THE BIAS OF THE ORDINARY LEAST SQUARES ESTIMATOR OF THE REGRESSION COEFFICIENT FOR A BIVARIATE POPULATION WHEN BOTH VARIABLES ARE SUBJECT TO CORRELATED MEASUREMENT ERRORS*

John J. Chai, Syracuse University

1. INTRODUCTION

In surveys of socio-economic status or in scientific experiments, the influence of measurement errors always exists.

A stock boy taking an inventory of certain products at a specified time and day by counting the number of items for every product is expected to make counting errors. These counting errors may or may not be correlated between different units he counts on that day. The counting errors may be correlated between the units which are counted in a given day, if the stock boy counts the items using a certain method in the morning, say, and then he finishes counting the items using another method in the afternoon. Reporting on the number of family members by self enumeration of a respondent or by an interviewer produces another type of measurement errors, response errors. It has been well known from past studies that response errors of this kind are correlated within an interviewer assignment area due to interviewer bias in consecutive interviews. Even when a self-enumeration method (where no interviewer bias is involved) is used, we expect correlated measurement errors. For example, consider a survey of price of houses in a community. Suppose that people in the community are asked to assess their own homes. If a person in that community had just sold his house at a certain price level, the assessment of the other houses in the community may be affected [10].

Correlated measurement errors may also be expected in measuring the length or weight of an object, or in consecutive readings of fluctuating temperature of an instrument in a chemistry laboratory, or in grading student papers. For example, consider the grading of papers by an instructor. If the instructor grades one group of papers at one time, rests, and then grades the remainder of the papers at another time, we expect errors in grading to be correlated.

Theory of measurement errors in sample or census surveys for univariate case has been developed for some basic survey conditions in the recent past [1, 2, 4, 8, 9, 10, 13].

So far, we have illustrated the cases in which a single variable is taken separately. Now consider an example for a bivariate case where the characteristics of interest are the height and weight of a person. Measurement errors in this case may be caused by either instruments or by the person (or persons) measuring height and weight or by both. The measurement errors associated with height may be correlated with the errors associated with weight, positively or negatively. Furthermore, measurement errors for each variable may be correlated between the units within a set of observations. Therefore, if a person is interested in studying the relationship between two variables, he should recognize the existence of measurement errors and their effects on the estimators of relevant parameters. For example, suppose that we measure a set of bivariate characteristics X and Y and that the measurement is made with errors. Let the observed values (what we actually measure) be denoted by x and y and the measurement errors associated with X and Y by d and e respectively. Suppose further that

$$x = X + d \qquad (1.1)$$

$$y = Y + e \qquad (1.2)$$

and that the relationship between the two variables when neither variable is subject to measurement errors can be described by a linear functional model[1/]

$$Y = \alpha + \beta X \qquad (1.3)$$

However, the observed variables are x and y, so the model (1.3) above becomes

$$y = \alpha + \beta(x-d) + e \qquad (1.4)$$

by (1.1) and (1.2). And if we let

$$\omega = e - \beta d \qquad (1.5)$$

Then (1.4) becomes

$$y = \alpha + \beta x + \omega \qquad (1.6)$$

One may be interested in estimating β to see whether there exists a statistical relationship between X and Y from a set of sample observations. Suppose that the person tried the ordinary least squares estimator from a sample of n observations

$$b = \frac{\sum^{n}(x-\bar{x})(y-\bar{y})}{\sum^{n}(x-\bar{x})^2} \qquad (1.7)$$

where $\bar{x}$ and $\bar{y}$ are sample means of x and y, respectively. It is well known [12] that the estimator given by (1.7) is biased and inconsistent, i.e., Plim b = $\beta/(1+\sigma_d^2/\sigma_X^2)$. Thus b underestimates β unless $\sigma_d^2 = 0$, even when the errors d e are mutually and serially independent with constant variances, σ_d^2 and σ_e^2 and are also independent of X and Y.

There has been a considerable amount of theoretical work done in the past in developing better estimators (e.g., consistent estimators) of β when both variables are subject to errors (e.g., [3], [6],[7],[11],[12],[13],[15]). However, the ordinary least squares estimator is used more than often in practice, whether the variables concerned are subject to measurement errors or not. And there are many survey or experimental situations, where the ordinary assumptions mentioned earlier (the mutual and serial independence of errors, etc.) may not be satisfied fully. For example, d and e may be correlated with each other and d and e each may be serially correlated.

The purpose of our study is, therefore, to shed some light on the effect of measurement errors on the ordinary least squares estimator of β for the model given by (1.3), (1.4), and (1.6) for a large finite bivariate population when both variables are subject to correlated measurement errors for a large-scale sample survey situation.

In this paper, we present the following:

A. The mathematical development for the two-variable linear model [see equations (1.3) and (1.6)] to derive the bias factor of the ordinary least squares estimator of the parameter β, when both variables are subject to correlated measurement errors. This is presented in Section 2.

B. Estimators and estimates of the bias factor for some selected housing characteristics. This is presented in Section 3. We use two sets of data for calculation of the estimates. They are (1) a probability sample of about 5,000 housing units located in approximately 2,500 area segments of the United States. This sample was used for reinterview purposes by the Census Bureau after the 1960 Census of Housing in order to evaluate the accuracy of the statistics of housing characteristics. This project is known as the Content Evaluation Study (CES). The interviews were made in October 1960, six months after the census and the CES results were published in May, 1964. A detailed description of CES data is given by references [17]. (2) The second sample of housing units was drawn from six cities (Six-city data). This sample was chosen by the Census Bureau primarily for the purpose of evaluating the quality of housing conditions (e.g., sound, deteriorating, etc.) and methods of appraising the quality of housing conditions in 1964-65. Table 7 and reference [18] provide a detailed account of this sample.

C. Sensitivity analysis for the bias factor. This is also given in Section 3. For a set of hypothetical estimates of the parameters of the bias factor, sensitivity of the bias factor is examined.

2. THE MATHEMATICAL MODEL

We first define a set of survey conditions and assumptions for which the model is developed. Second, we derive the bias factor of the least squares estimator of the parameter β for the survey conditions and assumptions stated. Development of the model follows the work of Hansen, et.al.[9].

2.1 Survey Conditions

We use the term "survey" to mean census or sample surveys. This interpretation follows Hansen, et.al. [10].

A. Survey Conditions and Assumptions.

(1) We consider a large population of N elementary units, which is divided into M geographical areas (e.g., census tracts, enumeration districts, blocks, etc.). Each geographical area contains N_i elementary units, and thus

$$\sum_{i=1}^{M} N_i = N, \; i=1,\ 2,\ \ldots,\ M.$$

(2) We postulate a simple random sample of n elementary units yielding n_i units from i-th area. Thus

$$\sum_{i=1}^{M} n_i = n$$

(3) We assume that $n_i \doteq \bar{n}$ for all i, where $\bar{n} = \frac{n}{M}$, the average sample size per geographical area.

(4) Each of M interviewers is assigned at random to one area and so every interviewer is responsible for $\bar{n}$ units.

(5) The process of collecting data by interviewers is conducted in such a way that measurement error is correlated within interviewer assignment areas, but is uncorrelated between the different interviewer assignment areas.2/

(6) We assume that the survey can be repeated independently under the same survey conditions.3/

(7) We assume that the ratios $\frac{n}{N}$ and $\frac{\bar{n}}{N_i}$ are small enough to ignore the finite population multipliers (i.e., $1-\frac{n}{N}$ and $1-\frac{\bar{n}}{N_i}$).

(8) We further assume that

$$\bar{n} - 1 \doteq \bar{n}$$

$$N_i - 1 \doteq N_i \text{ for all } i$$

2.2 Development of the Model

In this section, we introduce the definitions and notations first.

