Journal of Economic Perspectives—Volume 4, Number 4—Fall 1990—Pages 125–152

Distingushed Fellow
Arthur S. Goldberger and Latent Variables in Econometrics

Gary Chamberlain

In 1988, Arthur Goldberger was honored as a Distinguished Fellow of the American Economic Association. The citation notes that "his early applications and analysis of empirical models brought understanding to a whole generation of econometricians." Goldberger's later work on models with measurement error and latent variables led him into other fields, where he has had a profound influence: "His methodological and substantive findings, also in the area of behavioral genetics, stand as everlasting contributions of one who has successfully bridged the many gaps that separate the social disciplines."

This paper reviews one major element of Goldberger's work—the role of latent variables in econometrics. Even here I am quite selective, developing a few themes in those areas where I have felt his influence most keenly.

The systematic study of models with unobservable variables was resurrected by Zellner (1970) and Goldberger [8, 9] after a long period of neglect.[1] I begin by reviewing this and subsequent work on identification, organizing the discussion along the lines of Goldberger's "multiple-indicator, multiple-cause" model. Consider, for example, an ongoing survey in which some of the same questions are asked on different occasions; for example, the questions might refer to the individual's education and to family background variables such as parent's education and family wealth. This method yields multiple but mismeasured indicators of the underlying true variables. The multiple indicators for these variables may allow us to eliminate biases due to measurement error

[1]Griliches (1974) traces some of the main developments prior to these papers and provides historical perspective.

■ *Gary Chamberlain is Professor of Economics, Harvard University, Cambridge, Massachusetts.*

when we try to assess the relative importance of education and family background in determining an individual's earnings. Another source of identification arises when there are observables that affect (or predict) other observables only via a single latent variable. For example, it may be that education and ability serve to predict current consumption of various goods only via the latent variable "permanent income."

Goldberger's work on salary discrimination also illustrates the power of his multiple-indicator, multiple-cause framework. This work will be discussed in the second main section of this review. The conventional approach to assessing salary discrimination had been based on the direct regression of salary on observed qualifications (such as years of schooling and experience) and gender. This method was criticized because failure to measure all relevant qualifications would result in an "errors-in-variables" problem. This prompted an alternative approach of reverse regression, which is explained in that section. Goldberger made the fruitful distinction between viewing the measured qualification variables as indicators versus causes of the employer's assessment of productivity. A basic point here is that a measured qualification can be imperfectly correlated with productivity without satisfying the strict errors-in-variables model. Failure to recognize this elementary distinction led to confusion in the literature.

The third major section of this paper reviews Goldberger's work on the highly controversial issue of heritability; in fact, genes may rival permanent income as a stimulus for latent variable modeling. Goldberger's role in the "IQ debate" has largely been that of a critic who clarifies the limitations of the models and the associated identification problems. His work in this area is marked by the depth and perseverance of his scholarship, an aspect that is not adequately conveyed in my review.

In fact, any brief review can cover only a portion of Goldberger's wide-ranging contributions to the study of econometrics. But I hope to illustrate the aptness of a remark he made to me: "There is nothing like a latent variable to stimulate the imagination."

Latent Variables and Permanent Income

To understand the role of latent variables, consider an attempt to test Milton Friedman's (1957) permanent income hypothesis. The econometrician would like to run a regression of consumption as a function of permanent income, but permanent income is a latent variable that cannot be observed directly. On the other hand, current income can be observed, and presumably serves as an indicator of permanent income. To this point, the problem takes the familiar form:

$$y = \beta x^* + u, \qquad x = x^* + v,$$

Arthur S. Goldberger

where y = consumption, x^* = permanent income, and x = measured income. The random error terms are both uncorrelated with permanent income, and with each other. (Throughout this section, I shall reduce notational clutter a bit by dropping intercept terms and assuming that all variables have a mean of zero.)

However, there is an immediate problem here. It is well known that β is not identified if only y and x are observed. A good way to see this is to consider the variances and covariances of the observed variables:

$$E(x^2) = \sigma_*^2 + \sigma_v^2, \qquad E(xy) = \beta\sigma_*^2, \qquad E(y^2) = \beta^2\sigma_*^2 + \sigma_u^2.$$

We can estimate these three population moments using the sample moments. But even if we knew the population moments exactly, we would have only three equations to solve for four unknowns—it can't be done.

Given this lack of identification, the least-squares regression of consumption on measured income cannot provide a consistent estimate of β. As sample size grows, the least-squares estimator does converge to something, which we shall call the regression parameter. If the link between observable income and permanent income is weak, then the regression parameter will be close to zero. In fact, one can show that it equals $\beta\tau$, where τ is the fraction of the variance in measured income that is due to permanent income:[2] $\tau = \sigma_*^2/(\sigma_*^2 + \sigma_v^2)$.

[2]The least-squares estimator is

$$b = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \Big/ \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

The law of large numbers asserts that sample means converge (with probability one) to population

Before discussing how additional information might be used to identify β, we should be clear on why β is more fundamental than the regression parameter. In motivating his treatment of errors-in-variables, Goldberger comments [3, pp. 84–85] that it had been customary to use a least-squares regression of consumption on income to estimate the consumption function:

> An accumulation of evidence, however, made this treatment suspect. Parameter estimates turned out to vary widely and unpredictably across diverse samples of households. To some extent, more elaborate theories which introduced additional variables explicitly were able to reconcile the conflicting evidence. But the major breakthrough came with Milton Friedman's (1957) presentation of the permanent income hypothesis.

He goes on to argue that whereas β may remain constant over diverse samples and groups, the tightness of the link between measured income and permanent income is likely to vary, and so the regression parameter is likely to vary.

The idea that structural parameters are invariant across populations is further developed in the following passages [10, pp. 5–6]:

> To sustain the objection to least-squares regression as a tool for estimating a structural model, we must make the case that the structural parameters are somehow more fundamental than the regression parameters. The key to this case is the notion that the regression parameters are *mixtures* of the structural parameters; if one structural parameter changes, *all* regression coefficients may change.
>
> Suppose that the population that produced our data is the only relevant population; i.e., the mechanism that generated our sample will continue to generate all samples in the future. If so, over the relevant universe, all structural parameters will remain the same *and* all regression parameters will also remain the same. The later are just as fundamental as the former. But suppose the population that generated our data will not continue to generate all data in the future. Specifically, suppose that for the next population, one and only one of the structural parameters changes its value. Then that next population may well have all its regression parameters different. ...
>
> This line of argument leads to the conclusion that the search for structural parameters is a search for invariant features of the mechanisms that generate observable variables. Invariant features are those that remain stable—or vary individually—over the set of populations in which we are interested. When regression parameters have this invariance, they

---

means; so, for example, $(1/n)\Sigma_i x_i y_i$ converges to $E(xy)$, and b converges to $E(xy)/E(x^2) = \beta\sigma_*^2/(\sigma_*^2 + \sigma_v^2)$.

Table 1
Works by Arthur S. Goldberger

[1] *Topics in Regression Analysis*. New York: Macmillan, 1968.
[2] "Criteria and Constraints in Multivariate Regression," Social Systems Research Institute: Workshop Paper 7026, University of Wisconsin, Madison, 1970.
[3] "Econometrics and Psychometrics: A Survey of Communalities," *Psychometrika*, 1971, *36*, 83–107.
[4] "A Minimum-Distance Interpretation of Limited-Information Estimation," (with Ingram Olkin), *Econometrica*, 1971, *39*, 635–39.
[5] "Simultaneity and Measurement Error," (with Vincent J. Geraci), Social Systems Research Institute: Workshop Paper 7125, University of Wisconsin, Madison, 1971.
[6] "The Treatment of Unobservable Variables in Path Analysis," (with Robert M. Hauser). In Costner, H. L., ed., *Sociological Methodology 1971*. San Francisco: Jossey-Bass, 1971, pp. 81–117.
[7] "Factor Analysis by Generalized Least Squares," (with Karl G. Jöreskog), *Psychometrika*, 1972, *37*, 243–60.
[8] "Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables," *International Economic Review*, 1972, *13*, 1–15.
[9] "Structural Equation Methods in the Social Sciences," *Econometrica*, 1972, *40*, 979–1001.
[10] "Structural Equation Models: An Overview." In Goldberger, Arthur S., and Otis D. Duncan, eds., *Structural Equation Models in the Social Sciences*. New York: Seminar Press, 1973, pp. 1–18.
[11] "Asymptotics of the Sample Regression Slope," unpublished lecture notes, No. 12, University of Wisconsin, Madison, 1974.
[12] "Unobservable Variables in Econometrics." In Zarembka, Paul, ed., *Frontiers in Econometrics*. New York: Academic Press, 1974, pp. 193–213.
[13] "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," (with Karl G. Jöreskog), *Journal of the American Statistical Association*, 1975, *70*, 631–39.
[14] "Jensen on Burks," *Educational Psychologist*, 1976, *12*, 64–78.
[15] "Mysteries of the Meritocracy." In Block, N. J., and G. Dworkin, eds., *The IQ Controversy: Critical Readings*. New York: Pantheon, 1976, pp. 265–79.
[16] "On Jensen's Method for Twins," *Educational Psychologist*, 1976, *12*, 79–82.
[17] "Twin Methods: A Skeptical View." In Taubman, Paul, ed., *Kinometrics: Determinants of Socioeconomic Success Within and Between Families*. Amsterdam: North-Holland, 1977, pp. 299–324.
[18] "The Genetic Determination of Income: Comment," *American Economic Review*, 1978, *68*, 960–69.
[19] "Models and Methods in the IQ Debate: Part I," (revised), Social Systems Research Institute: Workshop Paper 7801, University of Wisconsin, Madison, 1978.
[20] "The Non-resolution of IQ Inheritance by Path Analysis," *American Journal of Human Genetics*, 1978, *30*, 442–45.
[21] "Pitfalls in the Resolution of IQ Inheritance," (with discussion). In Morton, N. E., and C. S. Chung, eds., *Genetic Epidemiology*. New York: Academic Press, 1978, pp. 195–222.
[22] "Heritability," *Economica*, 1979, *46*, 327–47.
[23] "Reverse Regression and Salary Discrimination," *Journal of Human Resources*, 1984, *19*, 293–318.
[24] "Comment," (on article by Dempster), *Statistical Science*, 1988, *3*, 165–66.
[25] *A Course in Econometrics*, unpublished manuscript, University of Wisconsin, Madison, 1989.
[26] "Economic and Mechanical Models of Intergenerational Transmission," *American Economic Review*, 1989, *79*, 504–13.

