ERRORS-IN-VARIABLES ESTIMATION USING EXTRANEOUS INFORMATION: AN APPLICATION OF WEIGHTED REGRESSION TO THE U.K. CONSUMPTION FUNCTION*

J.E.H. Davidson

NUMBER 78

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

ERRORS-IN-VARIABLES ESTIMATION USING EXTRANEOUS INFORMATION: AN APPLICATION OF WEIGHTED REGRESSION TO THE U.K. CONSUMPTION FUNCTION*

J.E.H. Davidson

NUMBER 78

November, 1975

* I am very grateful to Dr. David Hendry for suggestions and discussion, and also to Professor J. D. Sargan and Mr. Dennis Leech for their comments. The remaining errors are of course my own. I should also like to thank the S.S.R.C. Project on Econometric Methodology for financial support.

This paper is circulated for discussion purposes only and its contents should be considered preliminary.

Consistent estimation of linear relationships when there are errors of measurement in the independent variables is not generally possible without further information extraneous to the sample. In Part I of this paper we review the theory of Maximum Likelihood (ML) estimation when this information takes the form of knowledge of the variances and covariances of the measurement errors. While there are some good text-book accounts of the method of Weighted Regression, e.g. [9], Chapter 29, [10] Chapter 10, we wish to consider the consequences of more general assumptions about the model than usually made, specifically, that (i) the error covariance matrix is unrestricted, (ii) only a subset of the regressors are measured with error, and (iii) both errors in the variables and errors in the equation ("shocks") are present.

In Part II we investigate the possibility of obtaining information about error variances from the patterns of revisions to the published data over time, with an application to the consumption function. Weighted Regression estimates obtained using estimated revision variances as proxy for measurement error variances provide supporting evidence on the magnitude of measurement error bias in the least squares estimates.

I

(i) The single-equation linear regression model is conventionally written in the form

$$y = X_1\beta + u, \tag{1.1}$$

the "dependent" variable y being segregated on the left hand side to emphasize that, under the usual assumptions, it alone is correlated with the stochastic term, and X_1 is a matrix of non-stochastic variables. The usual interpretation of the disturbance u in (1.1) is as an aggregate of minor unobservable variables - errors in the specification of the equation - which are distributed independently of all variables except y. Such errors are observationally indistinguishable from, and may include without violating the assumptions, measurement errors in y.

In the presence of measurement errors in the independent variables the usual assumptions do not hold, and any number of the variables may be stochastic. It is then more appropriate to write the model in the form

$$Xa = u \tag{1.2}$$

which can when necessary be partitioned under the conventional normalization as

$$X = [y : X_1]$$
$$a' = [1 : -\beta']$$

It will be helpful for expository purposes to postulate the existence of an <u>exact</u> linear relationship between k theoretical variables, $\xi_t = (\xi_{1t}, \ldots, \xi_{kT})$, of the form

$$a'\xi_t = 0 \tag{1.3}$$

and we wish to consider quite generally the problem of estimating the parameters a of (1.3) from the set of observations

$$x_t = (x_{1t}, \ldots, x_{kt}) \qquad t = 1, \ldots, T$$

(We shall hereafter refer to x_{1t} as y_t, conventionally identifying the variable corresponding to the restricted coefficient.)

The x_t will in general be imperfect proxies for the true variables ξ_t, subject to measurement and definitional errors of various kinds, so that

$$x_t = \xi_t + v_t \qquad t = 1, \ldots, T.$$

It will be assumed at this point that the errors v_t have the properties

$$E(v_t) = 0$$

$$E(v_t v_t') = \Omega, \text{ a positive definite matrix.}$$

The latter assumption will be relaxed in due course. It is true that the assumption of zero mean may be unrealistic, since due to errors of definition, a variable can be consistently over- or under-estimated. But it is clear that if $E(x_t) \neq \xi_t$, there is very little that can be said about consistent estimation.

This is an appropriate formulation for the pure "errors-in-variables" regression problem, and is the most convenient for our purpose. A model of the form (1.3) is not, of course, generally appropriate in the context of economics, where relationships can never be so exactly specified; but when necessary, we may employ the simple device of redefining the dependent variable to include the unobservables. Thus, suppose we have the stochastic relationship between true variables $\xi_t^* = (\xi_{1t}^*, \xi_{2t}, \ldots, \xi_{kt})$,

$$a' \xi_t^* = \varepsilon_t$$

This equality can be written in the form (1.3) by defining $\xi_{1t} = \xi_{1t}^* - \varepsilon_t$. We can then without any ambiguity treat the errors in the equation, ε_t, as errors in the measurement of ξ_{1t}, provided the independence assumptions hold.

The error vector for the redefined model can then be separated into two components,

$$\begin{pmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{kt} \end{pmatrix} = \begin{pmatrix} v^*_{1t} \\ v_{2t} \\ \vdots \\ v_{kt} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where v^*_{1t} is the "real" error of measurement in y_t. The definitions are summarized in the equalities,

$$y_t = \xi_{1t} + v_{1t} = \xi_{1t} + v^*_{1t} + \epsilon_t = \xi^*_{1t} + v^*_{1t}$$

Assuming for simplicity - and quite plausibly - that the measurement errors and equation errors are uncorrelated, the covariance matrix Ω can also be resolved into additive components,

$$\Omega = \Omega_v + \Omega_\epsilon \tag{1.4}$$

Ω_v is the covariance matrix of the measurement errors proper, and Ω_ϵ, which can be written $\sigma^2_\epsilon \Omega^o_\epsilon$, has the form

$$\sigma^2_\epsilon \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{pmatrix} . \tag{1.5}$$

It may be noted that if, contrary to our present assumption, the equation is part of a simultaneous system, the error vector will contain equation-error components corresponding to the other endogenous variables, and hence Ω_ϵ will possess more than one non-zero element. The practical importance of the distinction we have elaborated is, as will become clear, that Ω_v is independent of the equation parameters, in the sense that it can in principle be estimated (if at all) strictly prior to the estimation of a. But this is not true of Ω_ε. It will be convenient to neglect the distinction for the time being, and treat the model as in effect a pure "errors-in-variables" model, but we shall wish to return to it at a later stage.

(ii) The model (1.3) can now be written

$$a'(x_t - v_t) = 0 \qquad t = 1, \ldots, T,$$

or in matrix form, where $\{\xi_{it}\}$, the matrix of true variables, is represented by $\bar{X}$, and $\{v_{it}\}$, the matrix of errors, by V,

$$\begin{aligned} \bar{X}a &= 0 \\ X &= \bar{X} + V \end{aligned} \qquad (1.6)$$

Which implies

$$Xa = Va = u$$

The matrices $\bar{X}$, V can be partitioned conformably with X as

$$\bar{X} = [\bar{y} : \bar{X}_1] = [\bar{X}_1\beta : \bar{X}_1]$$

$$V = [v_y : V_1] .$$

The unknown parameters of the model consist of the k-1 unrestricted elements of a, the $\frac{1}{2}k(k+1)$ distinct elements of Ω, and the $T(k-1)$ independent true variables $\bar{X}_1$. It is well known that none of these unknowns can be consistently estimated from the information contained in the observations X alone. To illustrate the problem, it will be helpful to write the structural form of (1.6) in the conventional simultaneous equations notation. (See [2] for a similar approach.)

Define

$$X^{+} = \begin{bmatrix} X & : & I_T \end{bmatrix} \qquad (T \times (k+T))$$

$$A = \underset{(k)}{\begin{bmatrix} B \end{bmatrix}} : \underset{(T)}{\begin{bmatrix} C \end{bmatrix}} \qquad (k)$$

$$= \begin{bmatrix} \multicolumn{2}{c}{a'} & \vdots & 0' \\ \hdashline 0 & \vdots\; I_{k-1} & \vdots & -\bar{X}_1' \end{bmatrix} \begin{matrix} (1) \\ (k-1) \end{matrix} \qquad (k \times (k+T))$$

$$(1) \quad (k-1) \quad (T)$$

and $V^{+} = \begin{bmatrix} u & : & V_1 \end{bmatrix}$, where $V_1 = (X_1 - \bar{X}_1)$.