A. Definitions and Notations

Let x_{ijt}, y_{ijt} be the observed values of the variables x and y for the j-th sample unit of the i-th geographical area, when measurement is obtained at the t-th trial. The conditional expected values of x and y given the j-th unit of the i-th area are, say,

$$\underset{t}{E}(x_{ijt}|i,j) = X_{ij} \quad (1)$$

$$\underset{t}{E}(y_{ijt}|i,j) = Y_{ij} \quad (2)$$

where the expectation is taken over trials. Following Hansen, et. al. [10], we define the response deviation for x and y variables given the j-th sample unit of the i-th geographical area as follows:

$$d_{ijt} = x_{ijt} - X_{ij} \quad (3)$$

$$e_{ijt} = y_{ijt} - Y_{ij} \quad (4)$$

We assume that each of the error terms, d and e, follows a probability distribution4/, and that the mean and variance of the distribution exist. Then, from equations (1) and (2), the conditional means, variances and covariance of d and e for a fixed j-th sample unit of the i-th geographical area as t varies are given by:

$$\underset{t}{E}(d_{ijt}|i,j) = 0 \quad (5)$$

$$\underset{t}{E}(e_{ijt}|i,j) = 0 \quad (6)$$

$$Var(d_{ijt}|i,j) = \underset{t}{E}(d^2_{ijt}|i,j) = \sigma^2_{d(ij)}, \text{ say} \quad (7)$$

$$Var(e_{ijt}|i,j) = \underset{t}{E}(e^2_{ijt}|i,j) = \sigma^2_{e(ij)}, \text{ say} \quad (8)$$

$$Cov(d_{ijt}e_{ijt}|i,j) = \underset{t}{E}(d_{ijt}e_{ijt}|i,j) = \sigma^2_{de(ij)}, \text{ say} \quad (9)$$

We further define the uncorrelated component of the response variance and covariance for a geographical area and for the entire population as follows[10].

$\frac{1}{N_i} \sum_j \sigma^2_{d(ij)} = \sigma^2_{d(i)}$, the simple variance for x for the i-th geographical area (10)

$\frac{1}{N_i} \sum_j \sigma^2_{e(ij)} = \sigma^2_{e(i)}$, the simple variance for y for the i-th geographical area (11)

$\frac{1}{N_i} \sum_j \sigma_{de(ij)} - \sigma_{de(i)}$, the simple response covariance for x and y for the i-th geographical area (12)

$\frac{1}{N} \sum_{i,j} \sigma^2_{d(ij)} = \sigma^2_e$, the simple response variance of the population for x (13)

$\frac{1}{N} \sum_{i,j} \sigma^2_{e(ij)} = \sigma^2_e$, the simple response variance of the population for y (14)

$\frac{1}{N} \sum_{i,j} \sigma_{de(ij)} = \sigma_{de}$, the simple response covariance of the population for x and y (15)

We next define the correlated component of the response variance and covariance in terms of the intra-class correlation coefficient $\delta_{d(i)}$ between d_{ijt} and $d_{ij't}$ for the i-th geographical area. The intraclass correlation coefficient of response deviations for the i-th geographical area. The intraclass correlation coefficient of response deviations for the i-th geographical area for x is defined by

$$\delta_{d(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E(d_{ijt} d_{ij't} | i,j,j')}{(N_i - 1) \sum_{j}^{N_i} \sum_t E\,(d^2_{ijt} | i,j)} \quad (16)$$

From (7) and (10), we obtain,

$$\delta_{d(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E\,(d_{ijt} d_{ij't} | i,j,j')}{(N_i)(N_i - 1)\, \sigma^2_{d(i)}} \quad (17)$$

hence

$$\delta_{d(i)} \sigma^2_{d(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E\,(d_{ijt} d_{ij't} | i,j,j')}{(N_i)(N_i - 1)} \quad (18)$$

We call the quantity given by equation (18) the correlated component of the response variance of x for the i-th geographical area. Similarly, the correlated component of the response variance of y for the i-th geographical area is

$$\delta_{e(i)} \sigma^2_{e(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E\,(e_{ijt} e_{ij't} | i,j,j')}{(N_i)(N_i - 1)} \quad (19)$$

The correlated component of the response covariance is defined similarly by defining the intraclass correlation of the response deviations for the i-th geographical area for x and y, i.e.,

$$\delta_{de(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E\,(d_{ijt} e_{ij't} | i,j,j')}{(N_i)(N_i - 1)\, \sigma_{d(i)} \sigma_{e(i)}} \quad (20)$$

By multiplying both sides of (2) above by $\sigma_{de(i)}\, \rho^{-1}_{de(i)}$, we obtain the correlated component of the response covariance for the i-th geographical area:

$$\delta_{de(i)} \sigma_{de(i)} \sigma^{-1}_{de(i)} = \frac{\sum_{j \neq j'}^{N_i} \sum_{t}^{N_i} E(d_{ijt} e_{ij't} | i,j,j')}{(N_i)(N_i - 1)} \quad (21)$$

where

$$\rho_{de} = \frac{\sigma_{de(i)}}{\sigma_{d(i)\ e(i)}} \quad (22)$$

is the correlation coefficient of d, e, for the i-th geographical area.

The average correlated component of the response variance per geographical area is defined by

$$\overline{\sigma_d^2\delta_d} = \frac{1}{n}\sum_i^M n_i\, \sigma^2_{d(i)}\delta_{d(i)} \tag{23}$$

$$\doteq \frac{1}{M}\sum_i^M \sigma^2_{d(i)}\delta_{d(i)} \tag{24}$$

since we assumed that $n_i \doteq \bar{n}$ for all i.

Similarly, the average correlated component of the response covariance per geographical area is

$$\overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}}$$

$$= \frac{1}{n}\sum_i^M n_i\, \sigma_{de(i)}\delta_{de(i)}\rho^{-1}_{de(i)} \tag{25}$$

$$\doteq \frac{1}{M}\sum_i \sigma_{de(i)}\delta_{de(i)}\rho^{-1}_{de(i)} \tag{26}$$

since $n_i \doteq \bar{n}$ for all i.

B. Bias of the Least Squares Estimator of β

As we stated earlier, our main objective is to study the effect of correlated measurement error on the ordinary least squares estimator in estimating β of the model given by (1.3), (1.4), and (1.6). For our survey situation we rewrite the model given by (1.3) and (1.6) by

$$Y_{ij} = \alpha + \beta X_{ij} \tag{27}$$

$$y_{ijt} = \alpha + \beta x_{ijt} + \omega_{ijt} \tag{28}$$

where α and β are parameters and the random variables X_{ij} and Y_{ij} are the conditional expected values of x_{ijt} and y_{ijt},

and $\omega_{ijt} = e_{ijt} - \beta d_{ijt}$ (29)

We observe, from a sample of n units, a set of values x_{ijt} and y_{ijt}; and estimate β using the ordinary least squares estimator b_t, which is defined by

$$b_t = \frac{\sum_i^M \sum_j^{N_i} (x_{ijt} - \bar{x}_t)(y_{ijt} - \bar{y}_t)}{\sum_i^M \sum_j^{N_i} (x_{ijt} - \bar{x}_t)^2}$$

$$= \frac{s_{xy(t)}}{s^2_{x(t)}} \tag{30}$$

where $\bar{x}_t$ and $\bar{y}_t$ are the sample means of x and y, respectively, for the t-th trial, i.e.,

$$\bar{x}_t = \frac{1}{n}\sum_i^M \sum_j^{n_i} x_{ijt} \tag{31}$$

$$\bar{y}_t = \frac{1}{n}\sum_i^M \sum_j^{n_i} y_{ijt} \tag{32}$$

And $s^2_{x(t)}$ and $s_{xy(t)}$ respectively are the sample variance and covariance for the t-th trial, i.e.,

$$s^2_{x(t)} = \frac{1}{n}\sum_i^M \sum_j^{n_i} (x_{ijt} - \bar{x}_t)^2 \tag{33}$$

$$s_{xy(t)}$$

$$= \frac{1}{n}\sum_i^M \sum_j^{n_i} (x_{ijt} - \bar{x}_t)(y_{ijt} - \bar{y}_t) \tag{34}$$

We are concerned with the bias of b_t. We derive the bias by taking expectation of b_t. The expectation is taken first over repeated trials for a fixed sample, and then over all possible samples. The ratio of expected values of the denominator and numerator of b_t is not necessarily equal to the expected value of b_t, since b_t is the ratio of two random variables. However, it is shown [4] that the differences between the expectation of the ratio and the ratio of expectation is small enough to ignore when the size of M is reasonably large for our survey conditions. Therefore, we evaluate the expectations of the denominator and numerator of b_t separately.

