COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 294

TIME SERIES REGRESSION WITH LINEAR CONSTRAINTS

E.J. Hannan and R.D. Terrell

April 24, 1970

TIME SERIES REGRESSION WITH LINEAR CONSTRAINTS*

by

E. J. Hannan and R.D. Terrell**

1. Introduction

We consider the regression model

$$y(n) = Bx(n) + u(n) , \quad n = 1, \ldots, N , \tag{1}$$

wherein $y(n)$, $x(n)$, $u(n)$ are vectors of, respectively, q , p and q components and B is a $q \times p$ matrix. We shall be considering the case where there are time series and shall later more fully specify their nature. We shall be concerned with the situation where B is, a priori, subjected to r linear constraints. The simplest such constraint is of the form

$$v'Bw = c$$

where v and w are known vectors and c is a known constant. However the most general form of linear constraint is of the form $tr(BA') = c$ where A is a $q \times p$ matrix of known constants. It is best to introduce a different notation and we introduce the vector β which is got from B

*The research described in this paper was carried out under grants from the National Science and from the Ford Foundation.

**Australian National University.

by putting the successive rows of B down a column of pq entries. Thus β_{ij} goes into row $(i-1)p+j$ in β. If the linear constraints are

$$\text{(2)} \qquad tr(BA(k)') = c_k \ , \quad k = 1, \ldots, r$$

then we may rewrite these in the form

$$\text{(2)}' \qquad \alpha(k)'\beta = c_k \ , \quad k = 1, \ldots, r$$

where $\alpha(k)$ is obtained from $A(k)$ in precisely the same way as β was obtained from B. We may as well assume the $\alpha(k)$ linearly indepencent, and shall do that.

We introduce the matrix F which projects onto the space spanned by the vectors $\alpha(k)$. Thus F is a symmetric, idempotent pq x pq matrix which may be obtained as follows. We replace the $\alpha(k)$ by r new ortho-normal vectors $\phi(k)$ i.e. so that

$$\phi(k)'\phi(\ell) = \delta_k^\ell \ , \quad k, \ \ell = 1, \ldots, r \ ,$$

and each $\phi(k)$ is a linear combination of the $\alpha(j)$. Thus the $\phi(k)$ may be got from the $\alpha(j)$ by taking those in some convenient order and ortho-normalizing them by the Gram-Schmidt process. Since the $\alpha(k)$ are likely to be rather simple vectors (consisting largely of zeros) this will not be a difficult procedure if r is not large. In any case onece the $\phi(k)$ are formed then

$$F = \sum_{k=1}^{r} \phi(k)\phi(k)' \ .$$

We now have

$$(2)'' \qquad \phi(k)'\beta = d_k , \quad k = 1, \ldots, r$$

where d_k is the same linear combination of the c_j as $\phi(k)$ is of the $\alpha(j)$. Then, let us say, $F\beta = e$ where $e = \Sigma d_k \phi(k)$.

We now rewrite (1) in the form

$$(1)' \qquad y = (I_q \otimes X)\beta + u$$

wherein (i) y has $y_j(n)$ in row $(j-1)N+n$ and u has $u_j(n)$ in the same place; (ii) $I_q \otimes X$ is the Kronecker (or tensor) product of the q rowed unit matrix and the matrix X which has $x_{ij}(n)$ in row n column j. By the Kronecker product, $A \otimes B$, of a $p \times q$ matrix A and an $r \times s$ matrix α we mean the matrix of $pr \times qs$ rows and columns with $a_{ij}b_{k\ell}$ in row $(i-1)r+k$, column $(j-1)s+\ell$. In particular $I_q \otimes X$ consists of the "block," X, repeated q times down the diagonal and with zeros elsewhere.

We may assume that the linear restrictions are homogeneous, i.e. e is null, for otherwise we form

$$(1)'' \qquad y - (I_q \otimes X)e = (I_q \otimes X)(\beta-e) + u$$

and rename the left side as y and the vector $(\beta-e)$ as β. Since $F(\beta-e) = e-e = 0$ the linear restrictions are now homogeneous. <u>We henceforth do this</u> so that we have the linear restrictions

