# A GENERAL PROCEDURE FOR OBTAINING MAXIMUM LIKELIHOOD ESTIMATES IN GENERALIZED REGRESSION MODELS

By W. Oberhofer and J. Kmenta[1]

This paper describes an iterative procedure for obtaining maximum likelihood estimates of the parameters of a generalized regression model when direct maximization with respect to all parameters is difficult. A proof of convergence and some interesting applications are provided.

## 1. INTRODUCTION

CONSIDER THE PROBLEM of maximizing a function $f$ with respect to two variables (or two sets of variables) $a_1$ and $a_2$ within some space $S$. Suppose it is difficult to maximize $f$ as a function of $a_1$ and $a_2$, but relatively easy to maximize $f$ as a function of $a_1$ given $a_2$ and as a function of $a_2$ given $a_1$. Such a case is frequently encountered in connection with maximum likelihood or quasi-maximum likelihood estimation of certain regression models.[2] In this case it is advantageous to adopt a zig-zag iterative procedure which is described below. This procedure was used by Sargan [3] for the purpose of estimating regressions with autoregressive disturbances, but it is amenable to a much more general class of estimation problems. In particular, it can be advantageously used in all cases involving the application of iterative Aitken methods.

The plan of the paper is as follows. In Section 2, we present the fundamental lemma and give a proof of convergence in the general case. In Section 3, we demonstrate the applicability of the lemma to the generalized regression model. The last section contains some specific applications that are of interest to econometricians.

## 2. THE FUNDAMENTAL LEMMA[3]

Let $f(a)$ be a function which is to be maximized with respect to $a$, and $a \in U$. Further, let $a$ be partitioned as $a = (a_1, a_2)$ with $a_1 \in U_1$ and $a_2 \in U_2$. The number of components in $a_1$ and $a_2$ is taken to be $n$ and $m$ respectively, i.e., $U_1 \subseteq R^n$ and $U_2 \subseteq R^m$. It is assumed that $f(a)$ has the following properties:
(i) There exists an $s$ such that the set

$$S = \{a | a \in U = U_1 \times U_2, f(a) \geqslant s\}$$

---

[2] An investigation of a similar problem in the context of general optimization procedures has been carried out by Warga [4, 5].

[3] This lemma is also given in Sargan [3].

is nonempty and bounded;

(ii) $f(a)$ is continuous in $S$; and

(iii) the parameter space $U$ is closed, or $U_2$ is closed and $U_1 = R^n$.

From these assumptions it follows that $S$ is compact. Now we define the following iteration:

(i) Let $a_1^{(0)}$ be a vector of initial values of $a_1$ such that $a_1^{(0)} \in U_1$ and such that there exists an $a_2 \in U_2$ for which $f(a_1^{(0)}, a_2) \geqslant s$.

(ii) We maximize $f(a_1^{(0)}, a_2)$ in $U_2$. Because of the compactness property, the maximum will be reached at $a_2 = a_2^{(0)} \in U_2$.

(iii) We suppose inductively that we have obtained $(a_1^{(j)}, a_2^{(j)})$ for all $0 \leqslant j \leqslant k$. Now we maximize $f(a_1^{(k)}, a_2)$ in $U_2$. The maximum will be reached at $a_2 = a_2^{(k+1)}$. Then we maximize $f(a_1, a_2^{(k+1)})$ in $U_1$. The maximum will be attained at $a_1 = a_1^{(k+1)}$.

In this way the iterative procedure is unequivocally defined.

