INTELLIGENCE 9, 15–21 (1985)

Familiality Estimates from Restricted Samples*

DANIEL R. VINING, JR.
Population Studies Center
3178 Locust Walk/CR
University of Pennsylvania
Philadelphia, PA 19104

In a recent paper here, Benbow, Zonderman, and Stanley (1983) report that the coefficient of regression of offspring IQ on parental IQ is much lower among the gifted than in the population at large. Thus, Benbow, Stanley, Kirk, and Zonderman conclude in a second paper, the gifted resemble their parents less than do people in general. In this paper, I show that this result is an artifact of the particular estimator of the regression coefficient employed by Benbow, Zonderman, and Stanley. The least-squares estimator, which they employ, is severely biased downward, if the sample on the *dependent* variable is restricted to the upper tail of the distribution, and this is precisely the nature of Benbow et al.'s sample. That is to say, in a bivariate normal distribution with constant regression coefficient, samples restricted to values of the dependent variable (here, child's IQ) above a certain value will always produce a lower regression coefficient than unrestricted samples drawn from the entire but *same* distribution. I introduce an unbiased estimator that can be calculated from the sample statistics reported in the Benbow, Zonderman, and Stanley article and find that the coefficient of regression of gifted child's IQ on parental IQ is, in fact, *higher* than the regression coefficients reported in the literature for unrestricted samples. That is, Benbow et al.'s data suggest that the gifted in fact resemble their parents *more* than do persons in general.

In their study of cognitive abilities in families of extremely gifted children, Benbow, Zonderman, and Stanley (1983) estimate the familiality of these cognitive abilities by calculating the regression of offspring on parents. They report "median values for the father, mother, and mid-parent regression coefficients" of .17, .09, and .11, respectively, and note that "these are considerably lower than [regression coefficients] reported for populations selected without regard to ability" (p. 157), which are in the .42 to .60 range, according to Benbow, Zonderman, and Stanley (p. 158). My purpose here is to sound a note of caution on these particular results of Benbow, Zonderman, and Stanley. I will report the results of a monte-carlo experiment which show that the particular estimator of the regression coefficient employed by Benbow, Zonderman, and Stanley is

*I am indebted to Ralph Ginsberg for bringing to my attention the central papers on regression in selected samples. It goes without saying that he bears no responsibility whatsoever for the contents of this paper.

Correspondence and requests for reprints should be sent to the author at the address listed above.

biased downward from the population regression coefficient (an analytical demonstration of this fact may be found in Goldberger, 1981). I will also present three other estimators which are unbiased and which are easily calculated from Benbow, Zonderman, and Stanley's data. One of these estimators, in fact, can be calculated from information given in their paper, and when this is done, a much higher regression coefficient or estimate of familiality is obtained—higher, in fact, than the regression coefficients reported for unrestricted samples. Thus, it is improper to conclude, as Benbow, Stanley, Kirk, and Zonderman (1983) do, that gifted children resemble "their parents to a lesser extent than less able children resemble their parents" (p. 151). The data presented by Benbow, Zonderman, and Stanley suggest exactly the opposite: Gifted children resemble their parents *more* than less able children resemble their parents.

Assume that IQs of parent and child have, for the national population as a whole, a bivariate normal distribution with known means and variances. For convenience and without loss of generality, we can assume these IQs to be scaled so that their means are both zero, their standard deviations are both unity, and their coefficient of correlation (and, since their standard deviations are equal, their coefficient of regression) is ρ. As Benbow, Zonderman, and Stanley note (p. 158), this last parameter, for nationally representative samples, tends to vary around 0.5. If we denote the IQ of parent as X and the IQ of child as Y, then

$$Y = \rho X + \sqrt{1 - \rho^2}\, U \tag{1}$$

where U is a standard normal random variable with mean 0 and standard deviation 1 and is independent of X. Conversely,

$$X = \rho Y + \sqrt{1 - \rho^2}\, V \tag{2}$$

where, again, V is a standard normal random variable and is independent of Y. Both (1) and (2) follow from the properties of the bivariate standard normal distribution; (1) holds regardless of restrictions of range on X, and (2) holds regardless of restrictions of range on Y.