$$(2)''' \qquad F\beta = 0 .$$

By way of introduction let us consider the estimation of β under three sets of circumstances. In each case we take the matrix X to be composed of fixed numbers and u to have null expectation, $\mathcal{E}(u) = 0$. The three sets of circumstances referred to are obtained by prescribing $\mathcal{E}(u(m)u(n)')$. They are obtained by taking this, successively, as (i) $\sigma^2\delta_m^n I_q$, (ii) $\delta_m^n G$, (iii) $\Gamma(n-m)$. Here G and the $\Gamma(n-m)$ are $q \times q$ matrices. The case (i) is rather unreal and is included for comparison only. The case (iii) corresponds to $u(n)$ being prescribed as generated by a stationary vector time series. We may now write down the BLUE (subject to the constraints) for each case. Before doing this we introduce the form of generalized inverse we shall use below. We need this only for symmetric matrices and indeed only for matrices of the form (ECE), for various non singular $pq \times pq$ matrices C, where $E = I_{pq} - F$. Then we put

$$(ECE)^{-1} = (ECE + F)^{-1} - F .$$

The matrix $ECE + F$ is non singular. It is not difficult to show that this result is the same as would be got by diagonalizing ECE by an orthogonal transformation, taking the reciprocal of each non zero diagonal element and reversing the diagonalization. Let us call Γ_q the $Nq \times Nq$ matrix of N^2 blocks of q rows and columns, the (m, n)th block being $\Gamma(n-m)$. Thus under (iii) $\Gamma_q = \mathcal{E}(uu')$. Now we have the BLUE as

(3.i) $$\{E(I_q \otimes X'X)E\}^{-1}(I_q \otimes X')y$$

(3.ii) $$\{E(G^{-1} \otimes X'X)E\}^{-1}(G^{-1} \otimes X')y$$

(3.iii) $$\{E(I_q \otimes X')\Gamma_q^{-1}(I_q \otimes X)E\}^{-1}(I_q \otimes X')\Gamma_q^{-1}y \ .$$

We do <u>not</u> mean these to be taken to be formulae from which to compute, and shall deal with computations later. Indeed (3.iii) involves a very large computation, if q and N are large, since Γ_q is then a very large matrix. In any case G and Γ_q are unknown. The formula (3.i) becomes trivial when $E = E_1 \otimes E_2$ where E_1 and E_2 are (respectively) $q \times q$ and $p \times p$ symmetric idempotents, for then our restrictions are $E_1BE_2 = B$ and we have

$$E_1y(n) = (E_1BE_2)x(n) + E_1u(n) \ .$$

We may now change to new x and y variables, some of each of which may then be eliminated, so that we reach an unconstrained situation and the BLUE is got rather trivially. As is shown in [9] the formula (3.ii) reduces to (3.i) if and only if $G \otimes I_p$ commutes with E and that the two are the same for all non negative G if and only if $E = I_q \otimes E_2$. Even then (3.iii) will not reduce to (i).

We shall in the next section deal with a procedure for estimating B under conditions (3.iii) which is computationally practicable and asymptotically efficient. (For more precise specifications see the next section.) Indeed we shall essentially reduce (3.iii) to (3.ii). These methods are large sample methods but we feel that they should be adequate for samples of 100 or even fewer observations under many circumstances. We discuss the point again later. Before going on to these considerations we quote the covariance matrices of the BLUE of β under conditions (3.i, ii, iii) respectively. These are

(i) $$\sigma^2\{E(I_q \otimes X'X)E\}^{-1} ,$$

(ii) $$\{E(G^{-1} \otimes X'X)E\}^{-1}$$

(iii) $$\{E(I_q \otimes X')\Gamma_q^{-1}(I_q \otimes X)E\}^{-1} .$$

2. The Estimation of β using Spectral Methods

One technique[1] which might be used to estimate B is the following. We model $u(n)$ as an autoregression, for example

$$u(n) = Ru(n-1) + \epsilon(n)$$

where $\mathcal{E}(\epsilon(m)\epsilon(n)') = \delta_m^n G$. Then

$$y(n) = Ry(n-1) + Bx(n) - RBx(n-1) + \epsilon(n) , \quad n = 2, \ldots, N .$$

Now R , B and RB are estimated by direct least squares regression. The constraints on B will have to be allowed for so that (3.ii) above will be used with $x(n)$ in (1) now replaced by a new set vector composed of $y(n-1)$, $x(n)$ and $x(n-1)$ and B in (1) replaced by a new matrix composed of R , B and -RB . The influence of the constraints on RB would be neglected. Of course a more general specification of $u(n)$ could also be used. If N is small (< 50) then this seems the best available method. We shall not go into further details but shall devote the remainder of this section to describing a computational procedure which we shall show leads to estimates with desirable properties in the next section. The basic idea is simple. We replace the original observations by their finite Fourier transforms. Thus we introduce