We find the expectation of the denominator of the b_t first.

$$\underset{st}{EE}\ s^2_{x(t)} = \underset{st}{EE}\frac{1}{n}\sum_i\sum_j(x_{ijt}-\bar{x}_t)^2$$

$$= \underset{st}{EE}\frac{1}{n}\sum_i\sum_j\left[(X_{ij}-\bar{X}) + (d_{ijt}-\bar{d}_t)\right]^2 \text{ by (3)} \quad (35)$$

where

$\bar{X} = \frac{1}{n}\sum_i^M\sum_j^{N_i} X_{ij}$, the sample mean of X_{ij}

$\bar{d}_t = \frac{1}{n}\sum_i^M\sum_j^{N_i} d_{ijt}$, the sample mean of d_{ijt} for the t-th trial.

For our Survey Condition, it can be shown that [4]

$$\underset{st}{EE}\ s^2_{x(t)} = \frac{n-1}{n}\left[\sigma^2_X+\sigma^2_d - \frac{1}{n(n-1)}\sum_i n_i\ (n_i-1)\sigma^2_{d(i)}\delta_{d(i)}\right] \quad (36)$$

$$\doteq \sigma^2_X+\sigma^2_d - \frac{1}{M^2}\sum_i \sigma^2{}_{d(i)}\delta_{d(i)}$$

$$= \sigma^2_X+\sigma^2_d - \frac{1}{M}\overline{\sigma^2_d\delta_d}$$

since we assumed that $n-1\doteq\bar{n}$, and

$\frac{1}{M}\sum \sigma^2_{d(i)}\delta_{d(i)} = \overline{\sigma^2_d\delta_d}$ by equation (24);

and where

$\sigma^2_X = \frac{1}{M}\sum_i^M\sum_j^{N_i}(X_{ij}-\bar{X}_{(p)})^2$, the variance of X_{ij}, (38)

$\bar{X}_{(p)} = \frac{1}{N}\sum_i^M\sum_j^{N_i} X_{ij}$, the population mean of X_{ij}, (39)

and σ^2_d and $\sigma^2_{d(i)}\delta_{d(i)}$ are defined by (13) and (18) respectively.

In a similar manner, we can show the expected value of the numerator of the b_t to be

$$\underset{st}{EE}\ s_{xy(t)} = \frac{n-1}{n}\left[\sigma_{XY}+\sigma_{de} - \frac{1}{n(n-1)}\sum_i n_i(n_i-1)\sigma_{de(i)}\delta_{de(i)}\rho^{-1}_{de(i)}\right]$$

$$\doteq \sigma_{XY}+\sigma_{de} - \frac{1}{M}\ \overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}} \quad (40)$$

where

$$\sigma_{XY} = \frac{1}{N}\sum_i^M\sum_j^{N_i} (X_{ij}-\bar{X}_{(p)})(Y_{ij}-\bar{Y}_{(p)}) \quad (41)$$

the covariance of X_{ij} and Y_{ij}; and σ_{de}, $\sigma_{de(i)}\delta_{de(i)}\rho^{-1}_{de(i)}$, and $\overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}}$ are given by (15), (21), and (26) respectively.

We denote the ratio of expected values of the numerator and denominator of b_t by β^*. Thus,

$$\beta^* = \frac{\underset{st}{EE}\ s_{xy(t)}}{\underset{st}{EE}\ s^2_{x(t)}} \doteq \frac{\sigma_{XY}+\sigma_{de} - \frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}}}{\sigma^2_X+\sigma^2_d - \frac{1}{M}\overline{\sigma^2_d\delta_d}} \quad (42)$$

For the sake of simplicity in writing, we define

$$\sigma^2_{d(T)} = \sigma^2_d - \frac{1}{M}\sum_i \overline{\sigma^2_d\delta_d} \quad (43)$$

$$\sigma_{de(T)} = \sigma_{de} - \frac{1}{M}\sum_i \overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}} \quad (44)$$

$$\sigma^2_{x(T)} = \sigma^2_X + \sigma^2_{d(T)} \quad (45)$$

$$\sigma_{xy(T)} = \sigma_{XY} + \sigma_{de(T)} \quad (46)$$

Using (43), (44), (45), and (46) above, we write

$$\beta^* = \frac{\sigma_{XY} + \sigma_{de(T)}}{\sigma^2_X + \sigma^2_{d(T)}} = \frac{\sigma_{XY(T)}}{\sigma^2_{X(T)}} \quad (47)$$

Factoring $\beta = \frac{\sigma_{XY}}{\sigma^2_X}$ out from (47) above, we have

$$\beta^* = (\beta)\,\frac{(1+\sigma_{de(T)}/\sigma_{XY})}{(1+\sigma^2_{d(T)}/\sigma^2_X)} \qquad (48)$$

The second factor of the right hand member of (48) above is defined to be the component bias factor of the b_t for our Survey Conditions. We denote it by E_1, i.e.,

$$E_1 = \frac{1+\sigma_{de(T)}/\sigma_{XY}}{1+\sigma^2_{d(T)}/\sigma^2_X} \qquad (49)$$

Hence, we have

$$\beta^* = (\beta)(E_1) \qquad (50)$$

or

$$\frac{\beta^*}{\beta} = E_1 \qquad (51)$$

or

$$\frac{(\beta^*-\beta)}{\beta} = (E_1-1) \qquad (52)$$

Assuming that M is so large that $\underset{st}{EE}\, b_t - \beta^* \doteq 0$ we can note from (49),(50), and (52) that $E_1 > 1$ indicates an over-estimation of β by b_t on the average; $E_1 < 1$ means an underestimation of β; and that when $E_1 = 1$, b_t is unbiased of β. And the bias factor E_1 is a function of uncorrelated components of response variance and covariance and the correlated components of response variance and covariance. In the following section, we estimate the bias factor E_1 from the two sets of sample data which we described earlier.

3. ESTIMATORS AND ESTIMATES

In this section, we discuss estimators and estimates of σ^2_d, σ_{de}, $\overline{\sigma^2_d\delta_d}$, and $\overline{\sigma_{de}\delta_{de}\rho^{-1}_{de}}$, first; and then we discuss estimators and estimates of $\sigma^2_{x(T)}$, $\sigma_{xy(T)}$, σ^2_d/σ^2_X, and σ_{de}/σ_{XY}; and finally we discuss the estimates of E_1.

3.1 The Estimators and Estimates of σ^2_d and σ_{de}.

As an estimator of σ^2_d Hansen, et.al. [9]more or less give

$$\frac{g}{2} = \frac{1}{2n}\sum_i^M \sum_j^{n_i} (x_{ijtG} - x_{ijt'G'})^2 \qquad (53)$$

where $g = \frac{1}{n}\sum_i \sum_j (x_{ijtG} - x_{ijt'G'})$ stands for "gross difference rate" [2,0], and t and t' respectively refer to t-th and t'th trials and G and G' respectively refer to G-th and G'th survey conditions. Following Hansen, et.al. [9] we can show that

$$\underset{st}{EE}\,\frac{g}{2} = \sigma^2_d \qquad (54)$$

if $\underset{s}{E}(x_{ijtG}) = \underset{s}{E}(x_{ijt'G'})$ and if repeated surveys are done independently so that the trial to trial covariance is zero. According to Bailer [2], the between-trial covariance is relatively small for the items she studied for a reinterview procedure for which the interviewers did not have access to the Census Data (original data) and reconciliation was not made after reinterview. The CES data we used in our study is obtained by the same interview procedure as the one just mentioned above, although the items she studied are not the same as the ones we studied. As for estimates of the bias $\underset{s}{E}(x_{ijtG} - x_{ijt'G})$, Bailer's study [2] did not seem to show any definite conclusion on the differences in estimates of the bias for different interview procedures. However, Bailer points out that, for a large sample, "a reinterview procedure which specifies that the reinterview be closer in time to the original interview" [P.60, Ref. 2] than the CES data (six months lag between original interview and reinterview) seem to have smaller bias.

In short, we are not sure about the magnitudes of the between-trial covariance and bias due to different survey procedures for the housing items included in our study. But the estimation of these terms are beyond the scope of our study. Instead, we assert that the assumptions and survey conditions stated at the outset hold so that $\frac{g}{2}$ is a good estimator of σ^2_d. Pritzker [16] (see also [2]) gives an estimator of σ_{de} by

$$\frac{h}{2} = \frac{1}{2n} \sum_{i}^{M} \sum_{j}^{n_i} (x_{ijtG}-x_{ijt'G'})(y_{ijtG}-y_{ijt'G'}) \quad (55)$$

where

$$h = \frac{1}{n} \sum_{i} \sum_{j} (x_{ijtG}-x_{ijt'G'})(y_{ijtG}-y_{ijt'G'})$$

Following Pritzker [16], we can show that

$$\underset{st}{EE} \frac{h}{2} = \sigma_{de} \quad (56)$$

if $\underset{s}{E} (x_{ijtG}) = \underset{s}{E} (x_{ijt'G'})$ and $\underset{s}{E} (y_{ijtG}) = \underset{s}{E} (y_{ijt'G'})$ and if independent repetitions of a survey are made.