Benbow et al., in effect, restrict their sample from the national bivariate parent–child distribution of IQs to the range, $Y > a$, where a is a high positive number, and seek to estimate the parameter, ρ, from this restricted sample. To facilitate discussion of this problem, let μ_X^* be the mean of X, when $Y > a$, or $E(X \mid Y > a)$, and μ_Y^* be the mean of Y, also when $Y > a$, or $E(Y \mid Y > a)$. Several "method-of-moments" estimators of ρ are suggested by the expression (2). For example, if we multiply both sides of (2) by Y and take expectations of both sides, we obtain

$$\begin{aligned} E(XY) &= E(\rho Y^2) + \sqrt{1 - \rho^2}\, E(VY) \\ &= \rho E(Y^2), \end{aligned} \tag{3}$$

since $E(VY) = 0$. Hence

$$\rho = \frac{E(XY)}{E(Y^2)} . \tag{4}$$

(4) suggests the following easily calculated estimator of ρ,

$$\hat{\rho}_1 = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} Y_i^2} , \tag{5}$$

where X_i, Y_i, $i = 1, \ldots, N$, are random pairs of IQs from the national bivariate normal distribution of parent and child IQs, under the restriction that $Y_i > a$, $i = 1, \ldots, N$.

A second estimator of ρ is suggested by the following. If we take expectations of both sides of (2) (again under the restriction that $Y > a$), we get

$$\begin{aligned} E(X) &= \rho E(Y) + \sqrt{1 - \rho^2}\, E(V) \\ &= \rho E(Y). \end{aligned} \tag{6}$$

Solving for ρ, we obtain,

$$\rho = \frac{E(X)}{E(Y)} = \frac{\mu_X^*}{\mu_Y^*} . \tag{7}$$

This suggests, as a second, again easily calculated, estimator of ρ,

$$\hat{\rho}_2 = \frac{\overline{X}}{\overline{Y}} = \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i} , \tag{8}$$

where $\overline{X} = \sum_{i=1}^{N} X_i/N$ and $\overline{Y} = \sum_{i=1}^{N} Y_i/N$, again for a random sample drawn from a national population of parent and child IQs, but restricted to the range, $Y_i > a$, $i = 1, \ldots, N$.

A third expression for ρ is

$$\rho = \frac{E(XY) - \mu_X^* \mu_Y^*}{E(Y^2) - \mu_Y^{*2}} . \tag{9}$$

This follows from (2) and (6). That is, remembering that $E(YV) = 0$, we have

$$\frac{E(XY) - \mu_X^* \mu_Y^*}{E(Y^2) - \mu_Y^{*2}} = \frac{E(\rho YY) - \rho \mu_Y^{*2}}{E(Y^2) - \mu_Y^{*2}} = \frac{\rho[E(Y^2) - \mu_Y^{*2}]}{E(Y^2) - \mu_Y^{*2}} = \rho .$$

(9) suggests the estimator,

$$\hat{\rho}_3 = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2} . \tag{10}$$

(10), in fact, is the standard least-square estimator of the coefficient of regression of X on Y. In sum, for a random sample of pairs of values, X_i, Y_i, from a standard bivariate normal distribution but restricted to that part of the distribution such that $Y_i > a$, the method of moments suggests three easily calculated estimators, (5), (8), and (10), as summarized in Table 1. Benbow, Zonderman, and Stanley employ yet another estimator,

$$\hat{\rho}_4 = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{N} X_i^2 - N\bar{X}} , \tag{11}$$

which is the least-squares estimator of the coefficient of regression of Y on X, given that $Y > a$.