---

[1] This technique is due to J. Durbin.

$$(4) \quad a_{yj}(t) = \sum_{t=1}^{N} y_j(n)\cos n\omega_t , \quad b_{yj}(t) = \sum_{t=1}^{N} y_j(n)\sin n\omega_t , \quad \omega_t = 2\pi t/N ,$$

$$t = 1, \ldots, [\tfrac{1}{2} N] .$$

Here $[a]$ is the largest integer not greater than a. We have excluded $t = 0$ from the set of values of t for the following reason. Among the $x_j(n)$ will be one which will be identically unity. If, as is rather likely, the linear constraints do not involve the elements of the corresponding row of B then we may proceed by working entirely with mean corrected data, using $\bar{y} - \tilde{B}\bar{x}$ to estimate the vector of constant terms in the system of regressions (where $\tilde{B}$ is our yet to be defined estimate). Working with mean corrected data is the same as excluding $t = 0$. We shall later describe how to carry over our procedure to the case where the constraints do involve the vector of constant terms.

Now we have, using $a_y(t)$ for the vector with $a_{yj}(t)$ in the j^{th} place, and similarly for $b_y(t)$,

$$(5) \qquad a_y(t) = Ba_x(t) + a_u(t) , \quad b_y(t) = Bb_x(t) + b_u(t)$$

where

$$(4)' \qquad a_{xj}(t) = \sum_{t=1}^{N} x_j(n)\cos n\omega_t , \quad b_{xj}(t) = \sum_{t=1}^{N} x_j(n)\sin n\omega_t$$

and $a_u(t)$ is, of course, similarly defined though unobservable. Moreover $a_u(t)$, $b_u(t)$ have, approximately, a simpler covariance structure so that any two components are approximately uncorrelated if they correspond to different values of t. For example

$$\mathcal{E}(a_{uj}(s)b_{uk}(t)) \approx 0\ , \quad s \neq t\ .$$

(These are not precise statements and are inserted here only to make intelligible the procedure, by heuristic arguments.) The covariance matrix of the $a_{uj}(t)$, $b_{uj}(t)$, for fixed t , will depend upon t but for N large and for a set of adjacent ω_t this dependence will be weak so that (5) may be used to estimate B , via (3.ii) above, for each of a number of sets ("frequency bands") of adjacent ω_t . The remaining problem is that of optimally combining these different estimates. To do this we need to estimate the covariance structure of the $a_u(t)$, $b_u(t)$ and for this we need an estimate of B . We may obtain this from (3.i) for example, or even by means of unconstrained least squares regression. We call $\hat{\hat{B}}$ the estimate got from (3.i). Then we put

$$\hat{a}_u(t) = a_y(t) - \hat{\hat{B}}a_x(t)\ , \quad \hat{b}_u(t) = b_y(t) - \hat{\hat{B}}b_x(t)\ .$$

We next group the ω_t into $(M+1)$ sets of adjacent values. We shall state a theorem in the next section concerning the asymptotic behavior of the estimate we shall construct. This will assume that the smallest number, m let us say, of ω_t in any set increases with N at a rate faster than $N^{\frac{1}{2}}$. Thus these sets cannot be too small. If the sets are too large the procedure will have reduced efficiency since a sub-optimal weighting will be used within the set because in fact the covariance structure of the $a_u(t)$, $b_u(t)$ will vary across the set. Thus a compromise must be reached. The sets need not be all of the same size and indeed one might wich to make them smaller near $t = 0$ since here it might be expected that this covariance structure will be varying relatively rapidly with t . One has in mind, as a rough guide drawn from limited experience, values such as $m = 6$ for $N = 50$, $m = 12$ for $N = 200$, $m = 15$ for $N = 500$. We shall use the symbol $\Sigma_{(j)}$ for summation over the j^{th} set of adjacent t values, there being m_j of the ω_t in that set. Then we define the following matrices. We allot to j the index values $j = 0, 1, \ldots, M$ with $j = 0$ corresponding to the lowest frequency set and $j = M$ to the highest.

$$H_x(j) = \Sigma_{(j)}\{a_x(t)a_x(t)' + b_x(t)b_x(t)'\} , \quad j = 0, 1, \ldots, M .$$

$\hat{H}_u(j)$ is the same as $H_x(j)$ but with $\hat{a}_u$, $\hat{b}_u$ replacing a_x , b_x .

