "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success." University of Pennsylvania, 1976, mimeo.
2. John Bishop. "Biases in Measurement of the Productivity Benefits of Human Capital Measurement." University of Wisconsin, 1974, mimeo.
3. Peter M. Blau and Otis Dudley Duncan. *The American Occupational Structure*. New York: John Wiley and Sons, Inc., 1967.
4. L. Cavalli-Sforza and W. Bodmer. *The Genetics of Human Populations*. San Francisco: W.E. Freeman and Co., 1971.
5. Zvi Griliches. "Estimating the Returns to Schooling—Some Econometric Problems." *Econometrica* (1976).
6. Zvi Griliches and William Mason. "Education, Income and Ability." *Journal of Political Economy* (May/June 1972 supp.).
7. S. Jablon and others. "The NAS-NRC Twin Panel: Methods of Construction of the Panel, Zygosity Diagnosis and Proposed Use." *American Journal of Human Genetics* (1967).
8. Jacob Mincer. *Schooling, Experience and Earnings*. New York: Columbia University Press, 1974.
9. William Sewell and Robert Hauser. *Education, Occupation and Earnings: Achievement in the Early Career*. New York: Academic Press, 1975.
10. S. Stauffer and others. *The American Soldier*, vol. 4. Princeton, N.J.: Princeton University Press, 1950.
11. Paul Taubman. "The Determinants of Earnings: Genetics, Family and Other Environments." *American Economic Review* (forthcoming).
12. ———. *Sources of Inequality of Earnings*. Amsterdam: North Holland Publishing Co., 1975.
13. Paul Taubman and Terence Wales. *Mental Ability and Higher Educational Attainment in the 20th Century*. New York: McGraw-Hill Book Co., 1972.
14. ———. *Higher Education and Earnings*. New York: McGraw-Hill Book Co., 1974.
15. R. S. Woodworth. *Heredity and Environment*. New York: Social Science Research Council, 1941.