Journal of Official Statistics
Vol. 3, No. 4, 1987, pp. 403–418

Covariances of Measurement Errors in Survey Responses

Willard L. Rodgers and A. Regula Herzog[1]

Abstract: Using data from administrative records, census counts, and maps to assess errors in responses to a wide range of survey items, we investigated discrepancies in survey answers from interviews with a sample of the adult population in a metropolitan area. We found evidence that commonly made assumptions about the independence of measurement errors are often incorrect. Specifically, we found: (1) substantial correlations between discrepancy scores on most of the survey items and the values from the validating sources; (2) frequent, though generally small, correlations between discrepancy scores on one variable and measured values on other variables; and (3) that discrepancy scores on different variables tend to be somewhat correlated with one another only if the variables are within a single topic area. Such violations of assumptions can have profound effects on estimates of population parameters from survey data; for example, there may be substantial biases in estimates of regression coefficients.

Key words: Measurement error; validation; matching; estimation.

1. Introduction

Assumptions about the relationships of errors in measurement are commonly made when analyzing survey data. Generally the only justification for these assumptions is that it is impossible to estimate the population parameters without a set of assumptions. For example, an assumption that is frequently made by psychometricians (e.g., Nunnally (1967, p. 182)), sociologists (e.g., Alwin and Jackson (1979, p. 70, but see discussion on p. 105 ff.)), econometricians (e.g., Johnston (1972, p. 289)), and statisticians (e.g., Cochran (1968, p. 639, but see discussion in remainder of article)), among others, is that measurement errors are uncorrelated with the true values of the concept being measured. One usually makes this assumption when one treats certain variables (typically the exogenous variables in a causal model) as fixed, for example, as in an experimental design. This assertion may subsequently be relaxed by making conclusions conditional on the observed values of the exogenous variables. Nevertheless the assumption that is made, often implicitly, is that those values are known with certainty – that is, without measurement error. Other types of

[1] Willard L. Rodgers is an Associate Research Scientist in the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106–1248 U.S.A., and A. Regula Herzog is an Associate Research Scientist in the Survey Research Center, Institute for Social Research, and the Institute of Gerontology, University of Michigan. This research was supported by grant AGO2038 from the National Institute on Aging. The authors acknowledge helpful comments by Frank Andrews, Barbara Bailar, Greg Duncan, Jersey Liang, Sandra Newman and anonymous reviewers on earlier versions of this paper.

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 (e.g., previous citations). If such assumptions are not justified, estimates of population parameters may be highly biased.

Consider a simple causal model in which one variable, ξ, has a linear effect on a second variable, η, given by the following expression:

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

Assume that, in the population of interest, the following equality holds:

$$B = \frac{S_{\eta\xi}}{S_{\xi}^{2}}\ , \tag{2}$$

(where S_ξ^2 and $S_{\eta\xi}$ are the population variance and covariance, respectively), and let B be treated as an estimate of β. Let each of these concepts be measured for a sample of the population by a single survey item which is related to the corresponding concept as follows:

$$X_i = \xi_i + \delta_i, \tag{3}$$

and

$$Y_i = \eta_i + \varepsilon_i\,. \tag{4}$$

(Expressions (3) and (4) are simplified by adopting the conventions that the concepts and their indicators have the same units and that all are expressed as deviation scores.)

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

$$b = \frac{s_{xy}}{s_x^2}\,. \tag{5}$$

The value of b has the following limiting value:

$$\text{plim}(b) = \frac{BS_\xi^2 + S_{\eta\delta} + S_{\xi\varepsilon} + S_{\delta\varepsilon}}{S_\xi^2 + 2S_{\xi\delta} + S_\delta^2}\,. \tag{6}$$

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

$$S_\delta^2 = 0\,, \tag{7}$$

(*no* errors in measurement of the exogenous variable, which implies $S_{\xi\delta} = S_{\eta\delta} = S_{\delta\varepsilon} = 0$); and

$$S_{\xi\varepsilon} = 0 \tag{8}$$

(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 partially 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 the LISREL computer program (Jöreskog and Sörbom (1984)). Such procedures, however, generally ignore the possibility that the measurement errors may covary with the true values or with one another; that is, they implicitly assume that:

$$S_{\xi\delta} = S_{\eta\delta} = S_{\delta\varepsilon} = S_{\xi\varepsilon} = 0. \tag{9}$$

From expression (6) it is clear that violation of any of the assumptions about measurement errors 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. It is not necessary to take up additional space to demonstrate this, since the point has already been made with the bivariate models. It suffices to mention that if there are multiple exogenous variables, non-zero covariances involving measurement errors with respect to any one of the variables may introduce biases into regression coefficients for *all* of the predictors.