$$H_{yx}(j) = \Sigma_{(j)}\{a_y(t)a_x(t)' + b_y(t)b_x(t)'\} , \quad j = 1, \ldots, M-1 .$$

$$K_x(j) = \Sigma_{(j)}\{a_x(t)b_x(t)' - b_x(t)a_x(t)'\} , \quad j = 1, \ldots, M-1 .$$

$\hat{K}_u(j)$ is the same as $K_x(j)$ but with $\hat{a}_u$, $\hat{b}_u$ replacing a_x , b_x .

$$K_{yx}(j) = \Sigma_{(j)}\{a_y(t)b_x(t)' - b_y(t)a_x(t)'\} , \quad j = 1, \ldots, M-1 .$$

Next put, taking $\hat{K}_u$ as null for $j = 0, M$,

$$\hat{C}(j) = m_j(\hat{H}_u(j) + \hat{K}_u(j)\hat{H}_u(j)^{-1}\hat{K}_u(j))^{-1} , \quad j = 0, 1, \ldots, M$$

$$\hat{Q}(j) = m_j\hat{H}_u(j)^{-1}\hat{K}_u(j)\hat{C}(j) , \quad j = 1, \ldots, M-1 .$$

We put $\hat{Q}(o) = \hat{Q}(M) = 0$.

We now form the matrices

$$W = \frac{1}{M} \sum_{j=0}^{M} \{\hat{C}(j) \otimes H_x(j) + \hat{Q}(j) \otimes K_x(j)\}$$

$$V = \frac{1}{M} \sum_{j=0}^{M} \{\hat{C}(j) H_{yx}(j) - \hat{Q}(j) K_{yx}(j)\}$$

again treating K_x , K_{yx} as null for $j = 0, M$. From V we form the vector v in exactly the same way as we formed β from B i.e. with v_{ij} in row $(i-1)p+j$. Finally we put

$$\tilde{\beta} = (EWE)^{-1} v \quad (6)$$

Then $\tilde{\beta}$ is our "efficient" estimate. We have already described how to invert EWE , namely as $(EWE + F)^{-1} - F$. We shall in the next section indicate a fairly general set of circumstances under which it will be possible, for N sufficiently large, to treat $\tilde{\beta}$ as normal with mean vector β and covariance matrix $N^{-1}(EWE)^{-1}$.

We close this section with a number of comments.

(a) If the linear constraints involve the constant term we may proceed by simply including (let us say) $x_1(n) \equiv 1$ among the $x_j(n)$ and $t = 0$ among the t values for $a_y(t)$, $a_x(t)$. (Of course $a_y(0) = N\bar{y}$, $a_x(0) = N\bar{x}$.) These enter only into the sums for $j = 0$ and they enter with weight 1/2, so that the additional term in $H_x(0)$ is, for example, $1/2\ N^2 a_x(0) a_x(0)'$.

(b) The matrix inversions involved are of matrices of q rows save for $(EWE + F)$ which is pq rowed. Thus the computational effort here is not large unless q is large. The major effort is in the computation of $a_y(t)$,

$b_y(t)$, $a_x(t)$, $b_x(t)$. Each component of these requires N operations of multiplication followed by addition and there are $(p+q)N$ components in all so that there are $(p+q)N^2$ such operations. It is well known that when N is highly composite this effort may be greatly reduced. In any case alternative procedures are available which reduce this effort to reasonable proportions. (See [10], Chapter V, Section 3.) We shall not go into details here as with the sample sizes occurring in practice in the present type of problem the techniques we have presented would not be too costly.

(c) There may be a need to test some or all of the linear restrictions. We take $\alpha(1), \ldots, \alpha(r_1)$ to define r_1 linear constraints whose validity is maintained so that it is those defined by $\alpha(r_1 + j)$, $j = 1, \ldots, r_2$, $r_1 + r_2 = r$, which we wish to test. We now form $\tilde{\beta}_1$ exactly as we formed $\tilde{\beta}$ but using $E_1 = I_{pq} - F_1$ in place of E , where F_1 is set from the first r_1 of the $\alpha(k)$. We next form

$$u_j = \alpha'(r_1 + j)\tilde{\beta}_1 , \quad j = 1, \ldots, r_2$$

$$u_{j,k} = \alpha'(r_1 + j)(E_1WE_1)^{-1}\alpha'(r_1 + k) , \quad j, k = 1, \ldots, r_1 .$$

We arrange the u_j in a column vector u and the u_{jk} in a square matrix U . Then our test statistic is

$$Nu'U^{-1}u .$$

Then under the conditions of the theorem in the next section this is, on the null hypothesis that the additional restrictions are valid, asymptotically distributed as chi-square with r_2 degrees of freedom and may be used to test the validity of these restrictions. The result follows immediately from that theorem. A particular case is, of course, that where $r_1 = 0$, $r_2 = r_1$, so that $E_1 = I_{pq}$.