> are proper objects of research, and regression is an appropriate tool. But when, as appears to be the case in many social science areas, regression parameters lack this invariance, the proper objects of research are more fundamental parameters; and statistical tools which go beyond conventional regression are required.

Now suppose that we have an additional "indicator" z for the true value x^*:

$$z = \gamma x^* + w,$$

where the random error term w is uncorrelated with x^*, u, and v. In the permanent income model, z and x could be observed income at different dates; under another interpretation, z and y could be consumption at different dates or different components of consumption. Now identification is possible. If we write out the covariance matrix of the observable variables, we find that the structural parameter β can be expressed in terms of the following population moments:[3] $\beta = E(yz)/E(xz)$.

Goldberger [1] developed an approach to estimation in this situation which he called the "analogy principle." The approach "proposes that population parameters be estimated by sample statistics which have the same property in the sample as the parameters do in the population" (p. 4).[4] Following the analogy principle, a consistent estimator can be obtained by replacing population moments with sample moments:[5]

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i z_i \Big/ \frac{1}{n}\sum_{i=1}^{n} x_i z_i.$$

Goldberger's Schultz Lecture [9] showed that this was the approach taken by Sewall Wright (1925, 1934) in his pioneering work. Wright's method of "path coefficients" is essentially an algorithm for relating structural parameters to the population moments of observable variables. Wright applied this idea to both formulate and solve the identification problem in the analysis of supply and demand, and he also applied it to obtain consistent estimators in models

[3]The covariance matrix of the observable variables is

$$E\begin{pmatrix} x^2 & & \\ yx & y^2 & \\ zx & zy & z^2 \end{pmatrix} = \begin{pmatrix} \sigma_*^2 + \sigma_v^2 & & \\ \beta\sigma_*^2 & \beta^2\sigma_*^2 + \sigma_u^2 & \\ \gamma\sigma_*^2 & \gamma\beta\sigma_*^2 & \gamma^2\sigma_*^2 + \sigma_w^2 \end{pmatrix}$$

[4]I obtained this quotation from the preface to Manski's (1988) monograph on *Analog Estimation Methods in Econometrics*.

[5]To turn this principle into a proof simply involves, in this case, invoking the law of large numbers: the sample moments converge to the population moments (with probability one), and so the ratio of sample moments converges to the ratio of population moments, which equals β.

with latent variables. The headnote to Goldberger's lecture reads: "This survey of the use of structural equation models and methods by social scientists emphasizes the treatment of unobservable variables and attempts to redress economist's neglect of the work of Sewall Wright." That a solution to the identification problem appeared well before the work of the Cowles Commission makes a good story and Goldberger tells it well. We shall encounter Sewall Wright again when we discuss heritability.

There is another source of identification, developed by Zellner (1970) and Goldberger [8], that appeals to observable "causes" of the latent variable. Suppose that $x^* = \lambda r + e$, where r is an observed variable that is uncorrelated with the errors u, v, w, and e. For example, temporarily setting aside the indicator z, we could have y = consumption; x^* = permanent income; x = measured income; and r = education. If we write out the covariance matrix of the observable variables, we find that the structural parameter β can be expressed in terms of the following population moments:[6] $\beta = E(yr)/E(xr)$.

With multiple causes in r, there is overidentification. For example, y and z could be different components of consumption, and the vector r could include years of schooling, measures of school quality, test scores, and family background characteristics such as parent's education. If we replace the latent variable x^* by its decomposition into the part predictable from r and the unpredictable error e, then we have

$$x = \lambda_1 r_1 + \cdots + \lambda_K r_K + e + v,$$

$$y = \beta\lambda_1 r_1 + \cdots + \beta\lambda_K r_K + \beta e + u,$$

$$z = \gamma\lambda_1 r_1 + \cdots + \gamma\lambda_K r_K + \gamma e + w.$$

Least-squares regressions of the indicators (x, y, z) on the causes $(r_1, \ldots, r_K)$ provide consistent estimates of the r coefficients, since the errors in this reduced-form representation are uncorrelated with r. We see that there are proportionality restrictions, with, for example, the coefficients on the causes in the y equation equal to β times the coefficients in the x equation. So various ratios of population regression coefficients should all be equal to β.[7]

[6]The covariance matrix is

$$E\begin{pmatrix} x^2 & & \\ yx & y^2 & \\ rx & ry & r^2 \end{pmatrix} = \begin{pmatrix} \sigma_*^2 + \sigma_v^2 & & \\ \beta\sigma_*^2 & \beta^2\sigma_*^2 + \sigma_u^2 & \\ \lambda\sigma_r^2 & \beta\lambda\sigma_r^2 & \sigma_r^2 \end{pmatrix}$$

[7]By population regression I mean the population (minimum mean-square error) linear predictor. For example, the coefficients in the linear predictor of x given r are obtained from the solution to $\min_\pi E(x - \pi' r)^2$, which is $\pi = (Err')^{-1}E(rx)$. The least-squares estimator solves the sample analog to this population prediction problem.

If the errors u, v, w, and e are uncorrelated with each other, then there are additional restrictions. We can form the residuals from the population regressions of the indicators on the causes, and the covariance matrix of these residuals provides yet another solution for the structural parameter β. We have β equal to the covariance between the y and z residuals divided by the covariance between the x and z residuals.[8] Since there are alternative ways to express β in terms of the population moments, the analogy principle provides more than one consistent estimator. Appendix 1 considers ways to combine these estimates in order to gain precision.[9]

Reverse Regression

Goldberger's 1984 paper on "Reverse Regression and Salary Discrimination" [23] provides a good illustration of his multiple-indicator, multiple-cause framework. The context was assessing salary discrimination against, for example, women. The conventional approach was based on the direct regression of salary on observed qualifications (such as years of schooling and work experience) and an indicator variable for gender. Critics of this approach claimed that direct regression estimates of discrimination are biased unless the measured qualifications fully capture productivity. The following citation from Roberts (1980, pp. 177, 186, 188) is illustrative: "There is good reason to expect... that the omission of variables... may have a biasing effect, tending to give the appearance of discrimination when none exists. ... It is a consequence of the fact that statisticians must work with (crude) proxies rather than true productivity. ... Underadjustment... was due to the fact that the... proxy can be thought of as an imperfect measurement of true productivity."

This prompted an alternative, reverse regression, approach. Whereas direct regression compares the salaries of males and females with similar qualifications, reverse regression compares the qualifications of males and females with similar salaries. Various claims regarding reverse regression had appeared in the literature. For example, Roberts (1980, pp. 177, 186–187): "Reverse regression can cope with this bias. ... The statistician need merely compare mean values of the proxy between males and females at each given salary level." And Birnbaum (1979, p. 719): "In order to demonstrate systematic sex discrimination, it must be shown not only that women earn less on the average than

[8]The residual covariance matrix is the same as the covariance matrix in fn. 3 except that σ_*^2 is replaced by σ_e^2; i.e., only the part of x^* that is not predictable from r contributes to the residual variances and covariances.

[9]Before leaving our discussion of identification, I should note that Goldberger also considered the possibility of allowing for latent variables in simultaneous-equations models, in [9], [12], and with Geraci [5]. An important conclusion is that [9, p. 996]: "Errors of measurement in exogenous variables need not destroy identifiability, provided that the model is otherwise overidentified. In effect, one can trade off overidentifying restrictions against the underidentifiability introduced by measurement error."

men of the same qualifications, but also that they are more qualified on the average than men receiving the same salary."

