1972

Measurement of Structural Change: An Application of Random Coefficient Regression Model

Balvir Singh

A. L. Nagar

Baldev Raj

20299

7203

MEASUREMENT OF STRUCTURAL CHANGE

An Application of Random Coefficient Regression Model

by

Balvir Singh, A. L. Nagar and Baldev Raj

March 29, 1972 Preliminary

Research Report 7203

MEASUREMENT OF STRUCTURAL CHANGE

An Application of Random Coefficient Regression Model

by

Balvir Singh, A. L. Nagar and Baldev Raj

March, 1972

MEASUREMENT OF STRUCTURAL CHANGE

(An Application of Random Coefficient Regression Model)

Balvir Singh, A. L. Nagar and Baldev Raj[1]
University of Toronto, Delhi School of Economics, University of Western Ontario

1. Introduction

The multiple linear regression model with constant coefficients is often employed to explain variation in a certain economic variable in terms of some explanatory variables. Let us write such a model in algebraic form as

$$(1.1) \qquad y(t) = \sum_{\lambda=1}^{\Lambda} \beta_\lambda x_\lambda(t) + u(t) ,$$

where $y(t)$ and $x_\lambda(t)$ are the t-th $(t=1,\ldots,T)$ observations on the dependent variable (y) and the λ-th $(\lambda=1,\ldots,\Lambda)$ explanatory variable x_λ, respectively, and $u(t)$ is the usual disturbance term.

In the specification (1.1) the regression coefficient

$$(1.2) \qquad \beta_\lambda = \frac{\partial y(t)}{\partial x_\lambda(t)} , \qquad \lambda=1,\ldots,\Lambda ,$$

is assumed to remain constant over the entire sample period. This implies that the response in $y(t)$ to per unit change in $x_\lambda(t)$ is the same for all $t=1,\ldots,T$. Further, the same functional form of relationship between the dependent and explanatory variables is assumed for all time points in the sample period. A sample satisfying these restrictions may be called 'structurally homogeneous.' In practical examples, this property refers to peacetime pattern of behaviour, where any abnormal happenings (war, etc.) are excluded.

The fact, however, remains that this holy property of structural homogeneity (in its strict sense) is simply a rare phenomenon. Even during

---

[1]The authors are grateful to Professors S. J. Turnovsky and N. K. Choudhry for helpful discussions.
The research of the first author was partially supported by a grant from the Humanities and Social Science Committee, University of Toronto.

peace time, y may respond differently to per unit change in x_λ, over different years. Equal per unit change in total income may produce different effects in the aggregate consumption over different years depending upon overall consumer expectations about the future income and consumers' attitudes.[1] Same level of sales in different years may induce businessmen to take different decisions with regard to investment in inventories for the reason of change in government policy and business conditions. Equal doses of labour and capital in a particular production process may yield different levels of output over different years in view of technical progress, change in labour efficiency and managerial ability that might take place during the course of sample period.[2] Same amount of fertilizers applied to a particular farm may yield different output as the temperature, the rainfall, managerial and professional efficiency of the farmer may vary over different years.[3] Clearly then the analysis carried out in terms of (1.1) may be of dubious value. In such cases one may be required to either postulate different forms of the structural equation for different observations, or specify different values to regression parameters or do both.

In this paper we assume that the same form of functional relationship between the dependent and explanatory variables holds for all points of time in the sample period. However, we do not assume the regression coefficients to remain constant over time.

In the Section 2 we formulate alternative hypotheses about the random character of the regression coefficients:

---

[1]Klein (1953), pp. 216-18

[2]Nerlove (1965), pp. 34-35

[3]Hildreth and Houck (1968), p. 584.

(A) $\beta_\lambda(t)$, at time t, is a random variable with constant mean $E\,\beta_\lambda(t) = \bar{\beta}_\lambda$ and $V\,\beta_\lambda(t) = \sigma_{\lambda\lambda}$,

(B) $\beta_\lambda(t)$, at time t, is a random variable with mean $E\,\beta_\lambda(t) = \bar{\beta}_\lambda + \bar{\alpha}_\lambda\, f_\lambda(t)$

and $\text{Var}\,\beta_\lambda(t) = \sigma_{\lambda\lambda}$.

According to the formulation (A) the response coefficients fluctuate around a constant; however, according to (B) they fluctuate around a trend--thus indicating a parameter shift in the model. One may test the hypothesis $\bar{\alpha}_\lambda = 0$ to establish whether there is a parameter shift or not. The form of $f_\lambda(t)$ to be selected is crucial no doubt. It may be advisable to specify various alternative forms of $f_\lambda(t)$ and choose the one corresponding to which the multiple correlation is found to be the highest.

Different methods of estimation of regression models with random coefficients have been discussed in Section 3 and then in Section 4 we consider an application of the technique to examine structural change in the consumption function relating to Canada, Finland, India, Japan, Netherlands, Philippines and U.K.

2. General Linear Regression Model with Random Coefficients

Let us write the general linear regression model with random coefficients as

$$y(t) = \sum_{\lambda=1}^{\Lambda} \beta_\lambda(t)\, x_\lambda(t), \qquad t=1,\ldots,T, \tag{2.1}$$

where $y(t)$ and $x_\lambda(t)$ are observations, at time t, on the left-hand dependent variable (to be explained) and the right-hand explanatory variables, respectively; and

$$(2.2) \qquad \beta_\lambda(t) = \frac{\partial y(t)}{\partial x_\lambda(t)}$$

measures the change in $y(t)$ per unit change in $x_\lambda(t)$, at time t. In this sense $\beta_\lambda(t)$ is interpreted as a 'response coefficient' and it is assumed to vary with t. We will assume that x's are nonstochastic and fixed in repeated samples and they are independent of any variation in $\beta_\lambda(t)$'s. As we have discussed in the preceding section the variation in response coefficients may be attributed to the exclusion of certain explanatory variables from the equation. Therefore, it is not really necessary to include a disturbance term explicitly in (2.1).[1]

The specification (2.1) has ΛT coefficients, $\beta_\lambda(t)$'s, to be estimated. As it is obvious one cannot estimate all of these coefficients with only T observations at one's disposal. We may adopt one of the two alternative procedures to overcome this difficulty. Either, employ some extraneous information along with the sample data, or, alternatively, make some simplifying assumptions about the probabilistic behaviour of the response coefficients and thus reduce the number of parameters to be estimated. In this paper we shall adopt the latter approach.