One could, of course, test the restrictions by using the unconstrained estimate and W in place of E_1WE_1 but this would clearly be an inferior procedure since the rise of the first r_1 constraints improves the efficiency of the estimate of β.

3. The Asymptotic Justification of the Estimation Procedure

We shall now state a theorem justifying the procedures of Section 2 in an asymptotic fashion. We make the following requirements.[1]

(i) The vector $x(n)$ is of the form

$$x(n) = \sum_{-\infty}^{\infty} G(j)\epsilon(n-j)\ , \qquad \sum_{-\infty}^{\infty} \|G(j)\| < \infty$$

where by $\|A\|$ we mean a norm for the matrix A (e.g. the square root of the greatest eigenvalue of $A'A$) and the $\epsilon(n)$ are independent and identically distributed random vectors with zero mean vector and $\mathcal{E}(\epsilon(n)\epsilon(n)') = I_q$. The matrix $\Gamma(n)$ introduced in Section 1 is related to the $G(j)$ by

$$\Gamma(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda)d\lambda\ , \tag{7}$$

$$f(\lambda) = \frac{1}{2\pi} h(\lambda)h(\lambda)^*\ , \quad h(\lambda) = \sum_{-\infty}^{\infty} G(j)e^{ij\lambda}$$

---

[1] The validity of unproved statements in this section is established in [10], Chapter VII.

where the star indicates transposition combined with conjugation. In (7) we mean the integral to be evaluated element by element of the matrix $f(\lambda)$. We assume that the determinant of $f(\lambda)$ is never zero in $[-\pi, \pi]$.

(ii) So far as the $x(n)$ are concerned we adopt specifications introduced in [7]. We call $D(N)$ the diagonal matrix with $d_j(N)$ in the j^{th} place, $j = 1, \ldots, p$, and

$$d_j(N) = \left\{ \sum_{n=1}^{N} x_j(n)^2 \right\}^{\frac{1}{2}} .$$

Then we require that the following limits exist and have the values shown

$$\lim_{N\to\infty} d_j(N) = \infty , \quad \lim_{N\to\infty} |x_j(N)|/d_j(N) = 0 ,$$

$$\lim_{N\to\infty} \left\{ \left(\sum_{n=1}^{N} x_j(m)x_k(m+n) \right) \Big/ \left(d_j(N)d_k(N) \right) \right\} = \rho_{jk}(n) .$$

Then

$$\rho_{jk}(n) = \int_{-\pi}^{\pi} e^{in\lambda} dm_{jk}(\lambda)$$

where $m_{jk}(\lambda)$ is a complex valued function whose real and imaginary parts are signed measures. We rewrite this relation as

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} dM(\lambda)$$

where $R(n)$ is a matrix with $\rho_{jk}(n)$ in the typical place and $M(\lambda)$ is similarly defined from the $m_{jk}(\lambda)$.

(iii) Let $m(N)$ be the minimum of the m_j used in the procedure of the last section. We require that

$$\lim_{N\to\infty} N^{\frac{1}{2}}/m(N) = 0 .$$

(iv) Let $M(N)$ be the maximum of the m_j used in the procedure. We may define $H_u(j)$, $K_u(j)$ in terms of $a_u(t)$, $b_u(t)$ in the same way as for $H_x(j)$, $K_x(j)$. Of course H_u , K_u are not computable. Then, for λ the midpoint of the interval in which the ω_t used in forming $H_u(j)$, $K_u(j)$, we may regard

$$\hat{f}(\lambda) = (2\pi N m_j)^{-1}\{H_u(j) - iK_u(j)\}$$

as m estimate of $f(\lambda)$. Now we keep λ fixed and, choosing m so that $m(N) \leq m \leq M(N)$ and m of the ω_t nearest to λ , we allow N , $m(N)$, $M(N)$ to increase. Then we require that

$$\lim_{N\to\infty} M(N)^{1/2}\|\mathcal{E}\{\hat{f}(\lambda)\} - f(\lambda)\| \leq a < \infty$$

where a is independent of λ . This condition imposes a restriction on the speed with which $M(N)$ (and hence $m(N)$) may increase. In order that this may not conflict with (iii) it is necessary that $f(\lambda)$ be reasonably smooth, for example differentiable. (For details see [10], Chapter V.)