Goldberger collected a number of these claims and showed how they relied upon a simple errors-in-variables model; the latent variable is the employer's assessment of productivity, and it is measured subject to an independent error by a single qualification variable. It followed that [23, p. 303]: "Many of these claims can be disposed of immediately once we recognize that the errors-in-variable specification . . . is not the only one which permits imperfect correlation between measured qualifications and productivity."

He then developed alternative models for the empirically relevant situation in which there are several measured qualifications. In Model A, the measured qualifications are causes of the employer's assessment of productivity. In Model B, the measured qualifications are indicators of the productivity assessment. I shall review these models along with a test that Goldberger developed to assess the appropriateness of reverse regression.

Model A: Multiple Causes

First consider a model with multiple causes:

$$y = p + \alpha z, \qquad p = \beta' x + w,$$

with $E(w|x, z) = 0$. Here y is salary, p is a latent variable representing the employer's assessment of productivity, and z is gender, coded as 1 for male and 0 for female, so $\alpha > 0$ indicates discrimination against women. The productivity assessment is based on a set of observed qualification variables x and a disturbance w. Goldberger interpreted w as follows [23, p. 308]: "That disturbance represents the additional information available to the employer but not to the statistician." The key assumption is that the additional information (that is, the part of the total information not predictable from x) is uncorrelated with gender. The means of the qualifications may differ by gender, and for illustrative purposes we shall assume throughout that the means of x and p are higher for males than females, and that the elements of β are all positive.

The population regression of salary on qualifications and gender is[10]

$$E(y \mid x, z) = \beta' x + \alpha z.$$

The direct least-squares regression of salary on qualifications and gender is the sample counterpart of this population regression. Hence it provides consistent estimates of β and α, even though the observed qualifications do not provide a perfect prediction of the employer's assessment of productivity. Goldberger made the point as follows [23, p. 304]: "Clearly the direct regression estimate of α is unbiased, *even though x is not a perfect correlate of p*. . . . Observe that x can be

[10]The conditional expectation functions are taken to be linear. This is not crucial to the analysis; nonlinear predictors could be used instead.

a proxy for p (in the imperfect correlate sense) without being a fallible measure of p (in the strict errors-in-variable sense). Confusion on this elementary distinction has prevailed in the recent literature."

Model B: Multiple Indicators

Goldberger also provided a multiple-indicator model for the same data:

$$y = p + \alpha z, \qquad x_j = \gamma_j p + v_j,$$

with $E(v_j \mid p, z) = 0$. The idea here is that each of the measured qualifications x_j is merely an indicator of the employer's productivity assessment, subject to a gender-free disturbance.

In this model, it is reverse regression that provides a consistent estimate of the discrimination coefficient α. To see this, we need to work out the population regression of a qualification on salary and gender. There is, by assumption, an exact relationship connecting productivity, salary, and gender ($p = y - \alpha z$); hence the regression of productivity on salary and gender is simply $y - \alpha z$. It follows that the regression of a measured qualification on salary and gender is $E(x_j \mid y, z) = \gamma_j(y - \alpha z) = \gamma_j y - \gamma_j \alpha z$. Note that we can obtain the discrimination coefficient α as minus the ratio of the gender and salary coefficients. The least-squares regression of a qualification on salary and gender is the sample counterpart to this population relationship, and so taking minus the ratio of the least-squares coefficients provides a consistent estimate of α (by the analogy principle). In fact, with several measured qualifications in x, there are proportionality restrictions connecting the population regressions of the various qualifications on salary and gender. For all of these regressions, the ratio of the gender coefficient to the salary coefficient should be the same (namely, equal to $-\alpha$).

So what may seem like a rather minor difference in the specification of these two models in fact has major implications for estimation. Model A calls for direct regression; Model B calls for reverse regression. Goldberger reviewed several empirical papers and concluded [23, p. 297]: "It seems fair to summarize the empirical results as follows: Reverse regression points to a lower estimate of salary discrimination (in favor of men, or of whites) than does direct regression. Indeed, it often suggests reverse discrimination (against men, or against whites). If so, the new approach has obvious attractions for defendants (employers) in discrimination suits and, indeed, has already been used in that context." As a counterpart to these empirical findings, Goldberger showed that reverse regression understates discrimination under Model A, and direct regression overstates discrimination under Model B.[11]

[11]More precisely, there is a version of reverse regression that is intended to handle the multiple qualification case. A composite reverse regression is formed by using a linear combination of the qualifications as the dependent variable, where the weights come from the coefficients on x in the direct regression. It is this composite reverse regression that leads to an underestimate of α under Model A (i.e., as sample size increases, the estimator converges with probability one to a value less

As for choosing between the models, Goldberger reaches one constructive conclusion [23, pp. 314–315]:

> The only known stochastic specification under which reverse regression provides a valid estimator of α is the multiple-indicator one, Model B. That model implies coefficient restrictions on the multivariate reverse regression system. In an empirical context, where sampling variability prevails, we can use the restrictions to test the validity of the model, and thus the validity of the reverse regression estimators. And if the model is valid, we can use the restrictions to obtain a single optimal estimator of α.[12]

The proportionality restrictions on the multivariate regression coefficients are of the type considered by Anderson (1951) and by Hauser and Goldberger [6]. Applying a likelihood-ratio test to Kamalich and Polachek's (1982) data, Goldberger obtained a strong rejection of Model B. He concludes [23, p. 316]: "By conventional standards of statistical inference, therefore, Kamalich and Polachek's reverse regressions are useless as assessments of salary discrimination against women in their sample."

There is no claim, however, that the rejection of Model B implies that Model A is true. Goldberger [23, p. 314] wrote: "The models developed above hardly exhaust the possibilities. It is easy enough to write down a general omitted-variable system in which the structural discrimination parameter is not identified. For such a system neither direct nor reverse regression will be appropriate."[13]

---

than α). The direction of bias in the individual reverse regressions is indeterminant. Goldberger obtained these results by working out the implications of Model A for the population moments, and then expressing the population regression coefficients in terms of these moments. The same methodology shows that the direct regression overstates discrimination under Model B. The special case of this result with a single qualification variable is familiar; we have the coefficient on the qualification biased towards zero due to classical measurement error, and the coefficient on gender is biased in the opposite direction since the qualification and gender are assumed to be positively correlated.

[12]We should note that the proportionality restrictions can hold without Model B being valid. Goldberger considered adding a disturbance u to the salary determination equation, with u independent of productivity, qualifications, and gender. He showed that the proportionality restrictions persist, but now all the reverse regression estimates of discrimination are biased downwards. See the exchange between Dempster (1988) and Goldberger [24] for additional discussion of the proportionality test.

[13]For example, Goldberger also considered a multivariate, errors-in-variables model:

$$y = p + \alpha z, \qquad p = \beta' x^*, \qquad x = x^* + v,$$

with $E(v|x^*, z) = 0$. Now the employer's productivity assessment is an exact function of a set of true productivity measures, whose means may differ by gender. The observed qualifications measure these true qualifications with error. If there were only a single component in x^*, this model and the multiple-indicator model would be equivalent. Indeed, Kamalich and Polachek (1982) claimed that reverse regression is appropriate in the multivariate, errors-in-variables model. Goldberger showed, however, that both direct and reverse regression estimates of α are biased in this model, with the sign of the bias indeterminant.

In particular, the key assumption in Model A—that the employer's additional information is uncorrelated with gender—does not imply testable restrictions. If the assumption fails to be true, then direct regression is biased, but reverse regression does not necessarily eliminate that bias.

Heritability

Goldberger [22, p. 327] introduces his 1979 paper on "Heritability" as follows:

> When we look across a national population, we see large differences in intelligence as measured by IQ tests. To what extent are those differences the result of differences in genetic make-up, and to what extent are they the result of differences in life experience? What proportion of the variance in IQ test scores is attributable to genetic variance, and what proportion to environmental variance? This question has fascinated mankind—or at least the Anglo-American academic sub-species—for several generations. The fascination, I suppose, arises from the notion that the answer has some relevance to social policy: if IQ variance is largely genetic, then it is natural, just and immutable; but if IQ variance is largely environmental, then it is unnatural, unjust and easily eradicated.

By 1972, a consensus had developed among experts that the genetic component of IQ variance was very high, about 80 per cent. In 1969, the psychologist Arthur Jensen published an article in the *Harvard Educational Review* asking "How much can we boost IQ and scholastic achievement?" His conclusion was "not much," based on the evidence he surveyed for the high heritability of IQ. In 1971, the psychologist Richard Herrnstein published an article "IQ" in the *Atlantic Monthly* magazine and the book *IQ in the Meritocracy*. In discussing the work of the biometrical geneticists J. L. Jinks and D. W. Fulker, Herrnstein writes (1971b, p. 171): "They show, as did Jensen, the convergence of the overwhelming body of data on measured intelligence to a figure in the vicinity of 80 per cent or higher for the genetic contribution of scores on standard IQ tests."

