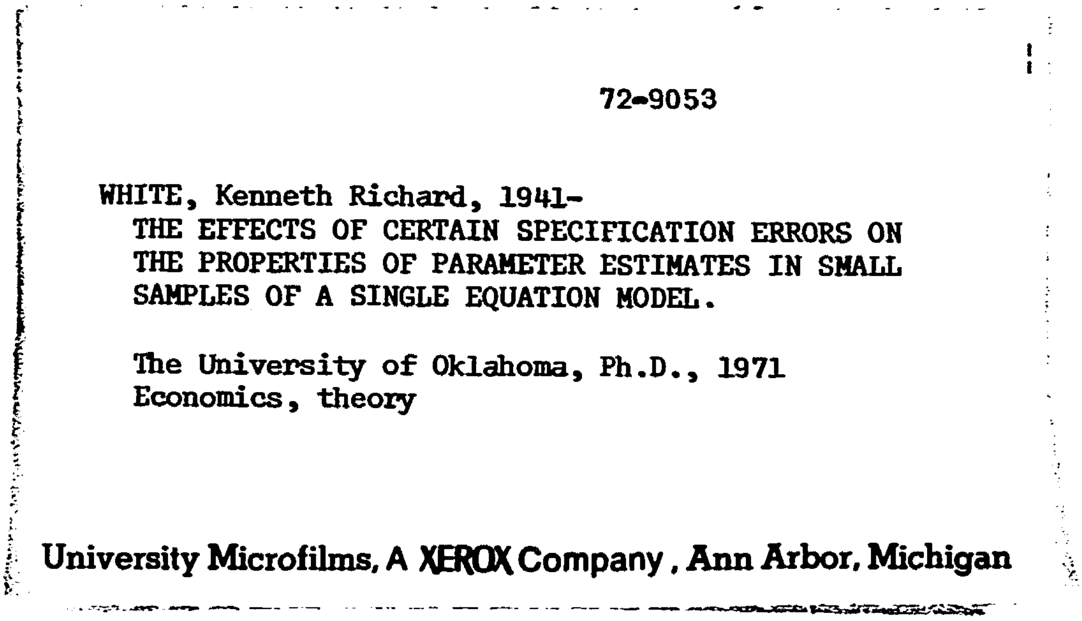


WHITE, Kenneth Richard, 1941-
THE EFFECTS OF CERTAIN SPECIFICATION ERRORS ON THE PROPERTIES OF PARAMETER ESTIMATES IN SMALL SAMPLES OF A SINGLE EQUATION MODEL.

The University of Oklahoma, Ph.D., 1971
Economics, theory

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE EFFECTS OF CERTAIN SPECIFICATION ERRORS ON THE PROPERTIES OF PARAMETER ESTIMATES IN SMALL SAMPLES OF A SINGLE EQUATION MODEL

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

KENNETH RICHARD WHITE

Norman, Oklahoma

1971

THE EFFECTS OF CERTAIN SPECIFICATION ERRORS ON THE PROPERTIES OF PARAMETER ESTIMATES IN SMALL SAMPLES OF A SINGLE EQUATION MODEL

APPROVED BY

Ed J. Brice Jr.

James E. Hibdon

Jack L. Robinson

A. J. Kondonassis

Oliver Benson

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I am indebted to many people for the successful completion of this Monte Carlo analysis. First, I would like to thank Dr. D. F. Crim, Jr., whose socratic method of teaching Econometrics first instilled an interest of mine in the field. It was he who ably guided my progress and pointed out flaws in earlier revisions. I also wish to thank the members of my committee, Dr. Oliver E. Benson, Dr. James E. Hibdon, Dr. Alexander J. Kondonassis, and Dr. Jack L. Robinson, for their constructive criticism and suggestions.

Various individuals cooperated to make this investigation possible. While the length of this list of contributors precludes naming them all, special recognition is due Mr. Y. Boo, computer analyst for the Bureau of Business and Economic Research, who helped design the computer program necessary for this study, and Dr. Frank L. Salzman, Department of Mathematics, Florida Technological University, who gave needed assistance in several of the proofs contained in this research paper.

The author is particularly grateful to his wife, Nancy, whose patience and encouragement made the end goal seem worthwhile.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

THE EFFECTS OF CERTAIN SPECIFICATION ERRORS ON THE PROPERTIES OF PARAMETER ESTIMATES IN SMALL SAMPLES OF A SINGLE EQUATION MODEL

CHAPTER I

INTRODUCTION - A FRAMEWORK WITHIN WHICH TO WORK

Econometrics

Econometrics is a science which deals with the problems of model-building and forecasting and ". . . may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference."[1] "Econometrics, the result of a certain outlook on the role of economics, consists of the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economics and to obtain numerical results."[2]

[1] Paul A. Samuelson, Tjalling C. Koopmans, and J. Richard N. Stone, "Report of the Evaluative Committee for Econometrics," Econometrica, XXII (April, 1954), 141.

[2] Gerhard Tintner, Methodology of Mathematical Economics and Econometrics, International Encyclopedia of Unified Science, II, No. 6 (Chicago: University of Chicago Press, 1968), p. 74.

"The main objective of econometrics is to give empirical content to a priori reasoning in economics."[3] The objective has two primary purposes: in the measurement of parameters which would promote a better understanding of true economic relationships and in the ability to use these parameters to be able to forecast future values of economic variables. In the area of the forecasting ability of parameters, "Perhaps more than anything else, we want equations that can forecast the future. For this purpose 'the future' should be interpreted to include anything unknown to the forecaster when he did his work; thus a person might 'forecast' some aspect of nineteenth century behavior by means of theory and data derived solely from the twentieth century. The latter sort of forecasting, that is, of temporally past data, can be useful for testing theories, but of course practical interest centers on forecasting the temporal future."[4]

The Method of Least Squares

Population Regression Function

This paper is primarily concerned with the analysis of two variable single equation model using the method of least squares. The method of least squares adheres to the principles

[3] Arthur S. Goldberger, Econometric Theory, (New York John Wiley & Sons, Inc., 1964), p. 1.
Lawrence R. Klein, An Introduction to Econometrics, (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962), p. 1.

[4] Carl F. Christ, Econometric Models and Methods, (New York: John Wiley & Sons, Inc., 1966), p. 5.

of statistical inference as its goal is to make correct inferences from a sample as to the true function form existing between certain variables. For example, suppose economists wanted to analyze the two variables consumption and income. Economists theorize that certainly these two variables must be related in some form. It might be logical to conclude that current spending is some function of current income. Using Y to represent consumption and X to represent income, then

$$Y_i = f\ (X_i) \qquad (1\text{-}1)$$

Since Y is some function of X, Y is considered the dependent variable, dependent on a particular value of income, (X_i). Further, theory may have specified the exact form in which Y_i is a function of X_i, and that this exact form is a linear one. Then

$$Y_i = a + bX_i \qquad (1\text{-}2)$$

It should not be expected that each individual within a defined income group would spend exactly the same amount as another individual in that same population. As a consequence, a stochastic element would have to be added to Equation 1-2 yielding

$$Y_i = a + bX_i + u_i \qquad (1\text{-}3)$$

where u denotes a stochastic variable which may take on positive or negative values. If each value of X_i represents a distinct population, then the mean value of Y_i would be $a + bX_i$, but the actual values of Y_i would be $a + bX_i + u_i$

for an individual unit within the population. Because regression models deal with such a variety of populations (one for each value of X_i), certain assumptions are made concerning the conditional probabilities of the Y_i values. These are that the values of Y_i are random variables statistically independent of each other and that the means of the Y_i distributions lie directly on the regression line. It is assumed that the density functions of Y_i, for each value of X_i, have a constant variance.

Usually the deviations of each Y_i from its expected value are described in terms of the stochastic error so that

$$Y_i = a + bX_i + u_i$$

where the u_i are independent random variables with a mean equal to zero and a constant variance. The distribution of u is just the distribution of Y transformed to a zero mean, and is pictured in Figure 1. However, no exact form of the density function need be specified.

A brief summary of the assumptions made concerning the error terms is appropriate. First,it was assumed that

$$E(u_i) = 0 \qquad (1\text{-}4)$$

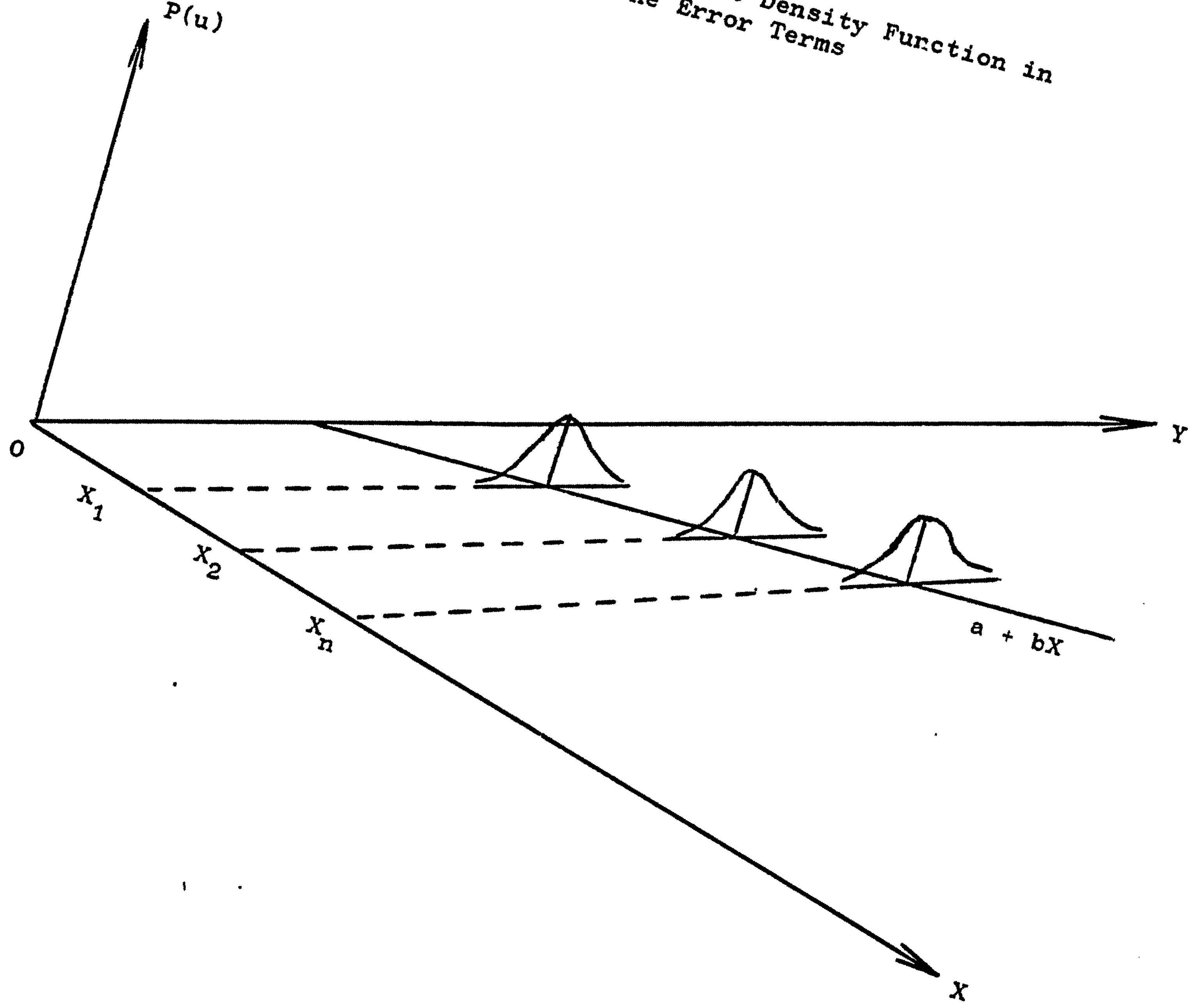
Second, it was assumed that Y_i is a random variable; therefore,

$$E(u_i u_j) = 0 \text{ when } i \neq j \qquad (1\text{-}5)$$

Third, the density functions of u_i is constant,given a particular value for the independent variable.

$$E(u_i)^2 = \sigma_u^2 \qquad (1\text{-}6)$$

Fig. 1.--A Constant Density Function in the Error Terms



Another assumption associated with the error term is that

$$E(X_i u_i) = 0 \qquad (1\text{-}7)$$

The actual insertion of the error term in Equation 1-3 may be justified on the basis of several considerations. First, u_i is simply an unpredictable element in human responses. Since there is no systematic component in the element, it should be expected to have a zero mean. For example, in the consumption function there are many factors which alter the spending habits of different families making the same income. These factors cause some families to spend more than the average of their income group and others to spend less. It might be expected that those who spend more will counteract the families who spend less. Second, it is possible that certain variables have been excluded from the original equation, and third, it is possible that there has been an error in observing the Y_i values. The influence of the latter two possibilities enter the error terms as deviations from their means, hence they always have a zero mean.[5]

When one of the assumptions of Equations 1-4 to 1-7 is violated a specification error occurs. For example, if

$$E(u_i u_j) \neq 0 \text{ when } i \neq j$$

then the error terms are autocorrelated and a specification error exists. An incorrect economic theory in this analysis

[5] See Potluri Rao and Roger LeRoy Miller, Applied Econometrics, (Belmont, Calif.: Wadsworth Publishing Company, Inc., 1971), pp. 7-8.

is not called a specification error.

