Staff Paper Series

STAFF PAPER P71-29 DECEMBER 1971

Generalized Least Squares With an Estimated Covariance Matrix – A Sampling Experiment

by
W.E. Griffiths

Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture
St. Paul, Minnesota 55101

STAFF PAPER P71-29 December 1971

GENERALIZED LEAST SQUARES WITH AN ESTIMATED COVARIANCE MATRIX -- A SAMPLING EXPERIMENT

W. E. GRIFFITHS

GENERALIZED LEAST SQUARES WITH AN ESTIMATED COVARIANCE MATRIX--
A SAMPLING EXPERIMENT*

by

W. E. Griffiths

I. INTRODUCTION

In the general linear model, $y = X\beta + u$, $Euu' = \alpha^2 I$, an assumption which may be overoptimistic for many economic applications is that of a vector of coefficients, β, which is constant for all observations. Over time conditions that influence an individual decision unit, a consumption unit or a production unit, are likely to change and this could be reflected by a change in the response coefficients. Alternatively, the diversity of regions and individual decision units, etc. makes it unlikely that each would have the same response coefficients. These considerations have led to the development of so called "random coefficient" models.

---

*Paper to be presented at the Winter meetings of the Econometric Society in New Orleans on December 27-29, 1971.

Hildreth and Houck [6] suggested the following model for situations in which response coefficients can differ with each observation:

$$y_t = \sum_{k=1}^{K} z_{tk}\, b_{tk}$$

$$= \sum_{k=1}^{K} z_{tk}\, (\beta_k + v_{tk})$$

(1.1) $$= \sum_{k=1}^{K} z_{tk}\, \beta_k + u_t,$$

where

(1.2) $$u_t = \sum_{k=1}^{K} z_{tk}\, v_{tk},$$

(1.3) $$b_{tk} = \beta_k + v_{tk},$$

$$z_{t1} \equiv 1,$$

and in each case $k = 1,2,\ldots,K$ and $t = 1,2,\ldots,T$.

The y_t are observed values of a random variable, the z_{tk} are known, nonrandom values of explanatory variables and v_{tk} are unobserved random variables satisfying the following assumptions:

(1.4) $$E\, v_{tk} = 0 \quad \text{for all } t \text{ and } k,$$

and

(1.5) $$E\, v_{tk}\, v_{sj} = \begin{cases} \delta_k & \text{for } t=s \text{ and } k=j, \\ 0 & \text{for } t \neq s \text{ or } k \neq j. \end{cases}$$

At observation t the response of the dependent variable to the k-th explanatory variable is given by b_{tk}. Since b_{tk} is a random variable with mean β_k and variance δ_k, this model relaxes the assumption of fixed response coefficients for each observation. Assumption (1.5) implies that response coefficients associated with different variables for the same observation, or for different observations, are uncorrelated.

Looking at (1.1) and (1.2) we can see that estimation of the mean response coefficients (the β_k) is identical to the problem of estimation of the coefficients in the general linear model where a special kind of heteroskedasticity in the disturbances exists. In this case the disturbances are given by the u_t and the variance of each is

$$(1.6) \qquad \theta_t = E\, u_t^2 = \sum_{k=1}^{K} z_{tk}^2\, \delta_k.$$

Before discussing different estimators it is convenient to put the model in matrix notation. Let y, β and u be column vectors of orders T, K and T respectively and let Z be a TxK matrix so that (1.1) can be written as

$$(1.7) \qquad y = Z\beta + u.$$

Let $V = \text{diagonal}\ (\theta_1, \theta_2, \ldots, \theta_T)$ so that

$$(1.8) \qquad Euu' = V,$$

and let $Z_k = \text{diagonal}\ (z_{1k}, z_{2k}, \ldots, z_{Tk})$, and $b_{.k} = (b_{1k}, b_{2k}, \ldots, b_{Tk})'$.

If the variances are known the best linear unbiased estimator (BLUE) of the mean response coefficients is given by the generalized least squares estimator

$$(1.9) \qquad \hat{\beta} = (Z' V^{-1} Z)^{-1} Z' V^{-1} y.$$

The BLUE of the actual response coefficients has been shown by Griffiths [4] to be

$$(1.10) \qquad \hat{b}_{.k} = \hat{\beta}_k \, \iota_T + \delta_k Z_k V^{-1} \hat{u}, \text{ for all } k,$$

where $\hat{u} = y - Z\hat{\beta}$ and $\hat{\beta}_k$ is the k-th element in $\hat{\beta}$.

For the usual case where the variances are unknown, Hildreth and Houck suggest several alternative consistent estimators for the δ_k, each of which can be used to obtain an estimated covariance matrix, $\hat{V}$, which can be used in place of the real V in (1.9) and (1.10). Under certain assumptions they show that not only are the suggested estimators for the δ_k consistent, but the $\hat{\beta}$'s obtained by substituting the resulting $\hat{V}$'s into (1.9) are also consistent.

Rubin [13] has derived the likelihood equations, which are highly non-linear, for the above model and Hurwicz [7] has given examples where a random coefficient model would be appropriate. The model has also been examined by Theil and Mennes [19] and again by Theil [18, p. 622]. The case in which we have a number of micro units, where each micro unit has a constant coefficient vector over time but across micro units the coefficients are random, and the implications for aggregation, have been studied by Zellner [20] and Theil [17]. Swamy [14] has provided an efficient estimation method for such a model.

The large number of estimators suggested by Hildreth and Houck, both of the variances and the generalized least squares estimators which result from using the estimated variances, have identical large sample properties in that they are all consistent. In terms of small sample properties, some have been shown to have lower mean square errors than others, thus providing some basis for choice. However, to my knowledge, these are the only analytical properties which have been derived for estimators in the Hildreth-Houck model. No basis exists for choosing which estimated covariance matrix, if any, should be used in (1.9) and (1.10) to estimate the mean and actual response coefficients respectively. Since determining analytical properties, which would provide such a basis, appears to be a formidable task, especially in small samples, it seems appropriate that a sampling experiment be conducted in the hope that the results will provide some basis for choice between estimators.

This paper presents the results of such a sampling experiment for a model which may be regarded as a special case of the Hildreth-Houck model. A number of the estimators suggested by Hildreth and Houck, some additional ones, and some test statistics are investigated in the experiment. Froehlich [2] has carried out a similar experiment using a slightly different model and placing more emphasis on estimators (and modifications of these) suggested by Theil and Mennes [19].

The model, estimators and tests which are considered are outlined in section 2. Section 3 outlines the details of the experiment, section 4 gives the results and section 5 contains conclusions and possibilities for future research.

2. THE MODEL, ESTIMATORS AND TESTS

2.1 The Model

The model to be investigated is the following:

(2.1) $$y_t = \sum_{k=1}^{K} z_{tk} \beta_{tk},$$

(2.2) $$\beta_{tk} = \sum_{j=1}^{J} \gamma_{kj} x_{tj} + v_{tk},$$

(2.3) $$E\, v_{tk} = 0,$$

(2.4) $$E\, v_{tk} v_{sj} = \begin{cases} \delta_k & \text{for } t=s \text{ and } k=j \\ 0 & \text{for } t\neq s \text{ or } k\neq j, \end{cases}$$

(2.5) $$z_{t1} \equiv 1, \quad x_{t1} \equiv 1,$$

where $t = 1,2,\ldots,T$; $k=1,2,\ldots,K$; and $j=1,2,\ldots,J$.

This model varies slightly from that of Hildreth and Houck, the difference being that the β_{tk}, instead of having a fixed mean, now have a mean which depends on the values of another set of exogenous variables, $x_{t1}, x_{t2}, \ldots, x_{tJ}$ and a set of fixed coefficients $\gamma_{k1}, \gamma_{k2}, \ldots, \gamma_{kJ}$.[1/]

[1/] This model is similar to one suggested by Lee, Judge and Zellner [11] to allow for transition probabilities which change over time depending on the values of a set of exogenous variables.

In the Hildreth-Houck model it was assumed the factors causing the response coefficients to change were unobservable. If they are observable they can be explicitly included using equation (2.2). Letting $x_{t1} \equiv 1$, for all t implies that when $\gamma_{kj} = 0$, for all k and j except j=1, the model reduces to that of Hildreth and Houck, and so, the Hildreth-Houck model could be regarded as a special case of (2.1) to (2.5). However, for estimation purposes, it will be seen that the above model is better viewed as a special case of the Hildreth-Houck model.