A. Constant Mean Response (CMR)

This is the most commonly used assumption about the stochastic behaviour of the response coefficients. We specify

$$(2.3) \qquad \beta_\lambda(t) = \bar{\beta}_\lambda + \epsilon_\lambda(t), \qquad t=1,\ldots,T,$$

---

[1]We exclude the disturbance term from the equation primarily for the sake of convenience. There is, of course, no loss of generality involved in doing so.

where

$$E\,\epsilon_\lambda(t) = 0, \quad \text{Var}\,\epsilon_\lambda(t) = \sigma_{\lambda\lambda} \text{ and } \text{Cov}\,[\epsilon_\lambda(t), \epsilon_{\lambda'}(t')] = 0, \text{ for } \lambda \neq \lambda', t \neq t'. \tag{2.4}$$

This specification implies

$$E\,\beta_\lambda(t) = \bar{\beta}_\lambda \quad \text{and} \quad \text{Var}\,\beta_\lambda(t) = \sigma_{\lambda\lambda}, \tag{2.5}$$

that is, $\beta_\lambda(t)$'s (for t=1,...,T) fluctuate around their mean $\bar{\beta}_\lambda$ with variance $\sigma_{\lambda\lambda}$. Then we may write (2.1) as follows:

$$y(t) = \sum_{\lambda=1}^{\Lambda} \bar{\beta}_\lambda x_\lambda(t) + w(t), \tag{2.6}$$

where

$$w(t) = \sum_{\lambda=1}^{\Lambda} x_\lambda(t)\,\epsilon_\lambda(t), \qquad t=1,\ldots,T; \tag{2.7}$$

and we observe that

$$E\,w(t) = 0, \quad \text{Var}\,w(t) = \sum_{\lambda=1}^{\Lambda} x_\lambda^2(t)\,\sigma_{\lambda\lambda} \equiv \varphi_t^2 \text{ and } \text{Cov}[w(t)\,w(t')] = 0, \text{ if } t \neq t'. \tag{2.8}$$

Thus we have to estimate parameters of the regression model (2.6) which has constant coefficients and heteroscedastic errors. We will be estimating the mean responses $\bar{\beta}_1,\ldots,\bar{\beta}_\Lambda$ and variances $\sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ specified in (2.4).

B. Variable Mean Response (VMR)/Mean Response with a Trend

In time series analysis it is more likely that the response coefficients $\beta_\lambda(t)$'s, for t=1,...,T, fluctuate around some trend rather than about a fixed value $\bar{\beta}_\lambda$. For instance, consumer's taste may vary over time and this may cause systematic shift in the marginal propensity to consume. More generally, some parameters may rise systematically in the upswing of the business cycle and fall during the downswing. Should the movement of the parameters in the two phases be symmetric, the path of the parameter over the complete cycle can be approximated by a parabolic trend, or some other suitable trend that can be specified. In such situations, it may be more appropriate to specify

$$(2.9) \qquad \beta_\lambda(t) = \bar{\beta}_\lambda + \bar{\alpha}_\lambda f_\lambda(t) + \epsilon_\lambda(t),$$

where $f_\lambda(t)$ is some function of t,

$$(2.10) \qquad E\,(\beta_\lambda(t)) = \bar{\beta}_\lambda + \bar{\alpha}_\lambda f_\lambda(t), \qquad t=1,\ldots,T,$$

$\bar{\beta}_\lambda$ being the intercept term and

$$(2.11) \qquad \bar{\alpha}_\lambda \frac{\partial f_\lambda(t)}{\partial t}$$

the slope of the trend around which $\beta_\lambda(t)$'s, $t=1,\ldots,T$, fluctuate with a disturbance $\epsilon_\lambda(t)$. We may assume that $\epsilon_\lambda(t)$'s have the same properties as stated in (2.4). The function $f_\lambda(t)$ represents a general form of the trend--linear, parabolic, exponential, etc., which may be specified differently for different parameters, viz., $\beta_\lambda(t)$'s for $\lambda=1,\ldots,\Lambda$. In this case (2.1) can be expressed as

$$(2.12) \qquad y(t) = \sum_{\lambda=1}^{\Lambda} \{\bar{\beta}_\lambda + \bar{\alpha}_\lambda f_\lambda(t) + \epsilon_\lambda(t)\}\, x_\lambda(t) ,$$

or

$$(2.13) \qquad y(t) = \sum_{\lambda=1}^{\Lambda} \bar{\beta}_\lambda x_\lambda(t) + \sum_{\lambda=1}^{\Lambda} \bar{\alpha}_\lambda f_\lambda(t)\, x_\lambda(t) + w(t)$$

$$= \sum_{\lambda=1}^{\Lambda} \bar{\beta}_\lambda x_\lambda(t) + \sum_{\lambda=1}^{\Lambda} \bar{\alpha}_\lambda x^*_\lambda(t) + w(t) ,$$

where

$$(2.14) \qquad x^*_\lambda(t) = f_\lambda(t)\, x_\lambda(t) \quad \text{and} \quad w(t) = \sum_{\lambda=1}^{\Lambda} \epsilon_\lambda(t)\, x_\lambda(t).$$

It is interesting to note that (2.6) can be obtained from (2.13) if we write $\bar{\alpha}_\lambda = 0$ in the latter equation. Thus the constant mean response (CMR) approach can be interpreted as a special case of the proposed variable mean response (VMR) approach. Further, if $f_\lambda(t)$ is specified as a linear or quadratic trend, then the systematic component of $\beta_\lambda(t)$ in (2.9) will be identical with the Stone's formulation for representing the dynamic Engel function.[1]

The problem in the present approach is mainly the determination of the form of $f_\lambda(t)$. Since intuitive guesses may not be necessarily appropriate, it is advisable to be guided by the sample information, i.e., estimate (2.13) with alternative forms for $f_\lambda(t)$ and then choose the one which explains the maximum variation in the dependent variable and yields highest numerical magnitude of the multiple correlation coefficient.

3. Estimation of the Linear Regression Model with Random Coefficients

As we have seen above a general linear regression model (2.1) with random coefficients can be reduced to a linear regression model with constant coefficients and heteroscedastic errors by making appropriate assumptions about the stochastic behaviour of the response coefficients. We considered the CMR and VMR approaches and arrived at the specifications (2.6) and (2.13), respectively. It was noted that (2.6) can be derived from (2.13) if we let $\bar{\alpha}_\lambda = 0$ in (2.13), thus CMR is a special case of VMR. In fact (2.13) can be treated as the most general formulation of linear regression models with random coefficients. Therefore, let us consider estimation of (2.13) as follows.

---

[1] Cf. Stone (1965) p. 276 and Stone (1966) p. 435.

In matrix notation (2.13) can be expressed as

$$(3.1) \qquad y = X\beta + w ,$$

where

$$(3.2) \qquad y = \begin{bmatrix} y(1) \\ \vdots \\ y(T) \end{bmatrix}$$

is the $T \times 1$ vector of observations on the left hand dependent variable, 'to be explained,' and

$$(3.3) \qquad X = \begin{bmatrix} x_1(1) & \cdots & x_\Lambda(1) & x_1^*(1) & \cdots & x_\Lambda^*(1) \\ \vdots & & & & & \vdots \\ x_1(T) & & x_\Lambda(T) & x_1^*(T) & & x_\Lambda^*(T) \end{bmatrix}$$

is the matrix of observations on the right hand explanatory variables in (2.13). Let us note that there are 2Λ columns in X, and the definition of $x_\lambda^*(t)$ is given in (2.14). The coefficient vector

$$(3.4) \qquad \beta = \begin{bmatrix} \bar{\beta}_1 \\ \vdots \\ \bar{\beta}_\Lambda \\ \bar{\alpha}_1 \\ \vdots \\ \bar{\alpha}_\Lambda \end{bmatrix}$$

has 2Λ elements as in (2.13), and

$$(3.5) \qquad w = \begin{bmatrix} w(1) \\ \vdots \\ w(T) \end{bmatrix}$$

is the disturbance vector with $w(t) = \sum_{\lambda=1}^{\Lambda} \epsilon_\lambda(t)\, x_\lambda(t)$ as defined in (2.14).