LEMMA 1: (i) *The sequence* $\{a^{(k)}\}$ *has at least one accumulation point* $a^*$ *in* $S$. (ii) *If* $a^*$ *and* $a^+$ *are two accumulation points of the sequence, then* $f(a^*) = f(a^+)$. (iii) *For every accumulation point* $a^* = (a_1^*, a_2^*)$

$$\max_{a_1 \in U_1} f(a_1, a_2^*) = \max_{a_2 \in U_2} f(a_1^*, a_2) = f(a_1^*, a_2^*) = f(a^*).$$

PROOF: Since $S$ is compact, the sequence $\{a^{(k)}\}$ has at least one accumulation point. This proves proposition (i) of the lemma. Let us now assume that there are two accumulation points $a^*$ and $a^+$ such that

(1)        $f(a^*) > f(a^+)$.

There must then be two sequences $\{a^{(k_v)}\}$ and $\{a^{(\ell_v)}\}$ such that

(2)        $\lim_{v \to \infty} a^{(k_v)} = a^*$     and     $\lim_{v \to \infty} a^{(\ell_v)} = a^+$.

But $f(a^{(k)})$ is monotonically growing with $k$ by construction. Because of (1) and (2) there is, therefore, a $v_0$ such that, for $k_{v_0} = \bar{k}$,

(3)        $f(a^{(\bar{k})}) \geqslant f(a^+)$.

Then there is a $v_1$ such that $\ell_{v_1} > k_{v_0}$. Therefore we can assert that, for $\ell_{v_1} = \bar{\ell}$,

(4)        $f(a^{(\bar{\ell})}) \geqslant f(a^{(\bar{k})})$.

On the other hand, because of (3) we have

(5)        $f(a^+) > f(a^{(\bar{\ell})})$.

From (3) and (5) it follows that

(6)         $f(a^{(\bar{k})}) > f(a^{(\bar{\ell})})$.

But (6) represents a contradiction of (4). Therefore (1) is not possible. Since $a^*$ and $a^+$ are interchangeable, we must have

(7)         $f(a^*) = f(a^+)$

for any two accumulation points $a^*$ and $a^+$. Therefore $f$ takes on the same value for all accumulation points. This proves proposition (ii) of the lemma. Finally, by construction $a_1^{(k_v)}$ has been determined so that

(8)         $f(a_1^{(k_v)}, a_2^{(k_v)}) \geqslant f(a_1, a_2^{(k_v)})$   for all $a_1 \in U_1$.

If we let $v \to \infty$, we obtain

(9)         $f(a_1^*, a_2^*) \geqslant f(a_1, a_2^*)$   for all $a_1 \in U_1$.

Correspondingly to (8) we have

(10)         $f(a_1^{(k_v)}, a_2^{(k_v + 1)}) \geqslant f(a_1^{(k_v)}, a_2)$.

If we choose an appropriate sequence $\{\bar{k}_v\}$ of $\{k_v + 1\}$, the compactness property leads to the convergence $a_1^{(\bar{k}_v)} \to a_1^+$ and $a_2^{(\bar{k}_v)} \to a_2^+$. In this way we obtain from (10)

(11)         $f(a_1^{(\bar{k}_v - 1)}, a_2^{(\bar{k}_v)}) \geqslant f(a_1^{(\bar{k}_v)}, a_2)$

for all $a_2 \in U_2$. For $v \to \infty$ this gives

(12)         $f(a_1^*, a_2^+) \geqslant f(a_1^*, a_2)$   for all $a_2 \in U_2$.

If in (8) we take the sequence $\{\bar{k}_v\}$ instead of the sequence $\{k_v\}$, then instead of (9) we obtain

(13)         $f(a_1^+, a_2^+) \geqslant f(a_1, a_2^+)$.

From (12) and (13) it follows that for $a_1 = a_1^*$

(14)         $f(a_1^+, a_2^+) \geqslant f(a_1^*, a_2)$.

We note that, by (ii) of Lemma 1, $f(a_1^+, a_2^+) = f(a_1^*, a_2^*)$ and, therefore, (14) can be written as

(15)         $f(a_1^*, a_2^*) \geqslant f(a_1^*, a_2)$   for all $a_2 \in U_2$.

The inequalities (9) and (15) give the proof of proposition (iii) of the lemma.