Investigations in which measurement errors have been explicitly evaluated raise doubts about the tenability of the types of assumptions that we have enumerated. The sheer frequency and magnitude of errors in survey measures have been found to be high enough to warrant skepticism concerning unsubstantiated assumptions about covariances involving those errors. For example, between one percent (for a question on workers' entitlement to paid vacation) and about sixty percent (for a question on the type of diagnosis during a hospital stay) of respondents reported information that did not agree with information gleaned from independent records (Duncan and Mathiowetz (1985); Herzog and Dielman (1985)). Moreover, Duncan and Mathiowetz (1985) and Radner (1982) found that, contrary to assumptions listed above, errors in measures of certain variables are significantly correlated with the true levels on those variables as well as with other variables that would often be treated as predictors in explanatory models. Presser (1984) also examined reporting errors across different variables. His findings suggest that reporting errors may be related across questions dealing with similar topics but not across questions dealing with different topics.

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. The operational definition of measurement error used in this paper is the discrepancy between a survey report and an external measure, and the covariances of those discrepancies with respect to a range of variables are examined using several different sources of validating information. We conclude with a brief consideration of some of the implications of the observed covariances on estimates of multivariate statistics from survey data.

2. Methods

2.1. *Design*

The data analyzed here were collected as two components of the Study of Michigan Generations project conducted by the Survey Research Center at The University of Michigan. Face to face interviews lasting an average of 90 minutes were conducted with 1 491 respondents in the Detroit metropolitan area. Independent information about many of the variables measured by survey responses 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.

2.2. *Sample and data collection*

The Sampling Section of the Survey Research Center (SRC) drew a multi-stage stratified area probability sample of households in the Michigan counties of Wayne, Macomb and Oakland, using procedures that would yield an oversample of older adults. Interviews were conducted with randomly selected respondents, with no proxy responses allowed. (Details about the sample design are available on request from the authors.)

The interviews were conducted from February through June, 1984, in the respondents' homes. The response rate was a rather low 58% but this is not out of line with other non-government surveys given the target population (major metropolitan area) and the length of the interview. Our objective in this paper, moreover, is to examine covariances involving response errors among survey respondents, not in the total population. (That is, we define the population of interest to include only those who would have responded to the request for an interview.)

2.3. *Measures*

The following is a list of survey questions and the independent information that was used to validate the answers to the survey questions. At least two staff members independently matched the records, with senior staff reviewing all "probable" matches. Information regarding the levels of agreement and bias in these measures may be found in a separate paper (Rodgers and Herzog (1987)).

Automobile. Respondents reported whether they had a vehicle registered in their own name, and if so, what the make and year of the (most frequently used) automobile was. We obtained validating information from the Michigan Secretary of State's office. The procedure for matching records was based on full name and address. If the records indicated the respondent had registered an automobile since the date of the interview, we treated absence of a match for the automobile reported in the interview as missing data rather than as an error.

Driver's license. Respondents reported on whether or not they had a valid Michigan driver's license. Validating information was obtained from the Michigan Secretary of State's office. The procedure for matching records was based on four criteria: (1) Michigan Driver's License number (as copied by interviewer); (2) first name (including initials) and last name (including search for misspellings); (3) street address and city; and (4) birth date (within 10 years if all other criteria matched).

Voting behavior. Three questions concerned whether the respondent voted in the 1980 presidential, the 1982 congressional, and the 1983 local school board elections, and another three questions concerned whether and at which address the respondent was registered to vote for these elections. Validating information was obtained from voter registration rolls in relevant city and township clerks' offices. The procedure for matching records was based on six criteria for both computer lists and hand-checking of registration rolls: (1) proper specific election; (2) ward and precinct; (3) name; (4) street address; (5) sex; and (6) birth date (again within 10 years).

Birth date. Respondents' reports of their date of birth were compared with the date recorded in the Michigan Driver's License and ID files, and also with the date recorded in voter registration rolls.

Distances to neighborhood facilities. Respondents judged the distances from their home to the nearest drug store, large grocery store, fire station, and general hospital using a scale with six bracketed categories ranging from "less than a quarter mile" to "more than five miles." The respondents also named each of these facilities and gave its street location. From this information, study staff were able to measure on a map the actual distance from the respondent's home to each facility, and to code these distances into the same six categories.