It should be noted that

$$(3.6) \qquad Ew = 0 \quad \text{and} \quad Eww' = \Phi = \begin{bmatrix} \varphi_1^2 & \cdots & 0 \\ \vdots & & \\ 0 & & \varphi_T^2 \end{bmatrix}$$

where $\varphi_t^2 = \text{var } w(t) = \sum_{\lambda=1}^{\Lambda} x_\lambda^2(t)\, \sigma_{\lambda\lambda}$ has been defined in (2.8).

If we assume that the elements of X are independently distributed of those of w and

$$(3.7) \qquad \text{Rank } X = 2\Lambda \leqq T ,$$

such that X'X is a nonsingular matrix, then the Least Squares estimator

$$(3.8) \qquad b = (X'X)^{-1} X'y$$

is a consistent estimator of β. However, it is well known that this will be an 'inefficient' estimator because the covariance matrix Φ, defined in (3.6), is not a scalar times an identity matrix as required by the Gauss-Markov theorem. The asymptotic covariance matrix of the least squares estimator b is

$$(3.9) \qquad E\,(b-\beta)(b-\beta)' = (X'X)^{-1} X'(Eww')\, X(X'X)^{-1} = (X'X)^{-1} X'\,\Phi X(X'X)^{-1}.$$

If Φ would be completely known one could apply the Generalized Least Squares procedure to (3.1), and then

$$(3.10) \qquad \tilde{\beta} = (X'\,\Phi^{-1} X)^{-1} X'\,\Phi^{-1} y$$

would be an efficient estimator of β. However, a straight forward application of the Generalized Least Squares is not possible in the present case because the elements of Φ depend upon $\sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ which are unknown parameters. It is, of course, possible to follow a two step procedure as proposed by Hildreth and Houck (1968), i.e., in the first step estimate $\sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ and obtain an estimate of Φ, then using this estimate of Φ apply generalized least squares to (3.1).

Let us outline the Hildreth and Houck procedure briefly as follows:

Apply Ordinary Least Squares to (3.1) to obtain b as defined in (3.8). This gives the Least Squares estimator of w as

$$(3.11) \qquad \hat{w} = y - X(X'X)^{-1} X'y ,$$

or,

$$(3.12) \qquad \hat{w} = Mw ,$$

where

$$(3.13) \qquad M = I - X(X'X)^{-1} X'.$$

It has been noted by Hildreth and Houck (1968) that

$$(3.14) \qquad E\,\dot{\hat{w}} = \dot{M}\dot{X}\sigma , \qquad \sigma = \begin{bmatrix} \sigma_{11} \\ \vdots \\ \sigma_{\Lambda\Lambda} \end{bmatrix} ,$$

where $\dot{\hat{w}}$ is the column vector of squared elements of $\hat{w}$ and similarly $\dot{M}$, $\dot{X}$ are matrices of squared elements of M, X, respectively.

Then apply Ordinary Least Squares to

$$(3.15) \qquad \dot{\hat{w}} = \dot{M}\dot{X}\sigma + \eta$$

where η is a column vector of disturbances, to obtain

$$(3.16) \qquad \hat{\sigma} = (\dot{X}'\dot{M}'\dot{M}\dot{X})^{-1} \dot{X}'\dot{M}'\dot{\hat{w}} .$$

Using the corresponding estimator $\hat{\Phi}$ of Φ, obtain

$$(3.17) \qquad \hat{\beta} = (X' \hat{\Phi}^{-1} X)^{-1} X' \hat{\Phi}^{-1} y,$$

and the asymptotic covariance matrix of $\hat{\beta}$ as

$$(3.18) \qquad E\,(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (X' \Phi^{-1} X)^{-1} .$$

It should be noted that $\hat{\sigma}$, as defined in (3.16), is a consistent but an inefficient estimator of σ. The asymptotic covariance matrix of $\hat{\sigma}$ is

$$E(\hat{\sigma}-\sigma)(\hat{\sigma}-\sigma)' = (\dot{X}'\dot{M}'\ddot{M}\dot{X})^{-1}\,\dot{X}'\dot{M}'\,(E\,\eta\eta')\,\ddot{M}X\,(\dot{X}'\dot{M}'\dot{M}\dot{X})^{-1}, \tag{3.19}$$

and as shown in the Appendix A we have

$$E\,\eta\eta' = 2\,\dot{\Psi} \tag{3.20}$$

where $\dot{\Psi}$ is the T x T matrix of the squared elements of $\Psi = M\Phi M$.

If we estimate $\sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ according to (3.16), the estimates so obtained will have a range $-\infty$ to $+\infty$. Thus there will always be a positive probability of arriving at negative estimates of σ's.[1]

Since

$$\operatorname*{plim}_{T\to\infty} \hat{\sigma} = \sigma, \tag{3.21}$$

the probability of getting negative estimates of the variances $\sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ tends to zero as the sample size increases indefinitely. On the contrary, this probability will be substantial in case of small samples because the estimates obtained in (3.16) have large standard errors. We can minimize the probability of getting negative estimates of σ's if we employ an efficient procedure of estimating σ from (3.15), instead of the Ordinary Least Squares as suggested by Hildreth and Houck. For example, one may adopt the two-step Aitken procedure as follows:

In the first step obtain $\hat{\sigma}$ according to the Ordinary Least Squares, as in (3.16), and obtain an estimate of $E\,\eta\eta'$ as defined in (3.20). Employ

---

[1] In their paper Hildreth and Houck (1968) have suggested that one may replace negative $\hat{\sigma}$'s by zeros as their closest approximation. Alternatively, one may apply nonlinear programming such that σ's are then constrained to be non-negative. It is obvious that the first approach is arbitrary, and if one follows the second approach there is no way of determining the precision of the estimates.

this estimated covariance matrix to obtain a fresh estimate of σ according to generalized least squares applied to (3.15).[1]

The efficient estimates of β can then be obtained as in (3.17). In a special case, Theil and Mennes (1959) have shown that the off-diagonal elements of (3.20) are of lower order of magnitude than the diagonal elements. In fact it is not difficult to see that this result holds good in general. Therefore, instead of using $E\,\eta\eta'$, as defined in (3.20), we may use the diagonal matrix obtained from (3.20) by replacing the off-diagonal elements by zeros, to apply the two-step Aitken procedure to (3.15).

Besides the Generalized Least Squares approach discussed above, one may follow the Maximum Likelihood procedure, to estimate the parameters of (3.1), as proposed by Rubin (1950). Obviously, then, one has to assume that the disturbances in (3.1) are normally distributed with means and variances as specified in (3.6).