Substituting (2.2) into (2.1) we have

$$(2.6) \qquad y_t = \sum_{k=1}^{K} \sum_{j=1}^{J} z_{tk} x_{tj} \gamma_{kj} + \sum_{k=1}^{K} z_{tk} v_{tk}.$$

If, (a) $i = j + (k-1) J$ and $N = KJ$ so that i is an index from 1 to N; (b) $g_{ti} = z_{tk} x_{tj}$; and (c) $u_t = \sum_{k=1}^{K} z_{tk} v_{tk}$, (2.6) can be written as

$$(2.7) \qquad y_t = \sum_{i=1}^{N} g_{ti} \gamma_i + u_t.$$

Since $x_{t1} \equiv 1$, for all t, the variables z_{tk} (k=1,2,...,K) are a subset of the variables g_{ti} (i=1,2,...,N). Thus (2.7) can be regarded as a special case of the Hildreth-Houck model where some of the δ_i are zero. The problem of estimating the γ_i is equivalent to that of estimating the β_k in (1.1).

Written in matrix form (2.7) becomes

$$(2.8) \qquad y = G\gamma + u$$

where G and γ are of orders TxN and Nx1 respectively. The disturbance vector u has the same properties as it did in the Hildreth-Houck model, namely, Euu' = V, where V is a diagonal matrix containing the elements given in (1.6).

The generalized least squares estimator for γ is given by

$$\hat{\gamma}_G = (G' V^{-1} G)^{-1} G' V^{-1} y, \tag{2.9}$$

and its covariance matrix is

$$V(\hat{\gamma}_G) = (G' V^{-1} G)^{-1}. \tag{2.10}$$

If ordinary least squares is applied to (2.8), the resulting estimator and its covariance matrix are given by

$$\hat{\gamma}_O = (G'G)^{-1} G'y, \tag{2.11}$$

and

$$V(\hat{\gamma}_O) = (G'G)^{-1} G' V G(G'G)^{-1}. \tag{2.12}$$

The problems discussed in section 1 in connection with model (1.7) are again relevant for (2.8). The estimator given in (2.9) is BLUE but typically V is unknown. It is possible to use (2.11) without any knowledge of V but this estimator is inefficient. There are a number of possible estimators for V each of which can be used in place of the real V in (2.9). However, apart from consistency, the analytical properties of the resulting estimators for γ have not been derived. The properties of these estimators and the variance estimators ($\hat{\delta}$'s) will be investigated in the sampling experiment.

2.2 Estimators

The estimators considered are outlined explicitly below. Each was suggested by Hildreth and Houck or is a modification of one of those suggested by Hildreth and Houck.

Let the ordinary least squares residuals be given by

$$\hat{u} = y - G\hat{\gamma}_o = Mu, \tag{2.13}$$

where $M = I - G(G'G)^{-1}G'$. The covariance matrix of these residuals is

$$E\hat{u}\hat{u}' = MVM. \tag{2.14}$$

Let w be a Tx1 vector containing the diagonal elements of $\hat{u}\hat{u}'$, $\dot{M}$ be the matrix obtained by squaring each element in M and θ be a Tx1 vector containing the diagonal elements of V. Then, the diagonal elements of (2.14) can be written as

$$Ew = \dot{M}\theta. \tag{2.15}$$

If $\dot{Z}$ is a matrix containing the squares of the observations on the z variables and δ is a Kx1 vector of the unknown variances, equation (1.6) can be rewritten as

$$\theta = \dot{Z}\delta. \tag{2.16}$$

Substituting this into (2.15) and letting $W = \dot{M}\dot{Z}$, we have

$$Ew = \dot{M}\dot{Z}\delta = W\delta \tag{2.17}$$

Since w and W are observable least squares can be applied to (2.17) to obtain an unbiased estimator of the variances

$$\hat{\delta}(1) = (W'W)^{-1}W'w. \tag{2.18}$$ [2/]

This is the first of the variance estimators to be investigated. For the case where u is normally distributed, the variance of $\hat{\delta}(1)$ has been shown [6] to be

$$V(\hat{\delta}(1)) = 2(W'W)^{-1}W'QW(W'W)^{-1}, \tag{2.19}$$

where Q is the matrix MVM with each of its elements squared.

An undesirable property of $\hat{\delta}(1)$ is that it may contain some negative estimates. A method of overcoming this problem is to set all negative estimates equal to zero. This leads to the following estimator:

$$\hat{\delta}(2) = (\hat{\delta}_1(2), \hat{\delta}_2(2), \ldots, \hat{\delta}_K(2))' \tag{2.20}$$

where

[2/] Although unbiased this estimator is not efficient. Generalized least squares cannot be applied to estimate δ because the covariance matrix of w depends on δ. Froehlich [2] is investigating a two step procedure in an attempt to see whether or not it improves efficiency.

(2.21) $$\hat{\delta}_k(2) = \text{Min } \{0, \hat{\delta}_k(1)\}, \text{ for all } k.$$

Since $\hat{\delta}(1)$ is obtained by minimizing a given sum of squares, a more appealing estimator which excludes the possibility of negative estimates is one which minimizes the same sum of squares subject to a non negativity restriction.

Define $\hat{\delta}(3)$ by

$$\hat{\delta}(3) \geqq 0$$

and

(2.22) $$(w-W\hat{\delta}(3))' \ (w-W\hat{\delta}(3))$$

$$\leqq (w-W\hat{\delta})' \ (w-W\hat{\delta}) \text{ for all } \hat{\delta} \geqq 0.$$

Calculation of this estimator is a quadratic programming problem and has been discussed by Judge and Takayama [8].

The estimators $\hat{\delta}(2)$ and $\hat{\delta}(3)$ are not unbiased, however Hildreth and Houck have shown that they have lower mean square errors than $\hat{\delta}(1)$ and that under certain assumptions they are consistent.

Another unbiased estimator for δ which is obtained by considering (2.17) and which, a priori, cannot be considered better or worse than $\hat{\delta}(1)$ on the basis of a mean square error criterion is

(2.23) $$\hat{\delta}(4) = (\dot{Z}' \dot{M} \dot{Z})^{-1} \dot{Z}'w.$$

When u is normally distributed its variance is

(2.24) $$V(\hat{\delta}(4)) = 2(\dot{Z}'W)^{-1}\ \dot{Z}'Q\dot{Z}(\dot{Z}'W)^{-1}$$

Rao [12] has introduced a term, MINQUE, to denote a variance estimation procedure. This procedure produces minimum norm quadratic unbiased estimators which have been shown by Rao to possess a number of desirable properties. Froehlich [2] and Griffiths [3] have demonstrated that $\hat{\delta}(4)$ is the MINQUE for δ in this model.

As was possible with $\hat{\delta}(1)$, negative estimates can arise using $\hat{\delta}(4)$. This can be overcome by using $\hat{\delta}(5)$ where

(2.25) $$\hat{\delta}_k(5) = \text{Min}\ \{0, \hat{\delta}_k(4)\}.$$

There is a quadratic programming estimator for δ which corresponds to $\hat{\delta}(4)$ just as $\hat{\delta}(3)$ corresponds to $\hat{\delta}(1)$. However this estimator was not examined in the experiment.

All the above estimators use as a first step the ordinary least squares estimator, $\hat{\gamma}_o$, to obtain residuals, $\hat{u}$, to use in estimating the variances. An alternative would be to employ some kind of initial approximation, or guess, of the covariance matrix V, apply generalized least squares and use the resulting residuals. If $\hat{V}_o$ is this initial approximation then one could obtain estimates

(2.26) $$\hat{\gamma}_s = (G'\hat{V}_o^{-1}G)^{-1}\ G'\hat{V}_o^{-1}y,$$

and derive the corresponding residuals,

(2.27) $$\hat{u}_s = y - G\hat{\gamma}_s = Pu$$

where $P = I - G(G'\hat{V}_o^{-1}G)^{-1} G'\hat{V}_o^{-1}$. Then, analogous to (2.14) we have

$$E\hat{u}_s\hat{u}_s' = PVP'. \tag{2.28}$$

The diagonal elements of $\hat{u}_s\hat{u}_s'$ and PVP' can now be used in a similar way to derive a whole new series of variance estimators. Although M is symmetric, P is not and so in some instances M is replaced by P and in others by P'.