Characteristics of neighborhood residents. There were four questions on the percentages of blacks, persons 60 years of age or over, families with incomes over $10 000 and over

$30 000 living in the respondent's neighborhood. Validating information was obtained from census information (from the 1980 U.S. Census Data tape) on block groups corresponding to SRC sampling segments. Unfortunately, respondents find it difficult to characterize their neighbors in percentage terms. During the pretesting phase of this study, we asked respondents to specify the proportion of their neighbors who were black, over age 60, and so on, but obtained so many negative comments from the interviewers about this question format, and so small a proportion of usable responses from the pretest respondents, that we opted for a set of verbal categories. To compare these responses to census block group data, which are in percentage terms, we made arbitrary assignments of the verbal categories to the midpoints of what we considered reasonable percentage ranges for each category. (Specifically, percentages were assigned to the verbal categories as follows: "None" = 0 %; "Almost none" = 10 %; "Less than half" = 35 %; "About half" = 50 %; "More than half" = 65 %; "Almost all" = 90 %; and "All" = 100 %.)

House value and property taxes. Homeowners were asked to estimate the value of their home and the amount of property taxes they paid in the preceding year. These values were compared to those recorded in city and township assessors' offices. The procedure for matching records was based on three criteria: (1) first and last name of respondent (and spouse); (2) address of housing unit; and (3) school district. Millage rate for district, state equalization factor and other fees added to tax bill were also coded. The assessed values were multiplied by two because the official assessment is intended to be 50 % of market value.

2.4. Data analysis procedures

We use correlation coefficients to evaluate three types of assumptions concerning the independence of measurement errors. Our first concern is with the assumption that errors in measuring a variable are unrelated to the true values of that variable. It cannot be assumed, of course, that the data from external sources are error-free. For example, Kilss and Alvey (1976) examined data from the Current Population Survey (CPS), the Social Security Administration (SSA), and the Internal Revenue Service (IRS) and found that, with wage income classified into $500 categories, 17 % of the records from IRS disagreed with those from the SSA. This is a considerably lower rate of disagreement, however, than that between the CPS and either of the administrative sources: using the same categorization, they observed 44 % disagreement between the CPS and the SSA, and 38 % disagreement between the CPS and the IRS (Kilss and Alvey (1976, Table 6)). This indicates that the records data, while imperfect, have lower error rates than the survey reports. Since we do not know the "true" values, we settle for examining the correlations of discrepancies between the two sources of information on a variable with the value obtained from the external source. We then consider the correlations between discrepancy scores with respect to one variable and values on *different* variables. Finally, we consider the correlations between discrepancy scores on different variables.

3. Correlations with True Scores

3.1. Observed correlations of measurement errors with true values on the same variables

To estimate the extent to which measurement errors are related to true values, we consider the discrepancies between each of 16 survey measures and the validation records for those same variables. The correlations of these discrepancies with the record values are shown in Table 1. A majority of these correlations (11 out of 17) have absolute values greater than .20, and 4 of them have absolute values

Table 1. Correlations of discrepancy scores with record values

	n^*	Correlation	Standard error
Age of automobile	811	–.226	.072
Driver's license	1 465	–.483	.033
Vote:			
in 1980	1 176	–.544	.034
in 1982	1 198	–.476	.025
in 1983	812	–.480	.027
Age:			
compared to Michigan driver's license or ID	1 029	.035	.039
compared to voter registration rolls	910	–.035	.078
Distance to:			
drug store	1 428	–.417	.082
grocery store	1 443	–.397	.081
fire station	1 317	–.285	.060
hospital	1 423	–.268	.041
Proportion of neighbors:			
60 years or older	1 425	–.099	.060
black	1 470	–.222	.066
family income over $10 000	1 311	–.024	.062
family income over $30 000	1 264	.004	.055
Assessed value of house	761	.242	.138
Property tax paid in 1983	677	.073	.045

* The entries in this column are the number of cases with non-missing data for each pair of items.

larger than .45, so the relationships are by no means trivial or unimportant.

With four exceptions (none of which is statistically significant), 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 artificially introduced by the measurement procedure.[2] 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

[2] The negative correlations may also be attributable in part to violations of the assumption that the records data are error-free. Observe that if both the survey reports, X_{1i}, and the records, X_{2i} are error-laden measures of the actual level, η_i:

$$X_{1i} = \eta_i + \delta_{1i}$$

$$X_{2i} = \eta_i + \delta_{2i}\,,$$

then the discrepancy score is:

$$\Delta_i = X_{1i} - X_{2i} = \delta_{1i} - \delta_{2i}\,.$$

The covariance of the discrepancy and the record scores, then, is:

$$S_{\Delta x_2} = S_{\eta\delta_2} - S_{\eta\delta_1} + S_{\delta_1\delta_2} - S^2_{\delta_2}\,.$$