Then, (1.6) can be written equivalently as

$$AX^{+'} = V^{+'}, \tag{1.7}$$

or

$$BX' + C = V^{+'}. \tag{1.8}$$

Premultiplying (1.8) by B^{-1} will be found to give

$$X' - \bar{X}' = B^{-1}V^{+'} = V'$$

so that $V^{+'} = BV'$, and it follows that, where $E(v_t^{+}v_t^{+'}) = \Omega^{+}$,

$$\Omega^{+} = B\Omega B' \tag{1.9}$$

Now assuming that the v_t, and hence the v_t^{+} are normally and independently distributed, the log-likelihood of the sample is

$$L(a, \Omega, \bar{X}_1 | X) = -\frac{Tk}{2}\log 2\pi$$

$$-\frac{T}{2}\log(\det \Omega^{+}) - \frac{1}{2}\,\mathrm{tr}\,(\Omega^{+^{-1}} AX^{+'}X^{+}A') \tag{1.10}$$

(note that $|\det B| = 1$)

Consider first the consequences of trying to maximize (1.10) with respect to all the unknown parameters. It is convenient to follow a procedure analogous to Limited Information Maximum Likelihood. (See Koopmans & Hood, [8], Chap. VI)

Choose the matrix

$$H = \begin{bmatrix} 1 & 0' \\ h_2 & H_2 \end{bmatrix} \begin{matrix} (1) \\ (k-1) \end{matrix}$$
$$\quad (1)(k-1)$$

so that

$$H\Omega^+H' = \begin{bmatrix} \omega_{11}^+ & 0' \\ 0 & I_{k-1} \end{bmatrix}$$

where ω_{11}^+ is the (1,1) element of Ω^+, and is unchanged by the transformation. Observe that by (1.9),

$$\omega_{11}^+ = a'\Omega a.$$

Also,

$$HA = A^* = \begin{bmatrix} a' \; \vdots \; 0' \\ \cdots\cdots\cdots \\ A_2^* \end{bmatrix} \begin{matrix} (1) \\ (k-1) \end{matrix}$$

In terms of A* it is then possible to write (1.10) as

$$L = -\frac{Tk}{2}\log 2\pi - \frac{T}{2}\log(a'\Omega a) + T\log|\det H|$$
$$-\frac{1}{2}\left(\frac{a'X'Xa}{a'\Omega a}\right) - \frac{1}{2}\,\mathrm{tr}\,(A_2^* X^{+'} X^+ A_2^{*'}) \qquad (1.11)$$

Partitioning A_2^* as $[B_2^* : C_2^*]$, the last term in (1.11) can also be written

$$-\frac{1}{2}\,\mathrm{tr}\left[(B_2^* : C_2^*)\begin{bmatrix} X'X & X' \\ X & I_T \end{bmatrix}\begin{pmatrix} B_2^{*'} \\ \cdots \\ C_2^{*'} \end{pmatrix}\right] \qquad (1.12)$$

The maximum of (1.11) with respect to C_2^* is found by differentiation of (1.12) which yields the solution

$$C_2^* = -B_2^* X' \qquad (1.13)$$

But it will be seen that substituting the solution into (1.12) causes the term to vanish identically. This is, of course, the consequence of there being no degrees of freedom for the estimation of the T(k-1) elements of C_2^*. However, the solution for the original unknown $\bar{X}_1'$ may be written

$$\bar{X}_1' = X_1' + H_2^{-1}h_2 a'X' \qquad (1.14)$$

from (1.13), using that $C_2^* = -H_2\bar{X}_1'$, and that $B_2^* = [h_2 : -h_2\beta' + H_2]$.

Thus, the likelihood concentrated with respect to C_2^* is simply (1.11) less the last term, or,

$$L^* = -\frac{Tk}{2}\log 2\pi - \frac{T}{2}\log(a'\Omega a) + T\log|\det H| - \frac{1}{2}\left(\frac{a'X'Xa}{a'\Omega a}\right) \qquad (1.15)$$

The next step should be to concentrate (1.15) again with respect to B_2^* - or, which is the same thing, with respect to H. But it is clear that (1.15) has no maximum with respect to H. Referring to (1.14), we see that $\bar{X}_1 = X_1$ at the points on the likelihood surface where |det H| approaches infinity. It must be concluded that no ML estimate of Ω - and hence, of a - can be obtained from the sample X alone.

To proceed further, it will be necessary to assume that Ω is known, and this we will now do, postponing the question of feasible methods until later. It will, in fact, only be necessary to assume that Ω is known up to a scalar multiple. We write $\Omega = \sigma^2\Omega^0$, and treat Ω^0 as a matrix of known constants.

Note that

$$\begin{aligned} T\log|\det H| &= T\log|\det HB| \\ &= T/2\log|\det HB\Omega B'H'| - \frac{T}{2}\log\det\Omega \\ &= T/2\log(a'\Omega a) - \frac{T}{2}\log\det\Omega. \end{aligned}$$

Hence, (1.15) can be written

$$L^*(a, \sigma^2|\Omega^0, X) = -\frac{Tk}{2}\log 2\pi - \frac{T}{2}\log\det\Omega - \frac{1}{2}\left(\frac{a'X'Xa}{a'\Omega a}\right)$$

$$= -\frac{Tk}{2}\log 2\pi\sigma^2 - \frac{T}{2}\log\det\Omega^o - \frac{1}{2\sigma^2}\left(\frac{a'X'Xa}{a'\Omega^o a}\right) \qquad (1.16)$$

Differentiating with respect to a gives the ML estimate of a as the solution to

$$-\frac{1}{\sigma^2(\hat{a}'\Omega^o\hat{a})}(X'X - \hat{\lambda}\Omega^o)\hat{a} = 0 \qquad (1.17)$$

where $\hat{\lambda} = \dfrac{\hat{a}'X'X\hat{a}}{\hat{a}'\Omega^o\hat{a}}$.

$\hat{\lambda}$ is an eigenvalue of $\Omega^{o-1}X'X$, and $\hat{a}$ the corresponding eigenvector. Since λ is a minimand, the smallest eigenvalue is appropriate. The conventional normalization of a is obtained by dividing through by the first element of the solution eigenvector, giving

$$\hat{a}' = (1 : -\hat{\beta}').$$

Differentiating with respect to σ^2 gives

$$-\frac{Tk}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\left(\frac{\hat{a}'X'X\hat{a}}{\hat{a}'\Omega^o\hat{a}}\right) = 0 \qquad (1.18)$$

or

$$\hat{\sigma}^2 = \frac{1}{Tk}\left(\frac{\hat{a}'X'X\hat{a}}{\hat{a}'\Omega^o\hat{a}}\right) \qquad (1.19)$$

The consistency of $\hat{a}$ can be shown as follows.* Define

$$M = \underset{T\to\infty}{\text{plim}}\left(\frac{X'X}{T}\right), \qquad \bar{M} = \underset{T\to\infty}{\text{Lim}}\left(\frac{\bar{X}'\bar{X}}{T}\right), \quad \text{so that } M = \bar{M} + \sigma^2\Omega^o, \text{ or,}$$

$$\Omega^{o^{-1}}M = \Omega^{o^{-1}}\bar{M} + \sigma^2 I \qquad (1.20)$$

We assume that (1.6) holds in the limit, so $\bar{M}a = 0$, and also that $\bar{M}$ has rank k-1, so that a is unique up to a scalar multiple. So, since $\bar{M}$ is a positive semi-definite k×k matrix, and $\Omega^{o^{-1}}$ has rank k, the smallest eigenvalue of $\Omega^{o^{-1}}\bar{M}$ equals 0, and hence, that of $\Omega^{o^{-1}}\bar{M} + \sigma^2 I$ equals σ^2.

---

* Essentially as in [10], p. 387

But by Slutsky's Theorem it follows from (1.20) and (1.17) that this eigenvalue is the probability limit of $\hat{\lambda}/_T$. Hence,

$$\text{plim}\ \frac{1}{T}\ \left[(X'X - \hat{\lambda}\Omega^o)\hat{a}\right] = (M - \sigma^2\Omega^o)\text{plim}(\hat{a}) = \bar{M}\ \text{plim}(\hat{a}) = 0$$

which implies that $\text{plim}(\hat{a}) = a$.