Only one of this new series, the estimator analogous to $\hat{\delta}(1)$, was investigated in the experiment. It is given by

$$\hat{\delta}(6) = (W_s'W_s)^{-1} W_s'w_s \tag{2.29}$$

where $W_s = \dot{P}\ \dot{Z}$ and w_s is a vector centering the diagonal elements of $\hat{u}_s\hat{u}_s'$.

The natural question which arises is what to use as an initial approximation for V. The initial approximation implied by the use of ordinary least squares is $\delta_1 > 0$ and all other variances zero, that is $\hat{V}_o = \delta_1 I$. An alternative which may be worth considering is to assume, at the outset, that all the variances are equal. This leads to the initial approximation of

$$\hat{V}_o = \delta\sum_{k=1}^{K} Z_k\ Z_k, \tag{2.30}$$

where δ is the common variance and Z_k is defined just below equation (1.8). If, instead, we assume that the variances are not equal but their ratios are known, that is, (2.31) $$\delta_1 = \lambda_2\, \delta_2 = \ . \ . \ . = \lambda_K\, \delta_K$$

where the λ's are known, then the initial approximation becomes

(2.32) $$\hat{V}_o = \delta_1 \sum_{k=1}^{K} 1/\lambda_k\, Z_k\, Z_k,$$

where $\lambda_1 = 1$.

From each $\hat{\delta}(i)$, $i = 1,2,\ldots,6$, an estimated covariance matrix $\hat{V}(i)$ can be constructed. The i-th generalized least squares estimator for γ which uses an estimated covariance matrix is then given by

(2.33) $$\hat{\gamma}(i) = (G'\hat{V}^{-1}(i)G)^{-1}\, G'\hat{V}^{-1}(i)y \qquad i = 1,2,\ldots,6.$$

Discussion of an initial approximation of V and the estimators for γ which correspond to alternative $\hat{V}$'s suggests the use of an iterative procedure. For any i, residuals can be obtained using the γ estimator given in (2.33), these residuals can be used to obtain another estimate for δ, which can be used to again estimate γ, and so on. The convergence properties of such estimators still need to be investigated. Some light should be shed on the properties by the results of the experiment. The estimators $\hat{\gamma}(7)$ and $\hat{\delta}(7)$ will refer to the estimates obtained after iteration, using initially $\hat{\delta}(1)$ and $\hat{\gamma}(1)$. Some iteration was also carried out using $\hat{\delta}(4)$ and $\hat{\gamma}(4)$ as the initial estimators, but, after some preliminary results, this estimator was abandoned. This is discussed in section 3.

2.3 "t Tests"

If the disturbances are normally distributed and V is known, $\hat{\gamma}_G$ (given in 2.9) is normally distributed with mean γ and covariance matrix $(G'V^{-1}G)^{-1}$ and it is relatively easy to test hypotheses concerning linear functions of γ. Since V is typically unknown it is of interest to see how these tests perform when an estimated V is substituted for the real V.

Under certain conditions[3/] $\hat{\gamma}_G - \gamma$ will have an asymptotic normal distribution with mean zero and covariance matrix $(G'V^{-1}G)^{-1}$. The added assumptions necessary to prove $\sqrt{T}\ (\hat{\gamma}(i) - \hat{\gamma}_G)$ converges in probability to zero still need to be investigated. If this can be shown $\hat{\gamma}(i)$ will have the same asymptotic distribution as $\hat{\gamma}_G$ and tests using $\hat{V}$ in place of V will be asymptotically justified.

The most commonly used test is that where each element of γ is tested to determine whether or not it differs significantly from zero. The statistic, whose performance in testing this hypothesis will be investigaged is given by

$$(2.34) \qquad t = \hat{\gamma}_i(j)/\sqrt{a_{ii}(j)} \qquad i = 1,2,\ldots,N,$$

where $a_{ii}(j)$ is the i-th diagonal element of $(G'\hat{V}^{-1}(j)G)^{-1}$. This test will be carried out in conjunction with all the estimators outlined above, that is, $j = 1,2,\ldots,7$.

[3/] See, for example, Theil [18, Ch. 6,8].

An idea of the probability of a type II error for a given, false, null hypothesis, namely, $\gamma_i = 0$, is obtained using (2.34). Also important is the probability of a type I error, which, under the significance level used, would be 0.05 if the test were an exact one. Since the real parameters are known when a sampling experiment is set up it is possible to investigate this probability using the statistic

$$(2.35) \qquad t = (\hat{\gamma}_i(j) - \gamma_i)/\sqrt{a_{ii}(j)} \;, \qquad i = 1,2,\ldots,N, \quad j = 1,2,\ldots,7.$$

Suppose the model was improperly specified so that it was assumed $Euu' = \delta I$, when in fact $Euu' = V$. It is of interest to determine how sensitive (2.34) and (2.35) are to this misspecification. Assume that the ordinary least squares estimator $\hat{\gamma}_o$ is employed. The denominator in the "t tests" would then be $\sqrt{a_{ii}}$ where a_{ii} is the i-th diagonal element of $\hat{\delta}(G'G)^{-1}$ and $\hat{\delta}$ is the usual estimate of the variance of u_t. The sensitivity of the tests was examined by looking at the performance of (2.34) and (2.35) under these assumptions.

2.4 Testing the Variances

Important hypotheses to test are those concerning the variances (δ's). If δ_1 is the only non zero variance then $\hat{\gamma}_o$ is BLUE. Alternatively, if (2.31) holds where the λ's are known, then a generalized least squares estimator using $\hat{V}_o$ of (2.32) is efficient. It is possible to test such hypotheses using BLUS residuals.

The concept of BLUS residuals was developed by Theil [15] and has subsequently been studied by Koerts [9], Koerts and Abrahamse [10], Abrahamse and Koerts [1] and Theil [16, 18]. A number of theorems concerning BLUS residuals have been developed in this literature. A very brief outline of the procedure is given below.

Suppose we wish to test the hypothesis

$$\delta_k = 0, \quad k = 2,\ldots,K, \tag{2.36}$$

or, in other words, $Euu' = V = \delta_1 I$. The vector u is unobserved and so cannot be used to test this hypothesis. Its estimate $\hat{u}$ given by (2.13) is observable, but, assuming the hypothesis is true, has a covariance matrix given by $E\hat{u}\hat{u}' = \delta_1 M$. Since the elements of $\hat{u}$ are correlated and have different variances ($\delta_1 M$ is not a scalar covariance matrix), it is difficult to use these elements to test hypotheses about the variances. If we transform the residuals using a (T-N)xT matrix B we have

$$\hat{u}_* = B\hat{u}, \tag{2.37}$$

and

$$E\hat{u}_*\hat{u}_*' = \delta_1 BMB'. \tag{2.38}$$

For $\hat{u}_*$ to be the BLUS residuals it is necessary, but not sufficient, that B satisfy the following properties:

(2.39) $$BG = 0,$$

(2.40) $$BMB' = I,$$

(2.41) $$BB' = I,$$

(2.42) $$B'B = M.$$

These conditions show that B is an orthogonal matrix which is also orthogonal to the space spanned by G and that the rows of B are the characteristic vectors corresponding to the unit roots of M. The estimator $\hat{u}_*$ is a linear function of y, an unbiased estimator of (T-N) elements of u and has a scalar covariance matrix because

(2.43) $$E\hat{u}_*\hat{u}_*' = \delta_1 BMB' = \delta_1 I.$$

This gives us the LUS from BLUS.