In the preceeding analysis it has been assumed that X_i is fixed. When X is a random variable the analysis does not break down if the conditional means $E(Y_i/X_i)$ are sought in relation to a given value of X_i. With X_i random, the statistical procedures appropriate for fixed values of X_i remain valid provided the same assumption as outlined in Equations 1-4 to 1-7 hold.[6]

Sample Regression Function

The analysis just completed is associated with the population linear regression model. Unfortunately, except for synthetic studies, a population's properties and parameters usuallyare unknown and have to be estimated from available data. The goal of a sample regression function should be to estimate accurately the parameters of the parent population. To accomplish this goal the assumptions made for the parent population regarding Y_i, X_i, and u_i should hold in the sample data.

The sample regression equation is

$$Y_i = \hat{a} + \hat{b}X_i + e_i \tag{1-8}$$

where e_i represents an observable residual term, and $\hat{a}$ and $\hat{b}$ represent estimates of a and b. Rewriting Equation 1-8 yields

$$e_i = Y_i - \hat{a} - \hat{b}X_i \tag{1-9}$$

in which it is desired that

[6] See Dennis J. Aigner, Basic Econometrics, (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1971), pp. 16-17.

$$\Sigma(e_i) = 0$$

Unfortunately, there are many estimates of a and b that will satisfy this condition. Therefore, it is necessary to specify a second condition, namely that the sum of the deviations of the residuals squared must be a minimum. Thus

$$\Sigma e_i^2 = \Sigma(Y_i - \hat{a} - \hat{b}X_i)^2 \qquad (1\text{-}10)$$

is to be minimized.

To minimize a function it must be set equal to zero. From Equation 1-10

$$(\partial/\partial\hat{a})\,\Sigma e_i^2 = -2\,\Sigma(Y_i - \hat{a} - \hat{b}X_i) \qquad (1\text{-}11)$$

$$(\partial/\partial\hat{b})\,\Sigma e_i^2 = -2\Sigma X_i(Y_i - \hat{a} - \hat{b}X_i) \qquad (1\text{-}12)$$

Substituting Equation 1-9 into 1-12

$$(\partial/\partial\hat{b})\,\Sigma e_i^2 = -2\,\Sigma X_i e_i \qquad (1\text{-}13)$$

which when set equal to zero

$$E(X_i e_i) = 0$$

The correlation between the independent variable and the residuals is equal to zero.

When Equations 1-11 and 1-12 are set equal to zero the result is the "normal equations"

$$\Sigma Y_i = a\,n + b\,\Sigma X_i$$

$$\Sigma X_i Y_i = a\,\Sigma X_i + b\,\Sigma X_i^2$$

which, when solved, yield

$$\hat{b} = \Sigma xy / \Sigma x^2 \qquad (1\text{-}14)$$

$$\hat{a} = \overline{Y} - \hat{b}\,\overline{X} \qquad (1\text{-}15)$$

where lower case letters represent deviations from the mean.

Violations of Assumptions

A Correlation between X and u

If a correlation exists between X and u, then from Equation 1-14

$$\hat{b} = \Sigma x(Y - \bar{Y}) / \Sigma x^2$$

$$= (\Sigma xY / \Sigma x^2) - (\bar{Y} \Sigma x / \Sigma x^2)$$

$$= \Sigma xY / \Sigma x^2$$

and since $Y = a + bX + u$

$$= \Sigma x(a + bX + u) / \Sigma x^2$$

$$= (a \Sigma x + b \Sigma xX + \Sigma xu) / \Sigma x^2$$

$$= b + \Sigma xu / \Sigma x^2 \qquad (1\text{-}16)$$

since $\Sigma x = 0$, and $\Sigma xX = \Sigma x^2$.

If a correlation exists between X and u then $\Sigma xu \neq 0$ causing $E(\hat{b}) \neq b$. The estimate of b is biased. Since

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

and $\hat{b}$ is biased, $\hat{a}$ will also be biased. Because of the influence the regression coefficient b exerts in a single-equation analysis, in comparison to a, subsequent proofs are to be limited to the estimation of b.

A $\Sigma xu \neq 0$ will not only have an effect on bias but also on consistency. As the sample size increases to infinity a consistent estimator will provide a perfect point estimate of the true parameter. The parameter $\hat{b}$ is consistent if

$$E(\hat{b}_i - b)^2 \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

Consistency is measured in terms of the mean-square error (MSE),

$$MSE = E(\hat{b}_i - b)^2 \qquad (1\text{-}17)$$

It must follow that

$$E(\hat{b}_i - b)^2 = E(\hat{b}_i - \bar{\hat{b}})^2 + (\bar{\hat{b}} - b)^2 \qquad (1\text{-}18)$$

or that the MSE is equal to the variance of the estimator plus the square of its bias. Consistency, then, requires that an estimator's bias and variance both approach zero as the sample size increases to infinity. It is observable from Equation 1-18 ". . . that a biased estimator may thus show a smaller mean-square error than an unbiased one if it more than compensates for its bias by having a smaller variance."[8]

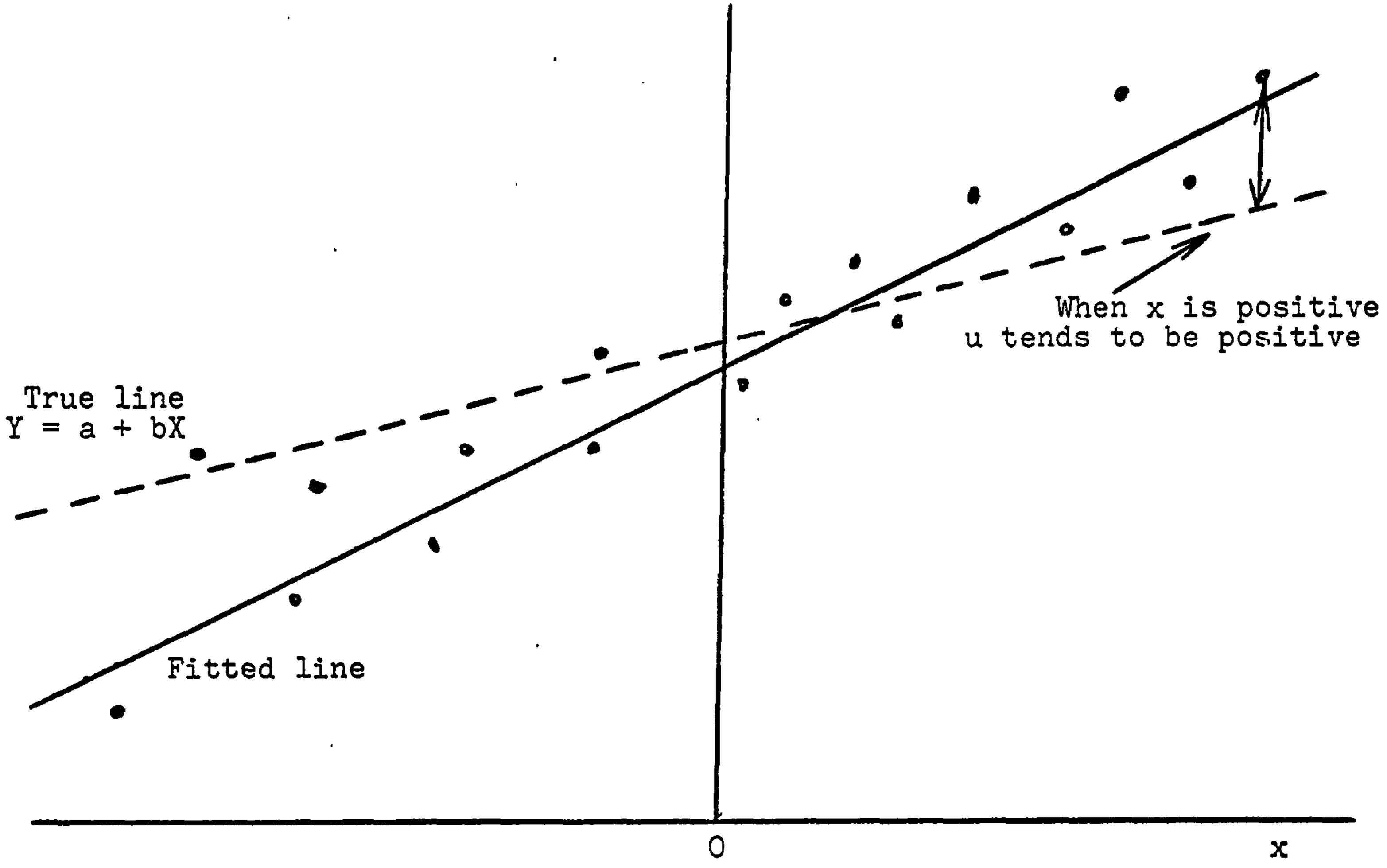
Often the square roots of the variance and MSE of an estimator are used, in which case they are termed the standard deviation and the root mean-square error (RMSE).

A correlation between X and u will cause an estimator to be inconsistent. The inconsistency can be seen in Figure 2. In Figure 2, X and u are positively correlated. The positive values of x are associated with positive values of u, and negative values of x are associated with negative values of u. The consequences of a positive correlation between X and u is that least squares produces a slope with an upwards bias and an intercept with a downward bias. The bias in b in Figure 2 will persist even for an infinitely large sample. The reason for the bias is that when X and u are correlated, some of the effects of u are wrongly attributed to X.[9]

[8] J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company, Inc., 1963), p. 277.

[9] See Ronald J. Wonnacott and Thomas H. Wonnacott, Econometrics, (New York: John Wiley & Sons, Inc., 1970), pp. 152-153.

Fig. 2.--How Correlation of X and u Makes b Biased and Inconsistent



In summary, if the covariance between Xu $\neq 0$, then biased and inconsistent estimators may result. The assumption of $E(Xu) = 0$ will be satisfied if either X is fixed or if X is a random variable distributed independently of u. The question still remains why the covariance between X and u is not equal to zero.

A dependence between the independent variable and the error term can be caused by errors of observation. The effects of such errors are as follows:

$$X' = X + u \tag{1-19}$$

$$Y' = Y + v \tag{1-20}$$

$$Y = a + bX \tag{1-21}$$

where X' and Y' are the observed values of X and Y, and u and v are the errors of observation. It follows that

$$\begin{aligned} Y' &= a + bX + v \qquad (1\text{-}22) \\ &= a + b(X' - u) + v \\ &= a + bX' + (v - bu) \qquad (1\text{-}23) \end{aligned}$$

The covariance of X' and (v - bu) is

$$\begin{aligned} E\left[(X' - X)(v - bu)\right] &= E\left[u(v - vu)\right] \qquad (1\text{-}24) \\ &= E(uv) - b\,E(u^2) \\ &= -b \text{ Var } (u) \qquad (1\text{-}25) \end{aligned}$$

Since the covariance is not equal to zero, a dependence exists between the error term (v - bu) and the explanatory variable

X'.[10]

If there were an observation error in X but not in Y, the results would be identical. Again assuming Equations 1-19 and 1-21, and that Y = Y', then

$$\text{Y or Y'} = a + bX \qquad (1\text{-}26)$$

$$= a + b(X' - u)$$

$$= a + bX' - bu \qquad (1\text{-}27)$$

The covariance of X' and (-bu) is

$$E\left[(X' - X)(-bu)\right] = E\left[u(-bu)\right]$$

$$= E(u) - b\,E(u^2)$$

$$= -b\,\text{Var}\,(u) \qquad (1\text{-}28)$$

Errors in observation are not the only causes of bias. Lagged variables may also cause a correlation between the independent variable and the error term. In the equation

$$Y_t = a + bY_{t-1} + u_t \qquad (1\text{-}29)$$

the independent variable is neither fixed nor randomly

[10] For a complete analysis of errors in observation, see Abraham Wald, "The Fitting of Straight Lines if Both Variables Are Subject to Error," Annals of Mathematical Statistics, XI (1940), 284-300.