This consensus turned into the "IQ debate." Goldberger [22, pp. 328–329] provides the following sketch:

> The consensus that IQ variation accounts for 80 per cent of IQ variance was rudely shaken by three publications: Christopher Jencks's book *In-*

equality (1972), which put the figure at 45 per cent; Leon Kamin's book *The Science and Politics of IQ* (1974), which showed how many of the empirical correlations had been distorted or misrepresented, while others were fictitious; and most recently and most sensationally Oliver Gillie's articles in *The Sunday Times* (24 and 31 October 1976), which gave additional grounds for discarding those kinship correlations that Cyril Burt had provided.

Jensen and others argued that the consensus figure holds up, even after Burt's data are discarded. Particular weight was placed on the work of two schools of biometrical genetics: the Birmingham school, represented by J. L. Jinks, D. W. Fulker, and L. J. Eaves; and the Honolulu school, represented by N. E. Morton and D. L. Rao.

Unraveling the effects of genes and environment would appear to provide ample opportunity for latent variable modeling. In fact, Goldberger played a major role in the IQ debate, but largely as a critic. I shall review his critical contributions in three areas: (1) identification—the models are identified only under implausible assumptions and in some cases are internally inconsistent; (2) replication—in a number of cases, empirical results cannot be replicated and secondary sources misrepresent the primary data; (3) policy implications—heritability estimates serve no worthwhile purpose.

Identification

The models in the IQ debate have their origins in Gregor Mendel's experiments with pea plants. He obtained lines of plants that bred true for a particular character or phenotype ("the form that is shown"); that is, all offspring obtained by self-pollinization or by crossing within this population produce the same form for the character. For example, one line had white flowers and another line had purple flowers. He used pollen from a white flowered plant to pollinate a purple flowered plant. All the plants resulting from the cross had purple flowers. He then took the progeny and self-fertilized them. Of the resulting plants, 705 were purple and 224 were white, for a ratio of 3.1 to 1.[14]

Mendel's explanation, in modern terms, was that each plant had two genes for the character being studied, and that each gene could assume two forms (alleles), which he labeled dominant (G) and recessive (g). The genotypes GG and Gg give rise to purple flowers; only gg gives rise to white flowers. His first generation plants were all Gg, and when they are self-pollinated, the genotype of the off-spring is the result of two independent draws, each of which is equally likely to be G or g. Hence the genotype frequencies in the second generation

[14]He also observed ratios close to 3 : 1 for several other characteristics, such as round vs. wrinkled seeds and yellow vs. green seeds.

are $\frac{1}{4}GG$, $\frac{1}{2}Gg$, and $\frac{1}{4}gg$. This gives a 3 : 1 prediction for purple (GG or Gg) relative to white (gg) flowers.

Now suppose that the Gg genotype does not produce a form identical to that of GG. In addition, we shall allow for environmental effects, so that there is a distribution of phenotypes corresponding to each genotype. Let the character be one, like height, that can be measured on a continuous scale (a quantitative trait). Choose the scale so that conditional on GG, the expected phenotype is a; conditional on gg, it is $-a$; and conditional on Gg, it is d. (Then $d = a$ corresponds to Mendel's case.) Since the frequency of the three genotypes is $\frac{1}{4}GG$, $\frac{1}{2}Gg$, and $\frac{1}{4}gg$, the mean measurement is $\frac{1}{4}a + \frac{1}{2}d - \frac{1}{4}a = \frac{1}{2}d$. Then the genetic component of variance is $\frac{1}{4}a^2 + \frac{1}{2}d^2 + \frac{1}{4}a^2 - (\frac{1}{2}d)^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2$, and the total phenotypic variance equals the genetic component plus an environmental component.[15] In the genetic component, $\frac{1}{2}a^2$ is referred to as the additive variance and $\frac{1}{4}d^2$ is the dominance variance.

The analysis can be extended so that the trait is influenced by several gene pairs with additive effects. In the case of two gene pairs, we have additive effects if the expected phenotype given G_1G_1 and G_2G_2 is $a_1 + a_2$, while given G_1G_1 and G_2g_2 it is $a_1 + d_2$, and so on. If the gene pairs segregate independently (so that, for example, the probability of G_1G_1 and G_2g_2 is the product of their separate probabilities, $\frac{1}{4} \times \frac{1}{2}$), then the genetic component of variance can still be decomposed into additive variance ($= \frac{1}{2}\Sigma_j a_j^2$) and dominance variance ($= \frac{1}{4}\Sigma_j d_j^2$).

This decomposition is important because it is only the additive variance that contributes to resemblance between parent and offspring. To see this in a simple case, suppose there is a single gene pair and $a = 0$. Then the covariance between parent and offspring is 0, whereas the covariance between siblings is $d^2/16$.[16]

The classical work in quantitative genetics, due to R. A. Fisher, J. B. S. Haldane, and Sewall Wright, builds on this Mendelian foundation to provide formulas for the correlations between various kinships. Fisher's (1918) model

[15]This corresponds to the analysis of variance decomposition in which total variance equals the variance of the conditional mean plus the mean of the conditional variance; here the conditioning is with respect to the genotype.

[16]Under random mating, the genotype frequencies remain at $\frac{1}{4}GG, \frac{1}{2}Gg, \frac{1}{4}gg$ for each successive generation (Hardy-Weinberg equilibrium). For example, the conditional probability of a GG offspring given a GG father and a Gg mother is $\frac{1}{2}$. The probability of these parental genotypes is $\frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$, and so the contribution of these parental genotypes to the probability of a GG offspring is $\frac{1}{2} \times \frac{1}{8} = \frac{1}{16}$. Summing over the contributions from all the parental genotypes gives $\frac{1}{4}$ as the probability of a GG offspring. When $a = 0$, the only nonzero contribution to the expected cross-product arises when parent and offspring are both Gg, and the probability of this event is $\frac{1}{4}$. Hence the covariance between parent and offspring is $\frac{1}{4}d^2 - (\frac{1}{2}d)^2 = 0$. Likewise, the probability that both siblings are Gg is $\frac{5}{16}$, and so the covariance between the siblings is $\frac{5}{16}d^2 - (\frac{1}{2}d)^2 = d^2/16$. For further development of the Mendelian basis of dominance, see Cavalli-Sforza and Bodmer (1971) and Falconer (1981). Goldberger [18, fn. 3] notes that "Taubman's interpretation, 'Nonadditive gene effects mean that the average [of two variables]... is not half their sum'... is of course incorrect."

assumes that marriage is assortative on the basis of phenotype, that relatives do not share common environments, and that the system is in equilibrium.[17] The individual's phenotype is y, and it is decomposed into additive variance (σ_a^2), dominance variance (σ_d^2), and environmental variance (σ_E^2): $\sigma_y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_E^2$. The kinship correlations simplify somewhat if we adopt the following notation: $c_1 = (\sigma_a^2 + \sigma_d^2)/\sigma_y^2$, the ratio of total genetic variance to phenotypic variance; $c_2 = \sigma_a^2/(\sigma_a^2 + \sigma_d^2)$, the ratio of additive genetic variance to total genetic variance; μ = the correlation between the phenotypes of the father and the mother; and $A = c_1c_2\mu$, which is the correlation between the additive genotypic values of the spouses induced by the assortative mating. The parameter c_1 is referred to as broad heritability, with narrow heritability defined as $c_1c_2 = \sigma_a^2/\sigma_y^2$. Narrow heritability is the main object of interest in quantitative genetics since it can be used to predict the success of a selective breeding program. We shall return to this point when we discuss policy implications.

Let y_f, y_s, and y_u denote the phenotypes of the individual's father, sibling, and uncle. Then Fisher's (1918) correlation formulas for these kinships are

$$\text{Cor}(y_f, y) = c_1c_2(1 + \mu)/2,$$

$$\text{Cor}(y_s, y) = c_1c_2(1 + A)/2 + c_1(1 - c_2)/4,$$

$$\text{Cor}(y_u, y) = c_1c_2[(1 + A)/2]^2 + Ac_1(1 - c_2)/8.$$

The classical model succeeds in expressing all kinship correlations in terms of the three parameters c_1, c_2, and μ (with $A = c_1c_2\mu$).

We can illustrate with Fisher's (1918) analysis of the inheritance of height, which was based on data collected by Pearson and Lee (1903) on English University students. The observed correlations were .507 between parent and offspring, and .543 between sibs. Setting $\mu = .280$ as the observed correlation between spouses, we can solve the first two equations above for c_1 and c_2. This gives an estimate of c_1 equal to 1.03, so Fisher set the broad heritability equal to one, with an estimate of narrow heritability equal to .79. We shall return to this example when we discuss policy implications.