### 3. MAXIMUM LIKELIHOOD ESTIMATION OF THE GENERALIZED REGRESSION MODEL

Consider a linear regression model

(16)         $y = X\beta + u$

where $y$ is a $T \times 1$ vector of observations on the dependent variable, $X$ is a $T \times K$ matrix of the values of the regressors, $\beta$ is a $K \times 1$ vector of the regression

coefficients, and $u$ is a $T \times 1$ disturbance vector. We make the following assumptions:

ASSUMPTION 1: $u$ *is normally distributed.*[4]

ASSUMPTION 2: $E(u) = 0$.

ASSUMPTION 3: $E(uu') = \Omega, det \, \Omega \neq 0$.

ASSUMPTION 4: $X$ *is a fixed matrix of full rank and* $T > K$.

ASSUMPTION 5: *The parameters in* $\beta$ *are independent from those in* $\Omega$.

Except for the irrelevant constant, the logarithmic likelihood function then is:

$$(17) \qquad L(\beta, \omega) = \tfrac{1}{2} \log \det \Omega^{-1} - \tfrac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta)$$

where $\omega$ represents a vector of all parameters in $\Omega$. The elements of $\beta$ and $\omega$ in general are restricted by a priori restrictions: $\beta \in V_1$ and $\omega \in V_2$. We assume that $V_1$ and $V_2$ are closed. In most cases we have $\beta$ and $\omega$ as functions of free parameters: $\beta = \beta(\alpha)$ and $\omega = \omega(\gamma)$. Our problem is to determine those values of $\beta \in V_1$ and $\omega \in V_2$ which maximize $L(\beta, \omega)$. Now, when we maximize $L(\beta, \omega)$ with respect to $\omega$ while considering $\beta$ as given, we obtain $\hat{\omega}(\beta)$. We assume that this function exists. When, on the other hand, we maximize $L(\beta, \omega)$ with respect to $\beta$ while considering $\omega$ as given, we obtain $\hat{\beta}(\omega)$. This function is obtained by minimizing the quadratic form $(y - X\beta)'\Omega^{-1}(y - X\beta)$. When the parameters in $\beta$ are unrestricted, $\hat{\beta}(\omega)$ will be obtained as the solution of

$$(18) \qquad [y - X\hat{\beta}(\omega)]'\Omega^{-1}X = 0.$$

By replacing $\omega$ by $\hat{\omega}(\beta)$, the system in (18) can be solved for the maximizing values of $\beta$. This solution is, however, rather difficult to obtain, at least in general. Therefore we resort to an iterative procedure whose convergence can be demonstrated by reference to Lemma 1 of the preceding section. For this we need the following definitions:

$$(19) \qquad U_2 = \{\omega | \det \Omega \geqslant \lambda > 0, \omega'\omega \leqslant M, \Omega \text{ nonnegative definite}\} \cap V_2$$

where $\lambda$ and $M$ are some arbitrarily chosen positive numbers. Further,

$$(20) \qquad U_1 = R^K \cap V_1$$

and, therefore, $U = U_1 \times U_2$. Finally,

$$(21) \qquad S = \left\{ \begin{pmatrix} \beta \\ \omega \end{pmatrix} \middle| \beta \in U_1, \omega \in U_2, L(\beta, \omega) > s \right\}$$

where $s$ is arbitrary.

---

[4] Without the assumptions of normality the procedure described below leads to quasi-maximum likelihood estimates.