If the error terms do not covary with the true scores or with one another, this simplifies to:

$$S_{\Delta x_2} = -S^2_{\delta_2} \leq 0\,.$$

negative direction. Similarly, those living five or more miles from a grocery store could, given the measurement scale used in this survey, err only in the direction of underreporting the distance, whereas those living within a quarter mile could err only by overreporting. Only variables scored on an open-ended scale, such as dollars or years, escape this artificial constraint, and so it is no surprise to observe that three of the four positive correlations between measurement errors and the recorded values are with respect to house worth, property taxes, and age. The fact that most of the negative correlations are at least introduced by the response scale, and thus predictable without the additional work of collecting the validation data, does not, however, diminish the importance of these correlations. Whatever the reason for their existence, these correlations indicate substantial and frequent violations of the assumption that measurement errors are independent of the true values.

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. These standard errors are not simple random sample estimates, but were obtained from a program (REPERR, which is part of the OSIRIS.IV package: University of Michigan, Survey Research Center Computer Support Group (1982)) which takes into account the sample design. Because the sample was highly clustered, and overrepresents older people (with compensating sampling weights applied in the analysis), the design effect (Kish (1965)) is substantial for most of these correlations.

3.2. *Data transformations*

Some of the correlations of discrepancy scores with record values indicated in Table 1 may be reduced or eliminated by judicious transformations of the measurement scales. Consider, for example, the correlations between discrepancy scores on distances to neighborhood facilities and the distances measured from maps. We wanted to know the extent to which the relationships of discrepancies on the distance variables are artificially, imposed by the use of a set of categories as the response scale for a continuous variable. To obtain an approximation of what the covariance would be in the absence of such a constraint, we deleted respondents who, according to the map measure, lived less than a quarter mile or more than five miles from a facility – that is, those who were classified into the extreme distance categories on the six-point scale. The average covariance of the discrepancies with the map distances dropped from – .67 to – .20, indicating that most of the covariances may well reflect only the use of bracketed response scales. This is based on a rather peculiar subsample of the population, however – those who lived neither very close to nor far from a facility – so we would not try to draw any general conclusion from this observation.

It is doubtful that dollars are the best unit in which to compare errors on the house value and property tax variables. An error of \$10 000 has a very different importance if it is with respect to the value of a house assessed at \$20 000 than if it is with respect to the value of a house assessed at \$200 000. A more reasonable assessment of error, then, might be the discrepancy between the respondent's report and the official value, as a proportion of the official value. The correlation between such proportional discrepancies and the assessed values is – .21, in the opposite direction from the .24 observed for the simple discrepancies. The sampling error of this correlation is also much smaller after the transformation to proportions, so that it is now significantly different from zero ($p<.05$). Transforming the discrepancies on the property tax reports to proportional discrepancies relative to the record values also changed the sign of the correlation from positive to negative, but its magnitude is even smaller than for the simple discrepancies and does not differ significantly from zero. (The distribution of these transformed scores would, of course, have to be considered before undertaking a substantive analysis.)

4. Correlations with Values on Different Variables

4.1. Effects on bivariate regression coefficients

We next consider the assumption concerning the lack of relationships between measurement errors with respect to one variable and values of different variables. We examine some examples involving variables of a type which are often included in causal models. Specifically, we have examined the correlation of discrepancy scores on each of the 16 variables to each of a set of five standard demographic variables. These correlations are shown in Table 2.

In one sense, the correlations shown in Table 2 are reassuring; 73 % have absolute values less than .10, and 72 % are not significantly different from zero ($p>.05$, based on t-tests using the estimated standard errors, given in parentheses, which as noted earlier were calculated using REPERR to take account of the sample design). At the same time, the fact that 27 % of the correlations exceed .10 in absolute value, and that 5 % exceed .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.

The implications of these correlations become more obvious by considering expression (6). Take as an example the correlation of .24 between household income and the discrepancy between the self-report and the records with respect to the amount paid in property taxes. If we assume that income was measured without error, the variance and covariance terms in expression (6) that involve δ all would have values of zero and drop out, allowing us to rewrite the limiting value of the OLS estimator for B as:

$$\text{plim}(b) = B + \frac{S_{\xi\varepsilon}}{S_{\xi}^{2}} .$$

Using the sample covariance between the discrepancies and the records as an estimate of $S_{\xi\varepsilon}$, and the sample variance of the record values as an estimate of S_{ξ}^{2}, the estimated bias in the OLS estimator, b, is 3.35. The estimated regression coefficient, based on the assessors' records for property taxes, is 3.13, so the bias is larger in magnitude than the true value – that is, the estimate of B, based on self-reported property taxes, is 6.47, more than twice as large as the estimate based on the assessors' records. If, as seems highly likely (cf. Radner (1982)), there are errors in the self-reports of income as well as property taxes, additional bias terms have been introduced into the estimate of this B (cf. expression (6)); for example, it is plausible that respondents who overstated their property taxes also tended to overstate their incomes, so that the covariance of the error terms, $S_{\delta\varepsilon}$, would be positive. This bias would, however, be at least partially offset by the term for the variance of the error in income reports, S_{δ}^{2}, in the denominator of expression (6).

4.2. Effects on multiple regression coefficients

We noted in the introduction 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. We have looked at a large number of regression analyses for evidence of such biases, but space limitations dictate that we describe only two of them. The

*Table 2. Correlations of discrepancy scores with demographic characteristics**

	Age	Education	Marital status	Income	Race
Age of automobile	−.016 (.073)	.005 (.042)	−.038 (.082)	−.047 (.070)	.047 (.116)
Driver's license	−.048 (.025)	−.001 (.028)	−.020 (.053)	−.019 (.039)	.068 (.051)
Vote:					
in 1980	−.122 (.046)	.116 (.043)	−.133 (.065)	−.050 (.060)	.190 (.061)
in 1982	−.047 (.052)	.107 (.040)	.008 (.059)	.009 (.060)	.102 (.066)
in 1983	.014 (.045)	−.012 (.048)	−.049 (.068)	−.041 (.073)	.303 (.075)
Age:					
compared to Michigan driver's license or ID	.029 (.036)	.012 (.028)	.064 (.044)	.041 (.030)	−.053 (.069)
compared to voter registration rolls	.011 (.070)	−.080 (.051)	−.182 (.055)	−.082 (.041)	.064 (.084)
Distance to:					
drug store	.082 (.036)	−.038 (.031)	−.104 (.061)	−.038 (.081)	.130 (.099)
grocery store	.049 (.035)	.009 (.036)	.074 (.042)	.077 (.037)	−.013 (.048)
fire station	−.023 (.040)	−.038 (.051)	−.022 (.050)	−.024 (.051)	.076 (.089)
hospital	.041 (.034)	.004 (.044)	.002 (.043)	.057 (.054)	.025 (.041)
Proportion of neighbors:					
60 years or older	.133 (.041)	−.078 (.037)	−.001 (.041)	−.042 (.039)	.031 (.049)
black	−.093 (.048)	.033 (.048)	.060 (.047)	.059 (.058)	−.153 (.072)
family income over \$10 000	−.065 (.043)	.184 (.042)	.150 (.047)	.266 (.045)	−.075 (.060)
family income over \$30 000	−.135 (.046)	.150 (.043)	.114 (.052)	.286 (.053)	−.137 (.056)
Assesed value of house	−.085 (.060)	.087 (.042)	.046 (.054)	.224 (.123)	.038 (.064)
Property tax paid in 1983	−.050 (.044)	.107 (.070)	.057 (.041)	.238 (.077)	−.084 (.026)

The standard errors are shown in parentheses.
* 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	
	Coeff.	Standard error	Coeff.	Standard error
Constant	645.12		616.65	
Age	–7.54	9.18	–2.40	2.99
Income	6.52	2.18	2.90	0.86
Missing data on income	71.04	448.74	394.68	282.39
Marital status	–580.20	292.08	–185.51	179.60
Education	20.07	88.76	24.24	33.00
Race	–770.66	242.94	–509.44	147.18

* The predictor variables are defined as described in 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.

regression models are oversimplified both with respect to data analysis and to substantive considerations.[3] We could have selected other examples which show smaller consequences, but the ones shown are not particularly unusual cases. Moreover, for these examples we consider only measurement error as assessed with respect to the dependent variables, whereas more typically one might be concerned with errors in measured values of the predictor variables as well.

The dependent variable in the first 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, and as predicted by the set of five demographic variables shown in Table 2 (plus a dummy variable indicating missing data on family income; these cases are also recoded to the mean level on the income variable). There are substantial differences in the estimated regression coefficients depending on which report is used, as is shown in Table 3. The coefficient for income is twice as large if based on respondent reports rather than official values. Although this is the only predictor variable for which the coefficients in the two regressions differ at a statistically significant level,[4] all of the regression coefficients are noticeably different between the two columns. The estimated coefficient for

[3] Our justification for these simplifications is that our objective in this paper is *not* to estimate true causal parameters (i.e., the βs in a multiple-predictor version of expression (1)), but rather to assess the consequences of measurement error on estimates of population values (i.e., the *B*s corresponding to a multiple-predictor version of expression (2)) from a sample survey. In substantive analyses it is necessary to assume that the model has been properly specified, including the functional forms of the relationships among the dependent and predictor variables (e.g., their linearity and additivity) and that the sampled population corresponds to the target population, but for purposes of this paper we do not get into the issues that underlie such assumptions.