However, it is apparent from (1.19) that

$$\text{plim}\ (\hat{\sigma}^2) = \frac{1}{k}\ \sigma^2\ ,$$

so the ML estimator of σ^2 is not consistent - although a consistent estimate <u>is</u> provided by $\hat{\lambda}/_T$.

Differentiating (1.17) a second time with respect to a and taking the probability limit (after dividing by T) yields

$$\text{plim}\ \frac{1}{T}\left(\frac{\partial^2 L}{\partial a \partial a'}\right) = \frac{-1}{\sigma^2(a'\Omega^o a)}\ \bar{M}\ . \qquad (1.21)$$

Partitioning this matrix by

$$\bar{M} = \begin{pmatrix} \bar{m}_{yy} & \bar{m}'_{y1} \\ \bar{m}_{y1} & \bar{M}_{11} \end{pmatrix} \begin{matrix} (1) \\ (k-1) \end{matrix}\ ,$$
$$\qquad (1) \qquad (k-1)$$

and discarding the first row and column which correspond to the restricted element of a gives the asymptotic variance matrix of the unrestricted elements as

$$V(\hat{\beta}) = \frac{1}{T}\ \sigma^2(a'\Omega^o a)\ \bar{M}_{11}^{-1} \qquad (1.22)$$

Estimates of $\bar{X}_1$ can also be obtained by substitution in (1.14), but note that these are <u>not</u> consistent. The number of unknowns of course tends to infinity with the sample size.

(iii) The assumption that Ω is of full rank has been made only for convenience of exposition, and is not restrictive; consider the case where some of the variables are measured without error - dummy variables, for example.

Then,

$$V = \overset{(k_A)}{[V_A} : \overset{(k_B)}{0]},$$

$$\Omega = \begin{bmatrix} \Omega_A & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} (k_A) \\ (k_B) \end{matrix}$$

The last k_B equations of (1.7) are now identities, and can be eliminated in the usual way. The observation matrix X is partitioned,

$$X = \overset{(k_A)}{[X_A} : \overset{(k_B)}{X_B]} = \overset{(1)}{[y} : \overset{(k_A-1)}{X_{1A}} : \overset{(k_B)}{X_B]}$$

and the matrix of true variables $\bar{X}$ partitionod conformably,

$$\bar{X} = [\bar{y} : \bar{X}_{1A} : \bar{X}_B]$$

such that $X_B = \bar{X}_B$. Similarly, we have $a = [a_A' : a_B']$.

Solving the second set of k_B equations for X_B and substituting in the first k_A equations is found to give

$$\begin{bmatrix} a_A' & \vdots & a_B'\bar{X}_B' \\ \cdots & \cdots & \cdots \\ 0 \ \vdots \ I & \vdots & -\bar{X}_{1A}' \end{bmatrix} \begin{bmatrix} X_A' \\ \cdots \\ I_T \end{bmatrix} = V_A^{+\prime} \qquad (1.23)$$

$$\begin{matrix} (1) & (k_A-1) & (T) \end{matrix}$$

for the equality comparable to (1.7). Setting up the likelihood function for this system as before leads to

$$L^*(a, \sigma^2 | \Omega_A^o, X) = -\frac{Tk_A}{2} \log 2\pi\sigma^2 - \frac{T}{2} \log \det \Omega_A^o$$

$$- \frac{1}{2\sigma^2} \left(\frac{(a_A'X_A' + a_B'X_B')(X_A a_A + X_B a_B)}{a_A'\Omega_A^o a_A} \right) \qquad (1.24)$$

which is concentrated unrestrictedly with respect to a_B, yielding the solution

$$a_B = (X_B'X_B')^{-1} X_B'X_A a_A \qquad (1.25)$$

Substituting for a_B in the likelihood function and maximizing with respect to a_A leads finally to

$$\frac{1}{\sigma^2(\hat{a}_A'\Omega_A^o\hat{a}_A)}(X_A'Q_BX_A - \hat{\lambda}\Omega_A^o)\hat{a}_A = 0 \qquad (1.26)$$

where $Q_B = I - X_B(X_B'X_B)^{-1}X_B'$, as the equation comparable to (1.17). The estimates of a_B can then be obtained from (1.24).

A very obvious special case of (1.26) is where $k_A = 1$, so that Ω^o has only a single non-zero element, say $\omega_{11} = 1$ with no loss of generality. a_A is then a scalar with fixed value, and (1.26) reduces to ordinary least squares. The condition that Ω must be known up to a scalar multiple clearly ceases to be restrictive in this case, and OLS requires no information extraneous to the sample - beyond the very important knowledge that X_1 is measured without error.

(iv) At this point we reintroduce the distinction referred to earlier between errors of measurement proper, and errors in the equation, and make use of the corresponding dis-aggregation of the covariance matrix. From (1.20), (1.21) and (1.4) we may write

$$(M - \Omega_v - \Omega_\varepsilon)a = 0. \qquad (1.27)$$

When Ω_ε has the form (1.5), defining the partition

$$\Omega_v = \begin{bmatrix} \omega_{yy} & \underline{\omega}_{y1}' \\ \underline{\omega}_{y1} & \Omega_{11} \end{bmatrix} \begin{matrix} (1) \\ (k-1) \end{matrix}$$
$$\qquad\qquad (1) \qquad (k-1)$$

enables us to write (1.27) in partitioned form as

$$\begin{bmatrix} m_{yy} - \omega_{yy} - \sigma_\varepsilon^2 & m_{y1}' - \omega_{y1}' \\ m_{y1} - \omega_{y1} & M_{11} - \Omega_{11} \end{bmatrix} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = 0 \qquad (1.28)$$

If Ω_v is a known matrix, the estimator

$$\tilde{\beta} = (X_1'X_1 - T\Omega_{11})^{-1}(X_1'y - T\omega_{y1}) \tag{1.29}$$

clearly converges in probability to the solution of (1.28), and is hence by previous arguments the maximum likelihood estimator. σ_ε^2 is estimated consistently by

$$\tilde{\sigma}_\varepsilon^2 = \frac{1}{T}\left[(y'y - T\omega_{yy}) - (y'X_1 - T\omega_{y1}')\tilde{\beta}\right] \tag{1.30}$$

The case of practical interest here is that where there exists an extraneous estimate $\hat{\Omega}_v$ of Ω_v. Substituting $\hat{\Omega}_v$ into (1.29) gives an estimate of β, $\tilde{\tilde{\beta}}$ say, which will be asymptotically equivalent to ML whenever $\hat{\Omega}_v$ is consistent for Ω_v. We must be cautious in considering the properties of $\tilde{\tilde{\beta}}$ since such a "two stage" estimator will be a function of two sets of sample observations, x_t, $t = 1,\ldots,T$ and r_s, $s = 1,\ldots,T'$, say, where the r_s's are observations of an as yet unspecified kind from which an estimate of Ω_v can be computed. While the r_s's may include the x_t's, the two sets of observations cannot be identical, since extraneous information is known to be required; it is also quite possible that $T \neq T'$. When we speak of the consistency of $\tilde{\tilde{\beta}}$, we refer to it's behaviour as both T and T' tend to infinity.

(i) Obtaining suitable extraneous information about measurement errors is a notoriously intractable problem. A widely advocated approach to consistent information is the use of instrumental variables (IV). If there exists a matrix of instruments Z, uncorrelated with V but related to $\overline{X}$ by the linear stochastic system

$$\overline{X}_1 = Z\Gamma' + W \tag{2.1}$$

then

$$X_1 = Z\Gamma' + W + V_1 \tag{2.2}$$

and $\overline{X}_1$ can be estimated by $\hat{X}_1 = Z\hat{\Gamma}' = Z(Z'Z)^{-1}Z'X_1 = QX_1$.