Because of the multiplicity of the roots of M there are an infinite number of B matrices satisfying (2.39) to (2.42). This leads to the criterion of "best". Since $\hat{u}_*$ is an estimate of only T-N elements of u a decision first must be made concerning which T-N elements are to be estimated. Assume u is partitioned, $u' = [u_0', u_1']$, and that $\hat{u}_*$ is to estimate u_1 where u_1 is, of course, the same order as $\hat{u}_*$. Given this partitioning $\hat{u}_*$ is BLUS when B is chosen such that $E(\hat{u}_* - u_1)'(\hat{u}_* - u_1)$ is a minimum. Assume for the remainder of the paper that B has in fact been chosen this way and that $\hat{u}_*$ is, therefore, a vector of BLUS residuals.[4/]

When $u \sim N(0,\delta_1 I_T)$, it follows that $\hat{u}_* \sim N(0,\delta_1 I_{T-N})$, and so it is relatively easy to use $\hat{u}_*$ to test hypotheses about the disturbances. If we wish to test (2.36) the following F test can be constructed. If T-N is even the following statistics, η_1 and η_2 each have x^2 distribution with (T-N)/2 degrees of freedom.[5/]

$$\eta_1 = 1/\delta_1 \sum_{i=1}^{(T-N)/2} \hat{u}_{*i}^{\,2}, \tag{2.44}$$

$$\eta_2 = 1/\delta_1 \sum_{i=(T-N)/2+1}^{T-N} \hat{u}_{*i}^{\,2}. \tag{2.45}$$

The statistic $F = \eta_1/\eta_2$ has the F distribution with ((T-N)/2, (T-N)/2) degrees of freedom and so can be used to test (2.36).

Two questions immediately come to mind. The BLUS residuals are only unique for a given partitioning of u and so a decision as to which T-N elements of u are to be estimated by $\hat{u}_*$ has to be made. Secondly, what about the power of the test? When the alternative hypothesis is one of heteroskedasticity where the variances are either increasing or decreasing Theil [15] suggests using $\hat{u}_*$ to estimate the first (T-N)/2

[4/]Further properties of BLUS residuals and a convenient way to calculate them are outlined in Theil [18, Ch.5].

[5/]If T-N is odd one of the η's can be calculated from (T-N-1)/2 residuals and the other from (T-N+1)/2.

and the last (T-N)/2 elements in u. If the alternative hypothesis is that the variances are decreasing, and it is true, the $\hat{u}_{*i}$'s in η_1 will be estimates of the u_i's with high variances and the $\hat{u}_{*i}$'s in η_2 will be estimates of the u_i's with low variances. It is hoped, therefore, that by choosing the middle N disturbances as those not to be estimated, putting η_1 in the numerator of the F test and η_2 in the denominator, the power of the test will be made as large as possible.[6/]

In this case the alternative hypothesis is

$$(2.46) \qquad \delta_k > 0 \text{ for at least one k from the set } k=2,\ldots,K.$$

The variances of the disturbances are

$$(2.47) \qquad Eu_t^2 = \sum_{k=1}^{K} z_{tk}^2 \, \delta_k, \; t = 1,2,\ldots,T,$$

and so the greater the values of the z's the greater will be the variance of the disturbance. Thus it is logical to derive $\hat{u}_*$ where it estimates the disturbances associated with the large and the small values of the z's, and omits estimates of the disturbances associated with N medium sized z's. The (T-N)/2 $\hat{u}_{*i}^2$'s corresponding to large z values can be placed in the numerator and the (T-N)/2 $\hat{u}_{*i}^2$'s corresponding to small z values, in the denominator of the F test.

This procedure is somewhat more arbitrary than the situation described by Theil where the variances are simply increasing or decreasing.

[6/] When T-N is odd there is not a unique middle N observations. In this case Theil suggests estimating the two possibilities and choosing between them on another criterion.

Since it is possible for more than one z variable to exist $(K>2)$ there can be no strict ordering of observations from those corresponding to high z values to those corresponding to low z values. However, in the experiment, an attempt was made to classify the observations into the three relevant groups, high, low and medium sized z's, and these were used in the way described above to test (2.36) against (2.46). More research into the optimal partitioning of u and the power of the test when used against alternatives such as (2.46) is needed.

It is straightforward to extend the above procedure to test the hypothesis

(2.48) $$\delta_1 = \lambda_2\delta_2 = \ldots = \lambda_K\delta_K$$

where the λ_k are known. Transform the original variables in the following way:

(2.49) $$y_H = Hy,\ G_H = HG,\ u_H = Hu$$

where H is a TxT diagonal matrix with t-th element given by

(2.50) $$h_t = \left(\sum_{k=1}^{K} 1/\lambda_k\ z_{tk}^2\right)^{-1/2}.$$

When (2.48) holds $Eu_Hu_H' = \delta_1 I$ and so one can proceed just as before using the transformed variables instead of the original ones. For given λ's the performance of this test was examined in the experiment.

3. THE SAMPLING EXPERIMENT

The size of the model and parameter values used in the experiment are given below:

$$y_t = \beta_{t1} + z_{t2}\,\beta_{t2} + z_{t3}\,\beta_{t3}$$

(3.1)

$$\beta_{t1} = 400 + 2.94\,x_t + v_{t1}$$

$$\beta_{t2} = -10.9 - 0.563\,x_t + v_{t2}$$

$$\beta_{t3} = 7.61 + 0.334\,x_t + v_{t3},$$

(3.2) $$v_{tk} \sim N\,(0,\ \delta_k) \quad k=1,2,3,$$

(3.3) $$\delta_1 = 36.0,\ \delta_2 = 1.21,\ \delta_3 = 0.49$$

These values were chosen so that the model could resemble a demand equation in which y_t represents quantity demanded, z_{t2} price, z_{t3} income and x_t is some kind of index of preferences which influences not only quantity demanded but also the response of quantity demanded to changes in income and price.

One hundred samples were generated for each of three sample sizes, T = 14, T = 30 and T = 60. The variables x_t, z_{t2}, z_{t3} were fixed in repeated samples but generated initially from the following distribution:

$$z_{t2} \quad N\ (40,\ 576)$$

$$z_{t3} \quad N\ (65,\ 2031.7)$$

(3.4) $$x_t \quad N\ (100,\ 6400)$$

$$\text{Cov}\ (z_{t2},\ x_t) = \text{Cov}\ (z_{t3},\ x_t) = 0$$

$$\text{Cov}\ (z_{t2},\ z_{t3}) = 101.05$$

The sample means and standard deviations of the values generated are given in Table 1. In Table 2 the population and sample correlation coefficients between all the variables in the G matrix are presented.7/

TABLE 1

Sample Means and Standard Deviations of z_2, z_3 and x

		z_2	z_3	x
Means	T=14	49.90	71.87	98.21
	T=30	44.63	66.37	99.89
	T=60	44.97	71.40	121.99
Standard Devia- tions	T=14	27.26	41.06	51.71
	T=30	23.42	46.91	67.00
	T=60	25.25	45.74	80.74

7/Simplified formulae for deriving population correlation coefficients between a single variable and a quadratic variable can be found in Griffiths [5].

TABLE 2
Correlation Coefficients for all Variables in the G Matrix

		$g_2=x$	$g_3=z_2$	$g_4=z_2x$	$g_5=z_3$
$g_3=z_2$	T=14	-0.143			
	T=30	-0.172			
	T=60	-0.047			
	Pop.	0.0			
$g_4=z_2x$	T=14	0.629	0.657		
	T=30	0.699	0.440		
	T=60	0.636	0.627		
	Pop.	0.721	0.541		
$g_5=z_3$	T=14	-0.022	0.044	0.045	
	T=30	-0.266	-0.054	-0.321	
	T=60	0.012	-0.112	0.000	
	Pop.	0.0	0.093	0.051	
$g_6=z_3x$	T=14	0.700	-0.033	0.526	0.601
	T=30	0.515	-0.202	0.209	0.612
	T=60	0.686	-0.042	0.464	0.525
	Pop.	0.669	0.054	0.531	0.580

Ideally a sampling experiment should be carried out for as many sets of parameters and explanatory variables as possible. This enables one to determine how sensitive the performance of different estimators is to different sets of parameters. It is unlikely that any one estimator will be best for the whole parameter space. In fact, once the parameters and explanatory variables have been set, one can choose between two of the variance estimators $\hat{\delta}(1)$ and $\hat{\delta}(4)$ by comparing their respective variances, (2.19) and (2.24). This does not involve the generation of any samples. Despite the limitations, the set of parameters and explanatory variables outlined above was the only set studied in this paper. It was decided to look at a fairly large number of estimators and tests at the expense of not examining sensitivity to the parameter space. Fortunately, the results seem to indicate a set of inadmissable estimators leaving a smaller number upon which future research can concentrate.