The log likelihood function of the parameters in (3.1) can be written as,

$$(3.22) \qquad L = -\frac{T}{2}\log 2\Pi - \tfrac{1}{2}\log|\Phi| - \tfrac{1}{2} w'\Phi^{-1} w$$

where

$$(3.23) \qquad |\Phi| = \prod_{t=1}^{T} \varphi_t^2 \quad \text{and} \quad \varphi_t^2 = \sum_{\lambda=1}^{\Lambda} x_\lambda^2(t)\,\sigma_{\lambda\lambda}$$

are obtained from (3.6) and (2.8), respectively.

The likelihood equations can then be written as follows:

$$(3.24) \qquad (X'\Phi^{-1}X)\,\tilde{\tilde{\beta}} = X'\Phi^{-1}y$$

and

---

[1] In fact this procedure can be repeated until the iterations converge to a stable $\hat{\sigma}$.

$$(3.25) \qquad \sum_{t=1}^{T} \frac{1}{\varphi_t^2} \left\{ 1 - \frac{w^2(t)}{\varphi_t^2} \right\} x_\lambda^2(t) = 0 , \qquad \lambda=1,\ldots,\Lambda,$$

where w(t) is the t-th element of the vector w defined in (3.5).

These equations are highly non-linear. We can solve them by employing various numerical methods which are essentially iterative in nature. Starting with some initial values for the parameters we obtain the final estimates by way of successive approximations. However, for the convergence of any iterative procedure it is necessary that the initial value of parameter lies within the 'domain of attraction'.[1] Since it is hard to find this domain, one has to depend upon educated guesses based on some prior evidence. Thus, the convergence may sometimes be jeopardized.

In addition, the likelihood equations may have multiple roots and corresponding multiple relative maxima of the likelihood function in which case the iterative procedure may converge to relative rather than the global maximum. One may as well face the problem of saddle point of the function, in which case the method may not converge at all. In general, therefore, a more important problem is to find an iterative method which is at least sure to converge rather than how fast it does.

In the present analysis we shall employ a suitably modified Gauss-Newton technique the details of which have been presented in the Appendix B.

4. Illustration: An Analysis of Structural Change in Consumption Function

To analyze the problem of structural change, as discussed in Section 2, let us consider the following consumption function:

---

[1] Cf. Beltrami (1970. The 'domain of attraction' is defined as a set of initial points, say β^o, for which β^ℓ, the ℓ-th iterate, converges to the true parameter value β. Moreover, the point β, being unique does not depend on the initial approximation chosen within the set of points attracted to β.

$$(4.1)\qquad C(t) = K(t)\, Y_p(t), \qquad t=1,\ldots,T.$$

where $C(t)$ and $Y_p(t)$ represent private consumption expenditure and permanent income, respectively, at time t.[1] The coefficient, $K(t)$, of permanent income is postulated to vary with time because this may depend upon, interest rate, ratio of wealth and income and a host of other variables which are not likely to remain constant.[2] The permanent income, $Y_p(t)$, is, in fact, an unobservable quantity. One may, however, define it as a weighted sum of current and all past years' incomes. To simplify matters, one may follow Cagan's (1956) or Koyck's (1954) weighting scheme and in that case (4.1) can be expressed as

$$(4.2)\qquad C(t) = \beta_1(t)\, Y(t) + \beta_2(t)\, C(t-1)$$

where $Y(t)$ is observed disposable income. We assume that

$$(4.3)\qquad \beta_1(t) = \bar{\beta}_1 + \bar{\alpha}_1\, t + \epsilon_1(t)$$

$$(4.4)\qquad \beta_2(t) = \bar{\beta}_2 + \bar{\alpha}_2\, t + \epsilon_2(t)$$

and note that it is a particular case of the specification (2.9) such that $f_1(t) = f_2(t) = t$. This is only a simplifying assumption. One can specify alternative forms for $f_\lambda(t)$ and determine the appropriate form according to the multiple correlation criterion.

Combining (4.2), (4.3) and (4.4) we have

$$(4.5)\qquad C(t) = \bar{\beta}_1\, Y(t) + \bar{\beta}_2\, C(t-1) + \bar{\alpha}_1\, Y^*(t) + \bar{\alpha}_2\, C^*(t-1) + w(t),$$

where

$$(4.6)\qquad Y^*(t) = t\,Y(t), \qquad C^*(t-1) = t\, C(t-1)$$

and

---

[1]Assuming that the transitory component of consumption is small we disregard it in the specification (4.1). Also see Evans (1969), p. 22.

[2]Cf. Friedman (1957), p. 229.

TABLE 1: ESTIMATION OF (4.5)

	Method of Estimation	$\bar{\beta}_1$	$\bar{\beta}_2$	$\bar{\alpha}_1$	$\bar{\alpha}_2$	σ_{11}	σ_{22}	R^2
CANADA - 1950-1968	I	0.44141**	0.57690**	0.00565	-0.00756			0.9992
		(0.09214)	(0.11061)	(0.01112)	(0.01330)			
	II	0.55566**	0.43140**	-0.01782	0.02075	-0.00052	0.00080	1.000
		(0.09154)	(0.11163)	(0.01299)	(0,01551)	(0.00032)	(0.00045)	
	III (a)	0.45288**	0.55933**	-0.00109	0.00061	0.00081*	0.00117**	
		(0.07899)	(0.09978)	(0.01235)	(0.01489)	(0.00026)	(0.00037)	
	III (b)	0.44165**	0.57257**	0.00033	-0.00106	-0.00068**	0.00099**	
		(0.09321)	(0.11458)	(0.01299)	(0.01557)	(0.00022)	(0.00031)	
	IV	0.36646**	0.66430**	0.01839**	-0.02236**	0.00029**	-0.00031**	
		(0.09654)	(0.11135)	(0.00578)	(0.00658)	(0.00000)	(0.00000)	
FINLAND - 1951-1968	I	0.79196**	0.16565*	-0.00834	-0.00663			0.9992
		(0.07227)	(0.08642)	(0.01457)	(0.01810)			
	II	0.82205**	0.13676**	-0.01849	0.01848	0.00039	-0.00043	0.9991
		(0.03831)	(0.04314)	(0.01096)	(0.01356)	(0.00096)	(0.00144)	
	III (a)	0.84692**	0.10625	-0.02026	0.02065	0.00057	-0.00055	
		(0.07019)	(0.08369)	(0.01578)	(0.01962)	(0.00050)	(0.00072)	
	III (b)	0.86179**	0.08816	-0.02125	0.02186	0.00041	-0.00032	
		(0.07313)	(0.08863)	(0.01565)	(0.01951)	(0.00047)	(0.00067)	
	IV	0.81310**	0.14612**	-0.01686	0.01659	0.00035**	0.00038**	
		(0.03857)	(0.04393)	(0.01056)	(0.01310)	(0.00010)	(0.00010)	
INDIA - 1952-1965	I	0.88368**	0.08768	0.00517	-0.00926			0.9980
		(0.12897)	(0.14458)	(0.00933)	(0.01087)			
	II	0.82080**	0.14868	0.00745	-0.01079	-0.00054	0.00090	0.9970
		(0.09685)	(0.10964)	(0.00655)	(0.00776)	(0.00049)	(0.00068)	
	III (a)	0.74751**	0.23093**	0.01574*	-0.02028*	-0.00095*	0.00151*	
		(0.06553)	(0.078095)	(0.00671)	(0.007119)	(0.00043)	(0.00062)	
	III (b)	0.72948**	0.25130**	0.01809*	-0.00230	-0.00095**	0.00150*	
		(0.05066)	(0.06272)	(0.00825)	(0.00901)	(0.00037)	(0.00053)	
	IV	SAME AS H-H ESTIMATES						