I have been unable to derive the sampling distributions and parameters of any of these four estimators nor have I been able to find their expressions in the statistical literature, though Goldberger (1981) provides a good general analytical treatment of the estimation of regression coefficients from restricted samples. However, it is relatively easy to perform a monte-carlo experiment, which, in my view, provides us with a quite good picture of the relative efficiencies and biases of these estimators. The specific monte-carlo experiment performed was the following: pairs of X and Y are drawn from a standard bivariate normal

TABLE 1
Monte-Carlo Derived *M* and *SD*s of 4 Estimators of the Regression Coefficient in a Restricted Sample from a Bivariate Normal Distribution

Regression Coefficient ρ	Estimator $\hat{\rho}$	Computed *M* of Estimator	Computed *SD* of Estimator
.5	$\frac{\Sigma X_i Y_i}{\Sigma Y_i^2}$	.505	.034
.5	$\frac{\Sigma X_i}{\Sigma Y_i}$	.506	.034
.5	$\frac{\Sigma X_i Y_i - N\bar{X}\bar{Y}}{\Sigma Y_i^2 - N\bar{Y}^2}$	.478	.462
.5	$\frac{\Sigma X_i Y_i - N\bar{X}\bar{Y}}{\Sigma X_i^2 - N\bar{X}^2}$	.061	.071

distribution with $\rho = 0.5$ until $N = 50$ pairs are generated such that $Y_i > 2.0$, i = 1, . . . , 50. $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, and $\hat{\rho}_4$ are then calculated from the 50 pairs of numbers so generated. The experiment is performed 50 times, and the means and standard deviations of the 50 $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, and $\hat{\rho}_4$ so generated are then computed. The results of one particular realization of this experiment are presented in Table 1. The simple BASIC program used to implement (on an IBM PC) the monte carlo experiment just described may be obtained from the author upon request.

Table 1 demonstrates two major points about these estimators. The first is that the estimator employed by Benbow, Zonderman, and Stanley has a very significant downward bias, mathematical proof of which, again, may be found in Goldberger (1981). Clearly, computing a least squares estimate of the coefficient of regression of child's IQ on parent's IQ for a random sample restricted to those pairs of IQs where the child's IQ exceeds some high positive value will give us an estimate of the familiality of this trait which is too low.[1] By contrast, the least-

[1]Benbow, Zonderman, and Stanley (1983, p. 158) argue that the closeness of the standard deviations of parental and children's IQs in their data shows that restriction of range cannot be causing *their* unusually low regression coefficient. But a low regression coefficient can be due to either a low ratio (below one) in the standard deviations of children's and parental IQs or a low coefficient of correlation between children's and parental IQs, or to both, and restriction of range biases downward both this ratio and this correlation coefficient. So a ratio of standard deviations near unity is not sufficient evidence that restriction of range has not caused a downwardly biased regression coefficient. The correlation coefficient must also be shown not to be biased downward. There is the further question of why the standard deviations of parental and child IQs should be so similar in the first place. A random sample restricted to one tail of a bivariate normal distribution should produce a much lower standard deviation for the restricted than for the unrestricted variable, as long as the variances for the unrestricted populations are equal, as they should be here.

squares estimator of the coefficient of regression of parent's IQ on child's IQ, for a sample again restricted to those pairs of IQs where the child's IQ exceeds some high positive value, gives us an unbiased estimate of the familiality of this trait, as measured by the regression coefficient, ρ (on this same point, see Vogler & De Fries, 1983, and Lord & Novick, 1968), but one with a quite high sampling variance (see also Reed & Rich, 1982, p. 542), particularly when compared to the estimators, $\hat{\rho}_1$ and $\hat{\rho}_2$, which appear also to be unbiased. The standard deviations of $\hat{\rho}_1$ and $\hat{\rho}_2$ are virtually identical and are estimated here to be approximately one tenth of the standard deviation of $\hat{\rho}_3$.