M.S. Bartlett, "The Fitting of Straight Lines When Both Variables Are Subject to Error," Biometrics, V (1949), 207-212.

Albert Madansky, "The Fitting of Straight Lines When Both Variables Are Subject to Error," Journal of the American Statistical Association, LIV (1959), 173-205.

Max Halperin, "Fitting of Straight Lines and Prediction When Both Variables Are Subject to Error," Journal of the American Statistical Association, LVI (September, 1961), 657-669.

distributed independently of Y_t and, hence, u_t. Therefore, a correlation exists between the independent variable and the error term.[11]

Following the same analysis as lagged variables, in the simultaneous equation model of

$$Y = a + bX + u \tag{1-30}$$

$$X = Y + Z \tag{1-31}$$

the covariance of X and u is not equal to zero. When u takes on a large value, Y becomes greater as a consequence of Equation 1-30; when Y is greater, X is greater (as illustrated in Equation 1-31), since Z is fixed; thus u and X are positively correlated.[12]

Another cause of bias might be the incorrect inference of the functional form existing between variables. If a linear function were used when a non-linear function was called for, there might exist a correlation between the independent variable and the error term.

Heteroscedasticity

If the conditional variances of the error terms are identical, as depicted in Figure 1, the error terms are said

[11] For the mathematical proof, see J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company, Inc., 1963), pp. 219-225.

[12] For a complete analysis of simultaneous equation models see E. Malinvaud, Statistical Methods of Econometrics, (Chicago: Rand, McNally & Co., 1966), pp. 497-613.

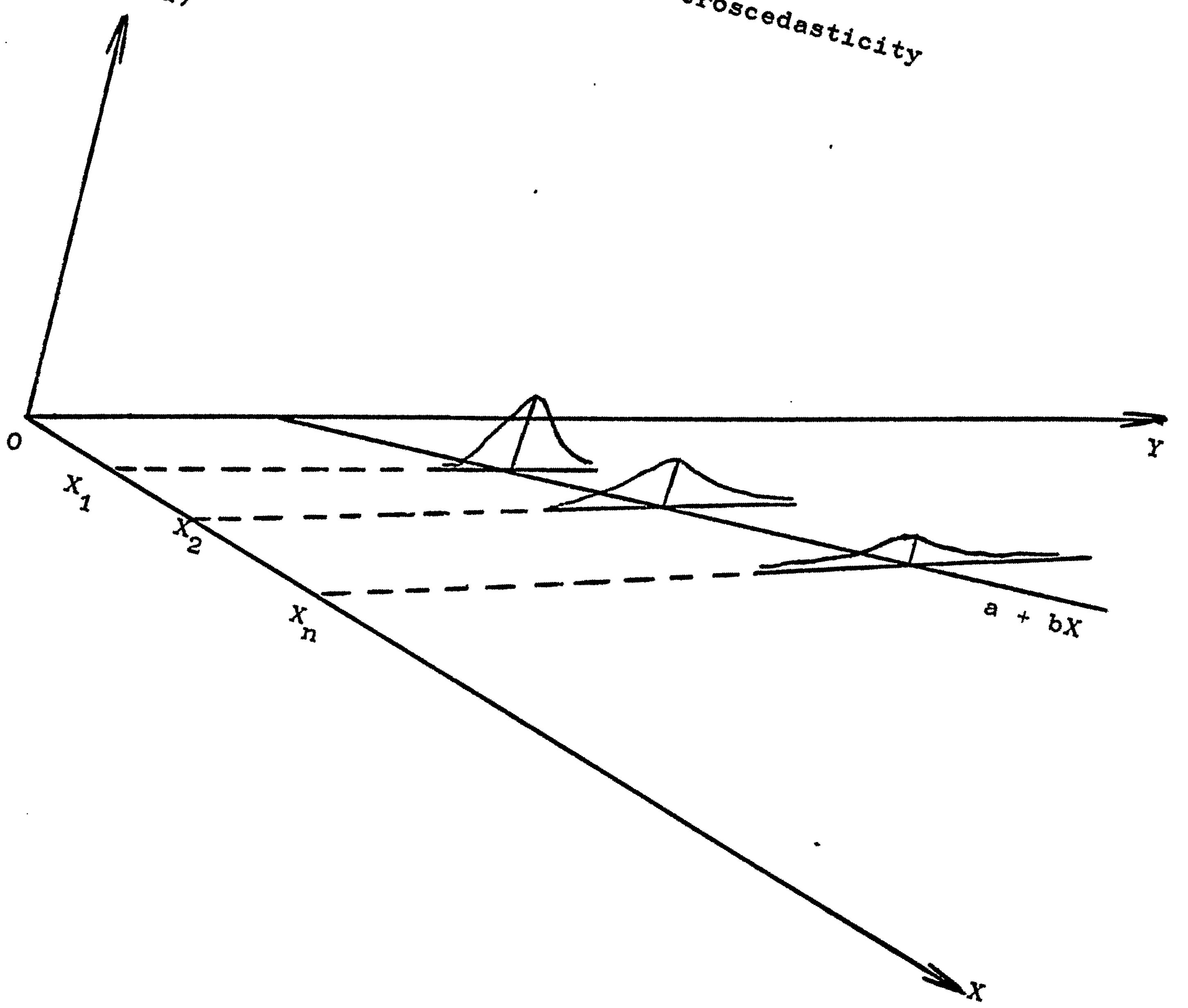
to be homoscedastic. When they appear as they do in Figure 3, the error terms are heteroscedastic. There can exist various forms of the probability density function of the error terms. The calculus involved in minimizing the density functions is awkward to develop, so assumptions usually are made as to the nature of the form. The most common assumption made is that the standard deviation of the error terms increases proportionally for each value of the independent variable.

The assumption that the density function of the error term is proportional to the independent variable is justified on the basis of empirical evidence concerning budgetary data. For example, in the consumption function there is empirical evidence to suggest that the variance of spending habits would be quite different between low- and high-income individuals. Among the high-income individuals, the spending pattern would vary greatly, but among the low-income individuals, a majority of their income would be spent on daily necessities.[13]

The effect of heteroscedasticity clearly cannot be on bias, because, from Equation 1-16, only the covariance between X and u controls the bias, which heteroscedasticity does not effect. The unbiasedness property of least squares is unaffected by the presence of interdependent disturbances in general. "From the point of view of estimation and hypothesis

[13] See David S. Huang, Regression and Econometric Methods, (New York: John Wiley & Sons, Inc., 1970), p. 146.

Fig. 3.--Heteroscedasticity



testing, the main effect of heteroscedasticity . . . is not on bias or consistency but on efficiency."[14] One unbiased estimator is said to be more efficient than another if its variance is smaller. From Figure 3, the standard error of an estimator would become unusually large because the density function of the error term changes for each value of X_i. In equation form,

$$\text{Var }(\hat{b}) = E(\hat{b} - b)^2 \qquad (1\text{-}32)$$

from Equation 1-16

$$= E\ (\Sigma xu/\Sigma x^2)^2$$

Assuming fixed X values, then Σx^2 is a constant, and $E(\Sigma xu/\ \Sigma x^2)^2$ becomes

$$= (1/\Sigma x^2)^2\ E(x_1{}^2u_1{}^2 + \ .\ .\ .\ + x_n{}^2u_1{}^2 + 2x_1x_2u_1u_2$$

$$.\ .\ .\ + 2x_{n-1}x_nu_{n-1}u_n) \qquad (1\text{-}33)$$

If the assumptions of Equations 1-4, 1-5, and 1-6 hold, then

$$E(x_i{}^2u_i{}^2) = x_i{}^2\sigma_u{}^2 \qquad (1\text{-}34)$$

and

$$\text{Var }(\hat{b}) = \sigma_u{}^2\Sigma x^2/(\Sigma x^2)^2$$

$$= \sigma_u{}^2/\Sigma x^2 \qquad (1\text{-}35)$$

If heteroscedasticity is present, then the complicated Equation 1-33 cannot be reduced to Equation 1-35.

Forecasting

Most econometric models are derived for their ability to forecast. Estimators having the properties of being unbiased, efficient, and consistent play a large role in a researcher's confidence in a given forecast. It is from the

[14] Edward J. Kane, Economic Statistics and Econometrics: An Introduction to Quantitative Economics, (New York: Harper & Row, Publishers, 1968), p. 363.

estimated parameters and their respective distributions that forecasts are to be made. If the specification error of a correlation between the independent variable and the error term is present, projections into the future could lead to incorrect conclusions. If projections were carried to future values of X_i in Figure 2, the forecasts would be much larger than the corresponding true values.

Heteroscedasticity also leads to difficulties when forecasting. The variance of the estimators will be so large as to yield interval estimates of the parameters and subsequent forecasts which will be unduly large.

CHAPTER II

THE MONTE CARLO METHOD

Introduction

Chapter 1 reviewed the individual effects of two specification errors in a single-equation model. The two specifications reviewed were those of a correlation between the independent variable and the error term, and heteroscedasticity. The individual effects of a correlation between the independent variable and the error term are that biased and inconsistent estimators may result. Heteroscedasticity yields estimates which have an unusually large variance, causing inefficiency. The individual effects of each specification error are known a priori.

It is not known a priori what the joint effects of the two specification errors are. "A striking weakness of the current state of econometrics is that the joint result of several complications cannot be inferred as the sum of their separate results."[1] It is the goal of this research paper to analyze what the combined effects of a correlation

[1] J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company, Inc., 1963), p. 216.

between the independent variable and the error term and heteroscedasticity may be.

The Monte Carlo Technique

The problem of using real data

To test the joint effects of two specification errors it is necessary to be able to isolate and distinguish them from the effects of other complications. Real data for which parameter values are generally unknown usually provide no basis for such an analysis. If a specification error is suspected in a model, a test for its presence usually is based on the assumption that other specification errors do not exist.

For example, the one test to determine the presence of autocorrelated error terms, the Durbin-Watson test, assumes normally distributed and homoscedastic error terms.[2] The most widely accepted test for the presence of heteroscedasticity, the Goldfeld-Quandt test, assumes that the error terms are non-autocorrelated.[3] It is clear that the individual presence of heteroscedasticity and autocorrelation cannot be known with certainty.

[2] J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. 1," Biometrika, XXXVII (December, 1950), 409-428.

J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. 11," Biometrika, XXXVIII (June, 1951), 159-178.

[3] Stephen M. Goldfeld and Richard E. Quandt, "Some Tests for Homoscedasticity," Journal of the American Statistical Association, LX (June, 1965), 539-547.

In a consumption-income analysis it might be expected that the variance of the error term is some function of income. If this is in fact true, then it might seem to follow that a correlation between the error term and the independent variable exists. For, if in the case of heteroscedasticity

$$\text{Var}\ (u) = f\ (X) \tag{2-1}$$

where X represents income, then

$$u = f\ (X) \tag{2-2}$$

The question that arises is which specification error caused the other, or are they both caused separately?

The purpose, then, of a Monte Carlo analysis would be to attempt to isolate the effects of the specification errors, individually and jointly. Such studies might suggest techniques of estimation which could be applied to real economic models.

Methodology of a Monte Carlo Analysis

The Monte Carlo method, with which so many econometric studies have been made, has cast a great deal of light on the small sample properties of estimators. The basic format of the method is to evaluate a given model by designing a synthetic system which matches it. In the real world economists usually have to make decisions based on a limited amount of data, especially in time-series analyses. Therefore, Monte Carlo models usually deal with small samples.

In Monte Carlo analyses econometricians first specify the model to be analyzed. A careful choice of the variables involved in the model must be made and the exact functional form existing between them determined. All properties, including the mean, variance, and covariance of the variables and the error terms, are specified. Such specifications provide the researcher with population parameters and the probability density functions involved, a priori.

Having specified the properties of the population, the researcher takes samples from it and compares the estimates, with their density functions, to the actual values. A relatively large number of samples, each of small size, is used to gain information about the estimates and their density functions. A relatively large number of samples is necessary to insure correct inferences and to allow analyses to be made of their density functions.

