IULTIPLE LINEAR REGRESSION VIEWPOINTS
/olume 11, Number 1, September, 1981

SPECIFICATION BIAS IN CAUSAL MODELS WITH FALLIBLE INDICATORS

BARBARA J. PATTESON AND LEE M. WOLFLE
VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

ABSTRACT

In the bivariate case, measurement error in the independent variable produces an attenuated estimate of the true regression coefficient. In the multivariate case, the bias which results from specifying, incorrectly, a model with no measurement error will produce biased estimates which are predictable in neither their direction nor magnitude. This paper demonstrates some of these biases in a causal model of educational attainment.

Educational researchers have known for a long time that measurement errors in independent variables cause regression estimates to be biased. In the bivariate case, measurement error in the independent variable produces attenuated regression estimates. In the multivariate situation, however, neither the size nor direction of the bias is predictable, unless one knows in advance the magnitude and nature of the errors. This paper examines the implications of measurement error in a socioeconomic model of educational attainment. Wolfle and Lichtman (1981) estimated models of educational attainment for whites, blacks, and Mexican-Americans using

This paper was presented at the annual meetings of the American Educational Research Association, Los Angeles, April 16, 1981. This research was supported in part by a grant from the National Center for Education Statistics (No. 300-78-0561).

estimates of structural parameters corrected for measurement error. Their report, however, concentrated on comparisons of interethnic differences in coefficients, and nowhere did they demonstrate that their LISREL-produced estimates differ in important ways from multiple regression estimates of the same model. Using the Wolfle and Lichtman (1981) model of educational attainment, this paper demonstrates that regression estimates will differ substantially from estimates that are corrected for the existence of measurement error.

The importance of bias created by ignoring measurement error is a point of some controversy. Jencks, et al. (1972, p. 336) concluded that the effects of random measurement error in a model of intergenerational mobility were relatively unimportant. In contrast, Bielby, Hauser and Featherman (1977) found that random measurement errors among nonblack men yielded regression estimates biased from 9 to 16 percent. For black men, however, Bielby, et al., found evidence of nonrandom errors, which yielded estimates whose biases were substantially larger than those for nonblacks. They concluded that, "because of the differing structures of response error among black and nonblack men, ignoring those structures leads to an exaggeration of black-nonblack returns to schooling and to understatement of racial differences in total and conditional inequality of occupational attainment" (Bielby, Hauser and Featherman, 1977, p. 1277). In addition, Wolfle (1979) has compared regression estimates in a model of educational attainment to LISREL-produced estimates corrected for measurement error using data from the National Longitudinal Study of the High School Class of 1972. Among whites, he found random measurement error produced regression estimates biased as much as 200 percent.

Concerned that differential levels of measurement bias would affect ›ir substantive conclusions about differences in the educational process ˊ whites, blacks, and Mexican-Americans, Wolfle and Lichtman (1981) ›d a general method of the analysis of covariance structures (Jöreskog d Sörbom, 1978) to generate structural parameter estimates free of asurement error bias in a model of educational attainment.

This paper examines the size and importance of measurement error ases in the Wolfle and Lichtman (1981) model as a demonstration of the ›sts involved in ignoring measurement error. In order to do this, the ırameters in the Wolfle and Lichtman model have been reestimated with ·dinary least squares regression. These new estimates have then been ›mpared with the LISREL (corrected) estimates reported by Wolfle and ichtman (1981).

IE MODEL

The basic model of educational attainment used in this analysis is hown diagrammatically in Figure 1. The variables of interest are shown ithin ellipses, and include father's occupational status, father's educa- ion, mother's education, number of siblings, sex, ability, academic ›reparation, college plans, and educational attainment. The arrows manating from the ellipses to mnemonic labels describe the measurement portion of the LISREL model, and are described in detail in Wolfle and Lichtman (1981). The variables used in the ordinary least squares regression are described below.