[4] To test whether the coefficients in the two columns are significantly different from one another, the LISREL computer program (Jöreskog and Sörbom (1984)) was used, constraining each pair of coefficients in turn to be equal. The test is approximate, since LISREL assumes a simple random sample design. An overall "design effect" was assumed, specifying the sample size to be only 127 instead of the actual 677 respondents from whom property tax reports were obtained, so that the average standard error for the regression coefficients as estimated by LISREL for the unconstrained model would be equal to the average standard error as estimated by REPERR. The statistical significance of the difference was evaluated by the chi-square goodness of fit statistic for the constrained model. For family income, the chi-square statistic is 28.4, with one degree of freedom, so is highly significant. For each of the other predictors in Table 3, constraining equality across the dependent variables did not produce a significant lack of fit; the chi-square statistic for each such constraint was less than 1.0.

Table 4. Logit regression coefficients: predicting voting behavior in 1980 from demographic characteristics

Predictor*	Respondent		Record	
	Coeff.	Standard error	Coeff.	Standard error
Constant	−4.934		−4.077	
Age	.050	.005	.050	.004
Income	.130	.056	.104	.045
Marital status	.455	.176	.883	.159
Education	.247	.035	.092	.028
Race	.768	.187	−.194	.164

* The predictor variables are defined as described in Table 2. Income is measured in hundreds of dollars.

marital status is three times larger when predicting the respondent reports than when predicting assessors' reports. As in previous tables, the standard errors were obtained using REPERR in order to take account of the sample design.

The dependent variable in the 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. (We used the DREG program in OSIRIS.IV for this analysis. The REPERR program does not handle logit regression, so the standard errors shown in parentheses in Table 4 assume a simple random sample and probably should be multiplied by a factor of about 2.) The coefficients in the first column are based on the respondents' own reports on whether or not they voted, while those in the third column are based on voting records. Again there are at least small differences in most of the coefficients, and three of the differences are substantial. If we use the official records, the estimated effect of marital status is considerably larger than if we were to use the respondent reports about voting. Education, on the other hand, is estimated to have a much smaller effect if based on the records instead of respondent reports. A similar pattern was reported by Silver et al. (1986), who found that several respondent characteristics, including education, which predict to voting behavior also predict to overreporting of voting behavior. The largest difference, however, 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. This difference in the estimated importance of race is consistent with previous reports which have found that blacks are more likely to overreport voting than whites (Katosh and Traugott (1981); Sigelman (1982)).

5. Intercorrelations of Measurement Errors on Different Variables

Table 5 shows correlations between discrepancy scores on different survey questions. Based upon previous work by Presser (1984) and preliminary inspection of the data, which suggest that intercorrelations may exist between measurement errors within but not between topic areas, correlations between pairs of discrepancy scores were averaged both within and between topic areas.

Our findings support those reported by Presser. Correlations within topic areas show stronger relationships than do those between topic areas, as indicated by the averages shown at the bottom of the table. Across all pairs of items, the average of the absolute values of the discrepancy scores is very small (.055). The average correlation between pairs of items within the same topic area, while small, is large enough to be of potential concern (.128), but the average correlation between items on different topics is not much higher than would be expected if all these discrepancies were independent. The entries in Table 5 are averages of the *absolute values* of the correlations. If a correlation between two variables has a population value of zero, the average observed (absolute) correlation in a sample of a thousand or so cases would be about .025 (based on simulated data generated from a wide range of distributions). The averages of the absolute values of the correlations between the discrepancy scores is .045, somewhat higher than the value expected by chance but not large enough to be a cause of great concern for most purposes. (The number of cases on which the correlations in Table 5 are based ranges from less than 700 to almost 1 500, with an average greater than 1 000, but the effective sample size, given the design effect due to clustering of respondents within households and of households within geographic areas, is probably smaller than 1 000.)