Then, $(I - Q)X_1 = (W + V_1)$. We may show that the IV estimator

$$\beta^* = (X_1'QX_1)^{-1}X_1'Qy \tag{2.3}$$

has the same probability limit as (1.29). In effect, Ω_{11} is estimated by $\frac{1}{T}(X_1'(I - Q)X_1)$ and ω_{y1} by $\frac{1}{T}(X_1'(I - Q)y)$. These estimates are inconsistent - since, for instance, $\text{plim}\,\frac{1}{T}(X_1'(I - Q)X_1) = \text{plim}\,\frac{1}{T}(W + V_1)'(W + V_1) \neq \Omega_{11}$ - but knowing by the standard result that IV is consistent, we may obtain from the probability limit of (2.3),

$$\text{plim}\,\frac{1}{T}(X_1'X_1 - V_1'V_1 - \overline{X}_1'(I - Q)\overline{X}_1)\ \text{plim}\,(\beta^*) =$$

$$\text{plim}\,\frac{1}{T}(X_1'y - V_1'v_y - \overline{X}_1'(I - Q)\overline{y})$$

$$\left(\text{using plim}\left(\frac{Z'V}{T}\right) = 0\right)$$

and hence,

$$\text{plim}\,(\beta^*) = \text{plim}\,(X_1'X_1 - V_1'V_1)^{-1}(X_1'y - V_1'v_y)$$

since $\overline{X}_1\beta = \overline{y}$.

If $W = 0$, so that the covariance estimates are consistent (1.29) and (2.3) are asymptotically equivalent. But in general, the asymptotic variance matrix of β^*, where $\sigma_*^2 = V(u) = \sigma^2(a'\Omega^o a)$, is

$$V(\beta^*) = \sigma_*^2 \operatorname{plim} \left(\frac{X_1'QX_1}{T} \right)^{-1}$$

$$= \sigma_*^2 \left(\bar{M}_{11} - \operatorname{plim} \left(\frac{\hat{W}'\hat{W}}{T} \right) \right)^{-1}$$

where $\hat{W} = (I - Q)\bar{X}$, and $\operatorname{plim}\left(\frac{\hat{W}'\hat{W}}{T}\right)$ is non-negative definite. So IV is always, as we should expect, less efficient asymptotically than ML with <u>known</u> Ω.

The practical difficulty with IV lies with the existence of suitable instruments. When the equation forms part of a simultaneous system, two stage least squares will give consistent estimates even though there are errors of measurement in the endogenous variables; but problems arise when there are measurement errors in the pre-determined variables, i.e., in the instrumenrs themselves. Consistent estimation then require a further set of instruments, which cannot exist if the model is already correctly specified. The problem extends a fortiori to the single equation model. Lagged values of the regressors are a popular choice of instrument, but will only be suitable if it is known that the measurement errors are serially independent, a condition which will often be found to be unduly strong.

Grouping methods such as those proposed by Wald [15] and Bartlett [1] are simply instrumental variable methods employing various kinds of dummy variables as instruments. It will be likely in these cases that the errors of W of (2.1) will be large, and the efficiency of these methods correspondingly poor.

A more sophisticated variant of the dummy-instrument approach would be to fit the columns of X to simple sinusoidal functions of t, and use the latter with the estimated parameters as instruments. The basic difficulty with all these methods is that when the instruments are chosen on the basis of observed good correlation with the empirical variables - rather than on theoretical grounds - then the better the correlation, the greater the likelihood that the instruments are correlated with the errors as well as the true variables.

These difficulties imply that IV cannot be a generally satisfactory solution to the measurement error problem. The alternative approach which is open to us is to seek methods of estimating Ω directly. There are two cases to be considered. If it is known that the model is of the pure errors-in-variables type, or at least that errors in the equation represent a sufficiently small proportion of the residual variance, then it is sufficient to know Ω up to a factor of proportionality, and (1.17) with estimated Ω^o is the appropriate estimator. If on the other hand the errors in the equation are not thought to be negligible, then it will be necessary to estimate the absolute values of the elements of Ω_v, to obtain a feasible analogue of (1.29). In the former case, it may be thought sufficient to employ a plausible subjective estimate of the relative precision of measurement of the variables, for example $\Omega^o = I$ (see, e.g. Casson [2] ; but for the general error model, quantitative estimates are necessarily required.

(ii) There can be no generally applicable approach to this problem, and in many cases it is probable that suitable information about the way in which the data is collected simply does not exist. The remainder of this paper is concerned with a method which may sometimes prove feasible, and it will be convenient to describe it in the context of the specific empirical problem which has been investigated, the estimation of a consumption function for the U.K. using quarterly data published by the Central Statistical Office. It is hoped that as the discussion proceeds, the scope for further applications will become apparent.

The data series in question are of consumption of non-durables (total consumption less expenditure on cars and motor-cycles, furniture and floor-coverings, and radio and electrical goods) and total disposable income, both revalued at constant 1963 prices; a total of 58 observations, from 1958(iii) to 1972(iv), drawn from the 1970, 1971 and 1972 October issues of Economic Trends.

The model under consideration has the form

$$\Delta_4 C_t = b_0 + b_1 \Delta_4 Y_t + b_2 \Delta_4 Y_{t-1} + \varepsilon_t \qquad (2.4)$$

where Δ_4 is the annual (four-period) difference operator. It goes beyond the scope of this paper to do more than justify (2.4) as a simple permanent income formulation,* but it may be noted that the differencing procedure is a convenient method for dealing with multiplicative seasonality. The differenced series show negligible seasonality or trend, and their variances are concentrated at the frequency of the business cycle.

The simple distributed lag (2.4) was chosen largely on grounds of parsimony, since longer lags were found to have generally small and insignificant coefficients. The lagged dependent variable, commonly included in such equations as approximation to a geometrically declining lag structure, was also found insignificant at the 5% level; it should

---

* A report on the work on consumption of which this study forms a part will, it is hoped, appear in due course.

be noted that the success of this explanatory variable in level formulations is largely due to its role as a proxy for the trend in the variables. In the difference formulation, the rate of linear trend is measured by the constant term b_0.

The parameter of principle interest is $b_1^* = b_1 + b_2$, which can be interpreted as the 'medium run' marginal propensity to consume; (2.4) can be written in a more illuminating form as

$$\Delta_4 C_t = b_0 + b_1^* \Delta_4 Y_t + b_2^* \Delta_1 \Delta_4 Y_t + \varepsilon_t \tag{2.5}$$

where $b_2^* = -b_2$. The equation is then interpreted as a linear relationship between annual increments in consumption and income, modified by a penalty term related to the rate at which income is changing. Incidentally, a log-linear version of (2.4) has been tried, and found to give almost equivalent results as regards goodness of fit and residual autocorrelation.

The feature of particular interest in relation to the measurement error problem is the use of differenced variables. The ordinary least squares estimates of (2.4) (see Table II) were found to be lower than expected, and the possibility of measurement error bias could not be ignored. (It is important to distinguish this problem from the "measurement error" interpretation of the Permanent Income model. In Friedman's [6] terminology, the "transitory" components of income are the deviations of actual current income from permanent income - whereas we are concerned with deviations of measured income from actual income - measurement errors proper.) The variance of the differences of a smoothly trending time series is characteristically much smaller than the variance of the original series; differencing has the advantageous effect of tending to cancel out relatively constant errors of measurement - omission of some little-changing component of the aggregate for example. But the relative size of random (serially independent) measurement errors will generally be amplified by differencing. For example, the mean of the disposable income series is £5776m., and the mean of the annual differences (which are almost all positive) is £161m. Taking these figures as representative values of the series, a serially independent measurement error of 1% in the level implies an error of around 30% in the corresponding difference.

For information on the reliability of the published series we turn first to the C.S.O. handbook, National Accounts Statistics: Sources and Methods, in particular to Chapter III. The information given there is understandably rather vague, and amounts to a categorization of the published figures into three reliability classes,

A :	Margin of error	±	less than 3%
B :	" " "	±	3% to 10%
C :	" " "	±	more than 10%,

with a subjective confidence level of about .9 for the estimate lying within the specified interval.

Both the income and the consumption series have an A rating, but as we have noted, the corresponding margin of error in the differences may be quite unacceptable. However, we quote the following passage from Sources and Methods:

"The gradings [A, B or C above] are applied to the absolute values of the various components. It is generally true to say that the absolute error in the change from year to year is likely to be less than might appear from the errors attached to the absolute values. Nearly always, when a figure is attributed to an item about which there is much uncertainty, consideration is paid to the probable change from the previous year. This implies that the error in the absolute figures, whatever it may be, is likely to be in the same direction in all years. The deviations between the estimates and the facts are likely to consist in part of a bias which is more or less constant from year to year, and partly of a more random element."