Again, we say that the survey conditions we assumed hold and $\frac{h}{2}$ is a good estimator of σ_{de}.

The sample stimates of σ_d^2 for some selected housing characteristics are given in Table 3. The sample estimates of σ_{de} for some selected housing characteristics are also calculated from the two sets (Census results and CES Data) of sample data in 1967 for the first time. These estimates are given in Table 4.

3.2 Estimators and Estimates of $\overline{\sigma_d^2 \delta_d}$ and $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$.

The Response Variance Study [1] conducted by Bailer at the Census Bureau shows estimates of "ρ_d, the interclass correlation between response deviations of different units assigned to the same interviewer" (which is comparable to δ_d in our study) for population and housing characteristics. The estimates in that study were made on the basis of an interpenetrated sample design using an estimator similar to the one given below:

$$\frac{1}{\bar{n}} (s_{xbt}^2 - s_{xwt}^2) \quad (56)$$

where

$$s_{xbt}^2 = \frac{1}{M-1} \left[\bar{n} \sum (\bar{x}_{it} - \bar{x}_t)^2\right] \text{, between-}$$

interviewers variance (57)

$$\bar{x}_{it} = \frac{1}{\bar{n}} \sum_{j}^{\bar{n}} x_{ijt} \text{, the sample mean of x for}$$

the i-th geographical area

$$s_{xwt}^2 = \frac{1}{M(\bar{n}-1)} \sum_{ij}^{M\bar{n}} \sum (x_{ijt}-\bar{x}_{it})^2 \text{, within-}$$

interviewers variance (58)

According to the response variance study by Bailer of the Census Bureau [1] the magnitudes of the estimates of the average correlated response variance decreases as the interviewer assignment areas increase. In fact,the study concludes:

> "Though the rate of decrease is not constant over all population and housing characteristics, it is reasonable to assume that the relationship between ρ_d and the size of the assignment area is described by an exponential function, ..." (pp.3-4, [1]).

The Census Bureau study [1] did not estimate δ_{de}, the intra-class correlation for d and e. Furthermore, the estimates of the housing characteristics in which we are interested in our study were not included in the Census Bureau study. Therefore, we estimated the average correlated component of response variance and covariance for some selected housing characteristics using the Six-City Data.

For estimating $\overline{\sigma_d^2\delta_d}$ we used the following estimator:

$$\frac{\bar{n}}{\bar{n}-1} (s_{\bar{x}t}^2 - s_{xt}^2/\bar{n}) \quad (59)$$

and for estimating $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ we used

$$\frac{\bar{n}}{\bar{n}-1}(s_{\bar{x}\bar{y}t} - s_{xyt}/\bar{n}) \quad (60)$$

where

$$s_{\bar{x}t}^2 = \frac{\bar{n}}{n} \sum_{i}^{M} (\bar{x}_{it} - \bar{x}_t)^2 \text{, sample between-}$$

area variance (61)

$$s_{\bar{x}\bar{y}t} = \frac{\bar{n}}{n} \sum_{i}^{M} (\bar{x}_{it} - \bar{x}_t)(\bar{y}_{it}-\bar{y}_t) \text{, sample}$$

between-area covariance (62)

and s_{xt}^2 and s_{xyt} are given by (33) and (34).

The estimator given by (56) is the same as the estimator given by (59) if M-1=M. And also, it can be shown that the estimators given by (56) and (59) are unbiased estimators of $\overline{\sigma_d^2 \delta_d}$ if our assumptions and survey conditions hold. Similarly, the estimator (60) is an unbiased estimator of $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$.

The estimates of $\overline{\sigma_d^2 \delta_d}$ and $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ are given by Tables 1A, 1B, and 1C.

3.3 Estimates of $\sigma^2_{d(T)}$ and $\sigma_{de(T)}$.

Tables 2A, 2B, and 2C show the sample estimates of σ_d^2, σ_{de}, $\frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$. From these estimates we note that, although $\sigma^2_{d(T)} = \sigma_d^2 - \frac{1}{M}\overline{\sigma_d^2\delta_d}$ and $\sigma_{de(T)} = \sigma_{de} - \frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$, in most cases $\frac{1}{M}\overline{\sigma_d^2\delta_d}$ and $\frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ are negligible compared with σ_d^2 and σ_{de}. For example, almost all of the estimates of $\frac{1}{M}\overline{\sigma_d^2\delta_d}$ for block-sized areas is zero (see Table 2A) and the largest value for the ratio of $\frac{1}{M}\overline{\sigma_d^2\delta_d}$ to σ_d^2 is only about 0.1 (see Table 2A, Shreveport, "Bath for exclusive use", tract size). The largest value for the ratio of $\frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ to σ_{de}, however, is about 0.3 (see Table 2C, Shreveport, "Owner occupied units," tract size). Nevertheless, most of the estimates given in Tables 2B and 2C indicate that $\frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}/\sigma_{de}$ is very small.

Therefore, at least from these estimates, we may conclude that

$$\sigma^2_{d(T)} \doteq \sigma_d^2 \tag{63}$$

$$\sigma_{de(T)} \doteq \sigma_{de} \tag{64}$$

Hence, from (45), (46), and (49), we have

$$\sigma^2_{x(T)} \doteq \sigma_X^2 + \sigma_d^2 \tag{65}$$

$$\sigma_{xy(T)} \doteq \sigma_{XY} + \sigma_{de} \tag{66}$$

$$E_1 \doteq \frac{1+\sigma_{de}/\sigma_{XY}}{1+\sigma_d^2/\sigma_X^2} \tag{67}$$

Let the sample variance of x for the t-th trial be denoted by s^2_{xt} and that for the t'th trial by $s^2_{xt'}$, and let the average of the two variances be denoted by $s^2_{x(T)}$, i.e., $s^2_{x(T)} = \frac{1}{2}(s^2_{xt} + s^2_{xt'})$ (68)

$s^2_{x(T)}$ is the estimator of $\sigma^2_{x(T)}$.[5/] Hence, $g/2s^2_{x(T)}$ will estimate $\sigma_d^2/\sigma^2_{x(T)}$.

Assuming that $\frac{g}{2}$ is an unbiased estimator of σ_d^2 and that $\sigma^2_{x(T)} \doteq \sigma_d^2+\sigma_X^2$ we can use, as an estimator of $\sigma_X^2/\sigma^2_{x(T)}$, $1 - \frac{g}{2s^2_{x(T)}}$. And furthermore, as (69) an estimator of σ_d^2/σ_X, we use

$$\left(\frac{g}{2s^2_{x(T)}}\right)\left(1 - \frac{g}{2s^2_{x(T)}}\right)^{-1} \tag{70}$$

The estimates of σ_{de}/σ_{XY} are similarly obtained. The sample estimates of σ_d^2/σ_x^2 and σ_{de}/σ_{xy} are given in Tables 3 and 4.

Table 3 shows that the range of the estimates of σ_d^2/σ_x^2 for the given housing characteristics is .1099 to 3.0388.

Assuming that (63) and (64) hold, (i.e., $\frac{1}{M}\overline{\sigma_d^2\delta_d}$ and $\frac{1}{M}\sigma_{de}\delta_{de}\rho_{de}^{-1}$ are negligible), we can see that

$$0 \le \frac{\sigma_d^2}{\sigma^2_{x(T)}} \le 1 \tag{71}$$

and

$$0 \le \frac{\sigma_d^2}{\sigma_X^2} \le \infty \tag{72}$$

In other words, the larger the estimates of $\sigma_d^2/\sigma^2_{x(T)}$, the larger the estimates of σ_d^2/σ_X^2. In fact, if an estimate of $\sigma_d^2/\sigma^2_{x(T)}$ is close to 1 (meaning a highly inaccurate measurement) a corresponding value of σ_d^2/σ_X^2 is very large.