The sets of variables listed in the top half of Table 5 not only are in the same topic area, the survey measures in most of these sets also share a common response scale: the three voting items are measured on simple yes-no scales; the neighborhood distances all use a set of bracketed distance categories; the questions about neighbors all use a set of verbal labels to refer to proportions; and homestead value and property taxes both are asked in terms of dollars. It is possible, then, that much of the observed covariation within each of these sets reflects a "methods effect" – that is, a contribution to total item variance from the method used to obtain the data – rather than reflecting that the questions are in the same topic area. Andrews (1984), using estimates of method effects obtained from analysis of multimethod multi-trait data, found that across a collection of social psychological variables the average proportion of variance explained by methods effects was about two or three percent. If the observed correlations within Table 5 are due entirely to methods effects, these items would appear to be subject to stronger methods effects than most of the items observed by Andrews, whereas we would have expected methods effects for these factual and behavioral items to be smaller than those for attitudinal and other subjective items of the type examined by Andrews.

In addition to the general patterns, considerable variation exists for correlations within the different topic areas. Errors in reporting participation in one election are relati-

Table 5. Average correlations between discrepancy scores

Topic areas	n*	Average correlation
Voting behavior (3 items)	1 062	.265
Neighborhood distances (4 items)	1 402	.081
Neighborhood characteristics (4 items)	1 402	.081
Homestead value and prop. tax (2 items)	719	.022
Average correlation *within topic* areas		.128
Average correlation *between topic* areas		.045
Average correlation for all questions		.055

* The entries in this column are average numbers of cases on which correlations in each topic area are based.

vely strongly related to errors in reporting participation in another election. Discrepancies on distances to various neighborhood facilities are also somewhat related to one another. Discrepancies in reports on housing and on neighborhood composition are only weakly interrelated, but still at a higher level than errors in reports across different topic areas.

6. Summary and Conclusions

In this paper we have examined commonly made assumptions concerning the independence of measurement errors in answers by survey respondents relative to true values of substantive variables and to each other. We have operationalized measurement errors as the discrepancies between responses to survey questions and data from administrative records, census counts, and maps. We have observed that the independence assumptions are often incorrect. There are substantial correlations between the values of discrepancies in survey measures of several of the variables considered and the values of those variables as measured from records and other outside sources. Correlations of discrepancy scores on one variable with a set of demographic variables (taken as examples of possible predictor variables in regression models) are mostly small, but may nevertheless introduce substantial biases into estimates of regression coefficients for which the true values may also be small. The correlations of discrepancy scores on different variables with each other are also generally small, but not always negligible, particularly when the variables are within the same topic area.

It is commonly thought that measurement error introduces a negative bias into estimates of relationships among variables, and that correcting for attenuation compensates for this bias. This would indeed be the case if measurement errors were truly random – if, that is, the various assumptions about independence of measurement errors were correct. If any of those assumptions are incorrect, however, the implications may be complex and difficult to predict without explicit knowledge about the correlates of measurement error. For example, if there is a positive correlation between errors in measures of two variables, the observed covariance between those measures will tend to overestimate the relationship between the underlying variables (if that relationship is also positive). But if it is the correlation coefficient rather than the covariance that is considered, this positive bias is countered by the increase in variance in the two observed variables due to measurement error, which tends to *reduce* the observed correlation.

With respect to bivariate and multivariate regression coefficients, the effects of measurement error become even more complex and hard to predict. Measurement error in the dependent variable, as long as it is uncorrelated with any of the predictor variables, does not cause a bias in the estimates of regression coefficients but does increase the sampling errors of such estimates. If, however, it is not justified to assume that measurement error in the dependent variable is independent of the predictors, *and* that it is independent of measurement error in measures of those predictors, more complex estimation techniques (e.g., the use of instrumental variables and two stage least squares) are necessary to avoid possible biases. Measurement error with respect to the predictor in a bivariate regression model results in a downward bias with respect to the absolute value of the regression coefficient, even if that error is uncorrelated with any other component of the model. Measurement error with respect to any of the predictors in a multivariate model has complex effects on the estimates of the coefficients. If, moreover, measurement errors with respect to any of the predictors in a multivariate model are correlated with any of the true values, the regression coefficients will be biased, but in a direction and to an extent that cannot be stated without specific knowledge about the correlations of the measurement errors.

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. Recent books by Sudman and Bradburn (1982) and by Schuman and Presser (1981) summarize much of the research into question wording and sequencing of questions within interviews. With respect to interviewing techniques, Cannell and his colleagues (see Cannell et al. (1981)) have long been engaged in a research program which suggests the importance of such procedures as obtaining the explicit commitment of respondents to supplying accurate information and positive feedback from the interviewer when the respondent responds in appropriate fashion to questions. A promising recent development is the collaboration between cognitive psychologists and survey methodologists in exploring the effects of cognitive factors (comprehension of the question, retrieval of the desired information from memory, and so on) on the accuracy of responses in surveys, with the goal of improving not only our understanding of the cognitive processes involved, but also enhancing the accuracy of survey data (see Jabine et al. (1984); Fienberg et al. (1985)).