The model is a fully recursive set of structural equations in which ability is dependent upon five exogenous variables plus a residual

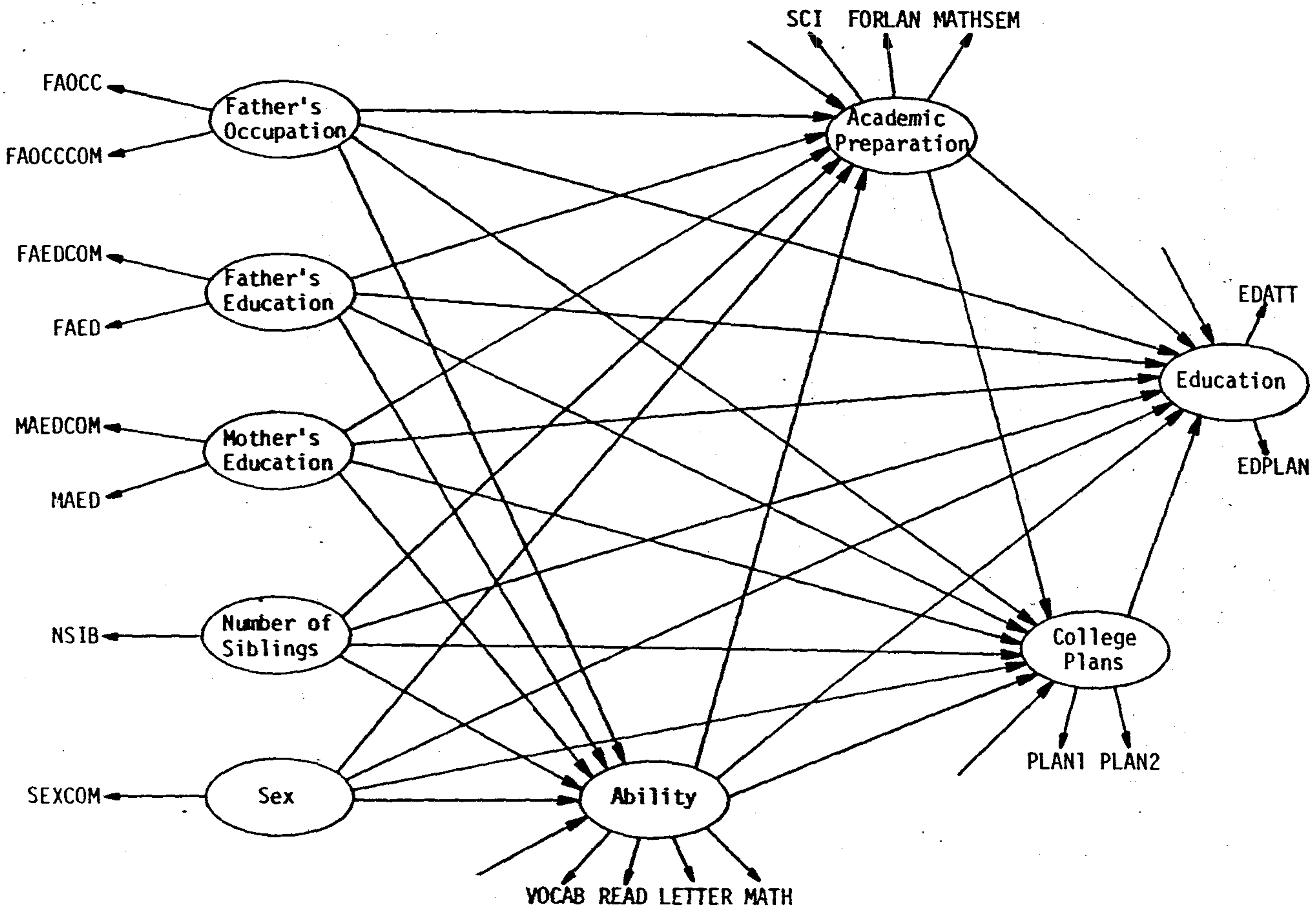

Figure 1. Structural Equation and Measurement Models of Educational Attainment

Among 1972 High School Graduates

sturbance term. Academic preparation is dependent upon ability, five ogenous variables, and a residual term. College plans is dependent upon ility, academic preparation, five exogenous variables, and a residual. inally, educational attainment is dependent upon all of the preceding ariables in the model, plus a residual term.