	Method of Estimation	β_1	β_2	α_1	α_2	σ_{11}	σ_{22}	R^2
JAPAN - 1959-1967	I	0.43771**	0.57275**	0.02098*	-0.03321**			0.9996
		(0.11097)	(0.13857)	(0.01028)	(0.01014)			
	II	0.38403**	0.65356**	0.02851**	-0.04461**	-0.00065	0.00145	1.000
		(0.11607)	(0.14838)	(0.00358)	(0.00486)	(0.00076)	(0.00146)	
	III (a)	0.48655*	0.53227*	0.03065*	-0.04948*	-0.00008	0.00056	
		(0.17240)	(0.21772)	(0.01434)	(0.01950)	(0.00022)	(0.00039)	
	III (b)	0.49092**	0.52718*	0.03080*	-0.04977*	-0.00004	0.00043	
		(0.16568)	(0.20914)	(0.01390)	(0.01889)	(0.00019)	(0.00034)	
	IV	SAME AS H-H ESTIMATES						
NETHERLANDS - 1950-1966	I	0.12387	0.94398**	0.03741*	-0.05032**			0.9985
		(0.15209)	(0.18126)	(0.01887)	(0.01544)			
	II	0.08496	0.98560**	0.03223*	-0.04272*	-0.00151	0.00280	0.9968
		(0.13605)	(0.16168)	(0.01191)	(0.01635)	(0.00086)	(0.00154)	
	III (a)	0.08581	(0.99077)**	0.03609**	-0.04842**	-0.00126	0.00238	
		(0.12361)	(0.14990)	(0.00581)	(0.00628)	(0.00071)	(0.00132)	
	III (b)	0.08582	0.99080**	0.03613**	-0.04848**	-0.00123	0.00233	
		(0.12306)	(0.14901)	(0.00633)	(0.00726)	(0.00659)	(0.00122)	
	IV	SAME AS H-H ESTIMATES						
PHILIPPINES - 1951-1968	I	0.82886**	0.14066	-0.04411*	0.04988*			0.9984
		(0.20807)	(0.23342)	(0.01762)	(0.02037)			
	II	0.88721**	0.08765	-0.04439**	0.04875**	-0.00010	0.00052	
		(0.15506)	(0.17508)	(0.01351)	(0.01574)	(0.00104)	(0.00149)	
	III (a)	0.89458**	0.07977	-0.04481**	0.04926**	-0.00019	0.00059	
		(0.14056)	(0.16421)	(0.01262)	(0.01473)	(0.00091)	(0.00128)	
	III (b)	0.89514**	0.07909	-0.04483*	0.04929**	-0.00019	0.00060	
		(0.14365)	(0.16265)	(0.01250)	(0.01459)	(0.00085)	(0.00120)	
	IV	0.90348**	0.06993	-0.04516**	0.04963**	-0.00029	0.00066	
		(0.13318)	(0.15112)	(0.01125)	(0.01316)	(0.00060)	(0.00085)	
UNITED KINGDOM - 1951-68	I	-0.22168	1.31470**	0.05505**	-0.06267**			0.9996
		(0.16816)	(0.18280)	(0.01457)	(0.01628)			
	II	-0.03906	1.10938**	0.02881*	-0.03467*	-0.00038	0.00051	0.9913
		(0.16983)	(0.18500)	(0.01461)	(0.01632)	(0.00028)	(0.00035)	
	III (a)	-0.24683	1.34021**	0.05273**	-0.05988**	-0.00047	0.00064	
		(0.19111)	(0.20813)	(0.01637)	(0.01828)	(0.00026)	(0.00033)	
	III (b)	-0.24804	1.34155**	0.05281**	-0.05995**	-0.00049	0.00066**	
		(0.19321)	(0.21045)	(0.01656)	(0.01850)	(0.00025)	(0.00031)	
	IV	-0.29629	1.39358**	0.05545**	-0.06276**	-0.00050	0.00067	
		(0.18281)	(0.19926)	(0.01500)	(0.01674)	(0.00032)	(0.00041)	

(4.7) $$w(t) = \epsilon_1(t)\, Y(t) + \epsilon_2(t)\, C(t-1)\ .$$

We shall analyze this consumption function for Canada, Finland, India, Japan, Netherlands, Philippines and the United Kingdom.[1] The parameters of the model have been estimated by the following methods:

I. Ordinary Least Squares (OLS)

II. Hildreth and Houck procedure

III. Modified Hildreth and Houck procedure:
 (a) Employing (3.20) to apply a two-step-Aitken Procedure to (3.15)
 (b) Employing (3.20) with off-diagonal elements replaced by zeros to apply a two-step Aitken Procedure to (3.15) and

IV. Maximum Likelihood Procedure-Employing, Modified Gauss-Newton Method discussed in the Appendix B.

Numerical estimates of the parameters of the model (4.5) obtained according to the above mentioned methods are displayed in Table 1. The significance of various parameter estimates at 5% and 1% level has been indicated by a single star and double stars, respectively. It turns out that in majority of the cases there is broad agreement between all the estimation procedures, so far as the significance of the parameter estimates is concerned. We may analyze these results in the light of the hypotheses implicit in (4.3) and (4.4) also discussed in subsection B of Section 2.

Statistically insignificant estimates of σ_{11} and σ_{22} for Japan, Netherlands, Philippines and U.K. ($\hat{\sigma}_{22}$ according to Method III(b) is although significant) indicate the absence of random character of the response coefficients in the model (4.2). However, in these cases, the statistical significance

---

[1]The data on both consumption and income are in current prices. The data for Netherlands have been discussed in Singh, Drost and Kumar (1971) and for the remaining countries in Singh and Drost (1971), Appendix, p. 333.

of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ indicates a definite trend (parameter shift) in the response coefficients of both the income and the lagged consumption. On the contrary, statistically significant $\hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$ in Canada supports the hypothesis about the randomness of the response coefficients. In this case, the estimates of $\bar{\beta}_1$ and $\bar{\beta}_2$ are statistically significant according to all the methods of estimation, although the significance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ is not entirely conclusive because of lack of unanimity among different methods. Keeping in view the conflicting evidence about α's, one might be tempted to conclude that the response coefficients $\beta_1(t)$ and $\beta_2(t)$ for Canada fluctuate around a constant rather than a trend.