For the zig-zag iterative procedure we need to maximize $L(\beta, \omega)$ with respect to $\beta \in U_1$ for given $\omega$, and with respect to $\omega \in U_2$ for given $\beta$. Our aim, however, is to determine $\hat{\omega}(\beta)$ by maximization of $L(\beta, \omega)$ with respect to $\omega$ in $V_2$ instead of maximization in $U_2$. If we denote the latter result by $\tilde{\omega}(\beta)$, we have to provide a guarantee that $\tilde{\omega}(\beta) = \hat{\omega}(\beta)$. For this purpose we extend the five assumptions stated at the outset of this section by the following:

ASSUMPTION 6: *If for a given $\beta^* \in U_1$ and $\omega^* \in U_2$ we have $L(\beta^*, \omega^*) \geq s$, then $\hat{\omega}(\beta^*) \in U_2$.*

In this case it follows that

$$L(\beta^*, \tilde{\omega}(\beta^*)) = \max_{\omega \in U_2} L(\beta^*, \omega) = \max_{\omega \in V_2} L(\beta^*, \omega) = L(\beta^*, \hat{\omega}(\beta^*)).$$

That means $\hat{\omega}(\beta^*) = \tilde{\omega}(\beta^*)$.

Since $S$ represents the space in which we search for the maximum of $L(\beta, \omega)$, we have to satisfy ourselves that at least one neighborhood of the true $\beta$ and $\omega$ lies in $S$. But this is guaranteed, because of the assumptions about $\Omega$, whenever we choose $\lambda$ and $s$ sufficiently small and $M$ sufficiently large. The matrices $\Omega$ have the additional property that there exist $\ell_1 > \ell_2 > 0$ such that all eigenvalues of every $\Omega$ with $\omega \in U_2$ lie in the interval $[\ell_2, \ell_1]$, that is,

$$(22) \qquad \ell_1 \geq \lambda(\Omega) \geq \ell_2 > 0,$$

where $\lambda(\Omega)$ stands for any eigenvalue of $\Omega$.[5] The inequality in (22) can be demonstrated as follows. The eigenvalues of $\Omega$ are always positive. The largest eigenvalue cannot be arbitrarily large because the trace of $\Omega$ (which is equal to the sum of all eigenvalues of $\Omega$) is restricted by the condition that $\omega'\omega \leq M$. The smallest eigenvalue also cannot be arbitrarily small because $\det \Omega \geq \lambda > 0$, and $\det \Omega$ is equal to the product of all eigenvalues of $\Omega$.

Now, the function to be maximized, $L(\beta, \omega)$, is continuous. Further, $U_2$ is closed. The only thing to be shown, then, is that $S$ is bounded. Let us assume the opposite, i.e., let there be a sequence

$$\begin{pmatrix} \beta^{(v)} \\ \omega^{(v)} \end{pmatrix}$$

in $S$ such that

$$(23) \qquad \lim_{v \to \infty} [(\beta^{(v)})'(\beta^{(v)}) + (\omega^{(v)})'(\omega^{(v)})] = \infty.$$

Since $\omega^{(v)} \in U_2$, it follows from (23) that

$$(24) \qquad \lim_{v \to \infty} (\beta^{(v)})'(\beta^{(v)}) = \infty.$$

Further, since $X'X$ is nonsingular by assumption, we also have

$$(25) \qquad \lim_{v \to \infty} (\beta^{(v)\prime}(X'X)\beta^{(v)}) = \infty$$

[5] Thus $U_2$ could also be defined as $U_2 = \{\omega | \ell_1 \geq \lambda(\Omega) \geq \ell_2\} \cap V_2$.

and, because of (22),

(26)        $\lim_{v \to \infty} (y - X\beta^{(v)})'\Omega^{-1}(y - X\beta^{(v)}) = \infty$.

Therefore, since det $\Omega \geqslant \lambda > 0$, (26) together with (17) leads to the conclusion that

(27)        $\lim_{v \to \infty} L(\beta^{(v)}, \omega^{(v)}) = -\infty$.

But equation (27) contradicts the fact that

$$\begin{pmatrix} \beta^{(v)} \\ \omega^{(v)} \end{pmatrix} \in S.$$

Therefore $S$ must be bounded. Thus all suppositions for the application of Lemma 1 are fulfilled.