The empirical work of the Birmingham school was based on the classical formulas for 14 different kinships, with the empirical correlations coming from Cyril Burt (1966). Fisher's model makes no allowance for relatives having correlated environments, and so the Birmingham school worked with two modifications of his model. In the first, the cross-sib formula was modified by adding a free parameter (e), which allowed for correlation between the environments of the sibs. In the second, a common-environment parameter (t) was added to both the cross-sib and parent-child formulas. The formulas for the

[17] In the Appendix, I present Goldberger's [19] model, which allows for a more general treatment of environmental components, and then follow him in obtaining Fisher's (1918) model as a special case.

other kinships were not changed. Then given a set of empirical correlations, the four unknown parameters (c_1, c_2, μ, and either e or t) are chosen so that the theoretical correlations match the empirical ones as closely as possible according to a suitable distance function.[18]

Goldberger criticized the Birmingham models for not being internally consistent. A hint of this can be seen in the formulas we have given to illustrate Fisher's model. We see that the dominance variance, which accounts for the excess resemblance of siblings relative to the parent-child correlation, appears again in diluted form in the equation for uncles. In fact, Goldberger showed (see Appendix 2) that the Birmingham formula for uncles should be modified by adding the term $Ae/2$. Similarly, the shared-by-sibs environment shows up, further diluted, in the phenotypic correlations for first and second cousins. There are related problems in the second Birmingham model, and Goldberger concludes [21, p. 196]: "But the Birmingham model is logically untenable because the consequences of the shared environment are not fully taken into account. It is essentially impossible to construct a causal model which will produce the Birmingham formulas for kinship correlations, if only because the conjunction of phenotypic assortative mating and environmental transmission from parents to children will generate gene-environment correlation."

The Honolulu school, building on the work of Sewall Wright (1931), did provide a coherent model of environmental transmission. But Goldberger [22, p. 336] questions the plausibility of the identifying assumptions:

> Ignorance of genetics need not deter us, because the models involve as much social science theorizing as genetic theorizing. How marriages take place, how adoption agencies operate, how parents raise their children, how brothers and sisters educate one another—all those processes are reflected in the biometrical–genetic models.
>
> I call attention to two such pieces of theorizing that are incorporated in the Birmingham and Honolulu models. (1) Identical twins share just as much IQ relevant experience as ordinary siblings do, no more. That happens despite the fact that identical twins are of the same age and sex, while ordinary siblings may differ in age and sex. (2) Adoption agencies place children in families randomly drawn from the population at large. That happens even though every adoption study shows that adoptive parents rank high on virtually every socioeconomic and psychological measure.
>
> ... If less restrictive, and hence more plausible, specifications were made, the number of unknown parameters would approach and soon exceed the number of observations.

[18] In fact Jinks and Eaves (1974) failed to impose the constraint that $A = c_1c_2\mu$, which led Goldberger to do some calculations of his own. This revealed, among other things, that they had misspecified the avuncular correlation, dividing $Ac_1(1 - c_2)$ by 2 instead of by 8.

Similar identification issues arise in the work of Taubman (1976) and Behrman, Taubman, and Wales (1977). They used data on 1,019 identical (monozygotic or MZ) twin pairs and 907 fraternal (dizygotic or DZ) twin pairs. The main variables of interest were education, initial occupation, current occupation (when the twins were about 50 years old), and earnings measured in logarithms.

We can illustrate the simplest version of the twin method by calculating the heritability of earnings. Specialize Fisher's formula for siblings by assuming that there is no assortative mating ($\mu = 0$) and that the dominance variance is zero ($c_2 = 1$). Extend the formula by allowing for a correlation (ρ) between the environments of the siblings. Then we have $\text{Cor}(y, y_s) = c_1/2 + \rho(1 - c_1)$. Fraternal twins are no more alike genetically than ordinary siblings, so we shall use this equation as the correlation between the earnings of the fraternal twins. Identical twins, however, have the same genes, and so $\text{Cor}(y, y_s^*) = c_1 + \rho^*(1 - c_1)$, where y_s^* is the earnings of the identical twin and ρ^* is the correlation between the environments of the individual and his identical twin.

Now if we assume that $\rho = \rho^*$, so that fraternal twins experience just as much environmental similarity as identical twins do, then we have the double-the-difference estimate of heritability: $c_1 = 2[\text{Cor}(y, y_s^*) - \text{Cor}(y, y_s)]$. Substituting the empirical correlations from Taubman's data gives $\hat{c}_1 = 2[.54 - .30] = .48$. As Goldberger notes [22, p. 340]: "Any reader who can subtract one number from another and multiply by two is now able to estimate the role of genetic factors in socioeconomic achievement." He goes on to note that by more or less the same technique, it has been recently demonstrated that in South Australia genetic factors explain 83 percent of the variance in knowledge of French, while in England genetic factors explain 34 percent of the variance in conservatism and about 60 percent of the variance in sexual satisfaction among young women.

Goldberger is particularly skeptical, however, about the assumption that $\rho = \rho^*$ [22, p. 341]: "Identical twins, as far as we know, are treated more similarly by parents and teachers, share more experience, and broadly speaking have more similar environments."[19] He considers an alternative specification (called the Goldberger, or naive, model) in which $c_1 = 0$ and ρ and ρ^* are unrestricted. This gives $\hat{\rho} = .30$, $\hat{\rho}^* = .54$, and 0 per cent of the variance of earnings is attributable to genetic differences. This naive model is no more complicated than Taubman's and fits the data equally well.

The issues in the simple twin method are clear enough. Part of Goldberger's contribution was to show that the same issues are central in the more complicated methods that were being used. Behrman, Taubman, and Wales (1977) develop a number of elaborate multivariate models for education, early occupa-

[19]"The famous human geneticist L. S. Penrose said, only half-jokingly, that if one looked at twin data uncritically, one might conclude that clothes are inherited biologically" [Cavalli-Sforza, 1977, p. 94].

tion, current occupation, and earnings. These four phenotypes are linearly related to unobservable genetic and environmental factors. Goldberger [17] reviewed the various models and how their identification was achieved. In none of the models (with the exception of a purely environmental one) was ρ^*/ρ itself identified. Furthermore, conclusions about heritability are sensitive to the value chosen for ρ^*/ρ.[20] Goldberger concludes [17, p. 315]: "It appears that heritability estimates produced by the twin methods are quite dependent upon assumptions about differential environmental correlations. We are left with the data which show MZs more highly correlated on observed variables than DZs, but have no basis for allocating these differences between genetic and environmental factors."

Replication

Jinks and Eaves (1974), working with Cyril Burt's data, assert that "whatever else may be said about the quality of the data, their quantity is such that our estimates are fairly precise and our test of the model fairly sensitive." Goldberger [19] argues that a reading of the primary sources might leave the reader less sanguine about the empirical material. He offers several quotations from Burt and Howard's writings. For example, consider Burt and Howard (1956, p. 122) and a year later (1957, p. 39): "The interview, the use of non-verbal tests, and the information available about the child's home circumstances usually made it practicable to allow for the influence of an exceptionally favorable or unfavorable cultural environment. ... having satisfied ourselves that by these means we can reduce the disturbing effects of environment to relatively slight proportions." Goldberger concludes [19, p. 33]: "It seems that Burt's observations are not correlations of IQ test scores, but rather estimated correlations of the genetic components of IQ test scores. If so, they are hardly suitable for estimating the relative contributions of heredity and environment to IQ test scores. One might say that the 17 percent ($= 100(1 - c_1)\%$) that is left to environment in Burt's data reflects only his failure to completely purify his figures."

Herrnstein (1971b) draws on a classic adoption study by Barbara Burks (1928) (which, in contrast to Burt's data collection, was very well documented) to support his argument that most of the variation in intelligence can be accounted for by genetic variation. Herrnstein says that the foster children's IQ's correlated more highly with the IQ's of their natural parents than with the IQ's of their foster parents. But Goldberger [15] points out that Burks had no data on the IQ's of the natural parents! Burks did report correlations of .45 and .46 between the IQ of "true child" and the IQ of father and mother, but

[20]Similar issues arise in the work of Taubman (1976) and Jensen (1975) that extends the classical twin method to allow for gene–environment correlation. That work is examined critically in [18] and [16].

these were for nonfoster children in the *control* group. Apparently Herrnstein mistook the control group figures for foster group figures.[21]

The Honolulu model distinguished between childhood and adult IQ, and provided separate heritability estimates for each. The estimates are much lower for adults (.21 versus .67 for children), which led Rao, Morton, and Yee (1976, p. 238) to conclude that "adult education of parents could... have greater effect on academic performance of children than preschool education of children."

But Goldberger's [20] attempt to replicate their results encountered difficulties. He eventually determined that they had mistakenly used their theoretical formula for the correlation between parent's IQ and the parent's environmental index to model the correlation between the parent's IQ and the child's environmental index. Furthermore, Goldberger realized that although the model provided formulas for eleven population correlations in terms of eight free parameters (such as heritabilities for child's IQ and adult's IQ), it could be reparameterized in terms of seven parameters (which are functions of the original eight), and so only these seven parameters are identified. In particular, the contrast between IQ heritability for children and adults is not identified, which led Goldberger to remark [21, p. 203]: "On this count alone, preschool educators may relax; their jobs are not in jeopardy."