In a Monte Carlo analysis there can be one or many models. Usually, models with and without specification errors, or with and without a second specification error, are compared to attempt to determine the particular effects a certain specification error or errors has on a particular model. Once the model has been specified appropriately for the task for which it was designed, the data must be generated. The data must be generated in such a manner that they contain certain properties, which are chosen by the researcher.

For a point of reference in the generation of synthetic data, econometricians use random numbers. The random numbers may contain a mean equal to zero and a finite variance to simulate the population error terms. From a matrix of random numbers with such properties, transformations of the matrix can be made by the premultiplication of another matrix. This matrix could be the population variance-covariance matrix. How the transformation is to take place depends upon what the researcher is trying to accomplish. For example, if a correlation between different independent variables is desired (multicollinearity), the random number matrix could be premultiplied by the population variance-covariance matrix, for which the covariance between the independent variables would not equal zero. From the parameters, error terms, and independent variables, the dependent variables can be generated and estimated parameters and density functions calculated and compared to the true structural parameters specified in the beginning.

In generating independent variables from a random number matrix the assumption is made that the independent variables are not fixed. The assumption that the independent variables are themselves random is justified on the grounds that any economic model is merely a part of some larger true structure of an economic system. Thus, a predetermined variable can only be considered as such when theorizing that a particular model is complete.[4] Certainly, the predetermined

[4] See Tjalling Koopmans, "Statistical Estimation of Simultaneous Economic Relations," Journal of the American Statistical Association, XL (December, 1945), 448-466.

variables of a small model may well be the dependent variable in a larger model, in which case they will not be fixed.

By a similar analysis the effects on forecasting can be analyzed. Dependent variables generated from the structural parameters (true values) can be compared to dependent variables generated from the estimated parameters (estimated values). Point and interval estimates can be compared and analyzed. Monte Carlo analyses have not been as effective in the area of forecasting as they have been on the small sample properties of estimators. "At the present time the number of computer simulation studies that can claim even a modicum of success in predicting the behavior of some economic system are meager indeed."[5] The reason for this inadequacy in forecasting is because computer simulation techniques are based on probability theory, not truth.

Monte Carlo Analyses

Econometric literature is abundant with Monte Carlo analyses. From several of these studies this researcher received valuable ideas in terms of model specification and data generation. A few significant articles will be reviewed to analyze some of the accomplishments made using the Monte Carlo technique.

W. A. Neiswanger and T. A. Yancey attempted to determine how a correlation between the predetermined variables

[5] Thomas H. Naylor, Joseph L. Balintfy, Donald S. Burdick, and Kong Chu, Computer Simulation Techniques, (New York: John Wiley & Sons, Inc., 1966), pp. 318-319.

and the error terms, and between different predetermined variables (multicollinearity), in a simultaneous equation model, affected parameters estimated from least squares and limited information single equation methods.[6] They compared models with and without a secular trend in the dependent variables not explained by the predetermined variables and the parameters of the structural equations. The authors defined this trend as autonomous growth.

The authors used 120 samples of 25 for each of their models, which included one model with a time component and one without. All of the structural parameters, error terms, and predetermined variables were specified, in addition to the specification of their respective properties. The means, variances, covariances, and correlations of the X's and u's were all specified.

The X's and u's, with their specified properties, were generated from a matrix of random numbers. The elements in the matrix had an expected value equal to zero and a variance of one. Through a transformation of the population variance-covariance matrix, the authors could specify any correlation between the X's and u's that they desired. This method seemed appropriate for the analysis in this research paper.

[6] W. A. Neiswanger and T. A. Yancey, "Parameter Estimates and Autonomous Growth," Journal of the American Statistical Association, LIV (June, 1959), 389-402.

The authors' conclusion was that the inclusion of time as an additional predetermined variable in a model is justified when time series data are used.

Potluri Rao and Zvi Griliches tested the efficiency of various two-stage estimation procedures in a model in which both the error terms and the independent variable were autocorrelated.[7] The authors started with the belief that the sampling variation in the autocorrelated coefficient negated much of the gain from efficient estimation techniques. The initial hypothesis, however, was rejected. In this analysis, least squares estimators proved to be inefficient when autocorrelation was present.

An analysis of the presence of observation errors in least squares was conducted by George W. Ladd.[8] He concluded that the errors of observation caused only little bias in least squares estimators but did increase the standard error of the estimators. The author also noticed that when least squares was applied directly to a simultaneous equation model, with a small non-zero covariance between the predetermined variables and the error terms, the result was not a very large bias. He further concluded that even in

[7] Potluri Rao and Zvi Griliches, "Small-Sample Properties of Several Two-Stage Regression Methods in the Conte t of Auto-Correlated Errors," Journal of the American Statistical Association, LXIV (March, 1969), 253-272.

[8] George W. Ladd, "Effects of Shocks and Errors in Estimation: An Empirical Comparison," Journal of Farm Economics, XXXVIII (May, 1956), 485-494.

instances where the covariance between X and u was large, there may be a regression coefficient where the least squares bias is negligible.

Guy H. Orcutt and Donald Cochrane analyzed the combined effects of autocorrelation and a correlation between the predetermined variable and the error terms in a system of equations by the least squares method.[9] Orcutt and Cochrane's analysis is similar to that done here, in that the joint effects of two specification errors dealing directly with the error terms are analyzed. The correlation between the predetermined variables and the error terms was caused by lagged variables. The effect of autocorrelation, like that of heteroscedasticity, is not on bias or consistency but on efficiency. Lagged variables do lead to biased results and, in the authors' model, produced a negative bias. The simultaneous presence of the two complications produced a substantial positive bias, a result which certainly could not be expected by the specification error's individual effects. The authors concluded that a prior knowledge must exist about the intercorrelation of the error term to be able to justify the using of least squares. Unless this is possible, transformations cannot be used to randomize the error terms.

[9] Guy H. Orcutt and Donald Cochrane, "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis," Journal of the American Statistical Association, XLV (September, 1949), 356-372.

In the same article, an attempt at short term point forecasting was made. Forecasts for the next period were made from knowledge of the independent variable and regression coefficient calculated from the series up to that point. The forecasts were compared to the actual values of the dependent variable obtained in the original generation of the data. The results were rewarding when the estimated variance of the errors of individual forecasts obtained from the analysis coincided with the actual values.

One of the most comprehensive Monte Carlo analyses thus far completed was accomplished by Robert Summers.[10] A simultaneous equation model was constructed which introduced a correlation between predetermined variables which were themselves autocorrelated. A comparison of least squares with four other estimating techniques suggested that least squares estimators had the greatest bias and the smallest variance. The minimum variance property of least squares caused a root mean square error comparable to all but one estimating technique. However, it is a minimum variance around a biased mean. In the overall ratings of the various estimating techniques, least squares placed last, except for the isolated samples where the model was misspecified.

In an analysis on forecasting, the expected values

[10] Robert Summers, "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, XXXIII (January, 1965), 1-40.

of the endogenous variables, for given values of the predetermined variables, was computed. The forecasts were compared on a pairwise basis for each of the estimating techniques involved. In almost every experimental design, least squares proved inferior to the other estimating techniques.

Another analysis on forecasting was made by Richard J. Foote and Frederick V. Waugh.[11] They compared forecasts obtained from least squares and another method of estimation with inconclusive results.

John S. Chipman analyzed the method of least squares when multicollinearity was present.[12] The conclusions differed from those in preceeding articles in that Chipman maintains that least squares yields a minimum variance estimator if, and only if, it is biased.

J. Durbin analyzed the properties of estimators when some of the predetermined variables were lagged values of the dependent variable.[13] His result was that when the error terms are normally distributed the method of least squares leads to optimum estimators. Durbin demonstrated that the properties of least squares are asymptotically the same as

[11] Richard J. Foote and Frederick V. Waugh, "Results of an Experiment to Test the Forecasting Merits of Least Squares and Limited Information Equations," Econometrica, XXVI (November, 1958), 607-608.

[12] John S. Chipman, "On Least Squares With Insufficient Observations," Journal of the American Statistical Association, LIX (December, 1964), 1078-1111.

[13] J. Durbin, "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, XXII (January, 1960), 139-153.

those of the least squares coefficients of regression models containing no lagged variables, whether or not the errors are normally distributed.

Conclusions

These Monte Carlo studies, as can be expected, cover a wide range of specification errors and methods of estimation. In reviewing many articles, this author noticed a conspicuous absence of Monte Carlo analyses covering the topic of heteroscedasticity. In fact, not one article can be found on the effects of heteroscedasticity in its combined presence with another specification error. In this paper an attempt will be made to analyze the effects of heteroscedasticity where it is combined with a correlation between the independent variable and the error term in a single equation model. .

CHAPTER III

THE MODEL

Introduction

The effects of two specification errors, each involving the error term, is analyzed to determine their combined effects on parameters and forecasting. The two specification errors are those of heteroscedasticity and a correlation between the independent variable and the error term. For simplicity, the correlation between the independent variable and the error term is to be defined as r; and when $r \neq 0$, then this specification error exists. A single equation Monte Carlo analysis will be used to determine the effects of heteroscedasticity and $r \neq 0$ in a least squares analysis. The individual effects of each specification error are known.[1] It is assumed in this analysis that there are no errors of observation.

Model Specifications

The model to be used is the simple linear function

$$Y_i = 3.64 + .9016(X_i) + u_i$$

[1] The individual effects of heteroscedasticity and $r \neq 0$ are reviewed in Chapter I.

where

$$\bar{X} = 287.3$$

$$\text{Var }(X) = 4638$$

$$\text{Var }(u) = 76$$

and

$$E(u_i) = 0$$

$$E(u_i u_j) = 0 \text{ when } i \neq j$$

$$E(u_i^2) \neq \sigma_u^2$$

$$E(X_i u_i) \neq 0$$

The parameters were obtained from a linear time-series regression of a consumption function using quarterly observations from 1948 to 1967.[2] By choosing parameters from a consumption function it may be possible that a close analogy to real data can be drawn in this analysis. As stated in Chapters I and II, heteroscedasticity might be expected when dealing with aggregative economic data on consumption. Spending and saving habits of individuals in different income brackets might be expected to vary. If the variance of spending habits is dependent upon income, then $r \neq 0$ may follow,[3] unless the form of heteroscedasticity is such to cause a zero correlation for r. In time

[2] Ralph D. Husby, "A Nonlinear Consumption Function Estimated from Time-Series and Cross-Section Data," The Review of Economics and Statistics, LX (February, 1971), 76-79. Linear functions are covered in this analysis.

[3] Review equations 2-1 and 2-2 and the subsequent analysis.

series data no particular form of hèteroscedasticity is to be expected ". . . since the variables are of similar orders of magnitude for the different observations . . . ."[4]

A random form of heteroscedasticity is not expected to cause $r \neq 0$; however, there are other factors present in aggregative economic data to cause a correlation between X and u. The correlation can easily be caused by errors of observation, lagged variables, and incorrect specification of a model. Therefore, the two specification errors being analyzed are relevent to current economic literature. However, because the data is synthetic it may not be possible to duplicate a real-world model with all of its complexities.

Two models are to be computed and analyzed. Henceforth they are to be defined as Model 1 and Model 11. Model 1 contains but one specification error, that of $r \neq 0$. The actual values chosen for r are +.9, +.5, +1, -.1, -.5, and -.9. For each value of r, 50 samples are drawn of size 20. For each r there will be 50 estimates of a and b from which the mean, standard deviation, and root mean square error will be computed. In addition, the estimated standard error of each individual $\hat{a}$ and $\hat{b}$ value will be obtained and analyzed in the following ways:

1. The number of times the estimated parameter values are greater than two standard errors, $(\hat{b} > 2 \text{ SE})$

[4] E. Malinvaud, <u>Statistical Methods of Econometrics</u>, (Chicago: Rand McNally & Company, 1966), p. 256.

and ($\hat{a} > 2$ SE).

2. The number of times the absolute values of the estimated parameters minus the true parameter are greater than two standard errors, $|\hat{b} - b| > 2$ SE and $|\hat{a} - a| > 2$ SE.

3. The number of times the absolute values of the estimated parameters minus the mean of the estimated parameters are greater than two standard errors, $|\hat{b} - \bar{\hat{b}}| > 2$ SE and $|\hat{a} - \bar{\hat{a}}| > 2$ SE.

The two standard errors represents a 95 percent confidence limit, or a 5 percent level of significance, based on the _t_ distribution. The critical value for a 5 percent level of significance on the _t_ distribution is 1.729. If an estimate of a parameter is greater than two standard errors, it is significantly different from zero, based on a 5 percent level of significance. When $|\hat{p} - p| >$ two standards errors, the estimated parameter would fall outside the confidence interval. the same analysis holds for the difference between $\hat{p}$ and $\bar{\hat{p}}$.