The iterative method proceeds as follows. We begin with some starting vector

$$\begin{pmatrix} \hat{\beta}^{(0)} \\ \hat{\omega}^{(0)} \end{pmatrix}.$$

For instance, $\hat{\beta}^{(0)}$ can represent the ordinary least square estimate of $\beta$, and $\hat{\omega}^{(0)}$ is obtained from $\hat{\omega}(\hat{\beta}^{(0)})$. If we now choose $s = L(\hat{\beta}^{(0)}, \hat{\omega}^{(0)})$, then it is guaranteed that for $\hat{\omega}^{(0)} \in U_2$ we have

$$\begin{pmatrix} \hat{\beta}^{(0)} \\ \hat{\omega}^{(0)} \end{pmatrix} \in S.$$

Using $\hat{\omega}^{(0)}$ we construct $\hat{\beta}^{(1)} = \hat{\beta}(\hat{\omega}^{(0)})$, which leads to $\hat{\omega}^{(1)}$, and so on. Because of Assumption 6 and the fact that, in the process of iteration, $L(\beta, \omega)$ gets always bigger, we are always confined to space $S$.

Now we are in the position to formulate the following theorem.

THEOREM 1 : (i) *The sequence*

$$\begin{pmatrix} \beta^{(v)} \\ \omega^{(v)} \end{pmatrix}$$

*has at least one accumulation point.*

(ii) *If*

$$\begin{pmatrix} \beta^* \\ \omega^* \end{pmatrix}$$

*is an accumulation point, then $L(\beta, \omega^*)$ taken as a function of $\beta$ has its absolute maximum in $S$ at $\beta = \beta^*$. Correspondingly, $L(\beta^*, \omega)$ taken as a function of $\omega$ has its absolute maximum in $S$ at $\omega = \omega^*$. If $\beta \in V_1$ is given by a differentiable function $\beta(\alpha)$ and $\omega \in V_2$ by a differentiable function $\omega(\gamma)$, where $\alpha$ and $\gamma$ are free parameters, then*

*it follows that*

$$\left.\frac{\partial L}{\partial \alpha}\right|_{\substack{\beta=\beta^* \\ \omega=\omega^*}} = 0 \quad and \quad \left.\frac{\partial L}{\partial \gamma}\right|_{\substack{\beta=\beta^* \\ \omega=\omega^*}} = 0$$

*providing only that*

$$\begin{pmatrix} \beta^* \\ \omega^* \end{pmatrix}$$

*is not a corner solution.*

(iii) *In all accumulation points $L(\beta, \omega)$ takes on the same value.*

PROOF: The proof follows immediately from Lemma 1.

The implication of Theorem 1 is that the iterative procedure always converges to a solution of the first-order maximizing conditions (which may or may not correspond to the *absolute* maximum of the likelihood function).

## 4. APPLICATIONS

In each of the applications it is supposed that Assumptions 1 through 5 of Section 3 are satisfied.

### Application 1: *Autoregressive Disturbances* [6]

Suppose the disturbance follows an autoregressive scheme of the form

$$(28) \qquad u(t) = \rho u(t-1) + v(t), \qquad t \geqslant 1.$$

The following assumptions are made: [7] $u(0) = 0$, $E[v(t)] = 0$, $E[v(t)]^2 = \sigma^2$, and $E[v(t)v(s)] = 0$ for all $t \neq s$.

Equation (28) can be written in matrix notation as

$$(29) \qquad v = Au,$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{bmatrix}.$$

[6] See Sargan [3].

[7] If $\rho^2 < 1$, a corner solution is more likely to occur.