Similarly, we can see that

$$-\infty \le \frac{\sigma_{de}}{\sigma_{xy(T)}} \le \infty \qquad (73)$$

and

$$-1 \le \frac{\sigma_{de}}{\sigma_{XY}} \le \infty \qquad (74)$$

since σ_{de} and σ_{xy} do not necessarily take the same sign. The actual estimates of $\sigma_{de}/\sigma_{xy(T)}$ for the selected housing characteristics shown in Table 4 range from -.1106 to .5493 and the estimates of σ_{de}/σ_{XY} vary from -.0996 to 1.2188.

3.4 Estimates of E_1

We first discuss direct estimates of E_1, which are calculated from the estimates of σ_d^2/σ_X^2 in Table 3 and σ_{de}/σ_{XY} in Table 4, and then show the sensitivity of E_1.

(1) Direct estimates of E_1.

Table 5 shows the estimates of E_1 along with estimates of β^* and β. As seen in Table 5, we note that more than half of the estimates of E_1 are greater than one. In other words, for more than half of the pairs of housing variables in Table 5, σ_{de}/σ_{XY} is greater than σ_d^2/σ_X^2 in magnitude. This means that an important contribution is made to the effect of measurement error on the estimator of β by the simple response covariance, σ_{de}. It is wrong, therefore, to assume in all cases that $\sigma_{de} = 0$ and say that the effect upon the least squares estimator of β due to measurement error is "attenuating."

It is interesting to note that there is a certain consistency in the variation of E_1 over different variables. First, we notice that the value of E_1 is greater than one for the variables, "owner occupied units" and "units with bath for exclusive use," in all three dependent variables ("sound units," "deteriorating units," and "dilapidated units"). This leaves the value of E_1 for the other two variables ("renter occupied units" and "units with shared or no bath") to be less than one.

The magnitudes of E_1 range from .5680 to 1.5164, and the effect of measurement error on the estimator of β for the housing characteristics shown in Table 5 is clearly seen from the estimates of E_1 or from the comparison of the estimates β^* and β. We can see that the attenuation of β was largest for the pair of variables, "sound units" vs. "units with shared or no bath" (i.e., the estimate of β is -.9602 whereas the estimate of β^* is -.5454) and the magnitude of the overestimates of β was greatest for the pair of variables, "sound units" vs. "units with bath for exclusive use" (the estimate of β is .2194 and the estimate of β^* is .3327). For the pair of variables, "deteriorating units" vs. "renter occupied", there is almost no effect of measurement error on the estimator of β.

(2) Sensitivity analysis of the estimates of E_1.

So far we have investigated the sample estimates of E_1 for some housing characteristics. In this section, we will study the sensitivity of the magnitudes of E_1 for different values of $\sigma_d^2/\sigma_{x(T)}^2$ and $\sigma_{de}/\sigma_{xy(T)}$ based on the largest magnitude of the sample estimates given in Tables 3 and 4.

From what we have observed in Tables 3 and 4, .8 is the largest estimate for $\sigma_d^2/\sigma_{x(T)}^2$ and $\sigma_{de}/\sigma_{xy(T)}$. Therefore, we take 25, 50, 75, and 100 percent of .8 and add to these different values -.2 for the smallest value of $\sigma_{de}/\sigma_{xy(T)}$ and zero for $\sigma_d^2/\sigma_{x(T)}^2$. And then we compute different values of E_1 for these different values of $\sigma_d^2/\sigma_{x(T)}^2$ and $\sigma_{de}/\sigma_{xy(T)}$. The results of these computations are shown in Table 6 and Figure 1.

Examining the sensitivity of the values of E_1, we can clearly see from Figure 1 and Table 6 that, for a given value of $\sigma_d^2/\sigma_{x(T)}^2$ the values of E_1 increase with increasing rate as the values of $\sigma_{de}/\sigma_{xy(T)}$ increase by .2 from -.2 to .8. We also notice from Figure 1 that the rate of increase E_1 over the different values of $\sigma_{de}/\sigma_{xy(T)}$ gets smaller as the value of $\sigma_d^2/\sigma_{x(T)}^2$ increases. On the other hand, the rate of increase in the values of E_1 for the different values of $\sigma_d^2/\sigma_{x(T)}^2$ in Figure 1 is smaller for $\sigma_{de}/\sigma_{xy(T)} \leq .4$ than for $\sigma_{de}/\sigma_{xy(T)} > .4$.

Furthermore, we can see from Figure 1 and Table 6 that the values of E_1 will be greater than one (i.e., $E_1>1$) when $\sigma_{de}/\sigma_{xy(T)} > \sigma_d^2/\sigma_{x(T)}^2$, and $E_1 < 1$ when $\sigma_{de}/\sigma_{xy(T)}<\sigma_d^2/\sigma_{x(T)}^2$. And of course, $E_1=1$ when $\sigma_{de}/\sigma_{xy(T)}=\sigma_d^2/\sigma_{x(T)}^2$, and $E_1<1$ when $\sigma_{de}/\sigma_{xy(T)}<0$ for all values of $\sigma_d^2/\sigma_{x(T)}^2$.

As we noted earlier, $E_2<1$ means underestimation of β, $E_1>1$ means overestimation of β, and $E_1=1$ means that β is neither underestimated nor overestimated.

4. SUMMARY

For the survey conditions assumed and the variables investigated in our study, we find that the major contribution to the response variances and covariance is from the uncorrelated components of response variances and covariance.

The sample estimates of E_1 for some housing characteristics given in Table 5 reveal that we estimate that the mean value of b_t underestimates β by as much as 43 percent and overestimates β by as much as 57 percent.

Table 6 and Figure 1 show the sensitivity of the values of E_1. As seen in Figure 1 the values of E_1 are more sensitive to $\sigma_{de}/\sigma_{xy(T)} \geq .5$ and $0 \leq \sigma_d^2/\sigma_{x(T)}^2 \leq .6$ than to $-.2 \leq \sigma_{de}/\sigma_{xy(T)} < .5$ and $.6 < \sigma_d^2/\sigma_{x(T)}^2 < 1$.

FOOTNOTES

1/ There are situations in which this exact functional relation may not hold. For example, for some cases, the linear model with disturbance term ε, i.e., $Y=\alpha+\beta X+\varepsilon$. But since the model (1.3) is a fundamental one and often discussed, our study is based on this model.

2/ Of course, if the interviewers are supervised by different persons, then the measurement error between different interviewer assignment areas may be correlated.

3/ Independent repetitions of a survey under constant survey conditions can hardly be achieved in practice, but the postulate given here will be the basis for defining the variance and covariance of the measurement errors. See Reference [2] for a case of a dependent reinterview situation.

4/ We do not assume any particular form of probability distribution for d and e.

5/ The Census Bureau used an estimator similar to $s_{x(T)}^2$ (i.e., $\frac{1}{2}(p_1q_1+p_2q_2)$, where p_iq_i is the sample variance for "0,1" variable for sample of size one for the i-th trial. See [17] for detail.)

Table 1A.--Estimates of the Average Correlated Component of Response Variance for Some Housing Variables.