The individual effects on bias can then be analyzed with various correlations between the independent variable and the error term. From the Neiswanger and Yancey and Summers articles it might be expected that estimated parameters would differ from the true parameters by more than two standard errors a large number of times.

In Model 11, the second specification error, heteroscedasticity, is to be introduced in addition to the specification error of $r \neq 0$. The values which r may take on are the same as those specified in Model 1. Again, 50 samples of size 20 are to be generated for each value of r. The same mean values and measures of dispersion computed in Model 1 are repeated.

Data Generation

The X's and u's have to be generated in such a manner that they maintain their respective variances and correlations. In order to accomplish this, specific covariances between X and u were specified. Given the variances and correlations desired, the covariances can be computed and are given in Table 1.

The method of actually generating the X and u values entailed the transforming of a matrix of random variates. This was accomplished by premultiplying the random variate matrix by a P matrix, which is a triangular matrix such that $PP' = M$ where M is the population variance-covariance matrix of the X's and u's.[5] That P is a triangular matrix means that for all elements of P (P_{ij}) will be equal to zero when $j > i$. The P matrix is actually defined as a lower case triangular matrix. An example, using a 2 by 2 matrix, is

[5] See W. A. Neiswanger and T. A. Yancey, "Parameter Estimates and Autonomous Growth," Journal of the American Statistical Association, LIV (June, 1959), 392.

Table 1.--The Population Variance-Covariance Values of X and u for Different Values of r

r	M_{11}	M_{12} and M_{21}	M_{22}
+.9	4638.	+534.33629	76
+.5	4638	+296.85349	76
+.1	4638	+ 59.37069	76
-.1	4638	- 59.37069	76
-.5	4638	-296.85349	76
-.9	4638	-534.33629	76

$$\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} P_{11} & P_{21} \\ 0 & P_{22} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

therefore,

$$\begin{bmatrix} P_{11}P_{11} & P_{11}P_{21} \\ P_{11}P_{21} & P_{21}P_{21} + P_{22}P_{22} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

and

$$P_{11} = M_{11}$$

$$P_{21} = M_{12}/\ P_{11} = M_{21}/\ P_{11}$$

$$P_{22} = \sqrt{M_{22} - P_{21}^{\ 2}}$$

since

$$P_{11}^{\ 2} = M_{11}$$

$$P_{11}P_{21} = M_{12} = M_{21}$$

$$P_{21}^{\ 2} + P_{22}^{\ 2} = M_{22}$$

In order that the elements of P have a single solution, M must be a symmetrical matrix; otherwise, this method of data generation could not be used. The proof that M is symmetrical follows. If

$$M = PP'$$

then

$$\begin{aligned} M' &= (PP')' \\ &= (P')'\ P' \\ &= PP' \\ &= M \end{aligned}$$

Since M is symmetrical, M = PP' has an orthonormal set of eigenvectors X_1, X_2, . . . , X_n, such that $X_iX_i' = 1$ and $X_iX_j' = 1$ and $X_iX_j' = 0$ where $MX_i' = \lambda_iX_i'\lambda_i$ is the eigenvalue corresponding to the eigenvector X_i. Now

$$\begin{aligned} X_iMX_i' &= X_i\lambda_iX_i' \\ &= \lambda_iX_iX_i' \\ &= \lambda_i \\ \lambda_i &= X_iPP'X_i' \\ &= (X_iP)(X_iP)' > 0 \end{aligned}$$

Note that X_iP is a row vector and for any row vector $Q \neq 0$, $QQ' > 0$. Since PP' = M is positive definite, all the roots (eigenvalues) of M must be positive.[6]

The elements of the X_iP vector are a linear equivalent to a weighted sum of the squares of the elements in $(X_iP)'$. If X_iP consists of normally and independently distributed variables with a zero mean and a constant variance (Equations 1-4, 1-5, and 1-6), then a transformation insures that the $(X_iP)'$ vector consists of the same properties.[7] Of course, the same holds true if one or more of the assumptions is violated.

[6] Dr. Frank L. Salzman of the Department of Mathematics, Florida Technological University, assisted in the proof that M was a positive definite matrix.

[7] See J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company, Inc., 1963), pp. 99-101.

Model 1

$$\begin{bmatrix} X(1). & . & .X(20) \\ u(1). & . & .u(20) \end{bmatrix} = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} S_{11}. & . & .S_{1,20} \\ S_{21}. & . & .S_{2,20} \end{bmatrix} + \begin{bmatrix} L_{11} \\ 0 \end{bmatrix}$$

The S matrix consists of random numbers with the specifications $E(S_{ij}) = 0$ and $SD\ (S_{ij}) = 5$. The elements of the S matrix change for every sample, whereas the elements of P remain fixed for each value of r. The elements of the P matrix, for each value of r, are given in Table 2. The L matrix simply contains the mean values for X and u, the latter being zero. The L matrix is actually a 2 by 20 matrix with each element of each row being identical.

After the X's and u's are generated with their specific properties, the Y's are generated and estimates of the parameters are obtained.

Model 11

To obtain the X's and u's for Model 11 the S matrix was altered. The elements S_{11} to $S_{1,20}$, the first row of the matrix, remained the same with a mean equal to zero and a standard deviation of five. The second row, elements S_{21} to $S_{2,20}$, had a mean equal to zero but had a standard deviation changing randomly from element to element. The standard deviation could take on values from 0.01 to 9.99 randomly, so that the expected value of the different standard deviations remained constant at 5. The variance of u changed for each and every value of X.

Table 2.--Values for the P Matrix such that PP' = M

r	P_{11}	P_{21}	P_{12}	P_{22}
+.9	2.72411	+.31384	0.00000	.15200
+.5	2.72411	+.17436	0.00000	.30199
+.1	2.72411	+.03487	0.00000	.34696
-.1	2.72411	-.03487	0.00000	.34696
-.05	2.72411	-.17436	0.00000	.30199
-.9	2.72411	-.31384	0.00000	.14200

The values for the P matrix are the same as those in Model 1 and are given in Table 2. The L matrix is the same as in Model 1. Again, after the X's and u's are generated with their specified properties the Y's are generated and estimators obtained.

Forecasting

An attempt was made here in the area of short term point forecasting. Forecasted values of the dependent variable (Y) were compared to the true values of Y. The true values of Y were obtained by extending the random variate matrix (the S matrix) so that five additional values of X and u could be obtained for each sample.

$$\begin{bmatrix} X(1). & . & .X(25) \\ u(1). & . & .u(25) \end{bmatrix} = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} S_{11}. & . & .S_{1,25} \\ S_{21}. & . & .S_{2,25} \end{bmatrix} + \begin{bmatrix} L_{11} \\ 0 \end{bmatrix}$$

From X(21), u(21); X(22), u(22); X(23), u(23); X(24), u(24); and X(25), u(25); values of Y(21), Y(22), Y(23), Y(24), and Y(25) were obtained. These are the true values of Y. The forecasted values of Y were obtained from the sample estimator and from the respective values for X.

$$\hat{Y}(21) = \hat{a} + \hat{b}X(21)$$

$$\hat{Y}(22) = \hat{a} + \hat{b}X(22)$$

$$\hat{Y}(23) = \hat{a} + \hat{b}X(23)$$

$$\hat{Y}(24) = \hat{a} + \hat{b}X(24)$$

$$\hat{Y}(25) = \hat{a} + \hat{b}X(25)$$

From these values the difference between the actual and forecasted values of Y were obtained.

$\hat{Y}(21) - Y(21)$

$\hat{Y}(22) - Y(22)$

$\hat{Y}(23) - Y(23)$

$\hat{Y}(24) - Y(24)$

$\hat{Y}(25) - Y(25)$

For each value of r there are 50 sets of such values from which the mean deviation is to be calculated. From this analysis it is hoped that it can be learned how the specification errors of heteroscedasticity and $r \neq 0$ effect short-term forecasts.

The extra five units of data will be used only in forecasting and they are not to be considered when the bias, SD, RMSE, and SE are computed for Models 1 and 11.

CHAPTER IV

THE ANALYSIS

Estimates of the Parameters

Each of the two Models (with and without heteroscedasticity) was used to obtain 50 estimates for each of the parameters. The estimates respective distributions for various correlations between the independent variable and the error terms were also obtained. The result of this analysis is given in Tables 3 to 15. The seven factors common to a majority of the tables are as follows:

(1) The mean value of each parameter estimated which was obtained from a sample of 50.

(2) The standard deviation of the estimated parameters.

(3) The mean of the standard errors obtained for each of the 50 samples.

(4) The root mean square error.

(5) The number of times a sample parameter was greater than two standard errors, ($\hat{p} > 2$ SE).

(6) The number of times the absolute value of a sample parameter minus the true value was

more than two standard errors, ($|\hat{p} - p| > 2$ SE).

(7) The number of times the absolute value of a sample parameter minus the mean value of that parameter's distribution was more than two standard errors, ($|\hat{p} - \bar{\hat{p}}| > 2$ SE).

In Tables 3 and 4, the results obtained from Models 1 and 11 where a high positive and negative correlation existed between X and u are revealed. The correlation in Tεble 3 is +.9 and in Table 4 the correlation is -.9. The results presented in these two ables indicates that the estimators were biased. For the positive correlation of .9, estimates of b were overestimated and estimates of a were underestimated as shown in Figure 2. For the negative correlation of .9 the estimates of b were underestimated and the estimates of a were overestimated. Because the bias in $\hat{b}$ in Tables 3 and 4 was of approximately the same magnitude, but in different directions, the standard deviation, mean of the standard errors, and root mean square error were almost identical between a positive and negative correlation in Models 1 and 11. However, the standard deviation, mean of the standard errors, and the root mean square error were considerably larger for Model 11, where heteroscedasticity was present, than they were for Model 1. Both the standard deviation and the mean of the standard errors increased over five times while the root mean square error increased by only about 17 percent. It does not appear that heteroscedasticity had any

Table 3.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between X and u is +.9

	a	b
Parameters p	+3.64	+.9016
<u>Model 1</u>		
Means of estimates	-29.3289	1.0165
S.D. of estimates	3.1800	.0109
Mean of S.E.	8.2321	.0286
RMSE	33.1289	.1155
$\hat{p} > 2$ S.E.	0	50
$\|\hat{p} - p\| > 2$ S.E.	50	50
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	0	0
<u>Model 11</u>		
Means of estimates	-28.2402	1.0136
S.D. of estimates	37.7594	.1307
Mean of S.E.	42.1176	.1465
RMSE	50.7706	.1771
$\hat{p} > 2$ S.E.	0	50
$\|\hat{p} - p\| > 2$ S.E.	4	4
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	2	2

Table 4.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between X and u is -.9

	a	b
Parameters p	+ 3.64	+ .9016
Model 1		
Means of estimates	36.8235	.7861
S.D. of estimates	3.1800	.0109
Mean of S.E.	8.2566	.0287
RMSE	33.3446	.1160
$\hat{p} > 2$ S.E.	50	50
$\lvert\hat{p} - p\rvert > 2$ S.E.	50	50
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	0	0
Model 11		
Means of estimates	37.9122	.7832
S.D. of estimates	37.7594	.1307
Mean of S.E.	42.3868	.1474
RMSE	53.0486	.1834
$\hat{p} > 2$ S.E.	7	50
$\lvert\hat{p} - p\rvert > 2$ S.E.	6	6
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	3	3

effect over bias since there was only a small difference between the estimated parameters of Models 1 and 11.

In Model 1, the estimated parameters exceeded the true parameters in each instance, indicating the difficulty in using least squares when a correlation between X and u is present. In Model 11 the estimated parameters exceeded the true parameters only 8 percent of the time. The result can be misleading because the reason for the low percentage was due to the very large standard error, and not due to the estimated parameters being closer to the true values when heteroscedasticity and $r \neq 0$ is present.

In both models, the estimate of b exceeded two standard errors 100 percent of the time. The $\hat{a}$ values were greater than two standard errors 50 times for Model 1 with a negative correlation, but failed to exceed it once for the positive correlation. The estimated parameters that exceeded their mean values by two standard errors, on the average, was 5 percent.