(page 40)

By this account, the differencing process will remove some proportion of the error, a matter of importance if the "more or less constant bias" is of very different proportionate magnitudes in income and in consumption. It is of course the "more random element" we are concerned about, and although encouraging, the remarks quoted still leave us with a very vague idea about actual reliability, and still less of what we should really like to have, a quantitative estimate of the covariance matrix of the measurement errors. (Ω_v in the notation of section I).

But there is one source of quantitative information to which we have direct access. The October issues of Economic Trends contain each year quarterly series of the principal macro-variables about twelve years in length. This means that each item of data eventually appears about twelve times in succeeding October issues. The first appearances of an item are generally provisional estimates, and are likely to be revised quite substantially as time goes on. What is perhaps rather surprising is the extent to which the figures are being revised continually over the period of their publication, even ten or more years after their first appearance. These revisions are explained in the handbook partly in terms of redefinitions and reassignments of various components of the aggregates - for instance, the quarter to which an item is assigned - and partly in terms of new information and the correction of errors.

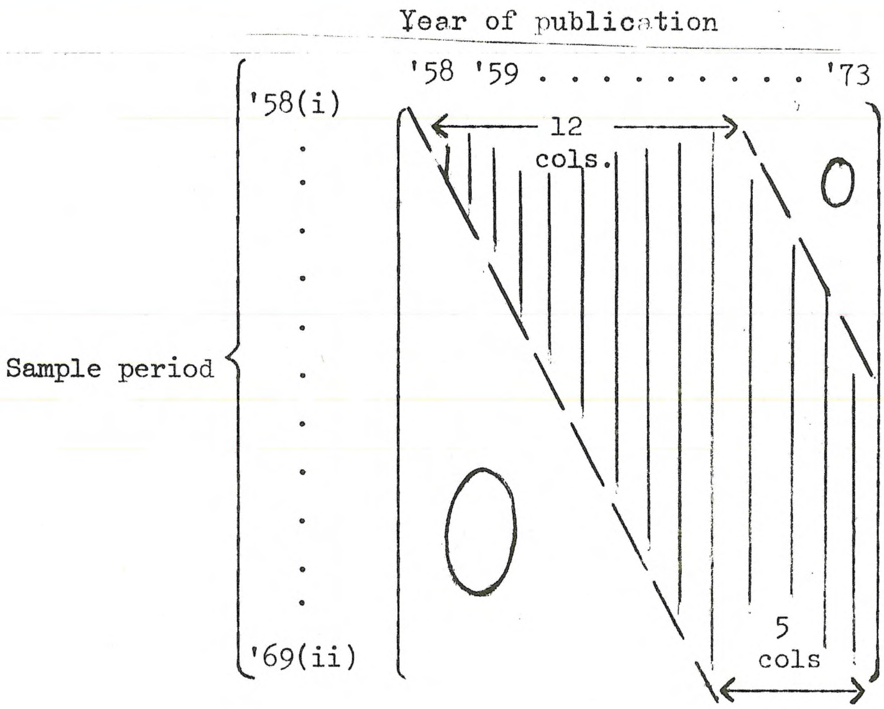
It seems a reasonable supposition that the magnitude and frequency of the revisions - i.e., the variance of the published estimates - is a sensible measure of reliability. It can be objected, of course, that to study the revisions is to use the properties of detected and corrected errors to infer the properties of undetected errors - a procedure of rather dubious validity. But on the other hand, the revisions are the only source of hard, quantitative information that we have. There is a logical problem involved here, since when we possess information about undetected errors, they cease to be undetected, and at least in principle, we can use it to improve the estimates, leaving us in just as much ignorance of the remaining undetected errors, if any. In any event, in the absence of any knowledge of the size and direction of undetected errors, we cannot do better than to assume them to be of a similar order of magnitude to detected ones.

Apart from the good general discussion by Morgenstern [13], chapter XIV, a number of studies of data errors and revisions have been published. For example, McDonald has fitted a Box-Jenkins model to the Residual Error of the National Accounts, [11], and also compared the estimated parameters of Box-Jenkins models fitted to revised and unrevised version of the same series, [12], and McDonald, Holden, and Denton & Kuiper [12], [7], [5], have compared regression results obtained using revised and unrevised series. As far as is known, though, there is no published study of the revisions of differenced series, nor any attempt to use all the information contained in the revisions to estimate the error variances.

In this case the following procedure was carried out. As many different estimates as possible of the series for disposable income and non-durable consumption, for the period 1958(i) to 1969(ii) were obtained and punched onto cards. This involved looking in the sixteen October issues of <u>Economic Trends</u>

for 1958 through to 1973, and yielded between twelve estimates of the earlier figures in the series down to five for the more recent. To obtain this number of estimates, it was necessary to use current price data, since the constant price data used in the actual estimation of the consumption function is subject to changes of the base year, and the number of comparable estimates is much smaller. The current price estimates were deflated using the retail price index - this being applied to total consumption, before subtraction of expenditure on durables deflated by the appropriate sectoral indices. For two reasons, this is a somewhat rough and ready procedure. First, constant price estimates are revalued prior to aggregation; but since it is intended to compare not different time periods, but different estimates of the same time period, it is sufficient just to scale the estimates appropriately. Second, and perhaps more serious, is that by using the current price data the influence of errors in the price indices is eliminated. This is unfortunate, but is a deficiency which can be remedied only at a much increased cost in data collection and computation: e.g., by deflating each years current price estimates by the corresponding years price indices. In view of the limitations already inherent in this method of investigation, such a refinement was felt to be not worthwhile at this stage, although there is clearly scope for further examination of the problem.

The figures obtained in this way for C and Y as defined can each be assembled into a 46×16 matrix, the rows corresponding to the time-period of the observations, and the columns to the "vintages" of the estimates; likewise, the differences $\Delta_4 C$ and $\Delta_4 Y$ yield a 42×15 matrix. The matrices have a roughly upper-triangular form:



To obtain an idea of the general pattern of the revisions, a simple programme was written to graph the rows of these data matrices. Figure 1 and Figure 2 show the plots for Y and $\Delta_4 Y$ for four representative periods. In Figure 1, the current and four-quarter lagged estimates are plotted together - the former by solid lines, the latter by broken lines. In Figure 2, the plots for the four-quarter differences for the same four periods are shown; i.e., the broken lines subtracted from the solid lines of Figure 1.

The variety of size and direction of the revisions is considerable, and the examples shown should be considered as unexceptional rather than typical. However, the earlier revisions tend to be larger than the later ones, as might be expected, and also the revisions in the levels tend to be upward, the errors in the provisional estimates usually being omissions. What is of most interest from our point of view is the high degree of correlation of the revisions in successive (i.e. four-quarters-apart) periods, resulting in the estimates of the annual differences having a comparatively small dispersion, and no discernible tendency for upward rather than downward revision. This is the pattern which the account in Sources and Methods has led us to expect, but it is of some interest to observe it directly.