City and variable	Estimates of $\overline{\sigma_d^2 \delta_d}$			City and variable	Estimates of $\overline{\sigma_d^2 \delta_d}$		
	Block	ED	Tract		Block	ED	Tract
Camden, New Jersey:				Louisville, Kentucky:			
Owner occupied units	.0694	.0498	.0238	Owner occupied units	.0824	.0764	.0632
Deteriorating units .	.0428	.0328	.0207	Deteriorating units .	.0353	.0326	.0192
Dilapidated units . .	.0104	.0075	.0038	Dilapidated units . .	.0092	.0075	.0037
Substandard units . .	.0230	.0178	.0138	Substandard units . .	.0371	.0366	.0286
Bath for exclusive use	.0204	.0132	.0108	Bath for exclusive use	.0351	.0320	.0252
Units built 1939 or earlier.	.0958	.0665	.0373	Units built 1939 or earlier.	.1331	.1200	.0982
Monthly rent less than $80.	.0395	.0259	.0177	Monthly rent less than $80.	.1047	.0993	.0867
Cleveland, Ohio:				Shreveport, Louisiana:			
Owner occupied units	.0672	.0623	.0526	Owner occupied units	.0776	.0748	.0632
Deteriorating units .	.0415	.0379	.0188	Deteriorating units .	.0391	.0371	.0196
Dilapidated units . .	.0092	.0084	.0046	Dilapidated units . .	.0178	.0154	.0048
Substandard units . .	.0241	.0226	.0159	Substandard units . .	.0764	.0688	.0550
Bath for exclusive use	.0157	.0137	.0096	Bath for exclusive use	.0738	.0634	.0514
Units built 1939 or earlier.	.0737	.0681	.0547	Units built 1939 or earlier.	.1265	.1135	.0810
Monthly rent less than $80.	.0672	.0632	.0532	Monthly rent less than $80.	.1139	.1076	.0899
Fort Wayne, Indiana:				South Bend, Indiana:			
Owner occupied units	.0501	.0463	.0375	Owner occupied units	.0523	.0485	.0342
Deteriorating units	.0290	.0207	.0102	Deteriorating units	.0192	.0154	.0072
Dilapidated units . .	.0044	.0027	.0011	Dilapidated units . .	.0050	.0018	.0006
Substandard units . .	.0105	.0095	.0059	Substandard units . .	.0176	.0118	.0079
Bath for exclusive use	.0115	.0067	.0042	Bath for exclusive use	.0184	.0092	.0067
Units built 1939 or earlier.	.1282	.1149	.0969	Units built 1939 or earlier.	.1536	.1385	.1180
Monthly rent less than $80.	.0889	.0774	.0634	Monthly rent less than $80.	.0891	.0789	.0661

Source: Six-city data, Bureau of the Census
(See Table 7 of this article)

Table 1B.--Estimates of the Average Correlated Component of Response Covariance for Some Housing Variables.

("Units Deteriorating" as Dependent Variables.)

City and variable	Estimates of $\sigma_{de}\delta_{de}\rho_{de}^{-1}$			City and variable	Estimates of $\sigma_{de}\delta_{de}\rho_{de}^{-1}$		
	Block	ED	Tract		Block	ED	Tract
Camden, New Jersey:				Shreveport, Louisiana:			
Owner occupied units	-.0094	-.0090	-.0139	Owner occupied units	-.0288	-.0258	-.0218
Substandard units .	.0143	.0123	.0123	Substandard units .	.0390	.0341	.0283
Bath for exclusive use.	-.0125	-.0099	-.0106	Bath for exclusive use.	-.0361	-.0319	-.0267
Units built 1939 or earlier.	.0228	.0170	.0136	Units built 1939 or earlier.	.0258	.0242	.0171
				Monthly rent less than $80	.0457	.0414	.0348
Cleveland, Ohio:							
Owner occupied units	-.0208	-.0197	-.0189	South Bend, Indiana:			
Bath for exclusive use.	-.0099	-.0097	-.0077	Owner occupied units	-.0150	-.0133	-.0113
Units built 1939 or earlier.	.0147	.0141	.0128	Bath for exclusive use.	-.0084	-.0065	-.0053
Monthly rent less than $80	.0248	.0243	.0224	Monthly rent less than $80	.0244	.0201	.0166
Fort Wayne, Indiana:				Louisville, Kentucky:			
Owner occupied units	-.0190	-.0149	-.0125	Owner occupied units	-.0176	-.0186	-.0192
Substandard units	.0100	.0090	.0058	Substandard units	.0239	.0229	.0181
Bath for exclusive use.	-.0051	-.0054	-.0041	Bath for exclusive use.	-.0213	-.0205	-.0164
Units built 1939 or earlier.	.0208	.0200	.0172	Units built 1939 or earlier.	.0308	.0294	.0250
Monthly rent less than $80	.0308	.0251	.0220	Monthly rent less than $80	.0324	.0210	.0143

Source: Six-city data, Bureau of the Census
(See Table 7 of this article)

Table 1C.--Estimates of the Average Correlated Component of Response Covariance for Some Housing Variables.

("Units Dilapidated" as Dependent Variable.)

City and variable	Estimates of $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$		
	Block	ED	Tract
Camden, New Jersey:			
Owner occupied units	-.0032	-.0045	-.0038
Substandard units .	.0121	.0090	.0058
Bath for exclusive use.	-.0082	-.0067	-.0046
Cleveland, Ohio:			
Owner occupied units	-.0061	-.0061	-.0059
Substandard units .	.0103	.0096	.0061
Bath for exclusive use.	-.0028	-.0028	-.0025
Units built 1939 or earlier.	.0035	.0033	.0028
Monthly rent less than $80	.0074	.0072	.0072
Fort Wayne, Indiana:			
Owner occupied units	-.0046	-.0041	-.0031
Substandard units .	.0056	.0036	.0018
Bath for exclusive use.	-.0036	-.0020	-.0010
Monthly rent less than $80	.0068	.0064	.0054
Louisville, Kentucky:			
Owner occupied units	-.0073	-.0074	-.0076
Substandard units .	.0124	.0113	.0080
Bath for exclusive use.	-.0093	-.0088	-.0069
Units built 1939 or earlier.	.0087	.0084	.0072
Monthly rent less than $80	.0108	.0103	.0098
Shreveport, Indiana:			
Owner occupied units	-.0118	-.0112	-.0071
Substandard units .	.0249	.0206	.0134
Bath for exclusive use.	-.0207	-.0171	-.0123
Units built 1939 or earlier.	.0104	.0085	.0043
Monthly rent less than $80	.0197	.0170	.0134
South Bend, Indiana:			
Owner occupied units	-.0037	-.0030	-.0023
Substandard units .	.0060	.0034	.0017

Source: Six-city data, Bureau of the Census
(See Table 7 of this article)

Table 2A.--Estimates of the Simple Response Variance and the Average Correlated Component of Response Variance Divided by M.

Numbers in the parentheses show the approximate number of geographical areas (i.e., M).*

City and housing characteristics	Estimates of σ_d^2 1/	Estimates of $\frac{1}{M}\overline{\sigma_d^2\delta_d}$ 2/		
		Block	ED	Tract
Camden, New Jersey:		(1,100)	(100)	(30)
Owner occupied units. . . .	.0230	.0001	.0005	.0008
Bath for exclusive use. . .	.0124	.0000	.0001	.0001
Deteriorating	.0872	.0000	.0003	.0007
Dilapidated	.0298	.0000	.0001	.0004
Cleveland, Ohio:		(4,400)	(1,000)	(200)
Owner occupied units. . . .	.0230	.0000	.0001	.0003
Bath for exclusive use. . .	.0124	.0000	.0000	.0001
Deteriorating	.0872	.0000	.0000	.0001
Dilapidated	.0298	.0000	.0000	.0000
Fort Wayne, Indiana:		(2,100)	(200)	(40)
Owner occupied units. . . .	.0230	.0000	.0003	.0009
Bath for exclusive use. . .	.0124	.0000	.0000	.0001
Deteriorating	.0872	.0000	.0001	.0003
Dilapidated	.0298	.0000	.0000	.0000
Louisville, Kentucky:		(3,000)	(500)	(100)
Owner occupied units. . . .	.0230	.0000	.0002	.0006
Bath for exclusive use. . .	.0124	.0000	.0001	.0003
Deteriorating	.0872	.0000	.0001	.0002
Dilapidated	.0298	.0000	.0000	.0000
Shreveport, Louisiana:		(2,000)	(200)	(40)
Owner occupied units. . . .	.0230	.0000	.0004	.0016
Bath for exclusive use. . .	.0124	.0000	.0003	.0013
Deteriorating	.0872	.0000	.0002	.0005
Dilapidated	.0298	.0000	.0001	.0001
South Bend, Indiana:		(1,700)	(200)	(30)
Owner occupied units. . . .	.0230	.0000	.0002	.0011
Bath for exclusive use. . .	.0124	.0000	.0000	.0002
Deteriorating	.0872	.0000	.0000	.0002
Dilapidated	.0298	.0000	.0000	.0000

* See Table 7 for the values of M.
1/ The estimates shown in this table are transcribed from Table 3.
2/ See Table 1A for the estimates of $\overline{\sigma_d^2\delta_d}$.

Table 2B.--Estimates of the Simple Response Covariance and the Average Correlated Response Covariance Divided by M.

("Units Deteriorating" as Dependent Variable.)