The estimators for δ compared in the experiment are those outlined in section 2, namely $\hat{\delta}(i)$, $i=1,2,\ldots,7$. Those estimators for γ which are evaluated are $\hat{\gamma}(i)$, $i=1,2,\ldots,7$, $\hat{\gamma}_O$, $\hat{\gamma}_S$ and $\hat{\gamma}_G$. All are compared on the basis of their mean square errors (MSE's). Since each estimator is a vector, its MSE is a matrix. However, just the diagonal elements of the MSE matrices were looked at, or, in other words, the MSE of each element in each estimator was estimated. For example, the MSE of each element in $\hat{\gamma}(j)$ is estimated by

$$(3.5) \qquad MSE(\hat{\gamma}_i(j)) = 1/100 \sum_{s=1}^{100} (\hat{\gamma}_{is}(j) - \gamma_i)^2$$

where $i=1,2,\ldots,6$; $j=1,2,\ldots,7$ and s refers to the estimate in the s-th sample.

It is possible to derive the MSE's of some of the estimators analytically and it is therefore unnecessary to estimate these using a formula such as (3.5). The MSE's of $\hat{\gamma}_G$, $\hat{\gamma}_O$, $\hat{\gamma}(1)$ and $\hat{\gamma}(4)$ are given by equations (2.10), (2.12), (2.19) and (2.24) respectively. The MSE's of $\hat{\gamma}_S$ and $\hat{\gamma}(6)$ are respectively

$$(3.6) \qquad (G'\hat{V}_o^{-1}G)^{-1} \; G'\hat{V}_o^{-1} \; V \; \hat{V}_o^{-1}G \; (G'\hat{V}_o^{-1}G)^{-1}$$

and

$$(3.7) \qquad 2(W_s'W_s)^{-1} \; W_s' \; Q_sW_s \; (W_s'W_s)^{-1},$$

where Q_s is the matrix PVP' with each of its elements squared. Before any samples were generated these MSE's were calculated using the above expressions. The MSE's of the remaining estimators and $\hat{\gamma}_o$ and $\hat{\delta}(4)$ were estimated using (3.5), or a similar expression. By comparing the estimated MSE's of $\hat{\gamma}_o$ and $\hat{\delta}(4)$ with the true ones given by (2.12) and (2.24) we can get an idea of the error involved when using this kind of sampling experiment.

For estimators $\hat{\delta}(6)$, $\hat{\gamma}_s$ and $\hat{\gamma}(6)$ we need an initial guess concerning the ratios of the variances. The initial guess used was

$$(3.8) \qquad \delta_1 = 100\delta_2 = 100\delta_3.$$

The estimators obtained using iteration are $\hat{\delta}(7)$ and $\hat{\gamma}(7)$ where the initial values in the iterative procedure are produced using $\hat{\delta}(1)$ and $\hat{\gamma}(1)$. At first, iteration was carried out using $\hat{\delta}(4)$ and $\hat{\gamma}(4)$ as the initial values. For each sample 10 iterations were allowed and if the estimates for γ had not converged after 10 iterations the estimates at the 10th iteration were used. The estimator $\hat{\gamma}$ was said to have converged if the third digit of each element in $\hat{\gamma}$ after t iterations did not differ by more than 1 from the third digit of the corresponding element of $\hat{\gamma}$ after t-1 iterations. However the results were far from satisfactory. When the estimates showed no signs of convergence they fluctuated violently, or exploded, with the result that the MSE's of the resulting estimator were huge. It was decided, therefore, to use $\hat{\delta}(1)$ and $\hat{\gamma}(1)$ as the initial values in the iterative procedure and to separate the results into two parts:

(a) estimates obtained using the results of every sample, that is, using the results after convergence or after 10 iterations whichever comes first; and

(b) estimates based only on those samples where γ estimates converged before ten iterations. Although the number of samples in which convergence occurred was no greater when using $\hat{\delta}(1)$ and $\hat{\gamma}(1)$, the estimates did not fluctuate as wildly when convergence did not take place, and, although still large, the MSE's obtained for part (a) were a more reasonable size.

The performance of "t tests" associated with each estimator was examined by calculating, using (2.35) and (2.34), the number of type I and type II errors in 100 samples. Since it is possible for $\hat{\delta}(i)$, i=1,4,6 and 7, to contain negative estimates it is also possible that the corresponding $\hat{V}(i)$, whose diagonal elements are given by $\hat{\theta}_t(i) = \sum_{k=1}^{K} z_{tk}^2\, \hat{\delta}_k(i)$, is not positive definite. Thus estimates of the variances of the elements of $\hat{\gamma}(i)$, i=1,4,6 and 7, could be negative. When this occurred it was recorded as both a type I and type II error.

The BLUS residuals are used to test the following two hypotheses:

$$\text{(3.9)} \qquad \text{H1:} \quad \delta_2 = \delta_3 = 0,$$

and

$$\text{(3.10)} \qquad H2: \quad \delta_1 = 100\delta_2 = 100\delta_3.$$

Since both these hypotheses are false the performance of the tests was examined by counting the number of samples in which each hypothesis was accepted, that is, the type II errors.

4. RESULTS OF THE SAMPLING EXPERIMENT

4.1 Variance Estimators

The means and MSE's of the variance estimators are presented in Table 3. Perhaps the most striking thing in Table 3 is the large means for $\hat{\delta}_1$ whose true value is 36. One might suspect $\hat{\delta}(4)$ is not unbiased when for δ_1, for different sample sizes, it produces means of -770, 541 and 96. However, because of the very large MSE's of $\hat{\delta}_1$, such values are quite possible. The large means for $\hat{\delta}_1$ using other estimators can be explained by their bias, caused by the preclusion of negative estimates.

We would expect $\hat{\delta}(2)$, $\hat{\delta}(3)$ and $\hat{\delta}(5)$ to be biased upward. This is so in all cases for δ_1, most cases for δ_3 but in only some cases for δ_2. This is perhaps because negative estimates were most frequent for δ_1, then δ_3 and least frequent for δ_2.

The estimator $\hat{\delta}(7)(a)$ stands out as unusual especially with respect to its MSE's. Recall that this is the estimator which includes all estimates obtained from the iterative procedure whether they had converged

TABLE 3

Means and MSE's of Alternative Variance Estimators

Estimator		$\hat{\delta}_1$		$\hat{\delta}_2$		$\hat{\delta}_3$	
		Mean	MSE[a/]	Mean	MSE	Mean	MSE
True Values		36.0		1.21		0.49	
T=14							
$\hat{\delta}(1)$	t[b/]	36.	4.402	1.21	2.9361	0.49	1.8094
$\hat{\delta}(2)$	e	1687.	1.389	1.269	2.3140	0.761	1.5224
$\hat{\delta}(3)$	e	583.	0.193	0.819	0.8106	0.583	0.7069
$\hat{\delta}(4)$	t	36.	4.432	1.21	2.9318	0.49	1.5932
$\hat{\delta}(4)$	e	-770.	5.481	1.184	2.3292	0.649	2.9788
$\hat{\delta}(5)$	e	1791.	1.379	1.234	2.1688	0.836	2.6520
$\hat{\delta}(6)$	t	36.	4.542	1.21	2.7697	0.49	1.6947
$\hat{\delta}(7)$ (a)	e	-18929.	1.16×10^4	17.87	73×10^3	-3.84	14×10^2
$\hat{\delta}(7)$ (b)	e	3072.	2.739	0.901	2.9853	0.025	0.4566
T=30							
$\hat{\delta}(1)$	t	36.	1.6454	1.21	2.0050	0.49	0.2928
$\hat{\delta}(2)$	e	1794.	0.7703	1.107	1.1744	0.523	0.2543
$\hat{\delta}(3)$	e	1264.	0.3863	0.952	0.8043	0.450	0.1670
$\hat{\delta}(4)$	t	36.	1.7746	1.21	2.0903	0.49	0.3032
$\hat{\delta}(4)$	e	541.	1.4278	1.038	1.4315	0.518	0.3225
$\hat{\delta}(5)$	e	1793.	0.7328	1.119	1.1702	0.543	0.2913
$\hat{\delta}(6)$	t	36.	2.4206	1.21	2.3112	0.49	0.3782
$\hat{\delta}(7)$ (a)	e	243.	20.4334	1.511	34.0737	0.405	7.4406
$\hat{\delta}(7)$ (b)	e	2451.	1.0673	0.686	0.7218	0.267	0.1817
T=60							
$\hat{\delta}(1)$	t	36.	0.8717	1.21	0.7781	0.49	0.1168
$\hat{\delta}(2)$	e	1085.	0.3144	1.246	0.5902	0.482	0.1065
$\hat{\delta}(3)$	e	1005.	0.2623	1.105	0.3867	0.430	0.0744
$\hat{\delta}(4)$	t	36.	1.1613	1.21	0.8733	0.49	0.1474
$\hat{\delta}(4)$	e	96.	0.8701	1.229	0.6169	0.479	0.1261
$\hat{\delta}(5)$	e	1114.	0.3324	1.232	0.6093	0.482	0.1230
$\hat{\delta}(6)$	t	36.	1.5468	1.21	1.0600	0.49	0.1885
$\hat{\delta}(7)$ (a)	e	420.	8.9512	1.031	4.4300	0.629	7.5643
$\hat{\delta}(7)$ (b)	e	1874.	0.6940	0.863	0.4987	0.317	0.1151