(v) There is a final requirement which calls for some explanation. This is that $I_q \otimes D(N)$ commute with E. This is the same as saying that $I_p \otimes D(N)$ commutes with F. The restriction seems to be a mild one for the following reason. So long as $d_j(N)/d_k(N)$ converges to a finite, non zero, limit we may always modify our definition of $D(N)$ so that $d_j(N) \equiv d_k(N)$ and the theorem stated below remains true. This modification has no effect on the computations in Section 2. If this can be done for all pairs (j, k) then $D(N)$ may be made into a scalar multiple of the identity matrix and the condition is always met. Since $D(N)$ could often be taken to be $N^{\frac{1}{2}} I_p$ a wide range of cases is already included. In general we may divide the columns of B into sets so that all pairs of columns j, k, in the same set have $d_j(N)/d_k(N)$ converging to a finite non zero limit. Then any linear restriction must refer only to elements in the same set of columns. Insofar as $d_j(N)/d_k(N)$ does not so converge, as would be the case with $x_j(n) \equiv 1$, $x_k(n) = n$, then it seems unlikely that restrictions would involve both simultaneously. It may be possible to modify the theorem stated below when $I_q \otimes D(N)$ does not commute with E but the statement will be more complicated and we have not attempted to do that.

Theorem: Under conditions (i), (ii), (iii), (iv) the asymptotic distribution of $(I_q \otimes D(N))(\tilde{\beta} - \beta)$ converges to the multivariate normal distribution

<u>with zero mean vector and covariance matrix</u>

$$(8) \qquad \left\{ E \int_{-\pi}^{\pi} \{2\pi f(\lambda)\}^{-1} \otimes dM(\lambda)' E \right\}^{-1} ,$$

<u>which is also</u>

$$\lim_{N\to\infty} (I_q \otimes D(N)) \{E(I_q \otimes X')\Gamma_q^{-1}(I_q \otimes X)E\}^{-1} (I_q \otimes D(N)) .$$

<u>The matrix (8) is consistently estimated by</u>

$$(9) \qquad N^{-1}(I_p \otimes D(N))(EWE)^{-1}(I_p \otimes D(N)) .$$

By the integral in (8) we mean the matrix of pq integrals obtained by taking each element of $\{2\pi f(\lambda)\}^{-1} \otimes dM(\lambda)'$. We shall not give a proof of this Theorem. It is not essentially different from the main theorem of [8]. (See also [10], Chapter VII, Theorem 10.) It provides an asymptotic justification for the procedures of Section 2 since using $(I_q \otimes D(N))(\tilde{\beta} - \beta)$ as normal with zero mean vector and covariance matrix estimated by (9) is the same as using $\tilde{\beta}$ as normal with mean β and covariance matrix $N^{-1}(EWE)^{-1}$.

4. Applications

System of Demand Equations

The estimation of a system of demand equations subject to the constraints derived from the theory of Consumer Demand [6] is a suitable area of application for the suggested methods. As an aid to exposition only we assume that the demand equations are linear after logarithmic transformation of of the variables and further that the system involves only two components of demand. A system containing a more realistic number of commodities would introduce much greater detail, which would, we believe, be less rather than more illuminating and would require an excessive amount of space to present.

The variables in the system are

$$y(n) = \begin{pmatrix} \log q_1(n) \\ \log q_2(n) \end{pmatrix}, \quad x(n) = \begin{pmatrix} \log p_1(n) \\ \log p_2(n) \\ \log O(n) \end{pmatrix}$$

where $q_i(n)$ is the quantity purchased of the i^{th} commodity, $p_i(n)$ is the price of the i^{th} commodity and $O(n)$ a measure of income or of total outlay in the system, in period n.

We will consider three sets of restrictions, homogeneity, aggregation (Cournot and Engel) and symmetry (see [6]). The main illustrative points arise in the implementation of each set, taken separately, but later a brief comment is given on combining any of the sets.