Next, in Canada and Japan, where both $\hat{\bar{\beta}}_1$ and $\hat{\bar{\beta}}_2$ are statistically significant, relatively high magnitude of the latter may indicate strong stickiness in consumer habits. This can be explained in terms of well developed consumer credit facilities in both the countries coupled with good social security system in Canada and traditional habit of thriftiness of Japanese.[1] Further, statistically significant estimate of $\bar{\beta}_2$ and not that of $\bar{\beta}_1$ for Netherlands and U.K. seems quite expected in view of generous social security system and typical Europe-mindedness (where traditions are very important) in both countries.[2] In addition, monetary equilibrium and healthy expansion in the Netherlands economy and high proportion of old age people in the British population[3] may also contribute towards the dependence of current consumption on past year's consumption in these countries. As a contrast, significant estimate of $\bar{\beta}_1$ and not that of $\bar{\beta}_2$ in case of the Philippines can be explained in terms of the absence of above factors, heavy dependence on agriculture and structural

---

[1]Broadridge (1966) pp. 20-21.

[2]Pen (1964) p. 179.

[3]Klein (1958) pp. 60-69.

disorders that existed in the country around 1950.[1]

As noted above, the response coefficients of both income and lagged consumption exhibit shift with time for Japan, Netherlands, Philippines and the U.K. Rising tendency in $\beta_1(t)$ and falling in $\beta_2(t)$ as for Japan, Netherlands and U.K. implies a change in consumer habits in the sense that the consumers tend to adapt to the changes in income more and more quickly. This indicates quite a significant structural change inasmuch as the countries known for maintaining traditional habits, exhibit a tendency of giving up their rigidities. Such a phenomenon can in part be due to (a) increasing industrialization and thus rising proportion of wage earners in the population in all the countries,[2] (b) the peculiar wage system in Japan under which the worker enjoys a bonus scheme of transient nature and is paid a very low wage,[3] (c) the experience of recession during 1952-58 and the unstable balance of payment situation during the fifties in the heavily export-dependent Dutch economy,[4] and (d) stagnation in the British economy resulting in the devaluation of the Pound first in 1949 and then in 1967, heavy unemployment, high prices and cut in the growth of real consumption.[5] Clearly, all these factors contribute towards weakening of consumer's confidence in the economy and thus tend to increase "urgency to consume" in the consumer which exhibit through shortening lags and increasing dependence on current income. Conversely, in the case of the Philippines, we find that the coefficient of income falls whereas that of lagged consumption rises.

---

[1]Roxas (1969). Also see Power (1969) and Power (1970).

[2]Singh and Drost (1971), pp. 331-332.

[3]Broadridge (1966).

[4]Pen (1964), pp. 179-180.

[5]Cripps and Reddaway (1971).

This indeed shows growing confidence among the consumers about their economic conditions which may strengthen stickiness in consumer habits. Such a phenomenon in a predominantly agricultural country like the Philippines where agriculture still depends on the vagaries of nature may arise due to the changing structure of the Philippine economy during the sample period. Vigorous industrial growth because of import and foreign exchange control, tax exemption, liberal credit, tariff protection during 1950-61[1] and substantial improvement in agricultural sector following devaluation and decontrol in 1962 notwithstanding its repressive effects on the growth of manufacturing seems to have increased confidence among the Filipinos. Moreover, the minimum wage law introduced in 1950 and in 1965, land reform code of 1964 and advanced welfare legislation may also reduce uncertainty and thus reinforce stickiness in consumer habits.[2]

Finally, in the case of Finland and India the estimates of $\overline{\beta}_1$ are significant by all methods, however, there is no consensus among the methods of estimation regarding other parameters.

---

[1]Bantegui, (1969), pp. 14-16.

[2]It may be for these reasons besides the effects of devaluation and decontrol on income distribution that Williamson (1967) finds an upward shift in the saving function after 1960, "even after controlling for the traditional measured disposable and personal income, permanent and transitory income, real interest rate and price changes."

Computational Note on the Maximum Likelihood Procedure

It may be noted that the determination of v' plays crucial role in the modified Gauss-Newton procedure described in the Appendix B. The table below indicates the domain of v and the number of iterations for different countries.

TABLE 2: The Number of Iterations and the domain of v for all the Countries

Countries	No. of Iterations	The Domain of v
Canada	26	$0 \leq v \leq 1$
Finland	22	$0 \leq v \leq 1$
India	-	$0 \leq v \leq 10^{-6}$
Japan	-	$0 \leq v \leq 10^{-6}$
Netherlands	-	$0 \leq v \leq 10^{-6}$
Philippines	36	$0 \leq v \leq 0.2$
United Kingdom	5	$0 \leq v \leq 1$

It may be noted that the original domain of v, (0,1), as indicated in the Appendix B, works satisfactorily only for Canada, Finland and the United Kingdom. For the Philippines we found the new domain for v to be as(0, 0.2). However, the value of v' computed according to (B.20) was found to be outside this domain. Consequently, we used 0.2 for v' in (B.13) and continued the iterative scheme. Small value of v' explains the reasons of slow convergence of the procedure in this case. In case of India, Japan, and Netherlands we did not find any value of v within the domain $(0, 10^{-6})$ for which the value of L would be larger than L(0). Since, for computational reasons, we could not narrow down the domain any further, we concluded that the Hildreth and Houck and the Maximum Likelihood estimates are identical within reasonable limits.

REFERENCES

Adenkule, J. O. (1968), "Rates of Inflation in Industrial, Other Developed and Less Developed Countries, 1949-65," International Monetary Fund Staff Papers, XV, pp. 531-559.

Beltrami, Edward J. (1970), An Algorithmic Approach to Non-Linear Analysis and Optimization, Academic Press, New York.

Bentegui, B. G. (1969), "National Income Accounting," The Philippine Economy Bulletin, Vol. VII, No. 6, pp. 1-18.

Broadridge, S. (1966), Industrial Dualism in Japan, Aldine Publishing Company: Chicago.

Brown, T. M. (1953), "The Habit Persistence and Lags in Consumer Behaviour," Econometrica, 20, pp. 355-371.

Cagan, P. (1956), "The Monetary Dynamics of Hyper Inflation" in M. Friedman (ed.) Studies in the Quantity Theory of Money, Chicago: University of Chicago Press.

Chew, Victor (1970), "Covariance Matrix Estimation in Linear Models," Journal of the American Statistical Association, Vol. 65, No. 329, pp. 173-181.

Cripps, T. F. and W. B. Reddaway (1971), "Economic Outlook for 1972--Great Problems, Some Achievements," London and Cambridge Bulletin, reprinted from The Times 30th and 31st December.

Evans, M. (1969), Macroeconomic Activity: Theory Forecasting and Control, An Econometric Approach, Harper and Row: New York.

Friedman, Milton (1957), A Theory of Consumption Function, Princeton University Press for NBER, Princeton, N.J.

Hartley, H. O. (1961), "The Modified Gauss-Newton Method for Fitting of Non-Linear Regression Functions by Least Squares," Technometrics, Vol. 3, No. 2, pp. 269-280.

Hildreth, C. and James P. Houck (1968), "Some Estimators for Linear Model with Random Coefficients," Journal of the American Statistical Association, Vol. 63, No. 321, pp. 584-595.

Huzurbazar, V. S. (1948), "The Likelihood Equations, Consistency and the Maxima of the Likelihood Functions," Annals of Eugenics, Vol. 14, pp. 185-200.

Klein, L. R. (1953), A Text Book of Econometrics, Row, Peterson and Co.; Evanston.

