COVARIANCES OF ERRORS IN SURVEY RESPONSES

Willard L. Rodgers and A. Regula Herzog, University of Michigan

1. Introduction

Assumptions about the relationships of errors in measurement are commonly made in analyzing survey data, generally with little justification beyond the pragmatic one that without such assumptions it is impossible to estimate the population parameters for which the survey was intended. For example, one type of assumption is that measurement errors are uncorrelated with the true values on the concept being measured (e.g., Alwin and Jackson, 1980, p. 70; Elashoff and Elashoff, 1978, p. 246; Johnston, 1972, p. 289; and Kenny, 1979, p. 71). Other types of assumptions that are frequently made concern the independence of errors in measurement of one variable relative to the true values of other variables, and the joint independence of errors in measures of different concepts. If such assumptions are not justified, estimates of population parameters may be biased.

Consider the simple causal model given by the following expression:

$$\eta_i = \beta\xi_i + \varsigma_i \ . \tag{1}$$

One variable, ξ, is assumed to have a linear effect on a second variable, η. A measurement model is given by expressions (2) and (3). The concept ξ is measured by a single survey item, X, and η is measured by another item, Y:

$$X_i = \xi_i + \delta_i \tag{2}$$

and

$$Y_i = \eta_i + \epsilon_i \ . \tag{3}$$

Standard practice is to estimate the effect of ξ on η by the ordinary least squares statistic:

$$b = s_{xy}/s_x^2 \ . \tag{4}$$

b is a consistent estimator of β only if a standard set of assumptions is correct with respect to the error terms for the two measures:

$$\sigma_\delta^2 = 0 \ , \tag{5}$$

and

$$\sigma_{\xi\epsilon} = 0 \tag{6}$$

Expression (5) says that there are *no* errors in measurement of the exogenous variable, and expression (6) says that measurement errors in the endogenous variable are uncorrelated with true values of the exogenous variable. Violations of the first of these assumptions are often recognized and taken into account by a procedure such as "correction for attenuation" or by simultaneous estimation of a measurement and a causal model using a procedure such as that implemented in Jöreskog's LISREL computer program (Jöreskog and Sörbom, 1984). With such procedures, however, come an additional set of implicit assumptions concerning the measurement errors in the exogenous variable, as given by expression (7):

$$\sigma_{\xi\delta} = \sigma_{\eta\delta} = \sigma_{\delta\epsilon} = 0 \ , \tag{7}$$

that is, such errors are assumed to be unrelated to the true values of both the exogenous and the endogenous variables and with errors on the endogenous variables. If these assumptions are not justified, the value of the OLS estimator in expression (4) has a limiting value that is given by expression (8):

$$\text{plim(b)} = (\beta\sigma_\xi^2 + \sigma_{\eta\delta} + \sigma_{\xi\epsilon} + \sigma_{\delta\epsilon})/(\sigma_\xi^2 + 2\sigma_{\xi\delta} + \sigma_\delta^2). \tag{8}$$

From this expression it is clear that violation of any of the assumptions given by expressions (5) through (7) may introduce bias into the standard OLS estimator of the effect of one variable on another. Without specific knowledge of the actual covariances of the error terms, it is impossible to improve on the standard estimator since the bias could be either positive or negative and nothing is known about its absolute value.

If we seek to estimate the parameters of a more complex model than the bivariate one given by expression (1), the potential effects of measurement error become even more difficult to predict. I will not belabor this point, but let me simply mention that if there are multiple exogenous variables, non-zero covariances involving measurement errors with respect to any one of them may introduce biases into regression coefficients for *all* of the predictors.

The few relevant investigations in which measurement errors have been explicitly evaluated (e.g., Duncan and Mathiowetz, 1985; Herzog and Dielman, 1985; Presser, 1984) raise doubts about the tenability of the types of assumptions that I have enumerated. The sheer frequency and magnitude of errors in survey measures of a variety of concepts have been found to be high enough to warrant skepticism concerning unsubstantiated assumptions about covariances involving those errors.

2. Methodology

The present paper investigates covariances of measurement errors with the true values of the concepts they are intended to represent, with measures of other concepts, and with one another. We operationalize measurement error as the discrepancy between a survey report and an external measure. In our research my colleagues and I have considered measurement error with respect to a range of variables and using several different sources of validating information.

The data were collected as two components of the Study of Michigan Generations project conducted by the Survey Research Center (SRC) at The University of Michigan. During the spring of 1984, SRC interviewers conducted face-to-face interviews lasting an average of 90 minutes with a probability sample of 1491 residents of the Detroit metropolitan area. Independent information about many of the variables measured by answers to survey questions was sought from existing, publicly accessible records. Such information may, of course, contain its own errors, although we made every effort to optimize the quality of this information.