Another method of reducing measurement errors, and in particular correlated measurement errors, is to obtain important bits of information from external records rather than from survey reports. This has been one impetus for the creation of exact matches between administrative and survey data files (for examples, see U.S. Office of Federal Statistical Policy and Standards (1980); Jabine and Scheuren (1986)). Administrative records have their own sources of error, of course, but in many cases it is safe to assume that measurement errors from these sources have smaller variances, and smaller covariances with other substantive variables, than do survey measures. Nevertheless, it may be impractical to gain access to administrative records for respondents to a survey, either because of privacy laws or because of the lack of adequate identification variables (such as Social Security numbers) that permit exact matching. In any event, only a small proportion of the variables obtained through surveys is available from any type of external records.

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 answered in 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 of these problems, 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. Ideally, such respondents – especially those in the extreme categories – would be assigned dollar amounts in accordance with the distributions observed in the answers to the open-ended question. Where closed-ended scales cannot be avoided, our own findings underline the importance of avoiding scales with just two or three answer categories (cf. Johnson and Creech (1983)).

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. We have seen examples in which logarithmic and other tranformations 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.

7. References

Alwin, D.F. and Jackson, D.J. (1979): Measurement Models for Response Errors in Surveys: Issues and Applications. In Schuessler, K.F. (ed.), Sociological Methodology 1980. Jossey-Bass, San Francisco, pp. 68–119.

Andrews, F.M. (1984): The Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach. Public Opinion Quarterly, 48, pp. 409–442.

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

Cochran, W.G. (1968): Errors of Measurement in Statistics. Technometrics, 10, pp. 637–666.

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

Fienberg, S.E., Loftus, E.F., and Tanur, J.M. (1985): Cognitive Aspects of Health Survey Methodology: An Overview. Milbank Memorial Fund Quarterly/Health and Society, 63, pp. 547–564.

Herzog, A.R. and Dielman, L. (1985): Age Differences in Response Accuracy for Factual Survey Questions. Journal of Gerontology, 40, pp. 350–357.

Jabine, T.B., Straf, M.L., Tanur, J.M., and Tourangeau, R. (Eds.) (1984): Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. National Academy Press, Washington, D.C.

Jabine, T.B. and Scheuren, F.J. (1986): Record Linkages for Statistical Purposes: Methodological Issues. Journal of Official Statistics, 2, pp. 255–277.

Johnson, D.R. and Creech, J.C. (1983): Ordinal Measures in Multiple Indicator Models: A Simulation Study of Categorization Error. American Sociological Review, 48, pp. 398–407.

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

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.

Katosh, J.P. and Traugott, M.W. (1981): The Consequences of Validated and Self-reported Voting Measures. Public Opinion Quarterly, 45, pp. 519–535.

Kilss, B. and Alvey, W. (1976): Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972. American Statistical Association, Proceedings of the Social Statistics Section, pp. 471–476.

Kish, L. (1965): Survey Sampling. Wiley, New York.

Nunnally, J.C. (1967): Psychometric Theory. McGraw Hill, New York.

Presser, S. (1984): Is Inaccuracy on Factual Survey Items Item-Specific or Respondent-Specific? Public Opinion Quarterly, 48, pp. 344–355.

Radner, D.B. (1982): Distribution of Family Income: Improved Estimates. Social Security Bulletin, 45(7), pp. 13–21.

Rodgers, W.L. and Herzog, A.R. (1987): Interviewing the Elderly: The Accuracy of Factual Information. Journal of Gerontology, 42, pp. 387–394.

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

Sigelman, L. (1982): The Nonvoting Voter in Voting Research. American Journal of Political Science, 26, pp. 47–56.

Silver, B.D., Anderson, B.A., and Abramson, P.R. (1986): Who Overreports Voting? American Political Science Review, 80, pp. 613–624.

Sudman, S. and Bradburn, N.M. (1982): Asking Questions: A Practical Guide to Questionnaire Design. Jossey-Bass, San Francisco.

University of Michigan, Survey Research Center Computer Support Group (1982): OSIRIS.IV User's Manual. Institute for Social Research, Ann Arbor, MI.

U.S. Office of Federal Statistical Policy and Standards (1980): Report on Exact and Statistical Matching Techniques. Statistical Policy Working Paper 5.

Received May 1987
Revised December 1987