Numbers in the parentheses show the approximate number of geographical areas (i.e., M).*

City and housing characteristics	Estimates of			
	σ_{de} 1/	$\frac{1}{M}\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ 2/		
		Block	ED	Tract
Camden, New Jersey:		(1,100)	(100)	(30)
Owner occupied units.	-.0027	.0000	-.0001	-.0005
Bath for exclusive use. . . .	-.0058	.0000	-.0001	-.0004
Cleveland, Ohio:		(4,400)	(1,000)	(200)
Owner occupied units.	-.0027	.0000	.0000	-.0001
Bath for exclusive use. . . .	-.0058	.0000	.0000	.0000
Fort Wayne, Indiana:		(2,100)	(200)	(40)
Owner occupied units.	-.0027	.0000	-.0001	-.0003
Bath for exclusive use. . . .	-.0058	.0000	.0000	-.0001
Louisville, Kentucky:		(3,000)	(500)	(100)
Owner occupied units.	-.0027	.0000	.0000	-.0002
Bath for exclusive use. . . .	-.0058	.0000	.0000	-.0002
Shreveport, Louisiana:		(2,000)	(200)	(40)
Owner occupied units.	-.0027	.0000	-.0001	-.0005
Bath for exclusive use. . . .	-.0058	.0000	-.0002	-.0006
South Bend, Indiana:		(1,700)	(200)	(30)
Owner occupied units.	-.0027	.0000	-.0001	-.0003
Bath for exclusive use. . . .	-.0058	.0000	.0000	-.0002

* See Table 7 for the values of M.
1/ The estimates shown in this table are transcribed from Table 4.
2/ See Table 1B for the estimates of $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$.

Table 2C.--Estimates of the Simple Response Covariance and the Average Correlated Response Covariance Divided by M.

("Units Dilapidated" as Dependent Variable.)

Numbers in the parentheses show the approximate number of geographical areas (i.e., M).*

City and housing characteristics	σ_{de} 1/	Estimates of $\frac{1}{M}\ \overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$ 2/		
		Block	ED	Tract
Camden, New Jersey:		(1,100)	(100)	(30)
Owner occupied units.	.0007	.0000	-.0000	-.0001
Bath for exclusive use. . . .	-.0045	.0000	-.0001	-.0002
Cleveland, Ohio:		(4,400)	(1,000)	(200)
Owner occupied units.	.0007	.0000	.0000	.0000
Bath for exclusive use. . . .	-.0045	.0000	.0000	.0000
Fort Wayne, Indiana:		(2,100)	(200)	(40)
Owner occupied units.	.0007	.0000	0000	-.0001
Bath for exclusive use. . . .	-.0045	.0000	.0000	.0000
Louisville, Kentucky:		(3,000)	(500)	(100)
Owner occupied units.	.0007	.0000	.0000	-.0001
Bath for exclusive use. . . .	-.0045	.0000	.0000	-.0001
Shreveport, Louisiana:		(2,000)	(200)	(40)
Owner occupied units.	.0007	.0000	-.0001	-.0002
Bath for exclusive use. . . .	-.0045	.0000	-.0001	-.0003
South Bend, Indiana:		(1,700)	(200)	(30)
Owner occupied units.	.0007	.0000	.0000	-.0001
Bath for exclusive use. . . .	-.0045	--	--	--

*See Table 7 for the values of M.

1/ The estimates shown in this table are transcribed from Table 4.

2/ See Table 1C for the estimates of $\overline{\sigma_{de}\delta_{de}\rho_{de}^{-1}}$.

-- Not Available

Table 3.--Sample Estimates of $\sigma_d^2, \sigma_{x(T)}^2$ and Ratios $\sigma_d^2/\sigma_{x(T)}^2$, $\sigma_X^2/\sigma_{x(T)}^2$, and σ_d^2/σ_X^2.

Housing Variables	Estimate of				
	σ_d^2	$\sigma_{x(T)}^2$	$\sigma_d^2/\sigma_{x(T)}^2$	$\sigma_X^2/\sigma_{x(T)}^2$	σ_d^2/σ_X^2
Owner occupied units	.0230	.2323	.0990	.9010	.1099
Renter occupied units.	.0230	.2323	.0990	.9010	.1099
Units with shared or no bath . . .	.0540	.1211	.4459	.5541	.8047
Units with bath for exclusive use	.0124	.1074	.1155	.8845	.1306
Sound.	.0783	.1530	.5117	.4883	1.0479
Deteriorating.	.0872	.1159	.7524	.2476	3.0388
Dilapidated.	.0298	.0520	.5731	.4269	1.3425

Source: Reference [17].

Table 4.--Sample Estimates of $\sigma_{de}, \sigma_{xy(T)}$, and Ratios, $\sigma_{de}/\sigma_{xy(T)}, \sigma_{XY}/\sigma_{xy(T)}$, and σ_{de}/σ_{XY}.

Housing variables (Dependent variable vs. independent variable)	Estimates of				
	σ_{de}	$\sigma_{xy(T)}$	σ_{de}/σ_{xyT}	σ_{XY}/σ_{xyT}	σ_{de}/σ_{XY}
Condition of Housing - Sound:					
vs. Owner occupied units	.0043	.01464	.2964	.7036	.4213
Renter occupied units.	.0014	-.03139	-.0446	1.0446	-.0427
Units with shared or no bath	-.0016	-.06600	.0244	.9756	.0250
Units with bath for exclusive use. . . .	.0149	.03573	.4167	.5833	.7144
Units with monthly rent less than $80. .	-.0003	-.02692	.0093	.9907	.0094
Units with monthly rent less than $60. .	.0004	-.02948	-.0136	1.0136	-.0134
Condition of Housing - Deteriorating:					
vs. Owner occupied units	-.0027	-.00497	.5493	.4507	1.2188
Renter occupied units.	.0014	.01662	.0818	.9182	.0891
Units with shared or no bath	.0010	.03275	.0302	.9698	.0311
Units with bath for exclusive use. . . .	-.0058	-.01973	.2955	.7045	.4194
Units with monthly rent less than $80. .	.0014	.01489	.0913	.9087	.1005
Units with monthly rent less than $60. .	.0007	.01538	.0481	.9519	.0505
Condition of Housing - Dilapidated:					
vs. Owner occupied units	.0007	-.00633	-.1106	1.1106	-.0996
Renter occupied units.	-.0010	.01476	-.0678	1.0678	-.0635
Units with shared or no bath	.0009	.03350	.0260	.9740	.0267
Units with bath for exclusive use. . . .	-.0045	-.01551	.2882	.7118	.4049
Units with monthly rent less than $80. .	.0004	.01253	.0295	.9705	.0304
Units with monthly rent less than $60. .	.0000	.01464	.0000	1.0000	.0000

Source: The sample of housing units used in the Content Evaluation Study of 1960 at the Bureau of the Census.

See reference [17] for further details.

Table 5.--Estimates of β, E_1 and β^*.

Housing variables (Dependent variable vs. independent variable)	Estimates of		
	β^*	E_1	β
	(1)	(2)	(3)=(1)÷(2)
Condition of Housing - Sound:			
vs. Owner occupied units	.0630	1.2806	.0492
Renter occupied units	-.1351	.8625	-.1566
Units with shared or no bath	-.5454	.5680	-.9602
Units with bath for exclusive use	.3327	1.5164	.2194
Condition of Housing - Deteriorating:			
vs. Owner occupied units	-.0214	1.0981	-.0195
Renter occupied units	.0715	.9813	.0729
Units with shared or no bath	.2704	.5713	.4733
Units with bath for exclusive use	-.1837	1.2554	-.1463
Condition of Housing - Dilapidated:			
vs. Owner occupied units	-.0272	.8112	-.0335
Renter occupied units	.0635	.8438	.0753
Units with shared or no bath	.2766	.5689	.4862
Units with bath for exclusive use	-.1444	1.2426	-.1162

Source:Tables 3 and 4.

Table 6.--The Values of E_1 for a Set of Values of the Ratio σ_d^2/σ_X^2 and σ_{de}/σ_{XY}.

(The number of the upper level in the column labels indicate $\sigma_d^2/\sigma_{x(T)}^2$ and the ones on the lower level indicate σ_d^2/σ_X^2).