Benbow, Zonderman, and Stanley (1983, pp. 158–159) present sufficient information to allow us to estimate $\hat{\rho}_2$. The mean IQ of the parents in their sample was about 150 (for both mothers and fathers), whereas the mean IQ of their children was 165. In standardized terms and under the assumption that the unrestricted populational means are, for both parents and children, 100, and the standard deviations, 15, we have $\overline{Y} = (165 - 100)/15 = 4\frac{1}{3}$, $\overline{X} = (150 - 100)/15 = 3\frac{1}{3}$, and $\hat{\rho}_2 = \overline{X}/\overline{Y} = .77$. This is much higher than Benbow, Zonderman, and Stanley own estimate of ρ and should be much closer to the actual value of this parameter. Note also that it is substantially higher than the ".42 to .60 range" found for unselected populations (Benbow, Zonderman, and Stanley 1983, p. 158). Thus, the available evidence suggests that exceptional children resemble their parents *more* than less able children resemble their parents, contrary to the statement by Benbow, Stanley, Kirk, and Zonderman (1983, p. 151).

McAskie and Clarke (1976) have also presented familiality estimates from a highly restricted sample, namely, that of Terman and Oden where the *parents* are the highly restricted group ($X > 2.5$, approximately). In this case, $\hat{\rho}_4$ provides an unbiased estimator of ρ, but one which is highly inefficient when compared to the ratio, $\overline{Y}/\overline{X}$, which McAskie and Clarke (1976, pp. 264–266) also report. According to their data, $\overline{Y}/\overline{X}$ is .64, whereas $\hat{\rho}_4 = .09$. Again, as with the Benbow, Zonderman, and Stanley results, the former is almost certainly closer to the true value of ρ, though the low value of $\hat{\rho}_4$ is in this case explained not by downward bias (as it is in the case of Benbow, Zonderman, and Stanley) but rather by the large sampling variation inherent in the least-squares estimator.

The results which I have presented here lack analytical rigor. However, they should serve to convince the reader that the least-squares estimator of the regression coefficient in a sample restricted to one tail of the bivariate normal distribution is a poor estimator of that coefficient and that better estimators exist when the unrestricted populational means and variances are known, as they should be to a good approximation in most IQ studies. If the so-called dependent variable, Y, is the one that is restricted, then the least-squares estimator will give us an estimate of the regression coefficient which is severely biased downward. If the so-called independent variable, X, is restricted, then the least-squares estimator is unbiased but highly inefficient. These results are based on a monte-carlo

experiment which anyone with access to a personal computer can perform. It would be much preferred, of course, to have analytical expressions for the means and variances of the four estimators of the regression coefficient presented here and much preferred still to have an analytical expression for the maximum likelihood (asymptotically minimum variance unbiased) estimator of ρ for restricted samples, an estimator that would seem to be provided by the so-called Tobit model (Amemiya, 1984; Maddala, 1983) and that could be calculated from Benbow, Zonderman, and Stanley's data, though not from the statistics presented in their paper. I would hope that this paper might stimulate those with greater analytical gifts than my own to treat this problem with the rigor it deserves. In the absence of such a treatment, the results here presented should at least convince workers in this field that some care should be exercised in the estimation of familiality from restricted samples. On the basis of the available statistical literature as well as of the monte-carlo results presented here, one can say with some confidence that least squares estimation of the regression coefficient in a restricted sample is generally inappropriate and always subject to large sampling variation.

REFERENCES

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics, 24,* 3–61.

Benbow, C., Stanley, J., Kirk, M., & Zonderman, A. (1983). Structure of intelligence in intellectually precocious children and in their parents. *Intelligence, 7,* 129–152.

Benbow, C., Zonderman, A., & Stanley, J. (1983). Assortative marriage and the familiality of cognitive abilities in families of extremely gifted students. *Intelligence, 7,* 153–161.

Goldberger, A. (1981). Linear regression after selection. *Journal of Econometrics, 15,* 357–366.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

McAskie, M., & Clarke, A. (1976). Parent–offspring resemblances in intelligence: Theories and evidence. *British Journal of Psychology, 67,* 243–273.

Maddala, G. (1983). *Limited dependent and qualitative variables in econometrics.* Cambridge, England: Cambridge University Press.

Reed, S., & Rich, S. (1982). Parent–offspring correlations and regressions for IQ. *Behavior Genetics, 12,* 535–542.

Vogler, G., & DeFries, J. (1983). Linearity of offspring–parent regression for general cognitive ability. *Behavior Genetics, 13,* 355–360.