The behavior of Models 1 and 11 with a positive and negative correlation of .5 is presented in Tables 5 and 6. The results on bias, standard deviation, mean of the standard errors, and root mean square error appeared to be similar to the results obtained in Tables 3 and 4. Again the positive bias was equal to the negative bias, leaving the standard deviation, mean of the standard errors, and the root mean square error almost identical between the models with the

Table 5.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between X and u is +.5

	a	b
Parameters	+ 3.64	+ .9016
Model 1		
Means of estimates	-14.5220	.9651
S.D. of estimates	6.3180	.0218
Mean of S.E.	8.2905	.0288
RMSE	19.2640	.0673
$\hat{p} > 2$ S.E.	0	50
$\lvert\hat{p} - p\rvert > 2$ S.E.	31	32
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	0	0
Model 11		
Means of estimates	-12.3590	.9593
S.D. of estimates	75.0162	.2597
Mean of S.E.	82.6017	.2874
RMSE	81.9351	.2854
$\hat{p} > 2$ S.E.	2	45
$\lvert\hat{p} - p\rvert > 2$ S.E.	2	2
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	2	2

Table 6.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between X and u is -.5

	a	b
Parameters p	+ 3.64	+ .9016
Model 1		
Means of estimates	22.2281	.8371
S.D. of estimates	6.3181	.0218
Mean of S.E.	8.31434	.0289
RMSE	19.6902	.0682
$\hat{p} > 2$ S.E.	44	50
$\|\hat{p} - p\| > 2$ S.E.	34	34
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	1	1
Model 11		
Means of estimates	24.3929	.8313
S.D. of estimates	75.0195	.2597
Mean of S.E.	82.7521	.2879
RMSE	83.1542	.2874
$\hat{p} > 2$ S.E.	2	41
$\|\hat{p} - p\| > 2$ S.E.	3	3
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	3	3

different correlations. Compared to the correlations of + and -.9, the root mean square error was small in Model 1 but considerably larger in Model 11. A smaller root mean square error should have resulted from a smaller correlation, but it is difficult to explain why it became larger for Model 11. The smaller root mean square error was probably caused by the fact that, even though the bias was smaller, the standard deviation became larger causing a larger root mean square error.

When heteroscedasticity was present, the minimum variance property of least squares was lost. In Model 11 the root mean square error was approximately equal to the mean of the standard error and not much smaller than the standard deviation of the estimates. The number of times an estimate was greater than two standard errors was smaller, on the average, than was the case in Tables 3 and 4 due to the smaller correlation involved between X and u. The reason for a never being greater than two standard errors in Model 1 for correlations between X and u of +.9 and +.5 is that the estimates of a are negative while the standard error is always positive.

Results similar to those above were obtained from Tables 7 and 8 where a relatively low correlation of + and - .1 existed between X and u. The bias for both models was smaller due to the lower correlation involved. It is peculiar that the standard deviation tended to become larger

Table 7.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between X and u is +.1

	a	b
Parameters p	+ 3.64	+ .9016
Model 1		
Means of estimates	.2098	.9138
S.D. of estimates	6.1737	.0250
Mean of S.E.	8.3279	.0289
RMSE	8.0423	.0279
$\hat{p} > 2$ S.E.	0	50
$\lvert\hat{p} - p\rvert > 2$ S.E.	2	1
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	0	0
Model 11		
Means of estimates	13.4743	.9071
S.D. of estimates	86.1908	.2255
Mean of S.E.	94.8523	.3300
RMSE	92.5066	.3205
$\hat{p} > 2$ S.E.	2	38
$\lvert\hat{p} - p\rvert > 2$ S.E.	3	3
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	3	2

Table 8.--Mean Values of Coefficients with Measures of Dispersion for 50 Samples of 20 Observations for Models 1 and 11. The Coefficient of Correlation between x and u is -.1

	a	b
Parameters p	+ 3.64	+ .9016
Model 1		
Means of estimates	7.5598	.8882
S.D. of estimates	7.2589	.0250
Mean of S.E.	8.3326	.0289
RMSE	8.3468	.0286
$\hat{p} > 2$ S.E.	6	50
$\|\hat{p} - p\| > 2$ S.E.	1	1
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	1	1
Model 11		
Means of estimates	10.0449	.8815
S.D. of estimates	86.1908	.2984
Mean of S.E.	94.8824	.3301
RMSE	92.7411	.3210
$\hat{p} > 2$ S.E.	2	39
$\|\hat{p} - p\| > 2$ S.E.	3	3
$\|\hat{p} - \bar{\hat{p}}\| > 2$ S.E.	3	3

as the correlation between X and u diminished. It can also be seen that Tables 7 and 8 contain the largest mean of the standard errors and root mean square errors of the data analyzed. The standard deviation of the estimates increased nine times with the introduction of heteroscedasticity, while the mean of the standard errors and the root mean square error increased approximately eleven times. As was expected, a smaller percentage of parameters exceeded two standard errors.

A comparison of the estimated parameter b of Models 1 and 11, for positive correlations between X and u can be seen in Table 9. The bias decreased as the correlation between X and u decreased, while the standard deviation increased as the correlation decreased. The mean of the standard error remained relatively constant for Model 1 but increased drasticall. with the introduction of heteroscedasticity. No discernable trend could be seen from the root mean square error of Model 1; however, in Model 11 the root mean square error became substantially larger as the correlation decreased. The number of estimated parameters which exceeded two standard errors became smaller as r approached zero.

Exactly the same analysis held for the negative correlations listed in Table 10. The bias decreased with a smaller correlation, while the standard deviation increased and the mean of the standard errors and root mean square error increased substantially in Model 11. The estimated parameters exceeded twice their standard error in each instance in Model 1, but appeared to decline to 78 percent for Model 11. In

Table 9.--Comparing the Estimate of b in Model 1 and Model 11 (shown in Parentheses) with Different Positive Correlations between X and u. The True Value for b is +.9016

Correlations	+.9	+.5	+.1
Mean value of $\hat{b}$	1.0165	.9651	.9138
	(1.0136)	(.9593)	(.9071)
S.D. of estimated b	.0109	.0218	.0250
	(.1307)	(.2597)	(.2255)
Mean of S.E.	.0286	.0288	.0289
	(.1465)	(.2874)	(.3300)
RMSE	.1155	.0673	.0279
	(.1771)	(.2854)	(.3205)
$\hat{b} > 2$ S.E.	50	50	50
	(50)	(45)	(38)
$\lvert\hat{b} - b\rvert > 2$ S.E.	50	32	1
	(4)	(2)	(3)
$\lvert\hat{b} - \bar{\hat{b}}\rvert > 2$ S.E.	0	0	0
	(2)	(2)	(2)

Table 10.--Comparing the Estimate of b in Model 1 and Model 11 (shown in Parentheses) with Different Negative Correlations between X and u. The True Value for b is +.9016

Correlations	-.9	-.5	-.1
Mean value of $\hat{b}$	.7861	.8371	.8882
	(.7832)	(.8313)	(.8815)
S.D. of estimated b	.0109	.0218	.0250
	(.1307)	(.2597)	(.2984)
Mean of S.E.	.0287	.0289	.0289
	(.1474)	(.2879)	(.3301)
RMSE	.1160	.0682	.0286
	(.1834)	(.2874)	(.3210)
$\hat{b} > 2$ S.E.	50	50	50
	(50)	(41)	(39)
$\lvert\hat{b} - b\rvert > 2$ S.E.	50	34	1
	(6)	(3)	(3)
$\lvert\hat{b} - \bar{\hat{b}}\rvert > 2$ S.E.	0	1	1
	(3)	(3)	(3)

addition, 4 percent of the estimates were within two standard errors of the mean of the estimates.

The phenomena that the standard deviation increased as the correlation between X and u decreased is analyzed further in Table 11. The results of Model 11 with a zero correlation between X and u are presented in Table 11, therefore, only the specification error of heteroscedasticity was present. From analyzing Tables 9, 10, and 11 the trend that the standard deviation, mean of the standard errors, and root mean square errors decreased on the average as r approached zero appeared. This trend was due to the method of data generation. The element P_{22} of the P matrix becomes larger as the covariance between X and u decreased. From the formula

$$P_{22} = \sqrt{M_{22} - P_{21}^{2}}$$

where

$$P_{21} = M_{12}/P_{11} = M_{21}/P_{11}$$

and M_{12} and M_{21} represent the population covariance between X and u, P_{21} increased as the covariance increased. This caused P_{22} to become smaller the greater the value of r. A smaller P_{22} value (for the higher value of r) was multiplied by the random number matrix causing a smaller standard deviation, mean of the standard errors, and root mean square error.

Two parameter frequency distributions were selected and analyzed to determine if the distributions were normal or skewed. Since the bias for the positive and negative

Table 11.--Mean Values of Coefficients with Measures of Dispersion and Forecasts for 50 Samples of 20 Observations for Model 11. The Coefficient of Correlation between X and u is 0.*

	a	b
Parameters p	+3.64	+.9016
Means of estimates	10.1484	.8825
S.D. of estimates	65.8424	.2635
Mean of S.E.	96.0895	.3346
RMSE	82.6954	.2873
$\hat{p} > 2$ S.E.	1	38
$\lvert\hat{p} - p\rvert > 2$ S.E.	2	2
$\lvert\hat{p} - \bar{\hat{p}}\rvert > 2$ S.E.	2	1

Forecasts

$\hat{Y}(21) - Y(21)$	-.0578
$\hat{Y}(22) - Y(22)$	-.1366
$\hat{Y}(23) - Y(23)$	+3.5106
$\hat{Y}(24) - Y(24)$	-.2020
$\hat{Y}(25) - Y(25)$	+2.9103

*The P matrix used to obtain the data necessary for this Table is

$$\begin{bmatrix} 2.72411 & 0.00000 \\ 0.00000 & 0.34871 \end{bmatrix}$$

correlations were equal, though in opposite directions, and since heteroscedasticity did not effect the bias, it was not necessary to analyze all of the parameter distributions involved. The percentage of times the 50 estimated parameters fell within a given interval in Model 1, with r = +.9 is shown in Table 12. There did not appear to be any skewnessto the distribution -- the parameters seemed to be normally distributed about their mean.

In Table 13 the estimated parameters' frequency distribution of Model 11, with r = +.1 are shown. Again, a skewness was not discernable and it does appear that the parameters are normally distributed. The result that the estimated parameters were normally distributed about their mean should have been expected, since the values for the independent variables and error terms, (hence dependent variables) were generated from normally distributed random numbers. The actual dispersion of the estimated parameters was much greater in Table 13 than it was in Table 12, due to the presence of heteroscedasticity.

In Table 14 the 50 estimated parameters and two standard errors are shown for Model 1 with a correlation between X and u of -.5. In 88 percent of the instances, the estimate of a exceeded two standard errors while $\hat{b}$ exceeded two standard errors 100 percent of the time. The estimates of b exceeded two standard errors generally by atleast ten times, however, a remained consistently

Table 12.--Frequency Distribution for the Estimated Parameters of Model 1, with a Correlation between the Independent Variable and the Error Term of +.9

$\hat{a}$ = -29.3289		$\hat{b}$ = 1.0165	
Interval	Percent	Interval	Percent
Less than -37.8289	2	Less than .9911	2
-36.8290 to -37.8289	0	.9911 to .9940	0
-35.8290 to -36.8289	0	.9941 to .9970	0
-34.8290 to -35.8289	4	.9971 to 1.0000	2
-33.8290 to -34.8289	4	1.0001 to 1.0030	4
-32.8290 to -33.8289	4	1.0031 to 1.0060	6
-31.8290 to -32.8289	6	1.0061 to 1.0090	10
-30.8290 to -31.8289	12	1.009 to 1.0120	12
-29.8290 to -30.8289	8	1.0121 to 1.0150	8
-28.8290 to -29.8289	14	1.0151 to 1.0180	12
-27.8290 to -28.8289	12	1.0181 to 1.0210	10
-26.8290 to -27.8289	10	1.0211 to 1.0240	10
-25.8290 to -26.8289	14	1.0241 to 1.0270	6
-24.8290 to -25.8289	4	1.0271 to 1.0300	6
-23.8290 to -24.8289	2	1.0301 to 1.0330	4
-22.8290 to -23.8289	2	1.0331 to 1.0360	2
-21.8290 to -22.8289	2	1.0361 to 1.0390	4
-20.8290 to -21.8289	0	1.0391 to 1.0420	0
Greater than -20.8290	0	Greater than 1.0420	2

Table 13.--Frequency Distribution for the Estimated Parameters of Model 11, with a Correlation between the Independent Variable and the Error Term of +.1