HE DATA

Data for this study were drawn from the National Longitudinal Study f the High School Class of 1972 (see Levinsohn, et al., 1978). The NLS, hich has been and continues to be supported by the National Center for Education Statistics, was designed to provide data on a large cohort of high school seniors, and to follow these students as they made the move from high school into their early years of adulthood. The data file included base-year survey and test-score data collected in 1972, along with follow-up surveys in 1973, 1974, and 1976 (the 1978 follow-up data are now available, but had not yet been made public at the time of our analysis). The analysis reported here is restricted to white NLS respondents. As with most other analyses of the process of socioeconomic achievement, pairwise present correlations were used to estimate the parameters of the model; the average number of whites in the analysis was 11,743.

In estimating the parameters of the model using ordinary least squares regression, in some cases only one manifest measure was used in place of the LISREL latent variable, and in other cases a simple summated scale was computed. The variable used to measure father's occupation was FAOCCCOM (V2468), a composite variable measured in terms of Duncan's (1961)

SEI scale as revised to match the 1970 census occupation classification (Hauser and Featherman, 1977). The variable was a composite of each individual's responses to base-year and first-year follow-up questionnaire items which asked respondents to indicate their father's main occupation.

The variables used to represent father's and mother's education were FAEDCOM (V1627) and MAEDCOM (V1628), respectively. These also were composite scores based on base-year and first follow-up questionnaire items. The category responses for these two variables were recoded to the number of years of schooling completed.

Values for the number of siblings were obtained by summing the values of four questions, which asked respondents to indicate the number of older brothers, younger brothers, older sisters, and younger sisters. Completing the specifications of the exogenous variables, sex was measured by the base-year and first follow-up composite variable, V1626. This dummy variable was coded zero for males and unity for females, so that positive coefficients emanating from this variable indicate higher values on the dependent variable for women.

The dependent variable, ability, was computed by taking the arithmetic average of four subtest scores administered to the respondents in the base year. The four measures were tests of vocabulary, reading, letter-groups, and mathematics. Academic preparation was also computed as an average of three indicators. The variables included in this computation were the number of semesters of science taken between July 1, 1969, and high school graduation (V0046), the number of semesters of foreign language completed in the same period (V0053), and the semesters of mathematics (V0074).

College plans were indexed by an NLS routing question (V0385), to :h respondents indicated how they intended to spend the largest part :heir time in the year after leaving high school. Anyone who planned ıttend a two-year or four-year college or university either full time ɔart time was given a value of unity on the college plans variable. other respondents were given a value of zero.

Finally, educational attainment (V1854) was measured from responses a question asked in the 1976 follow-up in which respondents were asked indicate the highest level of education or training they had received. category responses were recoded to years of completed schooling.

ECTS OF RANDOM MEASUREMENT ERROR

Measurement errors can be of many kinds. One kind exists when 'ors in one variable are correlated with the values of another variable. example, Mare and Mason (1980) have shown that women report their ther's occupation with greater error than men, apparently because the ther's occupation is more salient for young men than women. Another nd of reporting error exists when the errors in one variable are rrelated with errors in another variable. Bielby, Hauser and Featherman 977), for example, found that blacks overstate the consistency among eir own status characteristics and those of their fathers.

It is also possible for measurement errors to be uncorrelated with ything else. For example, Bielby, Hauser and Featherman (1977) found at a status attainment model for nonblacks with correlated errors ovided no better fit to the observed covariances than did a model with ll error covariances specified to be zero. They concluded that measurement errors of status characteristics for nonblack men were strictly random. Wolfle and Lichtman (1981) reached the same conclusion for whites regarding the randomness of measurement errors in their model of educational attainment; they found that the most likely candidates for correlated errors yielded a model only marginally better in its fit than a model with only random errors.

As a result, the comparison of the Wolfle and Lichtman (1981) parameter estimates to ordinary least squares regression estimates will indicate some of the biasing effects of random errors of measurement in multivariate analyses. If there exist other kinds of errors, one should expect to find structural parameter estimates substantially affected by the nature of such errors.