Klein, L. R. (1958), "The British Propensity to Save," Journal of the Royal Statistical Society, Vol. 121, Series A, pp. 60-96.

Koyck, L. M. (1954), Distributed Lags and Investment Analysis, North Holland Publishing Company.

Nerlove, Marc (1965), Estimation and Identification of Cobb-Douglas Production Functions, Rand McNally and Co.: Chicago.

Pen, J. (1964), "Economic Policy in Netherlands 1959-63," Scottish Journal of Political Economy, Vol. 11, pp. 179-180.

Power, John H. (1969), "Industrialization in the Philippines: The Need for New Directions," Discussion Paper No. 69-14, Institute of Economic Development and Research, School of Economics, University of Philippines.

Power, John H. and Gerardo P. Sicat (1970), "Industrialization in the Philippines," IEDR, Discussion Paper No. 70-11, School of Economics, University of Philippines.

Rao, C. R. (1965), "The Theory of Least Squares When the Parameters are Stochastic and Its Application to the Analysis of Growth Curves," Biometrica, Vol. 52, pp. 447-458.

Rao, C. R. (1970), "Estimation of Heteroscadastic Variances in Linear Models," Journal of the American Statistical Association, Vol. 65, No. 329, pp. 161-172.

Roxas, Sixto K. (1969), "Policies for the Private Sector," The Philippine Economic Journal, First Semester.

Rubin, H. (1950), "Note on Random Coefficients" in Statistical Inference in Dynamic Economic Models, edited by T. C. Koopmans, Cowles Commission Monograph 10 pp.

Singh, Balvir and Helmar Drost (1971), "An Alternative Econometric Approach to the Permanent Income Hypothes s: An International Comparison," Review of Economics and Statistics, Vol. LIII, No. 4, pp. 326-334.

Singh, Balvir, Helmar Drost and Ramesh C. Kumar (1971), "The Post-war Theories of Consumption Function - An Empirical Evaluation," presented at European Econometric Society Meeting at Barcelona (Spain), September.

Stone, Richard (1965), "Model for Demand Projections," University of Cambridge Reprint Series No. 236, Department of Applied Economics, University of Cambridge.

Stone, Richard (1966), "The Changing Pattern of Consumption," University of Cambridge Reprint Series No. 245, Department of Applied Economics, University of Cambridge.

Theil, H. and L. B. M. Mennes (1959), "Conception Stochastique de Coefficients Multiplicateurs dans L'Adjustment Linearire des Series Temporilles," Publications de l'Institute de Statistique de l'Universite de Paris, Vol. 8, pp. 1-16.

Theil, H. (1971), Principles of Econometrics, John Wiley and Sons, New York.

Williamson, J. G. (1967), "Determinants of Personal Savings in Asia: Long and Short Run Effects," Discussion Paper No. 67-11, Department of Economics, University of Philippines, Appendix B.

Williamson, J. G. (1968), "Personal Saving in Developing Nations: An Intertemporal Cross-Section from Asia," Vol. 44, pp. 194-210.

B. NUMERICAL SOLUTION OF LIKELIHOOD EQUATIONS

B.1. GAUSS-NEWTON METHOD

The log likelihood function and the likelihood equations defined in (3.22) - (3.25) are functions of the parameters $\bar{\beta}_1,\ldots,\bar{\beta}_\Lambda, \bar{\alpha}_1,\ldots,\bar{\alpha}_\Lambda, \sigma_{11},\ldots,\sigma_{\Lambda\Lambda}$ of the model (3.1). Suppose we represent these parementers by a colum vector γ such that

(B.1) $$\gamma' = (\gamma_1,\ldots,\gamma_K) \equiv (\bar{\beta}_1,\ldots,\bar{\beta}_\Lambda, \bar{\alpha}_1,\ldots,\bar{\alpha}_\Lambda, \sigma_{11},\ldots,\sigma_{\Lambda\Lambda}),\ K = 3\Lambda,$$

where γ' indicates the transpose of γ. Now according to the method of maximum likelihood we have to determine γ such that the likelihood function (3.22) is maximum, that is

(B.2) $$\nabla g = \frac{\partial L}{\partial \gamma} \equiv \begin{bmatrix} \frac{\partial L}{\partial \gamma_1} \\ \vdots \\ \frac{\partial L}{\partial \gamma_k} \end{bmatrix} = 0$$

and

(B.3) $$H = \frac{\partial^2 L}{\partial \gamma\, \partial \gamma'} \equiv \left(\!\!\left(\frac{\partial^2 L}{\partial \gamma_k\, \partial \gamma_{k'}} \right)\!\!\right),\quad k,k' = 1,\ldots,K,$$

is a $K \times K$ negative definite matrix.

As shown in (3.24) - (3.25), the equations in (B.2) are highly non-linear; however, we may solve them by numerical methods. Therefore, linearizing (B.2), we write

(B.4) $$\nabla g = (\nabla g)_{\gamma^o} + H_{\gamma^o} (\gamma - \gamma^o)$$

according to the Taylor series expansion. In the present case, the initial

A P P E N D I X

A. COVARIANCE MATRIX OF THE RESIDUALS IN THE REGRESSION (3.15)

Let us write

$$(A.1) \qquad E\,\eta\,\eta' = E\,(\hat{\dot{w}} - E\,\hat{\dot{w}})(\hat{\dot{w}} - E\,\hat{\dot{w}})'$$

$$= E\,\hat{\dot{w}}\,\hat{\dot{w}}' - (E\,\hat{\dot{w}})(E\,\hat{\dot{w}})'$$

$$= E\,\hat{\dot{w}}\,\hat{\dot{w}}' - \dot{M}\,\dot{X}\,\sigma\,\sigma'\,\dot{X}'\,\dot{M}' ,$$

and thus to evaluate the required covariance matrix we must evaluate $E\,\hat{\dot{w}}\,\hat{\dot{w}}'$.

Since

$$(A.2) \qquad \hat{w} = M\,w = \begin{bmatrix} \sum_{t=1}^{T} m_{1t}\, w(t) \\ \vdots \\ \sum_{t=1}^{T} m_{Tt}\, w(t) \end{bmatrix}$$

we get

$$(A.3) \qquad \hat{\dot{w}} = \begin{bmatrix} \sum_{t,t'=1}^{T} m_{1t}\, m_{1t'}\, w(t)\, w(t') \\ \vdots \\ \sum_{t,t'=1}^{T} m_{Tt}\, m_{Tt'}\, w(t)\, w(t') \end{bmatrix}$$

where $m_{tt'}$ is the element in the t-th row and t'-th column of the T x T matrix M defined in (3.13) and w(t) in the t-th element of w in (3.5).