We treat the survey and the records data as two sources of fallible information, and analyze any discrepancy between them as an indicator of measurement error in at least one of those sources.

3. Findings

The correlations between the signed values of the discrepancies with respect to each of sixteen survey measures and the validation records for those same variables are shown in the first column of Table 1. Time constraints prevent me from discussing the correlations for specific variables, but it is useful to consider the overall pattern of these correlations. A majority of the correlations have absolute values greater than 0.20, and six of them have absolute values larger than 0.45, so the relationships are by no means trivial or unimportant.

With one notable (and two minor) exceptions, the directions of these correlations are negative, which means that large values according to the records tend to be under-reported by the respondents, whereas below-average values tend to be over-reported. Most of these negative correlations are at least to some extent statistical artifacts of the measurement procedure. This is particularly true for dichotomously scored variables such as the reports of voting. A reporting error by someone who, according to the records, did not vote in an election can only be an error in one direction, here scored as a positive error, whereas a reporting error by someone who *did* vote can only be in the negative direction. Only variables scored on an open-ended scale, such as dollars or years, escape this artifactual constraint, and so it is no surprise to observe that the only positive correlations between measurement error and the recorded value are with respect to value of house, property taxes, and age.

The estimated standard errors of the correlations are shown in parentheses, and the corresponding t-tests (not shown) indicate that ten of the correlations differ significantly ($p<.05$) from zero—all ten, in fact, being significantly *less* than zero.

Another of the assumptions I enumerated at the outset concerned the lack of relationships between measurement errors with respect to one variable and values of different variables. The possibilities here are endless and it is only practical to examine some examples involving variables of a type which are often included in causal models. Specifically, we have examined the correlation of measurement errors on each of sixteen variables to each of a set of five standard demographic variables. These correlations are shown in Table 2.

In one sense, these correlations are reassuring: 76 percent of them have absolute values less than 0.10, and 75 percent are not significantly different from zero ($p>.05$). At the same time, the fact that 24 percent of the correlations exceed 0.10 in absolute value, and that 4 percent exceed 0.20, should give one pause. These correlations are large enough that they could introduce substantial biases into estimates of causal effects, particularly if the true causal effects are small.

I noted in my introductory remarks that measurement error correlated with any of a set of predictor variables in a regression model may result in biased estimates of the entire set of regression coefficients. An example of such biases is shown in Table 3. The dependent variable in this example is the property taxes paid by homeowners in the previous year, as reported in the first instance by respondents, in the second instance by local assessors' offices. The predictors are five demographic variables. There are substantial differences in the estimated regression coefficients depending on which report is used. The estimated coefficient for marital status is considerably larger when predicting the respondent reports than when predicting assessors' reports, and in the latter case is not statistically different from zero ($p>.05$). The coefficient for race is almost twice as large if based on respondent reports rather than official values.

The dependent variable in a second example is voting behavior in the 1980 election. Since this is a dichotomous variable, we used logit analysis rather than ordinary least squares, obtaining the regression coefficients shown in Table 4. The coefficients in the first column are based on the respondents' own reports on whether or not they voted, while those in the second column are based on voting records. Again there are at least small differences in all of the coefficients, and three of the differences are substantial. The largest of these differences is with respect to race. Based on analysis of the self-reports, it appears that blacks were much more likely to vote in the 1980 election than were non-blacks, but there is no evidence whatsoever—indeed, the estimate is in the opposite direction—for such a racial difference according to the records data.

The final type of assumption concerning measurement errors that we consider is that the errors for different items are uncorrelated with one another. Table 5 shows correlations between errors of responses to different survey questions. Across all pairs of items, the average absolute correlation of the algebraic discrepancy scores is very small, only .055. The average correlation between pairs of items within the same topic area is 0.14 which is large enough to be of potential concern. On the other hand, the average correlation between items on different topics is only 0.04, not much higher than would be expected if these discrepancies were all independent of one another.

4. Conclusions

In addition to sounding a note of caution with respect to assumptions about covariances involving measurement errors, our findings also suggest some data collection and analysis procedures that may be useful in reducing the extent to which such assumptions are violated. The most basic implication is the importance of developing and applying improved methods of data collection, which for personal interviews includes both the wording of questions and training of interviewers in techniques that elicit complete and accurate responses. Research on both of these aspects of survey research is appallingly deficient given their importance and complexity, but there is a growing body of literature (Cannell et al., 1981; Schuman and Presser, 1981; Sudman and Bradburn, 1982).