a/ The MSE's of $\hat{\delta}_1$ have been divided by 10^7.

b/ A "t" means these values are the true values, an "e" means they were estimated in the sampling experiment.

or not. It performs consistently worse than any of the others with respect to both variance estimation and estimation of γ. When the non convergent estimates are excluded, the resulting estimator, $\hat{\delta}(7)(b)$, produces results more comparable to the other estimators. Thus, the iterative procedure gives far from satisfactory results unless the estimates converge. The number of samples upon which the results for $\hat{\delta}(7)(b)$ and $\hat{\gamma}(7)(b)$ are based, (that is, the number of samples out of 100 in which convergence took place), are 11, 50 and 58 for T=14, 30 and 60 respectively.

A comparison of MSE's shows that there is insufficient evidence to provide a basis for choice between $\hat{\delta}(1)$ and $\hat{\delta}(4)$. Even if there was a great difference between the two no general conclusion could be drawn since the MSE's are known as soon as the parameters and explanatory variables have been set for the experiment. Also, a choice is not possible on the basis of the number of negative estimates each produces. These are given in Table 4.

TABLE 4

Number of Samples in Which Negative Variance Estimates Occurred

	$\hat{\delta}(1)$	$\hat{\delta}(4)$
T=14	88	87
T=30	70	73
T=60	45	46

An analytical result shown by Hildreth and Houck is borne out in the results. The MSE's of estimators precluding negative variances are always lower than their corresponding estimators in which negative estimates can occur. An interesting result which has not been shown is the quadratic programming estimator $\hat{\delta}(3)$ performs consistently better than any of the other except for $\hat{\delta}_3(7)(b)$ for T=14.

A result which one would not suspect is that MSE $(\hat{\delta}_1(3))$ for T=14 is lower than that for T=30 and T=60. A possible explanation is that the number of negative variance estimates was much greater for T=14, and the use of quadratic programming estimates therefore allowed a greater reduction in the MSE. Alternatively, the result could be due to chance. The difference between the estimated and true values for MSE $(\hat{\delta}_1(4))$ shows that the variances of the estimates of the MSE's could be quite high.

Based on these results it appears $\hat{\delta}(3)$ is "better" than any of the others. However, one may prefer $\hat{\delta}(2)$ or $\hat{\delta}(5)$ if the added computations are considered. The economist is more likely to be interested in the γ estimates obtained by using these variance estimates than the variance estimates themselves. Therefore, the choice of variance estimator may be based on the resulting γ estimator. These are considered next.

4.2 Estimators for γ

The estimated means of the γ estimators are presented in Table 5. Because $\hat{\gamma}_G$, $\hat{\gamma}_O$ and $\hat{\gamma}_s$ are unbiased, their means are known, a priori. However the mean of $\hat{\gamma}_O$ was still estimated and this estimate is included in Table 5.

Except perhaps for $\hat{\gamma}(7)(a)$ there is insufficient evidence to conclude that any of the estimators are biased. The estimated bias is large in some instances but in these cases there is a correspondingly high MSE which makes the observed bias quite probable. The MSE's are given in Table 6.

Turning to the MSE's we see that $\hat{\gamma}_G$ has a lower MSE than all other estimators. This is as one would expect unless one of the estimators was biased and had a greatly reduced variance. Ignoring $\hat{\gamma}_G$, which is typically unattainable, we can readily divide the estimators into two distinct groups based on their MSE's.

Group A	Group B
$\hat{\gamma}_O$	$\hat{\gamma}(1)$
$\hat{\gamma}_s$	$\hat{\gamma}(4)$
$\hat{\gamma}(2)$	$\hat{\gamma}(6)$
$\hat{\gamma}(3)$	$\hat{\gamma}(7)(a)$
$\hat{\gamma}(5)$	
$\hat{\gamma}(7)(b)$	

Within each group there is no estimator which has lower MSE than any of the others for all elements in γ and for all sample sizes.[8] Every

---

[8] $\hat{\gamma}_s$ is consistently better than the others in Group A except for $MSE(\hat{\gamma}_3(7)(b)) < MSE(\hat{\gamma}_{3_s})$ for T=60.

TABLE 5

Estimated Means for $\hat{\gamma}$'s

Estimator	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
True Values	400.	2.94	-10.9	-0.563	7.61	0.334
T=14						
$\hat{\gamma}_0$	413.8	2.802	-11.45	-0.5228	7.716	0.3280
$\hat{\gamma}(1)$	423.7	2.703	-11.82	-0.4534	7.817	0.2970
$\hat{\gamma}(2)$	402.9	2.889	-11.06	-0.5593	7.604	0.3413
$\hat{\gamma}(3)$	399.7	2.910	-10.95	-0.5702	7.567	0.3459
$\hat{\gamma}(4)$	426.7	2.626	-11.24	-0.5319	7.373	0.3612
$\hat{\gamma}(5)$	401.5	2.901	-11.05	-0.5615	7.615	0.3410
$\hat{\gamma}(6)$	440.3	2.541	-11.27	-0.5653	7.272	0.3892
$\hat{\gamma}(7)$ (a)	-6787.	85.73	113.3	-5.96	13.93	-6.81
$\hat{\gamma}(7)$ (b)	295.3	3.738	-10.21	-0.6086	8.349	0.2711
T=30						
$\hat{\gamma}_0$	413.0	2.887	-11.06	-0.5566	7.546	0.3346
$\hat{\gamma}(1)$	405.0	2.900	-10.24	-0.5751	7.207	0.3337
$\hat{\gamma}(2)$	412.5	2.890	-10.99	-0.5612	7.517	0.3369
$\hat{\gamma}(3)$	410.7	2.915	-10.96	-0.5659	7.522	0.3367
$\hat{\gamma}(4)$	495.4	4.061	-26.18	0.0318	16.839	-0.2801
$\hat{\gamma}(5)$	414.4	2.869	-11.03	-0.5573	7.511	0.3377
$\hat{\gamma}(6)$	605.7	0.841	-9.13	-0.5074	3.822	0.5904
$\hat{\gamma}(7)$ (a)	227.5	4.927	-9.55	-0.8474	9.249	0.1910
$\hat{\gamma}(7)$ (b)	394.0	2.974	-10.87	-0.5638	7.676	0.3277
T=60						
$\hat{\gamma}_0$	399.9	2.929	-10.88	-0.5631	7.598	0.3357
$\hat{\gamma}(1)$	455.3	2.414	-12.04	-0.4863	7.516	0.3688
$\hat{\gamma}(2)$	398.7	2.915	-10.86	-0.5629	7.592	0.3382
$\hat{\gamma}(3)$	398.4	2.919	-10.87	-0.5631	7.600	0.3375
$\hat{\gamma}(4)$	389.1	1.990	-9.46	-0.4635	5.749	0.4905
$\hat{\gamma}(5)$	398.2	2.920	-10.86	-0.5638	7.597	0.3377
$\hat{\gamma}(6)$	391.0	2.291	-11.55	-0.4190	8.235	0.3264
$\hat{\gamma}(7)$ (a)	392.3	2.969	-10.82	-0.5317	7.568	0.3183
$\hat{\gamma}(7)$ (b)	397.4	2.929	-10.91	-0.5587	7.597	0.3373