Previous investigations of systems of demand equations (see [1], [2], [3], [4], [5] and [11]) have either directly specified or implicitly assumed that the nature of $\mathcal{E}(u(m)u'(n))$ is of the form (i) or (ii) discussed in Section 1. We derive the projection matrix, E, associated with each <u>separate</u> set of restrictions and discuss its relevance for estimation under the different specifications for the variance covariance matrix for the disturbance vector. It is however, in the part E plays in the evaluation of $(EWE)^{-1}$ in (6) when the third specification is appropriate that we are most interested.

The homogeneity restrictions for the given situation are

$$\alpha_H'(1) = (1,\ 1,\ 1,\ 0,\ 0,\ 0)\ ,\quad \alpha_H'(2) = (0,\ 0,\ 0,\ 1,\ 1,\ 1)$$

$$c_1 = c_2 = 0$$

and so

$$\beta_H'(i) = (1/\sqrt{3})\alpha_H'(i)\ ,\quad i = 1,\ 2\ .$$

The projection matrix E to be used in the restricted estimator is then of the form, $E = I_2 \otimes E_1$ where the matrix E_1 is

$$E_1 = \begin{bmatrix} (2/3)\ , & (-1/3)\ , & (-1/3)\ , \\ (-1/3)\ , & (2/3)\ , & (-1/3)\ , \\ (-1/3)\ , & (-1/3)\ , & (2/3) \end{bmatrix} .$$

Because of this structure of E, when the homogeneity restrictions are employed alone, the simple least squares procedure in which $x(n)$ is replaced by $P(n)$ and B by BP' is efficient for all G (see specification (ii)). The matrix P is

$$P = \begin{bmatrix} 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & -1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

and is the orthogonal matrix such that PE_1P' is diagonal with units in the main diagonal save for a zero in the last place.

Under the more general third specification on $\mathcal{E}(u(m)u'(n))$ the efficient estimate is (3.iii). It has already been emphasized that the computation is best pursued through (6) and thus requires the evaluation of $(EWE)^{-1}$, which for $E = I_2 \otimes E_1$ and E_1 as specified above, becomes

$$\frac{1}{M}\sum_j\{\hat{C}(j) \otimes E_1 H_x(j)E_1 + \hat{Q}(j) \otimes E_1 K_x(j)E_1\} + I_2 \otimes \frac{1}{3}\, 1_3 1_3'\}^{-1} - I_2 \otimes \frac{1}{3}\, 1_3 1_3'$$

where 1_3 is a vector of three units.

If it is desired to impose and test the validity of the aggregation restrictions alone then we define the vectors,

$$\alpha_A'(1) = (w_1,\ 0,\ 0,\ w_2,\ 0,\ 0)\ , \quad \alpha_A'(2) = (0,\ w_1,\ 0,\ 0,\ w_2,\ 0)$$

$$\alpha_A'(3) = (0,\ 0,\ w_1,\ 0,\ 0,\ w_2)\ , \quad c_1 = -w_1\ , \quad c_2 = -w_2\ , \quad c_3 = 1$$

where $w_i(n) = (p_i(n)q_i(n)/0(n))$, $i = 1, 2$. An adjustment is first made to the y and β vectors by substracting $(I_q \otimes X)e$ and e respectively where e is simply

$$e = \sum_{k=1}^{3} d_k \phi(k)$$

where

$$\phi'(k) = (\alpha'(k))/h\ , \quad d_k(c_k)/h\ , \quad h = \sqrt{w_1^2 + w_2^2}\ , \quad k = 1,\ 2,\ 3\ .$$

Now the matrix E is of the form $L_2 \otimes I_3$, where the matrix L_2 is

$$L_2 = \frac{1}{h^2}\begin{bmatrix} w_2^2 & -w_1w_2 \\ -w_2w_1 & w_1^2 \end{bmatrix}.$$

and because E is of the general form $E_1 \otimes E_2$ then the B.L.U.E. under the first specification for $\mathcal{E}(u(m)u(n)')$ is simply obtained, as stated in Section 1, by unrestricted ordinary least squares regression after changing to new x and y variables. The estimate relevant to specification (ii) becomes

$$\{(L_2G^{-1}L_2)^{-1} \otimes (X'X)^{-1}\}(G^{-1} \otimes X')y$$

and the evaluation of $(EWE)^{-1}$ needed for the estimate (6) using specification (iii) is obtained from

$$\left(\frac{1}{M}\sum_j\{L_2\hat{C}(j)L_2 \otimes H_x(j) + L_2\hat{Q}(j)L_2 \otimes K_2(j)\} + (L_1 \otimes I_3)\right)^{-1} - L_1 \otimes I_3$$

where

$$L_1 = \frac{1}{h^2}\begin{bmatrix} w_1^2 & -w_1w_2 \\ -w_2w_1 & w_2^2 \end{bmatrix} .$$