A more specific implication of the present research concerns the use of bracketed categories to obtain information about variables which have a wide potential range of true values. We noted that most of the correlations between self-reports and record values are negative, and pointed out that such negative correlations could be largely a consequence of using closed-ended scales, so that persons with high true values could only have negative measurement errors while those with low true values could only have positive measurement errors. This shortcoming argues strongly against using closed-ended scales when they can easily be avoided, although this recommendation must be weighed against the advantages offered by using closed-ended scales. For example, questions about dollar amounts should probably be asked in terms of dollars, not in terms of a set of dollar ranges, despite the somewhat higher response rates observed in response to the latter type of question. To minimize both types of problem, such questions could first be asked open-endedly, then those unable or unwilling to answer in such terms could be offered a set of categories.

One method of dealing with measurement errors at the data analysis stage is to transform the scale on which a variable is measured in a way that minimizes its covariances with substantive variables. In analyses that I have not had time to discuss here, we have seen examples in which logarithmic and other transformations have effectively eliminated observed covariances involving measurement errors on particular variables. Without specific knowledge about measurement errors on a particular variable, however, any transformation might introduce or exacerbate covariances rather than eliminating or reducing such covariances. We are not in a position to do so now, but perhaps continued study of measurement errors will lead to generalizations about the most appropriate transformations to apply to various types of response scales in order to reduce biases in statistics estimated from survey data due to covariances involving measurement errors.

Note: Helpful comments on an earlier version of this paper were made by Frank Andrews, Institute for Social Research, University of Michigan.

REFERENCES

Alwin, D.F. and Jackson, D.J. (1980). Measurement models for response errors in surveys: Issues and applications. In K.F. Schuessler (Ed.), **Sociological Methodology 1980**. San Francisco: Jossey-Bass, pp. 68–119.

Cannell, C.F., P.V. Miller, and L. Oksenberg. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), **Sociological Methodology 1981**. San Francisco: Jossey-Bass, pp. 389–437.

Duncan, G. and Mathiowetz, N. (1985). A Validation Study of Economic Survey Data. Research Report. Ann Arbor, MI: Survey Research Center.

Elashoff, J.D. and Elashoff, R.M. (1978). Effects of errors in statistical assumptions. In W.H. Kruskal and J.M. Tanur (Eds.), **International Encyclopedia of Statistics**, Volume 1. New York: The Free Press, pp. 229–250.

Herzog, A.R. and Dielman, L. (1985). Age differences in response accuracy for factual survey questions. *Journal of Gerontology*, **40**, 350–357.

Johnston, J. (1972). **Econometric Methods**. New York: McGraw Hill.

Jöreskog, K.G. and Sörbom, D. (1984). **LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood, User's Guide**. University of Uppsala, Sweden.

Kenny, D.A. (1979). **Correlation and Causality**. New York: Wiley.

Presser, S. (1984). Is inaccuracy on factual survey items item-specific or respondent-specific? *Public Opinion Quarterly*, **48**, 344–355.

Schuman, H. and Presser, S. (1981). **Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context**. New York: Academic Press.

Sudman, S. and Bradburn, N.M. (1974). **Response effects in surveys: A review and synthesis**. Chicago: Aldine.

Table 1. Correlations of Measurement Errors with Record Values

	n	Algebraic values of errors	Absolute values of errors
Age of automobile	811	-.2114 (.0640)	-.1181 (.0803)
Driver license	1465	-.5086 (.0349)	-.4881 (.0417)
Vote:			
in 1980	1176	-.5520 (.0314)	-.5274 (.0357)
in 1982	1198	-.4798 (.0249)	-.2846 (.0437)
in 1983	812	-.4917 (.0272)	-.0899 (.0532)
Age:			
compared to Michigan driver license or ID	1029	.0214 (.0406)	.0292 (.0434)
compared to voting and registration rolls	910	-.0439 (.0755)	-.0294 (.0833)
Distance to:			
drug store	1428	-.4628 (.0734)	.4389 (.1007)
grocery store	1443	-.3854 (.0755)	.3994 (.0762)
fire station	1317	-.2742 (.0560)	.1803 (.0557)
hospital	1423	-.2788 (.0368)	.0735 (.0508)
Proportion of neighbors:			
60 years or older	1425	-.1015 (.0564)	-.0188 (.0569)
black	1470	-.2244 (.0591)	.2066 (.0466)
income over $10,000	1311	-.0236 (.0619)	-.2588 (.0642)
income over $30,000	1264	-.0239 (.0748)	.0767 (.0540)
Assessed value of house	761	.4711 (.3494)	.5302 (.3520)
Property tax paid in 1983	677	.0815 (.0432)	.0914 (.0358)

NOTES: The entries in the column labelled "n" are the numbers of cases with non-missing data for each pair of items.