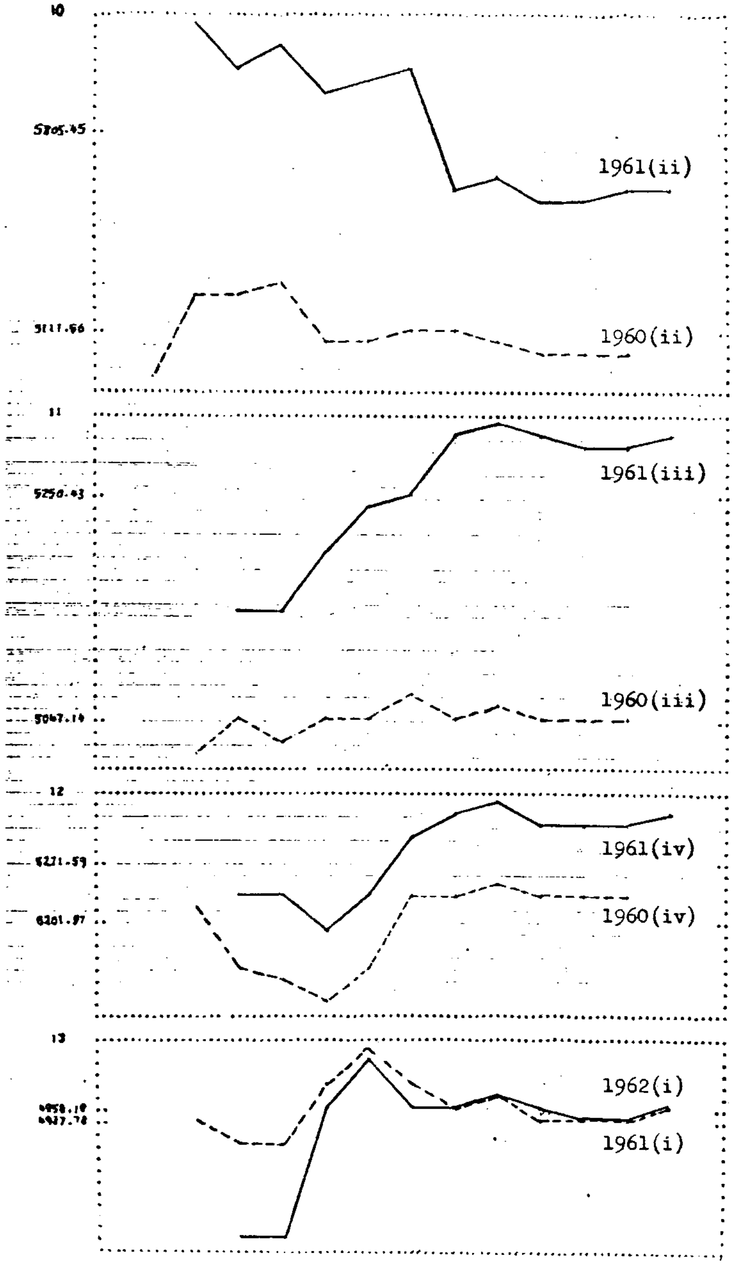

1961(ii)
1960(ii)
1961(iii)
1960(iii)
1961(iv)
1960(iv)
1962(i)
1961(i)

Fig. 1

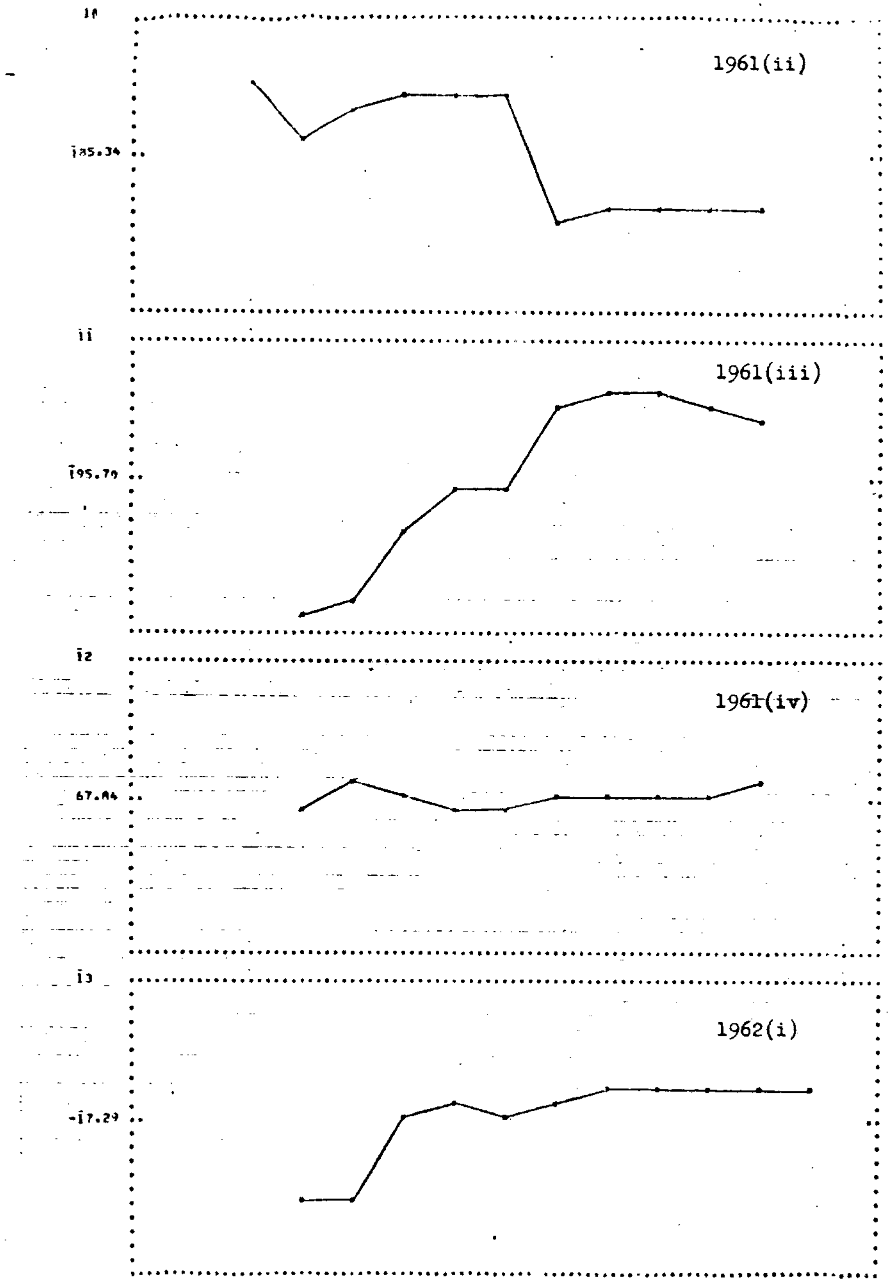

1961(ii)
1961(iii)
1961(iv)
1962(i)

Fig. 2

The next question is that of an appropriate summary statistic. The chosen procedure was to compute the ordinary sample variances of each row of the matrix, and then the column mean of these:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N_t-1}\sum_{i=1}^{N_t}(x_{it}-\bar{x}_t)^2\right]$$

$$\bar{x}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} x_{it} \qquad (2.6)$$

$$5 \leq N_t \leq 12$$

and similarly, the means of the covariances of the corresponding rows of the matrices for different variables and differences of variables,

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N_t-1}\sum_{i=1}^{N_t}(x_{it}-\bar{x}_t)(y_{it}-\bar{y}_t)\right] \qquad (2.7)$$

for $x, y = C, Y, Y_{-1}$ and $\Delta_4 C, \Delta_4 Y, \Delta_4 Y_{-1}$.

The covariance matrices with elements computed by (2.6) and (2.7) are shown in Table 1. Part (i) of the table shows the estimates for the levels of the variables, part (ii) the estimates for the differences, and part (iii) shows the estimated variances and covariances of the (differenced) variables themselves, for comparison with (ii);the ratios of the corresponding diagonal elements of (iii) and (ii) are approximately the "signal-noise ratios".

The results show the extent to which the revisions are serially correlated at the fourth order, bearing in mind that were they independent, the results for the differences, (ii), would be about twice the size of those for the levels, (i). We also have estimates of the first order autocovariances of the income revisions, for levels and differences, and in the latter case this quantity is quite small relative to the former.

Table I

(i) Revision Covariances (Levels)

C	1753.74		
Y	590.69	1729.78	
Y_{-1}	698.54	558.88	1729.78

(ii) Revision Covariances (differences)

$\Delta_4 C$	214.44		
$\Delta_4 Y$	126.20	678.70	
$\Delta_4 Y_{-1}$	136.50	81.74	678.70

(iii) $\frac{X'X}{T}$ (differences)

$\Delta_4 C$	2249.91		
$\Delta_4 Y$	4656.32	16877.66	
$\Delta_4 Y_{-1}$	4525.91	10158.71	17444.51

A point to be noted is that the covariances between C and Y and Y_{-1} will be large in part because of the common upward trend in the estimates according to vintage, and may therefore be interpreted as artefacts. In the differences, however, such trends are much less in evidence, and the size of the covariances may be judged more interesting.