It is conceivable that the only restrictions that an investigator may wish to impose are the symmetry restrictions. The system considered here has only one independent symmetry restriction defined by

$$\alpha_S'(1) = \{(1/w_2),\ 0,\ 1,\ (-1/w_1),\ 0,\ -1\}$$

and

$$\beta_S'(1) = (1/\ell)\alpha_S'(1)\ ,\quad \ell = \sqrt{2 + (1/w_1^2) + (1/w_2^2)}\ .$$

The associated projection matrix,

$$E = \frac{1}{\ell^2 w_1 w_2}\begin{bmatrix} w_1w_2\ell^2-(w_1w_2) & 0 & w_1 & -1 & 0 & -w_1 \\ 0 & \ell^2 w_1w_2 & 0 & 0 & 0 & 0 \\ w_1 & 0 & (\ell^2-1)w_1w_2 & -w_2 & 0 & -w_1w_2 \\ -1 & 0 & -w_2 & (w_1w_2\ell^2-(w_2/w_1)) & 0 & w_2 \\ 0 & 0 & 0 & 0 & \ell^2 w_1w_2 & 0 \\ -w_1 & 0 & -w_1w_2 & w_2 & 0 & (\ell^2-1)w_1w_2 \end{bmatrix}$$

can be clearly shown not to be of the form $E_1 \otimes E_2$ and therefore E as defined above is used in (3(i)), (3(ii)) or in (6), the computationally convenient form for (3(iii)), depending on the appropriate specification for $\xi(u(m)u'(n))$.

In practice some combination of these sets of restrictions will probably be employed. Suppose for example that both the homogeneity set and the aggregation set of restrictions were required. This combined set of restrictions only becomes linearly independent when one restriction is dropped; then the remaining restrictions are orthonormalized and the projection matrix F is of the form

$$F = L_1 \otimes I_3 + L_2 \otimes U ,$$

where

$$U = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} .$$

Thus E is $L_2 \otimes (I_3 - U)$ and is again of the form $E_1 \otimes E_2$ as was the case for the aggregate restrictions alone. The minor simplification of the proposed estimates under various disturbance specifications resulting from the fact that $E = L_2 \otimes (I_3 - U)$ is not pursued here because it follows analogous lines to those discussed for aggregation restrictions alone. A combination of the homogeneity set and the symmetry set of restrictions or of the aggregation set and the symmetry set of restrictions may also be posited. In both of these cases the combined set of restrictions is linearly independent and since E is not of the form $E_1 \otimes E_2$ the estimates are simply obtained from (3(i)), (3(ii)) and (6).

REFERENCES

[1] Barten, A.P., "Consumer Demand Functions under Conditions of Almost Additive Preferences," Econometrica, Vol. 32 (1964), pp. 1-38.

[2] ________, "Estimating Demand Equations," Econometrica, Vol. 36 (1968), pp. 213-251.

[3] Byron, R.P., "Methods for Estimating Demand Equations," Australian Economic Papers, Vol. 7 (1968), pp. 227-248.

[4] ________, "A Simple Method for Estimating Demand Systems under Separable Utility Assumptions," forthcoming Rev. of Ec. Studs., April 1970.

[5] Court, R.H., "Utility Maximization and the Demand for New Zealand Meats," Econometrica, Vol. 35 (1967), pp. 424-446.

[6] Frisch, R., "A Complete Scheme for Computing all Direct and Cross Demand Elasticities in a Model with Many Sectors," Econometrica, Vol. 27 (1959), pp. 177-196.

[7] Grenander, U., "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," Ann. Math. Statist., 25 (1954), pp. 252-272.

[8] Hannan, E.J., "Regression for Time Series," in Time Series Analysis, edited by M. Rosenblatt (1963), New York, John Wiley.

[9] ________, "Least-Squares Efficiency for Vector Time Series," J.R.S.S. Series B, Vol. 30 (1968), pp. 490-498.

[10] ________, Multiple Time Series, New York, John Wiley, 1970.

[11] Powell, A., "A Complete System of Consumer Demand Equations for the Australian Economy Fitted by a Model of Additive Preferences," Econometrica, Vol. 34 (1966), pp. 661-675.