Table 2. Correlations of Algebraic Measurement Errors with Demographic Characteristics

	Age	Education	Marital Status	Income	Race
Age of automobile	.0056	.0197	.0310	.0427	−.0422
	(.0732)	(.0424)	(.0785)	(.0608)	(.1021)
Driver license	−.0316	−.0107	−.0132	−.0415	.0731
	(.0263)	(.0278)	(.0501)	(.0286)	(.0516)
Vote:					
in 1980	−.1486	.0817	−.1437	−.0605	.1724
	(.0476)	(.0404)	(.0592)	(.0373)	(.0562)
in 1982	−.0741	.1001	−.0187	−.0131	.1004
	(.0512)	(.0373)	(.0571)	(.0386)	(.0609)
in 1983	−.0005	−.0471	−.0444	−.0722	.2901
	(.0480)	(.0422)	(.0686)	(.0531)	(.0744)
Age:					
compared to driver license or ID	.0154	.0141	−.0511	.0366	−.0378
	(.0382)	(.0233)	(.0461)	(.0241)	(.0602)
compared to voting and registration rolls	.0039	−.0585	−.1740	−.0672	.0783
	(.0702)	(.0532)	(.0659)	(.0380)	(.0781)
Distance to:					
drug store	.0473	−.0416	−.0933	−.0227	.1035
	(.0410)	(.0317)	(.0502)	(.0492)	(.0717)
grocery store	.0529	.0055	.0702	.0535	−.0095
	(.0363)	(.0309)	(.0369)	(.0338)	(.0452)
fire station	−.0236	−.0465	−.0293	−.0343	.0809
	(.0391)	(.0412)	(.0500)	(.0370)	(.0820)
hospital	.0363	.0434	.0135	.0864	.0097
	(.0349)	(.0468)	(.0402)	(.0514)	(.0372)
Proportion of neighbors:					
60 years or older	.1316	−.0561	.0134	.0331	.0429
	(.0432)	(.0349)	(.0361)	(.0848)	(.0487)
black	−.1139	.0343	.0298	.0195	−.1553
	(.0467)	(.0424)	(.0434)	(.0448)	(.0669)
income over \$10,000	−.0351	.1882	.1555	.2229	−.0742
	(.0424)	(.0447)	(.0392)	(.0658)	(.0550)
income over \$30,000	−.1204	.1595	.0971	.2342	−.1294
	(.0479)	(.0405)	(.0500)	(.0683)	(.0510)
Assessed value of house	−.0225	.0925	.0670	.0479	−.0067
	(.0634)	(.0530)	(.0383)	(.3446)	(.0506)
Property tax paid in 1983	−.0483	.1198	.0430	.1899	−.0897
	(.0524)	(.0587)	(.0366)	(.0419)	(.0235)

NOTES: The demographic variables are defined as follows:
Age: As reported by respondents.
Income: Family income for previous year as reported by respondents.
Marital status: Married or living together = 1, all others = 0.
Education: In years.
Race: Black = 1, others = 0.

Table 3. Multiple Regression Coefficients: Predicting Property Taxes from Demographic Characteristics

Predictor	Respondent	Record
Constant	534.446	815.658
Age	−7.359 (10.077)	−2.547 (3.603)
Income	5.898 (0.971)	4.222 (2.073)
Missing Data on Income	2.649 (330.145)	295.596 (206.809)
Marital status	−589.907 (247.976)	−407.414 (345.868)
Education	36.628 (66.851)	−9.058 (50.925)
Race	−785.721 (215.299)	−430.215 (198.080)

NOTES: Standard errors are given in parentheses. The predictor variables are defined as described in notes to Table 2. Income is measured in hundreds of dollars, and respondents with missing data on income are recoded to $25,000 (approximately the mean value). The indicator variable (labelled "Missing Data on Income") is set to 1 for those respondents with missing data, 0 for all others.

Table 4. Logit Regression Coefficients: Predicting Voting Behavior in 1980 from Demographic Characteristics

Predictor	Respondent	Record
Constant	−4.9698	−4.6156
Age	.0509 (.0052)	.0535 (.0047)
Income	.0108 (.0055)	.0081 (.0044)
Marital status	.4683 (.1820)	.9108 (.1645)
Education	.2541 (.0355)	.1286 (.0293)
Race	.6569 (.1925)	−.1818 (.1682)

NOTES: Standard errors are given in parentheses. The predictor variables are defined as described in notes to Table 2. Income is measured in thousands of dollars.

Table 5. Average Correlations Between Discrepancy Scores

Topic Areas	n	Algebraic Discrepancies	Absolute Discrepancies
Automobile characteristics (2 items)	720	—	.5255
Voting behavior (3 items)	1062	.2596	.2786
Neighborhood Distances (4 items)	1402	.0931	.0618
Neighborhood Characteristics (4 items)	1367	.1309	.0520
Homestead Value and Prop. Tax (2 items)	719	.0621	.0576
Average correlation within topic areas		.1365	.1236
Average correlation between topic areas		.0431	.0423
Average correlations for all questions		.0542	.0514

NOTES: The entries in the column labelled "n" are average numbers of cases on which correlations in each topic area are based.