Let us define

$$
R = A'A = \begin{bmatrix}
1 + \rho^2 & -\rho & 0 & \ldots & 0 & 0 \\
-\rho & 1 + \rho^2 & -\rho & \ldots & 0 & 0 \\
0 & -\rho & 1 + \rho^2 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & & & \\
0 & 0 & 0 & \ldots & -\rho & 1
\end{bmatrix}
$$

so that $\Omega^{-1} = R/\sigma^2$. Then

$$
(30) \qquad L(\beta, \omega) = -\frac{T}{2} \log \sigma^2 - \tfrac{1}{2}\sigma^{-2} \sum_{t=1}^{T} [v(t)]^2
$$

$$
= -\frac{T}{2} \log \sigma^2 - \tfrac{1}{2}\sigma^{-2} \sum_{t=1}^{T} [u(t) - \rho u(t-1)]^2 .
$$

From (30) we obtain, at least in the absence of corner solutions,

$$
(31) \qquad \hat{\rho} = \sum_{t=1}^{T} u(t)u(t-1) \Big/ \sum_{t=1}^{T} [u(t-1)]^2 ,
$$

and

$$
(32) \qquad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} [u(t) - \hat{\rho}u(t-1)]^2
$$

for a given $\beta$ and $u = y - X\beta$.

To satisfy Assumption 6, we need the following condition:

$$
(33) \qquad \inf_{\beta} \hat{\sigma}^2 > 0.
$$

Since $\det A = 1$ and, consequently, $\det R = 1$, it follows that $\det \Omega = \sigma^{2T}$. Therefore, because of (33), it is guaranteed that $\det \hat{\Omega} \geq \lambda > 0$ for a suitable value of $\lambda$ and all $\beta$. By replacing $\rho$ and $\sigma$ in (30) by $\hat{\rho}$ and $\hat{\sigma}$ from (31) and (32), we obtain

$$
(34) \qquad L(\beta, \hat{\omega}) = -\frac{T}{2} \log \hat{\sigma}^2 - \frac{T}{2}.
$$

Because $L(\beta, \hat{\omega}) \geq s$, it follows from (34) that $\hat{\sigma}^2$ is bounded. Further, by reference to (31), we know that by Schwarz' inequality

$$
(35) \qquad \hat{\rho}^2 \leq \sum_{t=1}^{T} [u(t)]^2 \Big/ \sum_{t=1}^{T} [u(t-1)]^2 .
$$

Now, let $X^*$ be a matrix obtained from $X$ by deleting the last row (i.e., by deleting the $T$th observation). Then it can be shown that the right-hand side of the inequality in (35) is bounded as long as $X^*$ is of full rank. This is a sufficient condition for the boundedness of $\hat{\rho}$. Therefore Assumption 6 is satisfied.

The iterative method then proceeds as follows. We start with an initial vector $\hat{\beta}^{(0)}$ (represented, e.g., by an ordinary least squares estimate of $\beta$) and obtain $\hat{\rho}^{(0)}$ by (31). Using $\hat{\rho}^{(0)}$ we construct the matrix $\hat{R}^{(0)}$ and calculate the first-round values of $\beta$ by maximizing $L(\beta, \hat{\omega}^{(0)})$ in (30). The result is

$$(36) \qquad \hat{\beta}^{(1)} = (X'\hat{R}^{(0)}X)^{-1}(X'\hat{R}^{(0)}y).$$

Note that this is equivalent to the ordinary least squares estimate of $\beta$ of the transformed regression equation

$$(37) \qquad y(t) - \hat{\rho}^{(0)}y(t-1) = \beta_1[x_1(t) - \hat{\rho}^{(0)}x_1(t-1)] + \beta_2[x_2(t) - \hat{\rho}^{(0)}x_2(t-1)]$$
$$+ \ldots + \beta_K[x_K(t) - \hat{\rho}^{(0)}x_K(t-1)] + v^{(0)}(t),$$

where $v^{(0)}(t) = u(t) - \hat{\rho}^{(0)}u(t-1)$. From (36) we obtain a new set of residuals

$$(38) \qquad \hat{u}^{(1)} = y - X\hat{\beta}^{(1)}$$

and calculate a new value of $\rho$, $\hat{\rho}^{(1)}$, and so on. This procedure, long known to the profession as the Cochran-Orcutt iterative method, represents a convenient way of calculating the maximum likelihood estimates for the autoregressive model.