Policy Implications

Heritability is not invariant across populations—it is not a structural parameter. R. A. Fisher's estimate of the (broad) heritability of height was 100 percent for English University students; that finding need not generalize to other populations with more diverse environments. Cavalli-Sforza and Bodmer (1971, p. 601) note that "in the last hundred years we have witnessed a change in average human stature of more than one standard deviation. This change is almost certainly the consequence of secular changes in environment, but measures of heritability are hardly informative in this respect."

Phenylketonuria (PKU) is a genetic disease, transmitted by simple Mendelian inheritance. The chemical phenylalanine, which is contained in the protein we eat, accumulates in individuals with PKU and, since it is toxic to the brain, causes severe mental retardation. In "normal" environments, the percentage of phenylalanine in blood plasma has high heritability.[22] But the phenotypic effects of the PKU genotype can be eliminated by the administration of an appropriate diet in the early years of life.

Behrman, Taubman, and Wales (1977, p. 36) write: "We think that a portion of the equity criteria within and between generations is related to the extent to which a person's earnings are due to his own efforts versus those of

[21]Goldberger [14] finds that Jensen's (1972, 1973) presentation of Burks's study is also quite unreliable.

[22]See Cavalli-Sforza and Bodmer (1971, p. 510).

the parents who bore and reared him." Goldberger [17, p. 318] responds: "But here again the distinction between genes and environment (between bearing and rearing) is irrelevant."[23]

Why then do quantitative geneticists measure heritability? Goldberger [18, p. 968] has the following view:

> Heritability analysis has been extensively developed and employed in plant and animal genetics. There it is used to predict the effectiveness of selective breeding programs under constant environmental conditions, not to set limits on the potential effectiveness of environmental improvements. In that context, h^2 (or rather its additive component, "narrow heritability") turns out to be the slope in the regression of offspring's phenotype on parent's phenotype.[24]

Goldberger [22, pp. 345–346] concludes:

> Surely the environmental–genetic dichotomy does not correspond to a high-cost–low-cost dichotomy. Some genetically based handicaps are remedied at low cost, some environmentally based handicaps are remedied at great cost if at all. Provision of eyeglasses to offset genetically deficient eyesight is a low-cost operation, while imposition of an inheritance tax to offset environmental advantages/disadvantages may well have severe disincentive effects. An allocation of earnings variance into environmental and genetic components tells us nothing about tradeoffs. Heritability analysis is just not a guide for policy, not a short-cut around the detailed cost–benefit analysis required for each specific policy proposal.

I want to close this section by recommending Goldberger's [22] paper on "Heritability" as a marvelous overview of the topic. A great deal of hard work led up to that paper—painstaking checking of original sources; evaluating the quality of the data and seeking out relevant data sets that were not being used; replicating results and examining their sensitivity to alternative samples and model specifications; assessing whether the procedures being used could be derived from a coherent model and working out the identification analysis. But

[23]The role of heritability in intergenerational mobility and in inequality within a generation is discussed in Goldberger [26]. One aspect of the inherited endowment (p. 507) "represents the capital that the child receives automatically and effortlessly from his parents—genes, reputation, culture, learning, skills, and goals that his parents provide at no cost." The separation of this endowment into genetic and environmental components plays no role.

[24]Using Fisher's formula for the correlation between the phenotypes of the father (or mother, y_m) and offspring, the population regression of y on the midparent value ($\bar{y} \equiv (y_f + y_m)/2$) is $\text{Cov}(y, \bar{y})/\text{Var}(\bar{y}) = [\frac{1}{2}c_1c_2(1 + \mu)]/[\frac{1}{2}(1 + \mu)] = c_1c_2$.

this stays in the background, and Goldberger writes with effortless authority and style.

Conclusion

In addition to omitting much of the substance in Goldberger's work, I have said nothing of his characteristic wit. I'll offer two examples, although there is no substitute for knowing Art. From "Heritability" [22, p. 337]:

> Professor Hans Eysenck was so moved by the twin study that he immediately announced to Hodgkinson that it "really tells the [Royal] Commission [on the Distribution of Income and Wealth] that they might as well pack up" (*The Times*, 13 May 1977). (A powerful intellect was at work. In the same vein, if it were shown that a large proportion of the variance in eyesight were due to genetic causes, then the Royal Commission on the Distribution of Eyeglasses might as well pack up. And if it were shown that most of the variation in rainfall is due to natural causes, then the Royal Commission on the Distribution of Umbrellas could pack up too.)

From his 1989 manuscript for *A Course in Econometrics* [25, chapter 23]:

> These and other texts devote many pages to the problem of multicollinearity in regression analysis, but say little about the closely analogous problem of small sample size in estimating a univariate mean. A possible explanation for that imbalance is that there is no fancy polysyllabic name for "small sample size." To remove that impediment, we introduce the term *micronumerosity*. . . .
>
> The extreme case, "perfect micronumerosity," arises when $n = 0$, in which case the sample estimate of μ is not unique. (Technically, there is a violation of the rank condition $n > 0$: the matrix (0) is singular.) That extreme case is easy enough to recognize, but "near micronumerosity" is more subtle, and yet very serious. It arises when the rank condition $n > 0$ is barely satisfied. Near micronumerosity is very prevalent in economic research. . . .
>
> Tests for the presence of micronumerosity require the judicious use of various fingers. Some researchers prefer a single finger, others use their toes, still others let their thumbs rule.

While this review is certainly not comprehensive, I hope it has conveyed something of the work of this remarkable man. His papers are meant to be read. Rereading them has brought pleasure and profit to me; they are a continuing source of stimulation.

Appendix 1
Latent Variables and Estimation

Some of the models in the Latent Variables and Permanent Income section of this paper are overidentified—there is more than one way to solve for the parameters of interest in terms of the population moments. Hence the analogy principle provides more than one consistent estimator, and we need a framework for combining these estimators. Furthermore, in addition to efficient point estimates, we need confidence intervals.

For a simple example, suppose that x and z are "parallel" measurements on the latent variable x^*, perhaps from a reinterview survey that asks the same question on different occasions, providing two error-ridden proxies for the same variable: $x = x^* + v$, $z = x^* + w$. With $y = \beta x^* + u$, and assuming that the errors u, v, and w are uncorrelated with x^* and with each other, we have the following covariance matrix for the observable variables:

$$E\begin{pmatrix} x^2 & & \\ zx & z^2 & \\ yx & yz & y^2 \end{pmatrix} = \begin{pmatrix} \sigma_*^2 + \sigma_v^2 & & \\ \sigma_*^2 & \sigma_*^2 + \sigma_w^2 & \\ \beta\sigma_*^2 & \beta\sigma_*^2 & \beta^2\sigma_*^2 + \sigma_u^2 \end{pmatrix}. \qquad \text{(A.1)}$$

(As earlier, I shall simplify notation by dropping intercept terms and assuming that all variables have a mean of zero.) These population moments are subject to the restriction that $E(yx) = E(yz)$. To impose this restriction, we can choose the underlying parameters $(\beta, \sigma_* \sigma_u, \sigma_v, \sigma_w)$ so that the implied covariance matrix is as close as possible to the sample covariance matrix.

This approach runs through a number of the Goldberger papers—the basic notion of dealing with overidentification by means of a fitting criterion relating sample moments to population moments. It is expressed in [10, pp. 13–14] as follows: "In the psychometric literature, as in the econometric literature, a large variety of procedures have been developed for efficient estimation of overidentified models. In some instances, these procedures are, in effect, devices for reconciling the conflicting sample equations by minimizing some measure of the difference between the sample moment matrix and its parametric structure."

This view is clearly articulated in the paper with Jöreskog [7] on "Factor Analysis by Generalized Least Squares." For a simple example of a factor-analysis model, consider J test scores, $y_1, \ldots, y_J$, observed for a random sample of n individuals from some population. The j^{th} score is decomposed into two components: $y_j = \gamma_j x^* + v_j$, where x^* is a general ability factor (IQ) that appears in all the scores, and the v_j are specific factors that are not correlated across the tests or with x^*. The sensitivity of a test to the general factor is measured by its "factor loading" γ_j. The model implies that the covariance

between the scores y_j and y_k is $\gamma_j\gamma_k\sigma_*^2$, which places restrictions on the population moments if the number of tests is sufficiently large.[25]

Goldberger and Jöreskog argue that a factor-analysis model implies that the population covariance matrix is a certain function of the parameters; so one could choose estimates of the parameters to minimize the distance to the sample covariance matrix in a suitable metric. Aitken's (1934–35) generalized least-squares (GLS) principle uses a quadratic norm with a weight matrix based on the inverse of the covariance matrix of the estimated moments.[26] They use an estimate of this matrix, appealing to the minimum- χ^2 (or minimum-distance) principle of Neyman (1949), Taylor (1953), and Ferguson (1958), and its application in Zellner (1962), Malinvaud (1966), and Rothenberg (1966). An important result of this theory is that replacing the weight matrix by a consistent estimate of it does not affect the asymptotic distribution of the minimum-distance estimates.