As merely descriptive statistics, (2.6) and (2.7) may be judged on their merits; but we must be extremely careful in relating, say, (2.6) to the theoretical measurement error variance, $E(v_t^2)$, where v_t is the undetected error. (2.6) would be an unbiassed estimator of $E(v_t^2)$ if the various estimates of the variables were random drawings from a single population of possible observations, with mean equal to the true value of the variable - which is obviously not the case. In fact, each estimate represents a drawing from a population of estimates of a given "vintage", corresponding to the number of times that the figure has been subject to re-examination and correction. The population variances of succeeding vintages, though not necessarily the mean squared errors, will be in general decreasing, as is suggested by the graphs of Figure II. But we must be very careful not to interpret the reduced variance of the later revisions as an improvement in precision. The revisions "home in" on a given final estimate as the compilers run out of new information, not necessarily because there are no more errors to be found. Moreover, $\bar{x}_t$ as defined in (2.6) is certainly not an unbiassed estimate of the true value of the variable, since for the levels at least, the early vintages are likely to be biassed downward.

However, it can be argued that (2.6) and (2.7) represent the best use of the available information, and the assumptions, although false, are not worse than any others that could be made. An alternative statistic based on the mean squared deviations about the most recent estimate of the previous estimates,

$$\frac{1}{T} \sum_{t=1}^{T} \left[\frac{1}{N_t - 1} \sum_{i=1}^{N_t - 1} (x_{it} - x_{N_t t})^2 \right] \qquad (2.8)$$

was rejected because of the increased possibility of upward bias. Another possibility is an unweighted summation of squared deviations from row means, divided by the appropriate number of degrees of freedom :

$$\frac{1}{\sum_{t=1}^{T} N_t - T} \sum_{t=1}^{T} \sum_{i=1}^{N_t} (x_{it} - \bar{x}_t)^2 \qquad (2.9)$$

The difference between (2.6) and (2.9) is that (2.6) weights the more recent observations more heavily, because there are fewer of them. In a test, the difference between the two was found to be small.

The estimated signal-noise ratios of $9\frac{1}{2} : 1$ for consumption, and $24 : 1$ for income are smaller than expected from the first inspection of the data, in particular for income, which is the variable we are worried about. Measurement errors in $\Delta_4 C$ are not a problem (as regards bias) except in lagged-dependent variable specifications. However, what is clear from a study of the plotted values is that data for estimation and testing purposes should not, if possible, be of too young a vintage. It takes the compilers three or four years on average to produce estimates which remain stable thereafter, and hence contain most of the available information. The sample graphs of Figure II are typical of the series as a whole in this respect.

A slightly delicate question which one should nevertheless not ignore is whether the good behaviour of the revisions of the differences is not in part due to "fudging"; - for example, an error detected and corrected in a given estimate might lead to the adjacent figures in the series being corrected by a similar amount for no better reason than the belief of the compilers that their original estimates of the changes were correct. The passage from "Sources and Methods" quoted above is a little ambiguous on this point. But it does seem reasonable to accept, in the present state of data collection technique, that the C.S.O. know all that there is to be known about their figures, and any "fudging" that occurs represents the best available, albeit subjective knowledge about the data; in short, if "fudging" improves the accuracy of the estimates, then any corresponding reduction in the revision variance is, at the least, not inappropriate.

(ii) The next step is to find out what measurement errors of the estimated magnitude would imply in terms of least squares bias, and to do this we compute Weighted Regression estimates, as described in section I. We shall not, of course, be able to claim even consistency for the estimator, without knowledge even of the asymptotic sampling properties of the variance estimation. Our continuing ignorance of the relationship between detected ans undetected errors, to which we drew attention in the last section, implies that the original problem of insufficient information is still with us; but it is hopefully reduced, and we can assert with some confidence that the estimates will be "better" (in the mean squared error sense) than ordinary least squares estimates, which imply the use of error variance estimates of zero.

We may use either an approximation to the true Weighted Regression estimator, as in (1.17) or (1.26), or the method of fixed weights, an approximation to (1.29). The former method being appropriate to the pure errors-in-variables model, we shall refer to this as the PEV method. The latter we shall call the EVE (errors in variables and equation) method. Note that the PEV method makes no use of the absolute estimated magnitudes of the error variances and covariances, but depends merely on their relative magnitudes. On the other hand, this method assumes that measurement errors are the only significant source of residual variance, and the EVE method is in principle more appropriate to the present model. Both estimates were computed, so that the results could be compared.

The PEV estimates were computed using a library generalized eigenvector routine, N.A.G. subroutine F02AEF [14], so that only the input-output, computation of sample moments etc. required special programming. The EVE estimates could then be computed easily from the sample moments already obtained.

Apart from the residual standard error, s, test statistics were not computed, since they would be difficult to interpret and possibly misleading when computed using error variance estimates having unknown statistical

properties. Note that s^2 is computed in each case as $(\frac{1}{T-k+1})\hat{a}'X'X\hat{a}$,*
and is therefore not an estimate of σ^2 as defined in (1.16) for the FEV estimator, but of $V(u) = \sigma_*^2 = \sigma^2 a'\Omega^o a$. For the EVE estimates, s_ε^2 was also computed as $s^2 - \hat{a}'\hat{\Omega}\hat{a}$, which is an estimate of σ_ε^2 as defined in (1.28).

As well as two estimation methods, two alternative formulations were adopted regarding the error covariance matrix. In (a), we use the matrix just as given in Table I(ii), but in (b), we assume that the estimated covariances of the errors in Δ_4Y and Δ_4C are spurious, and set them (the off-diagonal elements of the first row and column) to zero. Since it is difficult to gauge the amount of confidence that can be placed in these covariance estimates in particular, it is as well to know their importance for the outcome.

Equation (2.4) was estimated using all four methods, as well as by ordinary least squares, and the results are shown in Table II. The standard errors (in parentheses) and other test statistics given by the OLS regression programme are shown. The variable D is a dummy variable for the effects of the 1968 Budget announcement of tax increases, defined by

$$D = \begin{cases} 1 \text{ in } 1968(i) \\ -1 \text{ in } 1968(ii) \\ 0 \text{ elsewhere.} \end{cases}$$

The sample moments were then computed from the deviations from the corrected means, so that the moment matrix corresponds to $X_A'Q_BX_A$ in (1.26), where X_B has the form $(\underline{1} : D)$, a T×2 matrix where $\underline{1}$ is the constant dummy. Note that this is the matrix given in Table I(iii), divided through

---

* Since the estimators have only asymptotic validity, the bias correction for s^2 has no precise justification. It is employed to facilitate comparison with the OLS estimate.

by T-1.

	Regressand is $\Delta_4 C$				
	OLS	Weighted Regressions			
		(a)		(b)	
Variable		PEV	EVE	PEV	EVE
$\Delta_4 Y$	.1843 (.0396)	.1973	.1865	.2035	.1910
$\Delta_4 Y_{-1}$	.1521 (.0389)	.1382	.1496	.1603	.1553
Const	63.72 (7.019)	63.85	63.75	59.31	62.12
$\Delta_4 D$	53.20 (14.69)	51.27	52.86	51.63	52.66
s	27.47	27.52	27.47	27.67	27.50
s_ε			24.18		22.26
R^2	.7330				
d	1.728				

Table II

(iii) The most notable feature of these results is how little the various weighted regression estimates differ overall from the ordinary least squares estimates; and it is even more striking if we compare the estimates of $b_1^* = b_1 + b_2$, which are shown in Table III:

Table III:

	OLS	PEV(a)	EVE(a)	PEV(b)	EVE(b)
b_1^*	.3364	.3355	.3361	.3638	.3463

In particular, the results for specification (a) change hardly at all.

The fact that the estimated signal-noise ratios are quite large can be held to account for this in the case of the EVE estimates. Note that in (a), the equation errors are estimated to account for almost 80% of the residual variance. However, the similar results given by the PEV method indicate that the small absolute magnitudes of the variance estimates are not the only factor involved.