### Application 2: Heteroskedastic Disturbances

Let us write the regression equation as

$$(39) \qquad y_i(t) = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + u_i(t), \qquad i = 1, 2, \ldots, m;$$

$$t = 1, 2, \ldots, T_i; \qquad \sum_{i=1}^{m} T_i = T.$$

Suppose

$$E[u_i(t)u_j(s)] = \omega_i > 0 \quad \text{for } i = j \text{ and } t = s,$$
$$= 0 \qquad \text{for } i \neq j \text{ or } t \neq s.$$

This means that $\Omega$ is a diagonal matrix with the diagonal elements given by

$$\text{diag } \Omega = [\underbrace{\omega_1\omega_1 \ldots \omega_1}_{T_1} \quad \underbrace{\omega_2\omega_2 \ldots \omega_2}_{T_2} \ldots \underbrace{\omega_m\omega_m \ldots \omega_m}_{T_m}]'.$$

Then

$$(40) \qquad L(\beta, \omega) = -\frac{1}{2}\sum_{i=1}^{m} T_i \log \omega_i$$
$$-\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{T_i} [y_i(t) - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_K x_{iK}]^2/\omega_i$$

and

$$(41) \qquad \hat{\omega}_i(\beta) = \frac{1}{T_i}\sum_{t=1}^{T_i} [y_i(t) - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_K x_{iK}]^2.$$

Now we have to consider Assumption 6. To satisfy this assumption, we need

$$(42) \qquad \inf_{\beta} \hat{\omega}_i(\beta) > 0 \quad \text{for } i = 1, 2, \ldots, m.$$

This condition will guarantee that

$$\det \hat{\Omega}(\beta) = \prod_{i=1}^{m} [\omega_i(\beta)]^{T_i} \geqslant \lambda > 0$$

for a suitable value of $\lambda$ and for all $\beta$. By substituting $\hat{\omega}(\beta)$ for $\omega$ in (40) we get

$$(43) \qquad L[\beta, \hat{\omega}(\beta)] = -\frac{1}{2} \sum_{i=1}^{m} T_i \log \hat{\omega}_i(\beta) - \frac{T}{2}.$$

From (43) it follows that $[\hat{\omega}(\beta)]'[\hat{\omega}(\beta)]$ is bounded; otherwise, the condition that $L[\beta, \hat{\omega}(\beta)] \geqslant s$ would be violated. Therefore Assumption 6 is satisfied. The procedure then consists of first estimating $\beta$ by, say, the ordinary least squares method and then calculating $\hat{\omega}_i^{(0)}$ according to (41). In the next step we obtain $\hat{\beta}^{(1)}$ by minimizing

$$\sum_{i=1}^{m} \sum_{t=1}^{T_i} [y_i(t) - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_K x_{iK}]^2 / \hat{\omega}_i^{(0)}$$

with respect to $\beta$. Using $\hat{\beta}^{(1)}$ we calculate $\hat{\omega}_i^{(1)}$, and so on.

When $m = T$ (i.e., when we have only one observation on the dependent variable for each set of values of the regressors), maximum likelihood estimation under heteroskedasticity is possible only when $\Omega$ is subjected to further restrictions. For instance, it might be possible to assume that $\omega_i = f(x_{ik})\omega, 1 \leqslant k \leqslant K$, where $f$ is a suitably defined function. In this case we may again invoke Lemma 1 and apply the zig-zag iterative procedure for calculating the maximum likelihood estimates of the unknown parameters.