To obtain an estimate of the (GLS) weight matrix, Goldberger and Jöreskog assume that the observation vectors $y_i' = (y_{i1}, \ldots, y_{iJ})$ are a random sample from a J-variate normal distribution. This implies that the covariance matrix of the sample covariances just depends on $V(y)$, the covariance matrix of y. They estimate the "covariances of the sample covariances" by replacing $V(y)$ by its sample analog. Then they show that their estimator has the same asymptotic distribution as maximum likelihood, so there is no loss of efficiency when the normality assumption is correct. The plausibility of this result was noted earlier by Goldberger [3, p. 96], when he conjectured that an assertion in the factor-analysis literature that "maximum-likelihood estimates 'are known to be asymptotically superior to those provided by other solutions' is incorrect; 'superior' should be replaced by 'at least as good as'."

Given their setup, it is natural to ask what happens if we drop the normality assumption. The minimum-distance approach is not affected, except that we can no longer express the GLS weight matrix just as a function of $V(y)$. It is straightforward, however, to estimate the weight matrix directly. For example, with $z_{i1} = y_{i1}^2$ and $z_{i2} = y_{i1}y_{i2}$, we have

$$\operatorname{Cov}\left(\frac{1}{n}\sum_i z_{i1}, \frac{1}{n}\sum_i z_{i2}\right) = \frac{1}{n}\operatorname{Cov}(z_{i1}, z_{i2}),$$

which can be estimated using the sample covariance between z_1 and z_2; this amounts to replacing population fourth moments of the y's by their sample counterparts. Since this procedure always provides a consistent estimate of the GLS weight matrix, it does as well as maximum likelihood when normality

[25]Various connections between econometrics and psychometrics are developed in Goldberger [3], "Econometrics and Psychometrics: A Survey of Communalities."

[26]This generalizes the intuitive idea of using a weighted average to combine independent estimates, with the weight on an estimate proportional to the reciprocal of the estimate's variance.

is true, and it is generally more efficient when normality is false. Perhaps more important, this procedure continues to provide consistent confidence intervals when normality is false, whereas the conventional maximum-likelihood approach does not.[27]

This result has broad implications, since much of econometrics can be viewed as putting structure on a covariance matrix. Suppose, for example, that only x and y are observed, and we are interested in the regression parameter $\delta = E(xy)/E(x^2)$. Define the prediction error $\varepsilon = y - \delta x$ and parameterize the (x, y) covariance matrix as follows:[28] $E(x^2) = \sigma_x^2$, $E(xy) = \delta\sigma_x^2$, and $E(y^2) = \delta^2\sigma_x^2 + \sigma_\varepsilon^2$. There are no restrictions to impose, and the minimum-distance estimator is the sample analog:

$$\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \Big/ \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad \text{(the least-squares estimator)}.$$

It follows from the minimum-distance framework that $\sqrt{n}(\hat{\delta} - \delta)$ has a limiting normal distribution whose (asymptotic) variance is consistently estimated by

$$\hat{V}_a(\hat{\delta}) = \frac{1}{n}\sum_i \hat{\varepsilon}_i^2 x_i^2 \Big/ \left(\frac{1}{n}\sum_i x_i^2\right)^2, \tag{A.2}$$

with $\hat{\varepsilon}_i = y_i - \hat{\delta}x_i$. This formula is derived in Goldberger's [11] lecture notes.

Note that no assumption has been made that the prediction errors ε are homoskedastic. In fact the formula in (A.2) is the "heteroskedasticity-consistent" estimator that has seen increasing use in econometrics since White's (1980a, b) work. There is an interesting connection here between relaxing the normality assumption in a factor-analytic framework, and relaxing the homoskedasticity assumption in a regression framework. Here they amount to the same thing. If we assume that the bivariate distribution of (x, y) has the property that x^2 is uncorrelated with ε^2, then we get the asymptotic variance formula for the homoskedastic case, $V_a(\hat{\delta}) = E(\varepsilon^2)/E(x^2)$. The multivariate normality assumption has this implication since $\text{Cov}(x, \varepsilon) = 0$ implies that x is independent of ε.

Now consider estimation with multiple indicators and multiple causes. The proportionality restrictions are characteristic of the simultaneous-equations literature in econometrics. There is a masterful treatment of various fitting criteria in Goldberger [2]. Related issues, including a comparison between maximum-likelihood and minimum-distance estimation, and the role of canoni-

[27]MaCurdy (1982) used a quasi-maximum likelihood estimator in his study of the covariance structure of panel data on earnings. Although the likelihood function was based on multivariate normality, MaCurdy provided estimated standard errors that are consistent even if normality does not hold. Chamberlain (1982) used the minimum-distance framework to discuss the efficiency gains that are possible under nonnormality.

[28]As noted in fn. 7, δ solves the linear predictor problem $\min_\delta E(y - \delta x)^2$. Note that the prediction error ε is uncorrelated with x.

cal correlation, are developed in [12] and in the papers with Olkin [4] and with Hauser [6].

A distinctive feature of the multiple-indicator, multiple-cause model presented early in the second section of this paper is that β and γ appear both in the regression coefficients (through the proportionality restrictions) and in the residuals (as factor-loading coefficients). The additional restrictions implied by the factor-analytic structure are treated in the paper with Jöreskog [13] on "Estimation of a Model With Multiple Indicators and Multiple Causes of a Single Latent Variable." Jöreskog and Goldberger apply a maximum-likelihood algorithm by expressing the population covariance matrix of the observed variables (x, y, z, r) in terms of the underlying parameters. The likelihood function is based on a multivariate normality assumption, but their work on minimum-distance estimation in factor analysis could also be applied here, making it straightforward to relax the normality assumption.

Appendix 2
Kinship Correlations

This section lays out the classical model of quantitative genetics due to R. A. Fisher (1918). I shall first present Goldberger's [19] model, which allows for a more general treatment of environmental components, and then follow him in obtaining Fisher's model as a special case.

An individual's phenotype y is determined as $y = \sum_{k=1}^{K} \alpha_k x_k \equiv \alpha' x$, where x is a latent vector of components and α is a vector of constants. The parents of our individual, labeled f and m, have component vectors x_f and x_m, and hence phenotypes $y_f = \alpha' x_f$ and $y_m = \alpha' x_m$. Note the stationarity assumption that the relationship between phenotype and components is the same across generations. Likewise, the covariance matrix of the components is assumed to be the same across generations: $\Sigma \equiv E(xx') = E(x_f x_f') = E(x_m x_m')$. We simplify notation by assuming that all variables have a mean of zero.

The parents are matched solely on the basis of phenotypes, and any correlation between x_f and x_m is due to the correlation between the phenotypes y_f and y_m.[29] Hence, with the phenotypes scaled to have unit variance, we have

$$\Theta \equiv E(x_f x_m') = \mu \Sigma \alpha \alpha' \Sigma, \tag{A.3}$$

where $\mu = E(y_f y_m)$ is the covariance (= correlation) between the phenotypes of the father and mother.

[29]The residuals from the regression of the components of x_f on y_f are assumed to be uncorrelated with x_m, with a symmetric assumption that the residuals from the regression of x_m on y_m are uncorrelated with x_f.

The individual's component vector is determined linearly by the average of his parents' component vectors, subject to an additive disturbance vector:

$$x = B(x_f + x_m)/2 + w, \tag{A.4}$$

where B is a square matrix of constants, and w is uncorrelated with x_f and x_m (and uncorrelated with the component vectors of all other relatives except for the individual's siblings). The covariance matrix of w is $\Omega = E(ww')$.

The individual has a sibling, labeled s, and the sibling's vector is determined in the same fashion: $x_s = B(x_f + x_m)/2 + w_s$. We let $\Phi = E(w_s w')$ denote the covariance between the specific (= nontransmitted) parts of the siblings' component vectors.

Stationarity (or equilibrium) together with equation (A.4) implies that

$$\Sigma = E(xx') = B(\Sigma + \Theta)B'/2 + \Omega \equiv T + \Omega, \tag{A.5}$$

where T is the transmitted portion of Σ. Equation (A.5) determines Ω in terms of B, Σ, and Θ, and Θ is determined by equation (A.3) in terms of μ, Σ, and α.

Now the phenotype correlations for all kinships can be obtained in terms of α, B, Σ, Φ, and μ. The parent–child correlation is $E(y_f y) = \alpha' E(x_f x')\alpha = \alpha'(\Sigma + \Theta)B'\alpha/2$. The cross-sib correlation is $E(y_s y) = \alpha' E(x_s x')\alpha = \alpha'(T + \Phi)\alpha$. Our last example is the brother of our individual's mother (the cognate uncle, labeled u). Persons u and m are sibs, so that $E(x_u x_m') = E(x_s x') = T + \Phi$. The individual's father has a component vector x_f that is correlated with x_u only via its correlation with the mother's phenotype. Hence $E(x_u x_f') = (T + \Phi)\alpha\alpha'\Theta$. So we have $E(y_u y) = \alpha' E(x_u x')\alpha = \alpha'(T + \Phi)(\alpha\alpha'\Theta + I)B'\alpha/2$.