It will aid the interpretation if we derive the general expression for the asymptotic least squares bias. From the least squares formula,

$$\hat{\beta} = (X_1'X_1')^{-1}X_1'y$$

we obtain, assuming that measurement and equation errors are uncorrelated,

$$\text{plim}(\hat{\beta} - \beta) = (\bar{M}_{11} + \Omega_{11})^{-1} (\omega_{y1} - \Omega_{11}\beta) \qquad (2.10)$$

It is clear from (2.10) that the least squares estimator is unbiased if $\Omega_{11}^{-1} \omega y_1 = \beta$. Now, partitioning the estimated variance matrix given in Table I(ii) as

$$\begin{pmatrix} \hat{\omega}_{CC} & \hat{\omega}'_{CY} \\ \hat{\omega}_{CY} & \Omega_{YY} \end{pmatrix} \begin{matrix} (1) \\ (2) \end{matrix} ,$$

it is easily checked that

$$\hat{\Omega}_{YY}^{-1} \hat{\omega}_{CY} = \begin{pmatrix} .1641 \\ .1814 \end{pmatrix} ,$$

which is quite close to the OLS coefficient vector. Moreover, the sum of the elements is .3454, which is very close to the OLS estimate of b_1^*.

This result may be mere coincidence; but it seems reasonable that since the series for income and consumption are related by an accounting identity, the revisions in each would be kept approximately in step, so that the books balance. If the residual components of disposable income, expenditure on durable goods and savings, were never revised then of course the correspondence between the revisions in C and Y would be exact. What we cannot determine is whether this property of the revisions is also a property of the undetected errors, but the evidence is at any rate not unfavourable.

The results for error specification (b), where we set $\hat{\omega}_{CY} = 0$ on the assumption that the correlations between the revisions in C and Y are spurious as a guide to the behaviour of the errors, do as expected show somewhat larger changes in the coefficients, but in neither case of an order which suggests that measurement errors are a serious problem. The PEV estimator here weights the income variable to account for about three times as much of the residual variance as the consumption variable, which ignores equation errors and almost certainly overstates the problem. One explanation of the stability of these estimates can be given by reference to the lag structure of the model.

The bias expression (2.10) specialities in the present case, assuming $\omega_{CY} = 0$, to

$$\text{plim} \begin{pmatrix} \hat{b}_1 - b_1 \\ \hat{b}_2 - b_2 \end{pmatrix} = -\begin{pmatrix} m_{YY} & m_{YY_{-1}} \\ m_{YY_{-1}} & m_{YY} \end{pmatrix}^{-1} \begin{pmatrix} \omega_{YY} & \omega_{YY_{-1}} \\ \omega_{YY_{-1}} & \omega_{YY} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \qquad (2.11)$$

where $m_{YY} = \bar{m}_{YY} + \omega_{YY}$ is the asymptotic second moment of observed income, stationarity being assumed, and $m_{YY_{-1}} = \bar{m}_{YY_{-1}} + \omega_{YY_{-1}}$ is the first-order autocovariance. If we make the further simplification of assuming $\omega_{YY_{-1}} = 0$ – it is in any case estimated to be relatively small – then we obtain

$$\text{plim}\,(\hat{b}_1 - b_1) = \frac{\omega_{YY}\,(-\,b_1 m_{YY} + b_2 m_{YY-1})}{m_{YY}^2 - m_{YY_{-1}}^2}$$

$$= \frac{-\,b_1 + b_2\gamma}{(1+C)\,(1-\gamma^2)} \qquad (2.12)$$

where γ is the autocorrelation coefficient of observed income, and $C = \bar{m}_{YY}/\omega_{YY}$ is the signal-noise ratio. A similar expression can be derived for b_2.

It is an interesting and rather unexpected property of the model that the role of the lagged variable in damping the bias in b_1 - an effect which is general for two-regressor models provided b_1 and b_2 have the same (different) signs, and $\gamma > 0$ ($\gamma < 0$) - depends upon the fact that the lagged variable is also measured with error. For suppose that we contrive a relationship which is similar to (2.4) except that the lagged variable is replaced by a variable Z_t, say, which for the purposes of exposition is assumed to have the same coefficient in the equation, so that

$$\Delta_4 C_t = b_o + b_1 \Delta_4 Y_t + b_2 Z_t + \varepsilon_t \qquad (2.13)$$

and also assume that $m_{ZZ} = m_{YY}$, $\mathrm{Corr}(\Delta_4 Y_t, Z_t) = \gamma$. The only way in which Z_t is to differ from $\Delta_4 Y_{t-1}$ is that it is measured without error . Then, the bias in $\hat{b}_1$ becomes, instead of (2.12),

$$\mathrm{plim}\ (\hat{b}_1 - b_1) = \frac{-\ b_1}{(1 + C)(1 - \gamma^2)} \qquad (2.14)$$

Substituting in (2.12) the OLS estimates of b_1 and b_2 as approximations to the true values, and the estimate of γ obtained from the sample as .603, we obtain a value for the bias of $-\ .145(\frac{1}{1 + C})$. Corresponding substitutions in (2.14) yields a bias of $-\ .289(\frac{1}{1 + C})$. So the off-setting effect of the errors in the lagged variable reduces the bias by almost half, ceteris paribus. Notice that (2.13) has been specified so that in the absence of errors in $\Delta_4 Y_t$, least squares estimates of b_1 in (2.4) and (2.13) have the same probability limit.

While these results follow from the asymptotic properties of least squares, it has been found that Monte Carlo simulations performed on a similar model exhibit similar off-setting effects in quite small samples. [4]

It can be expected that these properties generalize to more complex distributed lag formulations, and provide the underlying justification for the time series specification of permanent income models. It is evident that even the simple model employed here is likely to be quite robust against the effects of measurement errors.

We have taken care to distinguish errors of measurement in current income from Friedman's transitory components of income, since naturally the data revisions contain no information regarding the transitory components. But it is also clear that both types of "error" will bias least squares estimates away from the population values of the theoretical parameters of interest - in this case, the marginal propensity to consume out of annual increments of permanent income, defined for an appropriate horizon. Both transitory components and measurement errors are characterised by a lower degree of autocorrelation relative to the permanent component, and when the model is written in the form (2.5) it will be seen that the role of the second order difference term $\Delta_1\Delta_4Y_t$ is to partial out these relatively random influences, which cannot in general be distinguished.

REFERENCES

1. Bartlett, M.S.(1949): "Fitting a straight line when both variables are subject to error". Biometrika, 31, 391.

2. Casson, M.C. (1974): "Generalised Errors-in-Variables Regression". Review of Economic Studies, Vol.CLI(3).

3. Central Statistical Office (1968): National Accounts Statistics: Sources and Methods. H.M.S.O.

4. Davidson, J.E.H.(1975): "A Monte Carlo Study of Measurement Error Bias". Warwick Economic Research Paper (forthcoming).

5. Denton, F.T., and Kuiper, J.(1965) "The Effect of Measurement Errors on Parameter Estimate and Forecasts: A Case Study Based on Canadian Preliminary National Accounts". Review of Economics and Statistics XLVII, 198-206.

6. Friedman,M.(1957): A Theory of the Consumption Function", Princeton University Press.

7. Holden, K.(1969): "The Effect of Pensions to Data on Two Econometric Studies". Manchester School, 37, 23-36.

8. Hood, W.C., and Koopmans, T.C.(1953): Studies in Econometric Method Cowles Foundation Monograph No.14. Yale University Press.

9. Kendall,M.G., and Stuart, A.(1973): The Advanced Theory of Statistics, Vol.II. 3rd Ed., Griffin.

10. Malinvaud, E.(1970): Statistical Methods of Econometrics 2nd Ed. North Holland.

11. McDonald,J. (1973) "An Analysis of Residual Error in the Quarterly National Accounts of the U.K.", Applied Statistics, Vol.22.

12. McDonald, J.(1974) "An Analysis of the Significance of Revisions of some Quarterly U.K. National Income Time Series". (unpublished Ms.).

13. Morgenstern, O. (1963) *On the Accuracy of Economic Observations*, 2nd Ed., Princeton University Press.

14. Nottingham Algorithms Group (I.C.L. 1900 System) Library Manual: FOZAEF, Document No. 379, May 1972.

15. Wold, A. (1940) "The Fitting of Straight Lines if Both Variables are Subject to Error", *Annals of Mathematical Statistics*, *11*, 284.