TABLE 6

MSE's for $\hat{\gamma}$'s

Estimator		$\hat{\gamma}_1$[b]	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
T=14							
G	t[a]	24.37	1.400	7.002	0.0582	2.225	0.0186
o	t	33.94	2.083	10.665	0.0925	2.369	0.0200
o	e	27.35	1.680	10.089	0.0896	2.341	0.0206
s	t	25.52	1.435	7.489	0.0618	2.257	0.0190
(1)	e	288.1	19.45	221.8	2.79	28.6	0.6534
(2)	e	25.42	1.436	9.951	0.0843	2.412	0.0217
(3)	e	26.33	1.469	10.639	0.0901	2.476	0.0230
(4)	e	150.6	11.77	45.59	0.4814	8.556	0.0501
(5)	e	25.80	1.456	9.883	0.0838	2.378	0.0213
(6)	e	356.6	26.74	82.7	0.8082	11.089	0.1110
(7) (a)	e	76×10^5	86×10^4	20×10^5	66×10^2	28×10^3	51×10^2
(7) (b)	e	28.30	1.743	11.04	0.1005	2.145	0.0335
T=30							
G	t	4.969	0.3966	1.0460	0.0102	0.5132	0.0032
o	t	7.712	0.5568	1.4294	0.0126	0.6978	0.0043
o	e	9.635	0.5980	1.7632	0.0136	0.8006	0.0047
s	t	5.320	0.4266	1.0856	0.0108	0.5540	0.0035
(1)	e	87.612	10.91	33.1	0.4440	11.1	0.0632
(2)	e	8.243	0.5654	1.5353	0.0135	0.7536	0.0047
(3)	e	9.521	0.7193	1.7789	0.0188	0.9000	0.0053
(4)	e	996.8	150.2	19×10^3	29.7	72×10^2	31.57
(5)	e	8.208	0.5539	1.5461	0.0133	0.7279	0.0045
(6)	e	13×10^2	129.4	250.	1.233	517.	2.25
(7) (a)	e	15×10^2	150.8	245.	4.989	98.	0.7186
(7) (b)	e	8.748	0.6945	1.3692	0.0144	0.8288	0.0058
T=60							
G	t	0.502	0.0304	0.4814	0.0024	0.0686	0.00043
o	t	1.350	0.0693	0.7048	0.0033	0.1376	0.0008
o	e	1.423	0.0863	0.7923	0.0043	0.1472	0.0007
s	t	0.523	0.0314	0.5061	0.0025	0.0704	0.00050
(1)	e	498.	34.	185.	0.8040	14.	0.1871
(2)	e	0.667	0.0459	0.5532	0.0033	0.0992	0.0005
(3)	e	0.636	0.0454	0.5570	0.0032	0.1034	0.0006
(4)	e	20.	47.	103.	0.4491	229.	1.4218
(5)	e	0.668	0.0478	0.5570	0.0033	0.1056	0.0006
(6)	e	29.	18.	12.	0.7761	20.	0.0646
(7) (a)	e	26.	1.1364	13.	0.2551	2.2078	0.0468
(7) (b)	e	0.757	0.0522	0.5052	0.0035	0.1229	0.0005

[a]/As in the variance estimators "t" refers to true and "e" to estimated.

[b]/The MSE of $\hat{\gamma}_1$ has been divided by 10^3.

estimator in Group A is better than every estimator in Group B for every sample size, and, the difference is considerable.

A factor that all estimators in Group B have in common is the use of an estimated covariance matrix, $\hat{V}$, which may not be positive definite. With the exception of $\hat{\gamma}(7)(b)$ all estimators in Group A use $\hat{V}$'s which must be positive definite. This would lead one to conclude that the use of a non positive definite $\hat{V}$ causes the MSE of the corresponding $\hat{\gamma}$ to explode. In the case of $\hat{\gamma}(7)(b)$, when variance estimates converged to negative quantities, they were usually very small. Thus, when the diagonal elements of $\hat{V}$ (given by $\hat{\theta}_t = \hat{\delta}_1 + z_{t2}^2\ \hat{\delta}_2 + z_{t3}^2\ \hat{\delta}_3$) were formed, it is quite likely, although not explicitly looked at, that the positive variance estimates outweighed the negative ones. This would result in a positive definite $\hat{V}$ for $\hat{\gamma}(7)(b)$.

It seems reasonable, on the basis of these results, to identify Group B as a set of inadmissable estimators and to consider in the future only estimators which use $\hat{V}$'s which must be positive definite.

Can anything be said with respect to choosing between the estimators in Group A? Based on one experiment it is difficult to give any general conclusions but it is worthwhile summarizing the results for this special case. In particular, are any of the $\hat{\gamma}$'s which use variance estimates better than ordinary least squares? It appears that when the sample size is large (T=60) it is worth the added computations to calculate variance estimates but when T=14 or 30 $\hat{\gamma}(2)$, $\hat{\gamma}(3)$ and $\hat{\gamma}(5)$ are not noticeably better than ordinary least squares. Also, no pattern emerges when $\hat{\gamma}(2)$, $\hat{\gamma}(3)$ and $\hat{\gamma}(5)$ are compared with each other.

The estimator which does perform much better than any of the others in Group A is $\hat{\gamma}_s$. In fact its MSE is quite close to that of the generalized least squares estimator. Unfortunately the relative position of this estimator depends on the initial guess made about the ratios of the variances. The guess made in this case was $\delta_1 = \lambda_2\delta_2 = \lambda_3\delta_3$ where $\lambda_2 = \lambda_3 = 100$ when in fact the real values are $\lambda_2 = 29.8$ and $\lambda_3 = 73.5$. Either this was a "good guess" or $\hat{\gamma}_s$ is not very sensitive to the values for λ_2, λ_3 selected. If the researcher has any a priori idea of the ratios of the variances he may be well advised to use this estimator. An interesting point to note is that although $\hat{\gamma}_s$ was a relatively good estimator for γ, the variance estimator $\hat{\delta}(6)$, obtained from using residuals $\hat{u}_s = y - G\hat{\gamma}_s$, was not better and in most cases worse than $\hat{\delta}(1)$ and $\hat{\delta}(4)$, derived from ordinary least squares residuals.

4.3 Results of the Tests

The results of the "t tests" indicate again that one should avoid any estimators which use a $\hat{V}$ which need not be positive definite. In Tables 7, 8 and 9 a large number of type I and II errors are shown for all Group B estimators and for every sample size. This suggests that negative elements in the diagonal of $(G'\hat{V}^{-1}(i)G)^{-1}$ may have occurred quite frequently.

For T=30 and T=60 the tests associated with estimators in Group A performed remarkably well. It is therefore quite possible that the tests are asymptolically justified and more research into this is needed.

TABLE 7

Performance of "t Tests" for T=14

Estimator	Error Type	Number of Errors					
		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
$\hat{\gamma}_0$	I	1	1	3	3	16	11
	II	54	71	11	70	0	23
$\hat{\gamma}(1)$	I	38	41	51	53	35	**36**
	II	65	66	42	64	30	62
$\hat{\gamma}(2)$	I	1	1	6	7	5	5
	II	55	58	17	60	8	50
$\hat{\gamma}(3)$	I	8	6	18	15	11	11
	II	39	42	7	42	4	33
$\hat{\gamma}(4)$	I	33	33	39	44	38	35
	II	64	67	38	63	29	61
$\hat{\gamma}(5)$	I	1	1	4	3	6	5
	II	57	58	14	63	7	44
$\hat{\gamma}(6)$	I	37	35	50	54	33	36
	II	67	59	40	66	29	59
$\hat{\gamma}(7)$ (a)	I	59	59	65	69	50	51
	II	72	75	62	81	47	75
$\hat{\gamma}(7)$ (b)[a/]	I	2	1	5	5	5	5
	II	7	4	4	7	4	8

[a/] These are the number of errors which occurred in 11 samples.