Now we can specialize to Fisher's model, in which x_1 = the additive genotypic value, x_2 = the dominance deviation, and x_3 = environment. We set $\alpha' = (1\ 1\ 1)$, so that the phenotype is the sum of the three components. The three components are assumed to be uncorrelated, so that $\Sigma = \text{diag}\{\sigma_a^2, \sigma_d^2, \sigma_E^2\}$. The only component transmitted from parent to child is the additive genotype, so that B is a matrix of zeros except for a one in the upper-left corner. The only nonzero covariance between w_s and w is between the sibs' dominance deviations, and they share one fourth of the dominance variance: $\phi_{22} = \sigma_d^2/4$. Then some straightforward algebra gives Fisher's three correlation formulas presented in the text.

The first Birmingham modification corresponds, in Goldberger's model, to setting $\phi_{33} = e$; then working through the model shows that the formula for uncles should be modified by adding the term $Ae/2$. The second Birmingham model corresponds to specifying a positive value in the third diagonal element of Goldberger's B matrix, to allow for environmental transmission from parent to child. But the Birmingham models maintain Fisher's assumption of no correlation between genes and environment. Then equation (A.5) forces Ω to have negative elements in the (1, 3) and (3, 1) positions if the assortative mating

correlation μ is positive. So we have the peculiar implication that the specific (= segregation) part of an individual's additive genotypic value is negatively correlated with his specific (= untransmitted) environment.

■ *I am grateful to Joshua Angrist, Christopher Cavanagh, Richard Freeman, Zvi Griliches, Charles Manski, Joseph Stiglitz, and Timothy Taylor for helpful comments.*

References

Aitken, A. C., "On Least Squares and the Linear Combination of Observations," *Proceedings of the Royal Society of Edinburgh*, 1934–35, *55*, 42–48.

Anderson, T. W., "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 1951, *22*, 327–51.

Behrman, J., P. Taubman, and T. Wales, "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success." In Taubman, P., ed., *Kinometrics: Determinants of Socioeconomic Success Within and Between Families*. Amsterdam: North-Holland, 1977, pp. 35–96.

Birnbaum, Michael H., "Is There Sex Bias in Salaries of Psychologists?" *American Psychologist*, 1979, 719–20.

Burks, Barbara S., "The Relative Influence of Nature and Nurture upon Mental Development: A Comparative Study of Foster Parent–Foster Child Resemblance and True Parent–True Child Resemblance." In *Twenty-seventh Yearbook: Nature and Nurture, Part 1, Their Influence on Intelligence*. Bloomington: National Society for the Study of Education, 1928, pp. 219–316.

Burt, Cyril, "The Genetic Determination of Differences in Intelligence: A Study of Monozygotic Twins Reared Together and Apart," *British Journal of Psychology*, 1966, *57*, 137–153.

Burt, Cyril, and M. Howard, "The Multifactorial Theory of Inheritance and its Application to Intelligence," *British Journal of Statistical Psychology*, 1956, *9*, 95–131.

Burt, Cyril, and M. Howard, "Heredity and Intelligence: A Reply to Criticisms," *British Journal of Statistical Psychology*, 1957, *10*, 33–63.

Cavalli-Sforza, L. L., *Elements of Human Genetics*. Menlo Park: Benjamin/Cummings, 1977.

Cavalli-Sforza, L. L., and W. F. Bodmer, *The Genetics of Human Populations*. San Francisco: Freeman, 1971.

Chamberlain, Gary, "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 1982, *18*, 5–46.

Dempster, Arthur P., "Employment Discrimination and Statistical Science" (with discussion), *Statistical Science*, 1988, *3*, 149–95.

Falconer, D. S., *Introduction to Quantitative Genetics*, (second edition). New York: Wiley, 1981.

Ferguson, T. S., "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 1958, *29*, 1049–62.

Fisher, R. A., "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh*, 1918, *52*, 399–433.

Friedman, Milton, *A Theory of the Consumption Function*. Princeton: Princeton University Press, 1957.

Griliches, Zvi, "Errors in Variables and other Unobservables," *Econometrica*, 1974, *42*, 971–98.

Herrnstein, Richard J., "IQ," *The Atlantic Monthly*, September 1971a, 43–64.

Herrnstein, Richard J., *IQ in the Meritocracy*. Boston: Little, Brown & Co., 1971b.

Jencks, Christopher, *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books, 1972.

Jensen, Arthur R., "How Much Can We

Boost IQ and Scholastic Achievement?" *Harvard Educational Review*, 1969, *39*, 1–123.

Jensen, Arthur R., *Genetics and Education*. New York: Harper & Row, 1972.

Jensen, Arthur R., *Educability and Group Differences*. New York: Harper & Row, 1973.

Jensen, Arthur R., "The Meaning of Heritability in the Behavioral Sciences," *Educational Psychologist*, 1975, *11*, 171–83.

Jinks, J. L., and L. J. Eaves, "IQ and Inequality," *Nature*, 1974, *248*, 287–89.

Kamalich, Richard A., and Solomon W. Polachek, "Discrimination: Fact or Fiction? An Examination Using an Alternative Approach," *Southern Economic Journal*, 1982, *49*, 450–61.

Kamin, Leon J., *The Science and Politics of IQ*. New York: Halstead Press, 1974.

MaCurdy, Thomas E., "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 1982, *18*, 83–114.

Malinvaud, Edmond, *Statistical Methods of Econometrics*. Chicago: Rand-McNally, 1966.

Manski, Charles F., *Analog Estimation Methods in Econometrics*. New York: Chapman and Hall, 1988.

Neyman, Jerzy, "Contribution to the Theory of the χ^2-Test," *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1949, pp. 239–73.

Pearson, Karl, and A. Lee, "On the Laws of Inheritance in Man. I. Inheritance of Physical Characters," *Biometrika*, 1903, *2*, 357–462.

Rao, D. C., N. E. Morton, and S. Yee, "Resolution of Cultural and Biological Inheritance by Path Analysis," *American Journal of Human Genetics*, 1976, *28*, 228–42.

Roberts, Harry V., "Statistical Biases in the Measurement of Employment Discrimination." In Livernash, E. R., ed., *Comparable Worth: Issues and Alternatives*. Washington D. C.: Equal Employment Advisory Council, 1980, pp. 173–95.

Rothenberg, Thomas, "Structural Restrictions and Estimation Efficiency in Linear Econometric Models." Cowles Foundation Discussion Paper No. 213, Yale University, 1966.

Taubman, Paul, "The Determinants of Earnings: Genetics, Family, and Other Environments; A Study of White Male Twins," *American Economic Review*, 1976, *66*, 858–70.

Taylor, W. F., "Distance Functions and Regular Best Asymptotically Normal Estimates," *Annals of Mathematical Statistics*, 1953, *24*, 85–92.

White, Halbert, "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 1980a, *21*, 149–70.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 1980b, *48*, 817–38.

Wright, Sewall, *Corn and Hog Correlations*. Washington: U.S. Department of Agriculture Bulletin 1300, 1925.

Wright, Sewall, "Statistical Methods in Biology," *Journal of the American Statistical Association*, Supplement, 1931, *26*, 155–63.

Wright, Sewall, "The Method of Path Coefficients," *Annals of Mathematical Statistics*, 1934, *5*, 161–215.

Zellner, Arnold, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 1962, *57*, 348–68.

Zellner, Arnold, "Estimation of Regression Relationships Containing Unobservable Independent Variables," *International Economic Review*, 1970, *11*, 441–54.

This article has been cited by:

1. Guillermina Jasso. 2012. Safeguarding Justice Research. *Sociological Methods & Research* **41**:1, 217-239. [Crossref]
2. Helen L. Berry, Peter Butterworth, Tanya M. Caldwell, Bryan Rodgers. 2008. Contemporary Australian Archetypes: Different People, Different Needs. *SSRN Electronic Journal* . [Crossref]
3. Pedro N. Teixeira. 2007. Great expectations, mixed results and resilient beliefs: the troubles of empirical research in economic controversies. *Journal of Economic Methodology* **14**:3, 291-309. [Crossref]
4. Harold J. Brumm, Dale O. Cloninger. 1996. Perceived risk of punishment and the commission of homicides: A covariance structure analysis. *Journal of Economic Behavior & Organization* **31**:1, 1-11. [Crossref]
5. Nancy J. Wulwick. 1996. Two Econometric Replications: The Historic Phillips and Lipsey-Phillips Curves. *History of Political Economy* **28**:3, 391-439. [Crossref]
6. G.S. Maddala, M. Nimalendran. 17 Errors-in-variables problems in financial models 507-528. [Crossref]
7. Wayne E. Ferson. Chapter 5 Theory and empirical testing of asset pricing models 145-200. [Crossref]
8. Wayne E. Ferson, Stephen R. Foerster. 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. *Journal of Financial Economics* **36**:1, 29-55. [Crossref]