The element in the τ-th row and τ'-th $(\tau, \tau' = 1, \ldots, T)$ column of $E\,\hat{w}\,\hat{w}'$ may then be written as

(A.4) $$E\left[\left\{\sum_{t,t'=1}^{T} m_{\tau t}\, m_{\tau t'}\, w(t)\, w(t')\right\} \times \left\{\sum_{t'',t'''=1}^{T} m_{\tau' t''}\, m_{\tau' t'''}\, w(t'')\, w(t''')\right\}\right]$$

$$= \sum_{t,t',t'',t'''=1}^{T} m_{\tau t}\, m_{\tau t'}\, m_{\tau t''}\, m_{\tau t'''}\, E\{ w(t)\, w(t')\, w(t'')\, w(t''') \}$$

$$= \left(\sum_{t=1}^{T} m_{\tau t}^2\, \varphi_t^2\right)\left(\sum_{t=1}^{T} m_{\tau' t}^2\, \varphi_t^2\right) + 2\left(\sum_{t=1}^{T} m_{\tau t}\, m_{\tau' t}\, \varphi_t^2\right)^2 .$$

Hence

(A.5) $$E\, \hat{\dot{w}}\, \hat{\dot{w}}' = \dot{M}\, \dot{X}\, \sigma\, \sigma'\, \dot{X}'\, \dot{M}' + 2\dot{\Psi}$$

or

(A.6) $$E\, \eta\, \eta' = 2\dot{\Psi}$$

where Ψ is a $T \times T$ matrix, the element in the τ-th row and τ'-th column of which is

(A.7) $$\sum_{t=1}^{T} m_{\tau t}\, m_{\tau' t}\, \varphi_t^2 ,$$

and $\dot{\Psi}$ is the matrix of squared elements of Ψ. We may also express

(A.8) $$\Psi = M\, \Phi\, M$$

where Φ has been defined in (3.6).

iterate vector γ^o has been obtained by the Hildreth and Houck procedure described above. Then (B.2) implies

$$(\nabla g)_{\gamma^o} + H_{\gamma^o} (\gamma - \gamma^o) = 0 , \tag{B.5}$$

and if H is nonsingular at $\gamma = \gamma^o$, we obtain

$$\gamma = \gamma^o - H^{-1}_{\gamma^o} (\nabla g)_{\gamma^o} \equiv \gamma^1 \text{ (say).} \tag{B.6}$$

Suppose we write

$$D = - H^{-1} \nabla g , \tag{B.7}$$

then

$$\gamma^1 = \gamma^o + D_{\gamma^o} \tag{B.8}$$

where D_{γ^o} is the value of D at $\gamma = \gamma^o$.

Then the Gauss-Newton iterative scheme can be outlined as

$$\gamma^{\ell} = \gamma^{\ell-1} + D_{\gamma^{\ell-1}} , \tag{B.9}$$

$D_{\gamma^{\ell-1}}$ being the value of D at $\gamma = \gamma^{\ell-1}$ and the superscript ℓ stands for the ℓ-th iteration. Further, we may interpret $D_{\gamma^{\ell-1}}$ as the correction to $\gamma^{\ell-1}$ to obtain γ^{ℓ} .

It should be noted that if the correction $D_{\gamma^{\ell-1}}$ is too large at any stage of iteration then there is a possibility of overshooting the maximum and the procedure does not permit retracing the step. In fact such a situation will arise if H becomes close to singularily.

B.2. MODIFIED GAUSS-NEWTON METHOD

In this part of the appendix we propose a modification of the Gauss-Newton method to overcome the difficulty mentioned above. The procedure is similar to the one proposed by Hartley (1961), yet it differs from him in as much as we, unlike him, require the second order derivatives of L.

We define the likelihood function at

(B.10) $\gamma^o + v \, D_{\gamma^o}$ $\qquad 0 \le v \le 1$,

instead of at

(B.11) $\gamma^o + D_{\gamma^o}$

as in the case of Gauss-Newton procedure. Then determine $v = v'$ such that

(B.12) $L \, (\gamma^o + v' \, D_{\gamma^o})$

is maximum in the interval (0,1) .

Next,

(B.13) $\gamma^1 = \gamma^o + v' \, D_{\gamma^o}$

and we may repeat the procedure to determine the value of v (say v'') which maximizes L in the subsequent iteration. Thus we have

(B.14) $\gamma^2 = \gamma^1 + v'' \, D_{\gamma^1}$

It follows that

(B.15) $L \, (\gamma^2) \ge L \, (\gamma^1) \ge L \, (\gamma^o)$

and that γ^2 and γ^1 lie in the interior of the closed convex set.

If we repeat the procedure, we shall finally have

(B.16) $\lim_{\ell \to \infty} L \, (\gamma^{\ell}) = L^*$

the maximum of L .

The estimate of asymptotic covariance matrix of the maximum likelihood estimate of γ is given by,

$$\text{(B.17)} \qquad - H^{-1}_{\gamma^{\ell}} .$$

B.3 DETERMINATION OF V WHICH MAXIMIZES L (V)

The function

$$\text{(B.18)} \qquad L(v) = L(\gamma^{\ell-1} + v D_{\gamma^{\ell-1}})$$

to be maximized may be specified to be parabolic as in Hartley (1961),

$$\text{(B.19)} \qquad L(v) = a + b v + c v^2$$

Theoretically we can determine the parameters a, b and c if we are given three points on the parabola. In the present analysis, however, we assume that the parabola passes through, $(0, L(0))$, $(\frac{1}{2}, L(\frac{1}{2}))$ and $(1, L(1))$. Then the value of v, say v', which maximizes L(v) can be given by

$$\text{(B.20)} \qquad v' = \frac{1}{4} + \frac{1}{2} \left\{ \frac{L(\frac{1}{2}) - L(0)}{2 L(\frac{1}{2}) - L(1) - L(0)} \right\}$$

Clearly v' expedites the convergence in that it represents the largest possible step in the required direction. Numerical values of v' indicate the nearness of the two successive iterates. Smaller is v' closer are the two successive iterates.

In special cases, however, where the function is not well-behaved, (B.20) may not work that satisfactorily and a number of problems may arise. For example, L(v) may have a maximum for v outside the interval (0,1). Nevertheless, in such a situation one may evaluate L(v) at $v = 0.0$, $\frac{1}{2}$ and 1.0 and choose that v for which the value of the function L is relatively large. This procedure would retard the convergence no doubt but it will still ensure movement in the right direction. In case L(0) is greater than $L(\frac{1}{2})$ and L(1.0), we get v' = 0 implying $\gamma^1 = \gamma^o$. This suggests that the maximum of L(v) is in a close neighbourhood of v = 0. Noting that v can not be negative, one may narrow down the range of v to, say, $0 - \frac{1}{2}$ and proceed as before. That is choose ξ_1 and ξ_2 in the range $(0,\frac{1}{2})$ and if

(B.21) $$L(0) < L(\xi_1) \text{ and/or } L(\xi_2) , \qquad \xi_1 < \xi_2$$

then compute v' according to (B.20) to lie between 0 and ξ_2. Should v' thus computed lie outside $(0,\xi_2)$, one may use

(B.22) $$\xi_1, \text{ if } L(0) < L(\xi_1) > L(\xi_2)$$

or

(B.23) $$\xi_2, \text{ if } L(0) < L(\xi_1) < L(\xi_2)$$

If, however, one finds

(B.24) $$L(0) > L(\xi_1) > L(\xi_2)$$

for ξ_1 and ξ_2 being sufficiently close to zero, then iteration would stop because the maximum of L is reached.