$\bar{\hat{a}}$ = +13.4743		$\bar{\hat{b}}$ - .9071	
Interval	Percent	Interval	Percent
Less than -156.5257	4	Less than .0572	0
-156.5257 to -136.5256	0	.0572 to .1571	4
-136.5257 to -116.5256	2	.1572 to .2571	0
-116.5257 to - 96.5256	6	.2572 to .3571	4
- 96.5257 to - 76.5256	0	.3572 to .4571	4
- 76.5257 to - 56.5256	6	.4572 to .5571	4
- 56.5257 to - 36.5256	10	.5572 to .6571	8
- 36.5257 to - 16.5256	12	.6572 to .7571	14
- 16.5257 to + 3.4743	8	.7572 to .8571	14
+ 3.4744 to + 23.4743	14	.8572 to .9571	8
+ 23.4744 to + 43.4743	6	.9572 tc 1.0571	18
+ 43.4744 to + 63.4743	10	1.0572 to 1.1571	10
+ 63.4744 to + 83.4743	6	1.1572 to 1.2571	0
+ 83.4744 to +103.4743	4	1.2572 to 1.3571	8
+103.4744 to +123.4743	2	1.3572 to 1.4571	0
+123.4744 to +143.4743	6	1.4572 to 1.5571	0
+143.4744 to +163.4743	0	1.5572 to 1.6571	0
+163.4744 to +183.4743	2	1.6572 to 1.7571	0
Greater than +183.4743	2	Greater than .7571	4

Table 14.--Estimated Parameters and Two Standard Errors of Each Parameter in Model 1 with a Correlation between X and u of -.5

Number	$\hat{a}$	2 SE	$\hat{b}$	2 SE
1	18.6172	18.1860	.8489	.0641
2	21.3054	15.3263	.8404	.0530
3	23.6205	16.4447	.8323	.0560
4	13.0115	16.8087	.8694	.0574
5	28.5944	23.1325	.8165	.0801
6	33.8044	17.8106	.7988	.0625
7	21.7991	16.5606	.8372	.0576
8	26.5126	19.5663	.8209	.0680
9	20.1504	17.7216	.8450	.0618
10	11.5805	18.3451	.8747	.0638
11	23.7221	16.4540	.0287	.0574
12	17.0758	12.9841	.0226	.0452
13	21.8191	13.6218	.0242	.0484
14	17.4675	10.2781	.0174	.0349
15	31.6005	18.5605	.0314	.0628
16	17.5445	15.3915	.0266	.0533
17	27.1559	19.7569	.0343	.0686
18	31.0476	18.9440	.0329	.0658
19	25.6833	16.8128	.0288	.0576
20	19.0924	13.0840	.0229	.0459
21	27.7900	15.9090	.8179	.0551
22	28.7864	16.1510	.8165	.0566

Table 14.--Continued.

Number	$\hat{a}$	2 SE	$\hat{b}$	2 SE
23	10.1634	13.3258	.8781	.0474
24	24.8768	15.7365	.8282	.0548
25	28.4906	18.4168	.8150	.0640
26	16.0799	14.9317	.8589	.0516
27	29.1677	20.1425	.8121	.0698
28	21.0682	19.0332	.8404	.0670
29	19.5819	13.6907	.8491	.0476
30	18.8845	12.7543	.8467	.0446
31	21.9441	[illegible]	.8381	.0531
32	20.60[illegible]	[illegible]	.8416	.0603
33	27.9[illegible]	[illegible]	.8155	.0679
34	3.9[illegible]	[illegible]	.9001	.0511
35	15.55[illegible]	[illegible]	.8593	.0602
36	24.577[illegible]	[illegible]	.8281	.0751
37	37.1172	[illegible]	.7854	.0841
38	25.0829	16.5268	.8243	.0568
39	10.2887	15.1377	.8791	.0526
40	26.0994	14.5585	.8227	.0505
41	23.0670	15.4933	.8342	.0541
42	25.3563	17.7345	.8269	.0605
43	21.5128	18.6325	.8392	.0648
44	28.7153	16.1951	.8154	.0572

Table 14.--Continued.

Number	$\hat{a}$	2 SE	$\hat{b}$	2 SE
45	14.7958	11.5673	.8625	.0401
46	24.6669	17.0041	.8308	.0600
47	21.3338	15.1373	.8407	.0511
48	29.5037	16.6757	.8120	.0579
49	14.3078	17.7733	.8658	.0618
50	18.9539	14.0435	.8495	.0481

Table 14.--Continued.

Number	$\hat{a}$	2 SE	$\hat{b}$	2 SE
23	10.1634	13.3258	.8781	.0474
24	24.8768	15.7365	.8282	.0548
25	28.4906	18.4168	.8150	.0640
26	16.0799	14.9317	.8589	.0516
27	29.1677	20.1425	.8121	.0698
28	21.0682	19.0332	.8404	.0670
29	19.5819	13.6907	.8491	.0476
30	18.8845	12.7543	.8467	.0446
31	21.9441	15.2195	.8381	.0531
32	20.6012	17.3746	.8416	.0603
33	27.9155	19.0420	.8155	.0679
34	3.9466	14.5679	.9001	.0511
35	15.5527	17.2937	.8593	.0602
36	24.5773	21.2094	.8281	.0751
37	37.1172	24.2691	.7854	.0841
38	25.0829	16.5268	.8243	.0568
39	10.2887	15.1377	.8791	.0526
40	26.0994	14.5585	.8227	.0505
41	23.0670	15.4933	.8342	.0541
42	25.3563	17.7345	.8269	.0605
43	21.5128	18.6325	.8392	.0648
44	28.7153	16.1951	.8154	.0572

Table 14.--Continued.

Number	$\hat{a}$	2 SE	$\hat{b}$	2 SE
45	14.7958	11.5673	.8625	.0401
46	24.6669	17.0041	.8308	.0600
47	21.3338	15.1373	.8407	.0511
48	29.5037	16.6757	.8120	.0579
49	14.3078	17.7733	.8658	.0618
50	18.9539	14.0435	.8495	.0481

close to the value of two standard errors.

The absolute differences between an estimated parameter and the parameter minus two standard errors, in Model 1 with $r = -.5$ is shown in the second and third columns of Table 15. The difference between the parameter and its estimate was obtained by subtracting $|\hat{p} - p|$ from two standard errors of the estimate. Therefore, a negative value represented the number of times $|\hat{p} - p|$ was greater than two standard errors. For both estimates of a and b, $|\hat{p} - p|$ was greater than two standard errors 68 percent of the time.

The absolute differences between an estimated parameter and the mean of the estimated parameters, minus two standard errors, also given in Table 15 in columns four and five. In only one instance out of fifty did $|\hat{p} - \overline{\hat{p}}|$ exceed two standard errors.

In summary, it appeared that the combined effects of heteroscedasticity and a correlation between X and u tend to be related to their individual effects. Heteroscedasticity had no effect on bias but did decrease the efficiency of the estimator. The Cochrane-Orcutt article, reviewed in Chapter II, yielded different results than those obtained here. Cochrane and Orcutt found that the joint effects of lagged variables and autocorrelation were, in fact, different than the sum of their individual effects. Of course, Cochrane and Orcutt were analyzing different specification errors.

Table 15.--Absolute Differences between an Estimated Parameter minus the Parameter and Minus Its Sample Mean in Model 1 with a Correlation between X and u of -.5

Number	$2SE - \lvert\hat{a} - a\rvert$	$2SE - \lvert\hat{b} - b\rvert$	$2SE - \lvert\hat{a} - \bar{\hat{a}}\rvert$	$2SE - \lvert\hat{b} - \bar{\hat{b}}\rvert$
1	-3.0646	-.0103	14.6059	.0518
2	.4133	.0004	16.6059	.0581
3	-5.2335	-.0181	12.4370	.0439
4	14.2613	.0496	-2.7959	-.0095
5	5.3810	.0180	11.5359	.0403
6	.2720	.0016	17.9426	.0637
7	-9.2080	-.0319	8.4625	.0301
8	-4.9160	-.0204	12.7544	.0417
9	8.4890	.0301	4.1158	.0129
10	-7.9008	-.0283	9.7696	.0338
11	-8.2404	-.0285	10.6079	.0367
12	-8.9953	-.0283	9.8536	.0368
13	6.8024	.0240	1.0003	.0056
14	-5.5002	-.0185	13.3487	.0467
15	-6.3337	-.0224	12.5151	.0428
16	2.4917	.0090	8.5227	.0290
17	-5.3857	-.0195	13.4638	.0456
18	1.6049	.0058	17.6124	.0629
19	-2.2511	-.0047	10.7836	.0349
20	-2.4901	-.0102	9.1499	.0342
21	-3.6280	-.0129	15.9528	.0549

Table 15.--Continued

Number	$2SE - \lvert\hat{a} - a\rvert$	$2SE - \lvert\hat{b} - b\rvert$	$2SE - \lvert\hat{a} - \bar{\hat{a}}\rvert$	$2SE - \lvert\hat{b} - \bar{\hat{b}}\rvert$
22	- .4517	-.0013	6.8391	.0240
23	- 4.5573	-.0149	12.2200	.0439
24	- 3.5494	-.0114	4.5247	.0134
25	- 9.4000	-.0337	10.1808	.0341
26	1.4869	.0049	9.7151	.0339
27	- 3.7589	-.0112	15.8219	.0566
28	- 8.4636	-.0288	11.1172	.0390
29	- 5.2305	-.0193	14.3503	.0484
30	- 2.3683	-.0094	8.9555	.0333
31	3.2138	.0115	14.8991	.0537
32	- 2.3391	-.0080	14.7326	.0511
33	- 3.5358	-.0132	14.7233	.0499
34	7.4371	.0253	7.9211	.0263
35	- 1.8218	-.0048	16.4372	.0582
36	-12.3537	-.0401	5.9054	.0229
37	- 1.5984	-.0067	16.4607	.0564
38	- 3.3063	-.0125	14.9528	.0505
39	1.2112	.0052	15.9729	.0552
40	10.4045	.0370	8.0265	.0275
41	-3.9336	-.0131	14.6477	.0506
42	-3.9818	-.0140	14.5995	.0498
43	.7596	.0025	17.9240	.0633

Table 15.--Continued

Number	$2SE - \lvert\hat{a} - a\rvert$	$2SE - \lvert\hat{b} - b\rvert$	$2SE - \lvert\hat{a} - \bar{\hat{a}}\rvert$	$2SE - \lvert\hat{b} - \bar{\hat{b}}\rvert$
44	- 8.8802	-.0288	9.7011	.0349
45	.4115	.0011	4.1417	.0153
46	- 4.0227	-.0106	14.5586	.0531
47	- 2.5565	-.0096	14.2499	.0481
48	- 9.1879	-.0316	9.3933	.0322
49	7.1054	.0261	9.8598	.0337
50	- 1.2704	-.0034	10.7760	.0362

Forecasting

In Tables 16 and 17 and five period forecasts obtained from the two models are presented. The forecasts obtained from a specific correlation between X and u, regardless of the sign, were almost identical. The only discernable trend was that heteroscedasticity did have an adverse effect on the forecasts. The results obtained from Model 11 are consistently larger than were their counterparts of Model 1.

Table 16.--Comparing the Forecasted Values of the Dependent Variable to the True Values for Different Positive Correlations between X and u for Model 1 and Model 11 (shown in Parentheses).

Correlations	+.9	+.5	+.1
The mean value of:			
$\hat{Y}(21) - Y(21)$	-.0780	-.1550	-.1781
	(-2.8872)	(-5.697)	(-6.5449)
$\hat{Y}(22) - Y(22)$	.0809	.1607	.1847
	(-.6522)	(-1.2958)	(-1.4887)
$\hat{Y}(23) - Y(23)$	.1019	.2024	.2325
	(.4298)	(.8538)	(.9808)
$\hat{Y}(24) - Y(24)$	-.1038	-.2062	-.2369
	(-.0591)	(-.1174	(-.1349)
$\hat{Y}(25) - Y(25)$	.1172	.2277	.2616
	(1.3861)	(-2.7539)	(-3.1641)

Table 17.--Comparing the Forecasted Values of the Dependent Variable to the True Values for Different Negative Correlations between X and u for Model 1 and Model 11 (shown in Parentheses)

Correlations	-.9	-.5	-.1
The mean value of:			
$\hat{Y}(21) - Y(21)$	-.0781	-.1551	-.1782
	(-2.8673)	(-5.6966)	(-6.5449)
$\hat{Y}(22) - Y(22)$	.0808	.1607	.1846
	(-.6523)	(-1.2957)	(-1.4887)
$\hat{Y}(23) - Y(23)$	.1018	.2024	.2325
	(.4297)	(.8538)	(.9809)
$\hat{Y}(24) - Y(24)$	-.10384	-.2062	-.2369
	(-.0592)	(-8.2857)	(-.1349)
$\hat{Y}(25) - Y(25)$	.1146	.2277	.2616
	(-1.3862)	(-2.7540)	(-3.1641)

CHAPTER V

SUMMARY AND CONCLUSION

Summary of Model

Purpose

The purpose of this research paper was to evaluate the combined effects of heteroscedasticity and a correlation between the independent variable and the error term in a single equation model through a Monte Carlo analysis. The correlation between the independent variable and the error term was changed from sample to sample to see whether a high or low correlation, in conjunction with heteroscedasticity, had any abnormal effects over the properties of the sample estimators. The effects of both specification errors were also sought in the area of forecasting.