Even with random errors, biases can sometimes be severe. For example, suppose both father's occupation and education were reported with random error. If these two fallible indicators were then to be included in a regression equation predicting variation in ability, the estimates of their structural effects would be biased toward zero, thus underestimating the dependence of ability upon these two social background variables. Moreover, it is unlikely that both father's occupation and education are measured with equal reliability, and to the extent that these errors of measurement are not equal, the regression estimates of ability on father's education and occupation will be either inflated or deflated. With unequal reliabilities, therefore, the bias in regression estimates is not necessarily toward zero (as in the bivariate case); the direction and magnitude of the bias depends on the relative reliabilities of the independent variables, and unless one knows what these are in advance, the biasing effects are unpredictable.

Consider now the regression of academic preparation on ability and exogenous variables. If the exogenous variables are reported with om errors, the joint dependence of both ability and academic prepara- on the exogenous variables will be underestimated. As a result, dependence of academic preparation on ability will be overstated, use the effects of the exogenous variables are understated.

It is reasonable to suspect that all the variables in this model reported with random error. Moreover, it is also reasonable to ose that their reliabilities differ. As a result, all the regression mates are likely to be biased, but it is unreasonable to suppose that magnitude of bias is necessarily either small or consistently in one ction.

ILTS

Table 1 shows two sets of estimated structural parameters in a model ducational attainment. For each dependent variable, the top row of fficients are corrected (LISREL) estimates as reported by Wolfle and htman (1981); the second row of coefficients for each dependent variable, wn in parentheses, are ordinary least squares (OLS) regression estimates. king first at the proportion of explained variance for each dependent iable, note that the method of least squares understates the true lained variance by 19 to 30 percent. This occurrence results from the mination of a considerable amount of random error in the corrected REL estimates of the variances and covariances among the latent variables.

Examining the effects of the five background characteristics on ility, as predicted the OLS estimates underestimate the effects

Table 1. Estimates of Parameters in a Model of Educational Attainment: White 1972 High School Graduates (N = 11,743)

Dependent Variable	Predetermined Variables								R^2
	Father's Occup.	Father's Educ.	Mother's Educ.	Number Siblings	Sex	Ability	Acad. Prep.	College Plans	
Ability	.047	.381	.572	-.262	.185				.16
	(.029)	(.374)	(.462)	(-.245)	(.534)				(.13)
Academic Preparation	.006	.011	.012	-.034	-.510	.119			.45
	(.004)	(.022)	(.024)	(-.036)	(-.405)	(.105)			(.32)
College Plans	.003	.008	.013	-.014	.031	.017	.090		.44
	(.001)	(.017)	(.013)	(-.016)	(.001)	(.017)	(.074)		(.31)
Educational Attainment	.001	.032	.035	-.025	-.029	.030	.115	1.786	.68
	(.001)	(.040)	(.048)	(-.042)	(-.004)	(.036)	(.134)	(1.379)	(.50)

Note: Corrected LISREL estimates are shown without parentheses. Ordinary least squares regression estimates are shown in parentheses.

of these variables. For example, the OLS estimate for the effect of father's occupation on ability is .029, whereas the corrected estimate is .047. This is a negative bias of about 38 percent. The OLS estimates of the effects of father's education, mother's education, and number of siblings on ability are also negatively biased by about 2 percent, 19 percent, and 6 percent respectively. The only OLS coefficient for ability which is not negatively biased is the effect from sex, but this may be due to the different ways in which the ability variable was constructed. LISREL gave the greatest weight to a mathematics test score in the construction of the latent ability variable, which when balanced against three manifest measures of verbal ability yielded estimates that suggest there is no sex effect on ability. The construction of the ability measure for the OLS equations gave these four subtests equal weight, giving verbal expression, on which women excell, more weight than math. Thus, the regression estimate shows a positive effect.

Looking at the effects on academic courses taken in high school, it is seen once again that the OLS estimate of father's occupational status is negatively biased by about one-third. OLS estimates of parental education are, however, positively biased, both by about 100 percent. This positive bias causes the dependence of academic preparation on parental education to be overstated in the OLS analysis; as a result, the effect of ability on academic preparation is understated by about 12 percent.