TABLE 8

Performance of "t Tests" for T=30

Estimator	Error Type	Number of Errors					
		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
$\hat{\gamma}_0$	I	12	9	14	7	12	9
	II	1	3	0	0	0	0
$\hat{\gamma}(1)$	I	27	26	26	19	22	22
	II	9	7	9	7	6	5
$\hat{\gamma}(2)$	I	8	9	6	5	10	10
	II	0	4	0	1	0	0
$\hat{\gamma}(3)$	I	12	15	11	11	13	12
	II	0	3	0	0	0	0
$\hat{\gamma}(4)$	I	23	24	31	24	26	23
	II	11	16	9	13	7	7
$\hat{\gamma}(5)$	I	7	7	6	4	9	8
	II	0	5	0	1	0	0
$\hat{\gamma}(6)$	I	28	25	26	27	30	27
	II	15	13	5	12	11	8
$\hat{\gamma}(7)$ (a)	I	36	34	30	32	37	37
	II	23	22	14	17	20	20
$\hat{\gamma}(7)$ (b)[a]	I	5	5	5	5	6	6
	II	1	1	0	0	0	0

[a] These are the number of errors in 50 samples.

TABLE 9

Performance of "t Tests" for T=60

Estimator	Error Type	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
		Number of Errors					
$\hat{\gamma}_0$	I	0	5	6	11	7	3
	II	0	0	0	0	0	0
$\hat{\gamma}(1)$	I	35	35	22	26	35	29
	II	25	20	6	5	11	10
$\hat{\gamma}(2)$	I	3	3	3	9	8	8
	II	0	0	0	0	0	0
$\hat{\gamma}(3)$	I	3	6	6	9	9	9
	II	0	0	0	0	0	0
$\hat{\gamma}(4)$	I	29	29	21	27	30	23
	II	21	16	4	6	12	6
$\hat{\gamma}(5)$	I	2	3	3	9	11	9
	II	0	0	0	0	0	0
$\hat{\gamma}(6)$	I	39	35	28	29	40	28
	II	27	18	7	6	17	8
$\hat{\gamma}(7)$ (a)	I	32	33	29	30	39	28
	II	26	22	14	12	18	15
$\hat{\gamma}(7)$ (b)[a]	I	5	5	4	6	10	4
	II	2	2	0	0	1	1

[a] These are the number of errors in 58 samples.

Another interesting point is that for T=30 and 60 tests based on $\hat{\gamma}_0$ were quite insensitive to the misspecification. One could make an incorrect assumption about the model and the covariance matrix of $\hat{\gamma}_0$ and still, in a high number of cases, come to the correct decision when testing hypotheses.

The performance of the tests using BLUS residuals to test (3.9) and (3.10) performed very badly. As shown in Table 10 in almost no cases were the false hypotheses rejected. If this test, or a similar one using BLUS residuals, is to be used to test such hypotheses its power needs to be more thoroughly investigated.

TABLE 10

Number of Samples in Which Type II Errors Occurred When Testing the Variances

Hypothesis	T=14	T=30	T=60
$\delta_k=0$ $(k\neq 1)$	89	98	99
$\delta_1=100$ $\delta_2=100$ δ_3	89	98	97

5. CONCLUSIONS

A general conclusion one can make is that when using the above model and estimators, negative variance estimates, if permitted, occur quite frequently. These estimates have a very adverse effect on the MSE of an estimator of response coefficients which uses them in an attempt to improve efficiency. They also make testing hypotheses about the response coefficients a difficult task.

When negative variances are excluded it is difficult to derive general conclusions when only one point in the parameter space is considered. However, the results of the study indicate that little efficiency is to be gained over ordinary least squares in small samples. In larger samples non negative variance estimates may be worth the added computation. If one is interested in γ only, $\hat{\gamma}(2)$, where negative estimates are changed to zero, gives added efficiency with little extra computation. However the quadratic programming estimator $\hat{\delta}(3)$ may be worth the extra calculation if one is also interested in a "good" variance estimate.

In terms of MSE the MINQUE estimator and estimators associated with it do not perform any better than the others.

The iterative procedure was far from satisfactory since convergence did not occur in a large number of cases. However, when convergence did occur the results were quite reasonable. An iterative estimator which does not permit negative estimates might be worth considering in the future. The negative variances do appear to be the cause of many problems and it is quite possible that changing these to zero will increase the number of times convergence occurs and decrease the MSE of these estimates which do converge.

There are also a large number of other unanswered questions. More comparisons of estimators in Group A using analytical properties and/or simulation is needed. The conditions necessary for $\sqrt{T}(\hat{\gamma}(i)-\hat{\gamma}_G)$ to converge to zero in probability also need to be investigated. When this condition holds $\hat{\gamma}(i)$ is asymptotically normally distributed and the "t tests" will be justified in large samples. Since the tests using BLUS residuals did not give satisfactory results an investigation of the optimal partitioning of the disturbance vector and the power of the test would be advantageous.

REFERENCES

[1] Abrahamse, A.P.J. and J. Koerts: "A Comparison between the Power of the Durbin-Watson Test and the Power of the BLUS Test", Journal of the American Statistical Association (September, 1969), 938-948.

[2] Froehlich, B.R.: "Estimation of a Random Coefficient Regression Model", Unpublished doctoral dissertation, University of Minnesota, 1971.

[3] Griffiths, W.E.: "Estimation of Regression Coefficients which Change Over Time", Unpublished doctoral dissertation, University of Illinois, 1971.

[4] Griffiths, W.E.: "Estimation of Actual Response Coefficients in the Hildreth-Houck Random Coefficient Model", University of Minnesota, Department of Agricultural and Applied Economics Staff Paper P 71-12, July, 1971.

[5] Griffiths, W.E.: "Multicollinearity in Regression with Quadratic Regressors", University of Minnesota, Department of Agricultural and Applied Economics Staff Paper P 71-13, August, 1971.

[6] Hildreth, C. and J.P. Houck: "Some Estimators for a Linear Model with Random Coefficients", Journal of the American Statistical Association (1968), 584-595.

[7] Hurwicz, L.: "Systems with New Additive Disturbances", Ch. 18 in Koopmans, T.C., editor, Statistical Inference in Dynamic Economic Models. New York: John Wiley and Sons, Inc., 1950.

[8] Judge, G.G. and T. Takayama: "Inequality Restrictions in Regression Analysis", Journal of the American Statistical Association (1966), 166-181.

[9] Koerts, J.: "Some Further Notes on Disturbance Estimates in Regression Analysis", Journal of the American Statistical Association (1967), 169-183.

[10] Koerts, J. and A.P.J. Abrahamse: "On the Power of the BLUS Procedure", Journal of the American Statistical Association (December, 1968), 1227-1236.

[11] Lee, T.C., G.G. Judge and A. Zellner: Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data. Amsterdam: North Holland, 1970.

[12] Rao, C.R.: "Estimation of Heteroscedastic Variances in Linear Models", Journal of the American Statistical Association (1970), 161-172.

[13] Rubin, H.: "Note on Random Coefficients". Ch. 19 in Koopmans, T.C., editor, Statistical Inference in Dynamic Economic Models. New York: John Wiley and Sons, Inc., 1950.

[14] Swamy, P.A.V.B.: "Efficient Inference in a Random Coefficient Regression Model", Econometrica (March, 1970), 311-323.

[15] Theil, H.: "The Analysis of Disturbances in Regression Analysis", Journal of the American Statistical Association (1965), 1067-1089.

[16] Theil, H.: "A Simplification of the BLUS Procedure for Analyzing Regression Disturbances", Journal of the American Statistical Association (March, 1968), 242-251.

[17] Theil, H.: "Consistent Aggregation of Micromodels with Random Coefficients", University of Chicago Center for Mathematical Studies in Business and Economics, Report 6816, 1968.

[18] Theil, H.: Principles of Econometrics. New York: John Wiley and Sons, Inc., 1971.

[19] Theil, H.: "Multiplicative Randomness in Time Series Regression Analysis", Netherlands School of Economics, Econometric Institute Report 5901, 1959.

[20] Zellner, A.: "On the Aggregation Problem: A New Approach to a Troublesome Problem", University of Chicago Center for Mathematical Studies in Business and Economics Report 6628, 1966.