Model Specifications

The model used was

$$Y_i = +3.64 + .9016(X_i) + u_i$$

where

$$\overline{X} = 287.3$$

$$\text{Var }(X) = 4638$$

$$\text{Var }(u) = 76$$

and

$E(u_i) = 0$

$E(u_i u_j) = 0$ where $i \neq j$

$E(u_i^2) \neq \sigma_u^2$

$E(X_i u_i) \neq 0$

Two models were analyzed and compared. Model 1 contained only one specification error that of $r \neq 0$. The values for r chosen to be sampled were +.9, +.5, +.1, -.1, -.5, and -.9. For each value of r, 50 samples were drawn of size 20. From these samples, the estimated parameters were computed. In Model 11 the second specification error, heteroscedasticity, was introduced, in addition to the specification error of $r \neq 0$. For the same values of r specified in Model 1, 50 samples of size 20 were again computed.[1] The estimated sample parameters were again computed and compared to those of Model 1.

Data Generation

The X's and u's were generated by transforming a matrix of random numbers (S) whose elements (S_{ij}) had a mean equal to zero and a standard deviation of five. The S matrix was premultiplied by a P matrix, such that PP' = M, where M was the population variance-covariance matrix of the X's and u's. Heteroscedasticity was introduced into Model 11 by randomly changing the standard deviation of the second row of the S matrix; the result of which was that the variance of u_i changed for each of every value of X_i.

[1] Model 11 also contains a run where r = 0.

After obtaining the X's and u's it was a simple task for the computer to generate the Y's, the sample estimates and their distributions involved.

Forecasting

An attempt was also made in the area of short term point forecasting. Forecasted values of the dependent variable (Y_i) were compared to the true values obtained for Y_i. The true values of Y were obtained by extending the S matrix so that five additional values of X and u were obtained. With these additional five values of X and u, five additional values of Y were obtained and were the true values of Y were obtained from the sample estimators and from the respective values of X_i.

$$Y_i = a + bX_i$$

where i = 21, 22, 23, 24 and 25.

From these values the differences between the actual and forecasted values of Y were obtained and compared.

Summary of Findings

The analysis suggests that the joint effects of heteroscedasticity and a correlation between the independent variable and the error term may be inferred from their individual effects. Heteroscedasticity had no discernable effect on bias as indicated in Tables 3 to 10. However, heteroscedasticity did effect the dispersions about the biased parameters. The dispersions increased over five times with the inclusion of heteroscedasticity.

The correlation between X and u behaved as expected -- the bias was positive when the correlation between X and u was positive, and was negative when the correlation was negative. For the same correlation, both positive and negative, the absolute value of the bias was identical. Also, the bias became greater as the correlation between X and u increased.

The experiment in forecasting indicated that heteroscedasticity adversely effects the forecasts. The difference between the estimated and actual values of the dependent variable were considerably larger in the model with heteroscedasticity present, even in the absence of a correlation between the independent variable and the error term.

Future Analyses

It would be appropriate at this time to analyze some of the problems related to heteroscedasticity and a correlation between the independent variable and the error term that this research paper did not attempt to analyze. These problems would have to be analyzed prior to knowing the over all effects of the two specification errors covered in this analysis.

First, least squares was the only method used to obtain parameters. Further, only a single equation model was analyzed. It might prove beneficial, if simultaneous equation models were used, with their various methods of estimating sample estimators. These might include indirect least squares, two stage least squares, three stage least squares, least squares with no restrictions, the limited information single equation method, and the full information maximum likelihood method.

Secondly, the parameters were specified a priori in addition to specifying the means of the independent variables and the error term, and the variance-covariance matrix of X and u. Therefore, all of the conclusions obtained from Tables 3 to 17 refer only to specific parameters. Of course, this would require the use of more elaborate procedures.

Thirdly, the analysis was applied only to a linear model. Some form of non-linear analysis might be applied to determine whether the results of this research paper are still valid. The non-linear forms might include some higher form of polynomial than a straight line or even an exponential form.

Fourth, the heteroscedasticity in Model 11 was introduced in a random manner. The variance of the error term changed randomly for each and every value of the independent variable. It might be appropriate to test some other form of heteroscedasticity such as the case in which variance of the error term is directly proportional to the value of the independent variable. Such an analysis would determine whether or not the form of heteroscedasticity had any influence on the results obtained in Tables 3 to 14.

Fifth, Tables 16 and 17 reveal that forecasts were made only five periods into the future. Since most models are made for their ability to forecast, longer duration forecasts might be analyzed to determine the joint effects of heteroscedasticity and a correlation between the independent variable and the error term.

Future analyses might attempt to determine the effects of heteroscedasticity in conjunction with other specification errors, even those which do not directly related to the error terms. Includes might be autocorrelation, multicollinearity, and errors of observation. Also, an attempt should be made to analyze more than two specification errors simultaneously. Such an attempt would be a monumental task to handle mathematically; however, there can be no doubt that this would lead to more insight into the areas of the properties of small estimators and their forecasting ability.

Conclusion

The author started this analysis with the intuitive feeling that heteroscedasticity, in conjunction with a correlation between the independent variable and the error term, would effect bias. The author's intuition proved to be incorrect. Heteroscedasticity did not discernably alter bias but did alter the efficiency of the estimates.

Heteroscedasticity and a correlation between the independent variable and the error term might be expected to behave as the sum of their individual effects.

SELECTED BIBLIOGRAPHY

Books

Aigner, Dennis J. Basic Econometrics. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1971.

Balsley, Howard L. Quantitative Research Methods for Business and Economics. New York: Random House, 1970.

Christ, Carl F. Econometric Models and Methods. New York: John Wiley & Sons, Inc., 1966.

Chu, Kong. Principles of Econometrics. Scranton, Penn.: International Textbook Company, 1968.

Dhrymes, Phoebus J. Econometrics: Statistical Foundations and Applications. New York: Harper & Row, Publishers, 1970.

Goldberger, Arthur S. Econometric Theory. New York: John Wiley & Sons, Inc., 1964.

Hadley, G. Linear Algebra. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1961.

Huang, David S. Regression and Econometric Methods. New York: John Wiley & Sons, Inc., 1970.

Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.

Kane, Edward J. Economic Statistics and Econometrics: An Introduction to Quantitative Economics. New York: Harper & Row, Publishers, 1968.

Klein, Lawrence R. An Introduction to Econometrics. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1962.

Kooros, A. Element of Mathematical Economics. Boston: Houghton Mifflin Company, 1965.

Lang, Serge. Linear Algebra. Reading, Mass.: Addison-Wesley Publishing Company, 1966.

Leser, D. E. V. Econometric Techniques and Problems. London, England: Charles Griffin & Company Limited, 1966.

Malinvaud, E. Statistical Methods of Econometrics. Chicago: Rand McNally & Company, 1966.

Naylor, Thomas H. Computer Simulation Experiments With Models of Economic Systems. New York: John Wiley & Sons, Inc., 1971.

Naylor, Thomas H., Balintfy, Joseph L., Burdick, Donald S., and Chu, Kong. Computer Simulation Techniques. New York: John Wiley & Sons, Inc., 1966.

Rao, Potluri, and Miller, Roger LeRoy. Applied Econometrics. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1971.

Tintner, Gerhard. Methodology of Mathematical Economics and Econometrics. Chicago: University of Chicago Press, 1968.

Tintner, Gerhard, and Millham, Charles B. Mathematics and Statistics for Economists. 2nd. ed. New York: Holt, Rinehart and Winston, Inc., 1970.

Walters, A. A. An Introduction to Econometrics. London, England: MacMillan and Co. Ltd., 1968.

Williams, E. J. Regression Analysis. New York: John Wiley & Sons, Inc., 1959.

Wonnacott, Ronald J., and Wonnacott, Thomas H. Econometrics. New York: John Wiley and Sons, Inc., 1970.

Zellner, Arnold. Readings in Economic Statistics and Econometrics. Boston: Brown and Company, 1968.

Articles

Bartlett, M. S. "The Fitting of Straight Lines When Both Variables Are Subject to Error." Biometrics, V (1949), 207-212.

Chipman, John S. "On Least Squares With Insufficient Observations," Journal of the American Statistical Association, LIX (December, 1964), 1078-1111.

Chipman, John S., and Rao, M. M. "The Treatment of Linear Restrictions in Regression Analysis," Econometrica, XXXII (January-April, 1964), 198-209.

Durbin, J. "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, XXII (January, 1960), 139-153.

Durbin, J., and Watson, G. S. "Testing for Serial Correlation in Least Squares Regression. I," Biometrika, XXXVII (December, 1950), 409-428.

Durbin, J., and Watson, G. S. "Testing for Serial Correlation in Least Squares Regression. II," Biometrika, XXXVIII (June, 1951), 159-178.

Fisher, Franklin M. "The Place of Least Squares in Econometrics: Comment," Econometrica, XXX (July, 1962), 565-567.

Foote, Richard J., and Waugh, Frederick V. "Results of an Experiment to Test the Forecasting Merits of Least Squares and Limited Information Equations," Econometrica, XXVI (November, 1958), 607-608.

Goldfeld, Stephen, M., and Quandt, Richard E. "Some Tests for Homoscedasticity," Journal of the American Statistical Association, LX (June, 1965), 539-547.

Haavelmo, Trygve. "Statistical Implications of a System of Simultaneous Equations," Econometrica, XI (January, 1943), 1-12.

Halperin, Max. "The Fitting of Straight Lines and Prediction When Both Variables Are Subject to Error," Journal of the American Statistical Association, LVI (September, 1961), 657-669.

Husby, Ralph D. "A Nonlinear Consumption Function Estimated from Time-Series and Cross-Section Data," The Review of Economics and Statistics, LX (February, 1971), 76-79.

Klein, Lawrence R. "Estimating Patterns of Savings Behavior from Sample Survey Data," Econometrica, XVIV (October, 1951), 438-454.

Kuh, Edwin, and Meyer, John R. "How Extraneous Are Extraneous Estimates," Review of Economics and Statistics, XXXIX (November, 1957), 380-393.

Ladd, George W. "Effects of Shocks and Errors in Estimation: An Empirical Comparison," Journal of Farm Economics, XXXVIII (May, 1956), 485-495.

Madansky, Albert. "The Fitting of Straight Lines When Both Variables Are Subject to Error," Journal of the American Statistical Association, IX (March, 1959), 173-205.

Neiswanger, W. A., and Yancey, T. A. "Parameter Estimates and Autonomous Growth," Journal of the American Statistical Association, LIV (June, 1959), 389-402.

Orcutt, Guy H., and Cochrane, Donald. "A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis," Journal of the American Statistical Association, XLV (September, 1949), 356-372.

Rao, Potluri, and Griliches, Zvi. "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," Journal of the American Statistical Association, LXIV (March, 1969), 253-272.

Samuelson, Paul A., Koopmans, Tjalling C., and Stone, J. Richard N. "Report of the Evaluative Committee for Econometrica," Econometrica, XXII (April, 1954), 141-146.

Suits, Daniel B. "Forecasting and Analysis With an Econometric Model," American Economic Review, LII (March, 1962), 104-132.

Summers, Robert. "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, XXXIII (January, 1965), 1-41.

Wald, Abraham. "The Fitting of Straight Lines if Both Variables Are Subject to Error," Annals of Mathematical Statistics, XI (1940), 284-300.

Waugh, Frederick V. "The Place of Least Squares in Econometrics," Econometrica, XXX (July, 1961), 386-396.

Wold, H., and Faxer, P. "On the Specification Error in Regression Analysis," Annals of Mathematical Statistics, XXVIII (March, 1957), 265-267.