A similar pattern seems to develop when looking at the effects on college plans. The structural coefficients for father's occupation and academic preparation are negatively biased by 67 percent and 18 percent, respectively, but father's education is positively biased by about 100 percent.

Turning to the dependent variable of primary interest, educational attainment, the OLS estimates of the effects of parental education are once again positively biased, thus overstating the effect of educational background on respondent's educational attainment. The sizes of these biases for father's and mother's education are 25 percent and 37 percent, respectively. If the previous pattern were to be followed, one might expect the OLS estimates of the effects of ability and academic preparation on educational attainment to be an underestimate of the true effect, due to the overestimation of the dependence of educational attainment on parental education. Such is not the case. The OLS estimate for ability is positively biased by about 20 percent, and academic preparation is positively biased by nearly 17 percent. Such fluctuation is to be expected in a multivariate model. The nature of measurement error causes OLS estimates not to be well behaved, so that the extent of bias becomes unpredictable in both magnitude and direction. Finally, the effect of college plans expressed in high school on educational attainment is underestimated by its OLS coefficient by about 23 percent.

CONCLUSION

This paper has examined the extent of bias inherent in ordinary least squares regression estimates when the presence of measurement error is ignored. While these results were based on but one population of high school students, and one structural model of educational attainment, the implications are much more widespread. It seems unwise to assume social variables are measured without error. This paper has demonstrated that the ordinary least squares estimates will be biased if measurement errors re incorrectly specified (e.g., assumed to be zero by choosing to ignore them). In some cases, the nature of bias can be predicted, but the more usual situation is that measurement error bias is unpredictable. Biases may be offsetting, but are just as likely to be additive. Mason, et al. (1976) were wise to suggest that:

> Since the errors may be large or small, and their effects may be additive or offsetting, there is no way to access the biases in naive (uncorrected) models of achievement processes without first investigating the separate and joint effects of each type of measurement error (Mason, et al., 1976, p. 444).

Researchers would be well advised to heed such advice. This paper has found that biases may exceed 100 percent of the corrected estimates. While this degree of bias is serious, indeed, it pales against the extent of bias possible with correlated errors of measurement. In applicable situations, educational researchers should avail themselves of new analytical techniques which allow for the assessment of, and correction for, measurement error in models of educational processes.

REFERENCES

Bielby, W. T., Hauser, R. M., and Featherman, D. L. Response errors of black and nonblack males in models of the intergenerational transmission of socioeconomic status. American Journal of Sociology, 1977, 82, 1242-1288.

Duncan, O. D. A socioeconomic index for all occupations. In A. J. Reiss, Jr. (ed.), Occupations and Social Status. New York: The Free Press of Glencoe, 1961.

Hauser, R. M., and Featherman, D. L. The Process of Stratification: Trends and Analyses. New York: Academic Press, 1977.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. Inequality: A Reassessment of the Family and Schooling in America. New York: Basic Books, 1972.

Jöreskog, K. G., and Sörbom, D. LISREL IV: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Chicago: National Educational Resources, 1978.

Levinsohn, J. R., Henderson, L. B., Roccobono, J. A., and Moore, R. P. National Longitudinal Study: Base Year, First, Second and Third Follow-Up Data File Users Manual. Washington, D.C.: National Center for Education Statistics, 1978.

Mare, R. D., and Mason, W. M. Children's reports of parental socioeconomic status: A multiple group measurement model. Sociological Methods & Research, 1980, 9, 178-198.

Mason, W. M., Hauser, R. M., Kerckhoff, A. C., Poss, S. S., and Manton, K. Models of response error in student reports of parental socioeconomic characteristics. In W. H. Sewell, R. M. Hauser, and D. L. Featherman (eds.), Schooling and Achievement in American Society. New York: Academic Press, 1976.

Wolfle, L. M. Unmeasured variables in path analysis. Multiple Linear Regression Viewpoints, 1979, 9, 20-56.

Wolfle, L. M., and Lichtman, M. Educational attainment among whites, blacks, and Mexican-Americans. Paper read at American Educational Research Association, Los Angeles, 1981.