### Application 3: Seemingly Unrelated Regressions [8]

Consider a set of seemingly unrelated regressions

$$(44) \qquad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_M \end{bmatrix}$$

or

$$(45) \qquad y = X\beta + u$$

---

[8] For Monte Carlo results, see Kmenta and Gilbert [2]. Dhrymes [1] discusses the equivalence of the maximum likelihood and the iterative Aitken estimates but assumes convergence without proof.

where

$$\Omega = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1M}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & \cdots & \sigma_{2M}I_T \\ \vdots & & & \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \cdots & \sigma_{MM}I_T \end{bmatrix}$$
$$= \Sigma \otimes I_T,$$

where $\Sigma$ is nonsingular.

The maximum likelihood estimates are given as[9]

$$(46) \qquad \hat{\beta}(\omega) = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y)$$

and

$$(47) \qquad \hat{\Omega}(\beta) = \hat{\Sigma}(\beta) \otimes I_T,$$

where the typical element of $\hat{\Sigma}(\beta)$, say $\hat{\sigma}_{ij}(\beta)$, is

$$(48) \qquad \hat{\sigma}_{ij}(\beta) = \frac{1}{T}(y_i - X_i\beta_i)'(y_j - X_j\beta_j).$$

The concentrated log-likelihood function is

$$(49) \qquad L[\beta, \hat{\omega}(\beta)] = -\tfrac{1}{2}\log \det \hat{\Omega}(\beta) - \frac{T}{2}.$$

Since $L[\beta, \hat{\omega}(\beta)] \geqslant s$ for all $\beta$, it follows that $\det \hat{\Omega}(\beta) \leqslant C$, where $C$ is some arbitrarily chosen positive number. To satisfy Assumption 6 we need

$$(50) \qquad \inf_{\beta} \hat{\sigma}_{ii} \geqslant \lambda_0 > 0$$

and

$$(51) \qquad \det \hat{R} \geqslant \lambda_1 > 0 \quad \text{for all } \beta,$$

where $\hat{R}$ is a $M \times M$ matrix with a typical element $\hat{\rho}_{ij} = \hat{\sigma}_{ij}/\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}$, and $\lambda_0$ and $\lambda_1$ are some arbitrarily chosen positive numbers.

We have

$$(52) \qquad C \geqslant \det \hat{\Omega} = \left(\prod_{i=1}^{M} \hat{\sigma}_{ii}\right)^T (\det \hat{R})^T \geqslant \prod_{i=1}^{M} \hat{\sigma}_{ii}^T \lambda_1^T \geqslant \lambda_1^T \lambda_0^{MT}.$$

Now, from (52) it follows that for all $\beta$,

$$(53) \qquad C/\lambda_1^T \geqslant \prod_{i=1}^{M} \hat{\sigma}_{ii}^T$$

---

[9] See Kmenta and Gilbert [2].

and

$$(54) \qquad \det \hat{\Omega} \geqslant \lambda^T \lambda_0^{MT}.$$

We note that, by (53), all $\hat{\sigma}_{ii}$ and therefore all $\hat{\sigma}_{ij}$ are bounded. Therefore, taking into account (53) and (54), Assumption 6 is satisfied. The iterative procedure here is the same as that described in Kmenta and Gilbert [2].

*University of Regensburg*
*and*
*University of Michigan*

### REFERENCES

[1] DHRYMES, P. J.: "Equivalence of Iterative Aitken and Maximum Likelihood Estimators for a System of Regression Equations," *Australian Economic Papers*, 10 (1971), 20–24.
[2] KMENTA, J., AND R. F. GILBERT: "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions," *Journal of The American Statistical Association*, 63 (1968), 1180–1200.
[3] SARGAN, J. D.: "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P. E. Hart, G. Mills, and J. K. Whitaker, eds., *Econometric Analysis for National Economic Planning*. London: Butterworths, 1964.
[4] WARGA, J.: "A Convergent Procedure for Convex Programming," *SIAM Journal on Applied Mathematics*, 11 (1963), 579–587.
[5] ———: "Minimizing Certain Convex Functions," *SIAM Journal on Applied Mathematics*, 11 (1963), 588–593.