$\frac{\sigma_{de}}{\sigma_{xy(T)}}$	$\frac{\sigma_{de}}{\sigma_{XY}}$	$\sigma_d^2/\sigma_{x(T)}^2$ and σ_d^2/σ_X^2					
		.0000	.2000	.4000	.5000	.6000	.8000
		.0000	.2500	.6667	1.0000	1.5000	4.0000
-.2000	-.1667	.8333	.666	.5000	.4167	.3333	.1667
.0000	.0000	1.0000	.8000	.6000	.5000	.4000	.2000
.2000	.2500	1.2500	1.0000	.7500	.6250	.5000	.2500
.4000	.6667	1.6667	1.3333	1.0000	.8333	.6667	.3333
.5000	1.0000	2.0000	1.6000	1.2000	1.0000	.8000	.4000
.6000	1.5000	2.5000	2.0000	1.5000	1.2500	1.0000	.5000
.8000	4.0000	5.0000	4.0000	3.0000	2.5000	2.0000	1.0000

Table 7.--Selected 1960 Census Data for Six Cities

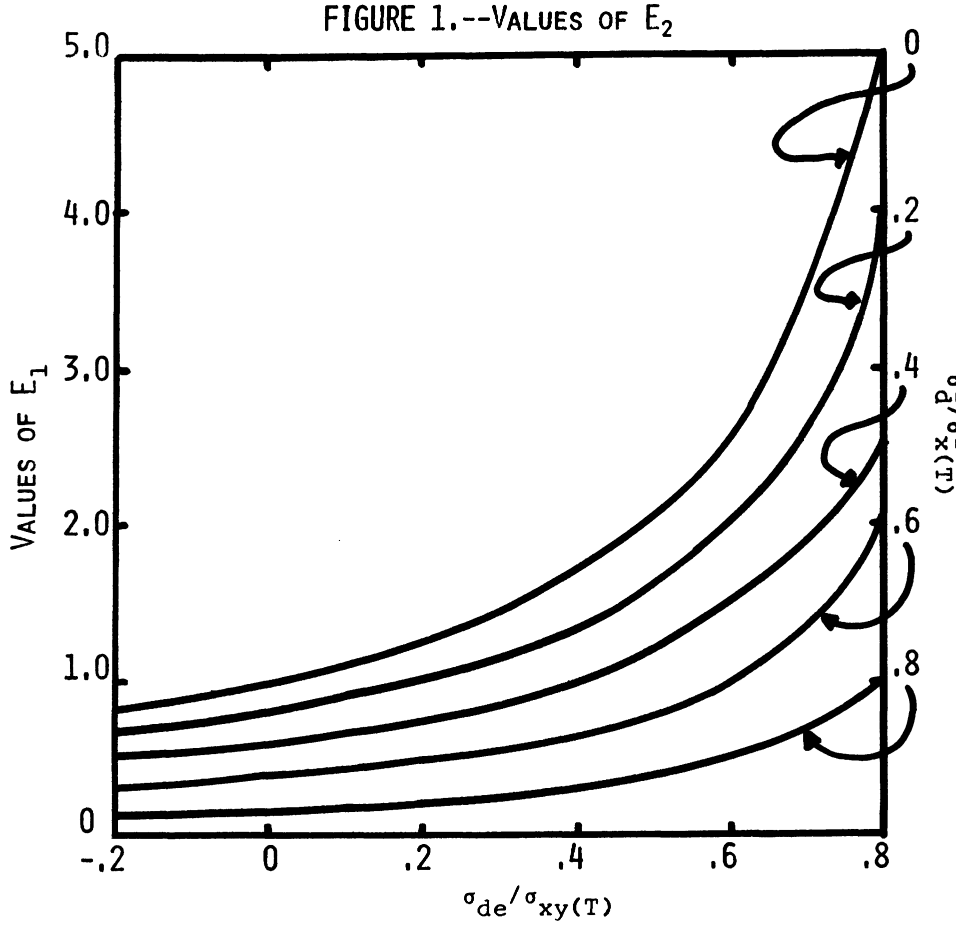
Subject	Camden, N.J.	Cleveland, Ohio	Ft. Wayne, Ind.	Louisville, Ky.	Shreveport, La.	South Bend, Ind.
1. Population	117,159	876,050	161,766	390,639	164,372	132,445
2. Housing units	37,015	282,893	53,002	128,238	54,191	42,590
3. Condition						
a. Percent deteriorating	15.3	14.1	11.8	14.7	16.4	10.5
b. Percent dilapidated	3.4	3.1	2.5	3.9	5.7	2.2
c. Percent substandard 1/	9.4	9.1	7.3	15.3	20.6	7.6
4. Area						
a. Census tracts 2/	27	203	39	111	40	34
b. Enumeration districts 2/	110	1,031	203	472	190	215
c. Blocks 2/	1,083	4,389	2,075	3,042	2,038	1,740

1/ Not an official Census classification. Used by other agencies. Includes "Sound" and "Deteriorating" units lacking one or more of these facilities; piped hot water, flush toilet for private use, bathtub or shower for private use, plus all "Dilapidated" units.

2/ Excludes areas in which there were no occupied housing units.

Source: p. 47 of Reference [18].

FIGURE 1.--VALUES OF E_2



REFERENCES

*The author gratefully acknowledges the useful comments given by Professor John Neter of the University of Minnesota, and Dr. Benjamin Tepping and Mrs. Barbara Bailer of the Bureau of the Census. Views expressed in this paper are those of the author and not necessarily those of the Bureau of the Census.

[1] Bailer, B.A., "The Effect of Varying the Size of the Interviewer's Assignment Area on ρ_d, the Interclass Correlation Between Response Deviations of Different Units Assigned to the Same Interviewer." Unpublished article, Center for Research in Measurement Methods, U.S. Bureau of the Census, September 28, 1967.

[2] Bailer, B.A., "Recent Research in Reinterview Procedures." Journal of the American Statistical Association, 63 March 1968.

[3] Berkson, Joseph, "Are there two regressions?", J. of Amer. Stat. Assoc., 45 (1950), pp. 164-180.

[4] Chai, John J., A Study of Effects of Correlated Measurement Errors and Grouping on the Ordinary Least Squares Estimator for the Regression Coefficient and Ordinary Estimator of the Product-Moment Correlation Coefficient of A Finite Bivariate Population. Ph.D. Thesis. University of Minnesota, 1968.

[5] Cochran, W.G., Sampling Techniques, Second Ed. New York: John Wiley and Sons, Inc., 1963.

[6] Durbin, J., "Errors in Variables," Rev. of Inter. Stat. Inst., 22 (1954), pp. 22-23.

[7] Goldberger, A.S., Econometric Theory, New York: John Wiley and Sons, Inc. 1964.

[8] Hansen, M.H., Hurwitz, W.N., and Madow, W.G., Sample Survey Methods and Theory, Vols. I and II. New York: John Wiley and Sons, Inc. (1953).

[9] Hansen, M.H., Hurwitz, W.N., and Pritzker, Leon, "The estimation and interpretation of gross differences and the simple response variance," Contributions to Statistics Presented to Professor P.S. Mahalanobis on the Occasion of His 70th Birthday. Calcutta, India: Pergamon Press(1964)

[10] Hansen, M.H., Hurwitz, W.N., and Bershad, M.A., "Measurement errors in censuses and surveys," Bulletin of the Inter. Stat. Inst., 38 (1961), pp. 359-374.

[11] Jessop, W.N., "One line or two?", Applied Stat., 2(1952),pp.131-317.

[12] Johnston, J., Econometric Methods. New York: McGraw-Hill Book Co., Inc. 1963.

[13] Mahalanobis, P.C., "Recent Experiments in Statistical Sampling in the Indian Statistical Institute", J. of the Royal Stat. Soc., 109(1946) pp. 18-55.

[14] Mandansky, A. "The fitting of straight lines when both variables are subject to error," J. Roy.Stat. Soc., 54(1959), pp. 173-205.

[15] Nair, K.R. and Banerjee, K.S., "A note on fitting of straight lines if both variables are subject to error," Sankhyā, Vol. 5(1943), p.331.

[16] Pritzker, Leon, "Response Response Research Branch Report No. 66", April 27, 1966, U.S. Bureau of the Census.

[17] U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Housing Characteristics, Series ER 60, No. 3. Dept. of Commerce, Bureau of the Census, Washington, D.C., (1964).

[18] U.S. Bureau of the Census, Measuring the Quality of Housing, An Appraisal of Census Statistics and Methods, Working Paper No. 25, Department of Commerce, Bureau of the Census, Washington, D.C. 1967.