

THE ITERATED MINIMUM DISTANCE ESTIMATOR AND THE QUASI-MAXIMUM LIKELIHOOD ESTIMATOR

BY P. C. B. PHILLIPS¹

A multiple equation nonlinear regression model with serially independent disturbances is considered. The estimation of the parameters in this model by maximum likelihood and minimum distance methods is discussed and our main subject is the relationship between these procedures. We establish that if the number of observations in a sample is sufficiently large, the iterated minimum distance procedure converges almost surely and the limit of this sequence of iterations is the quasi-maximum likelihood estimator.

1. INTRODUCTION

IN THIS PAPER we consider the model

$$(1) \quad y_t = g_t(\alpha_0) + u_t \quad (t = 1, 2, \dots)$$

where y_t is an $m \times 1$ vector of observable random functions of discrete time (t) and g_t is a vector of known functions which depend on a $p \times 1$ vector of unknown parameters whose true value is denoted by α_0 . In general, g_t is a function of a number of exogenous variables as well as α_0 so that the function is time dependent. The last component of the model is the vector u_t of additive stochastic disturbances.

The non-linear regression model (1) has recently been the subject of discussion by various authors [2, 3, 4, 5, 6, and 7]. The purpose of these investigations has, in the main, been to suggest procedures for estimating the unknown parameters in (1) and to derive the asymptotic properties of these estimators under varying assumptions about the stochastic properties of u_t . The aim of the present paper is to consider, in particular, the iterated minimum distance estimator (MDE) proposed by Malinvaud [5] and to discuss the relationship between this estimator and the quasi-maximum likelihood (QML) estimator when the disturbances in (1) are serially independent.

We start with the following assumption:

ASSUMPTION 1: (i) The parameter vector α_0 lies in a compact set ϕ in p -dimensional Euclidean space R^p . (ii) The disturbance vectors $\{u_t: t = 1, 2, \dots\}$ are stochastically independent and identically distributed with zero mean and positive definite covariance matrix Ω . (iii) The elements of g_t are continuous functions on ϕ .

Any vector $\alpha_T(S)$ in ϕ which minimizes the quadratic form

$$R_T(\alpha) = T^{-1} \sum_{t=1}^T (y_t - g_t(\alpha))' S (y_t - g_t(\alpha)),$$

given the observations $\{y_t: t = 1, \dots, T\}$ and some positive definite matrix S , is called a MDE of α_0 . The fact that $\alpha_T(S)$ exists under Assumption 1 as a measurable

¹ I wish to thank a referee for his helpful comments and suggestions on earlier drafts of this paper.

function of y_1, \dots, y_T follows directly from Lemma 2 of Jennrich [3]. On the other hand, the QML estimator $\hat{\alpha}_T$ of α_0 is obtained by maximizing with respect to α in ϕ what would be the likelihood function if the u_t were normally distributed. Concentrating the likelihood function with respect to α , we find that $\hat{\alpha}_T$ minimizes

$$\log \det \left\{ T^{-1} \sum_{t=1}^T (y_t - g_t(\alpha))(y_t - g_t(\alpha))' \right\}.$$

It has been suggested [5 and 7] that if we iterate the MDE $\alpha_T(S)$ by replacing S at each iteration with the inverse of the moment matrix of residuals $\{u_t^* = y_t - g_t(\alpha_T(S)): t = 1, \dots, T\}$ from the previous iteration, we will arrive eventually at the QML estimator $\hat{\alpha}_T$. This conjecture raises two questions: (i) whether the iteration converges; (ii) whether the outcome of the iteration (given that it does converge) depends on the choice of S used at the start of the iteration. We will study both questions and show that, at least for large enough T , the iteration does converge and the limit point, which is independent of S , is $\hat{\alpha}_T$.

2. STRONG CONVERGENCE OF $\alpha_T(S)$ AND $\hat{\alpha}_T$

Our notation is based on Jennrich [3]. Thus, if $\{x_t, y_t: t = 1, \dots, T\}$ are sequences of real vectors, then we represent $T^{-1} \sum_{t=1}^T x_t y_t'$ by $(x, y)_T$ and if the elements of $(x, y)_T$ converge to finite limits as $T \rightarrow \infty$ the limit matrix is denoted by (x, y) and is called the matrix tail product. Moreover, if $\{f_t(\alpha): t = 1, \dots, T\}$ is a sequence of real valued vector functions on ϕ and $(f(\alpha), f(\beta))_T \rightarrow (f(\alpha), f(\beta))$ uniformly for all α and β in ϕ , then we say that the matrix tail product of f exists in ϕ . The tail product in this case is a matrix function on $\phi \times \phi$ and its elements are continuous if the functions f_t are continuous for all t . We now add the following assumption:

ASSUMPTION 2: (i) *The matrix tail product $(g(\alpha), g(\beta))$ exists for all (α, β) in $\phi \times \phi$.*
 (ii) *The matrix tail product $(g(\alpha) - g(\alpha_0), g(\alpha) - g(\alpha_0))$ is positive definite for all $\alpha \neq \alpha_0$ in ϕ .*

Part (ii) of Assumption 2 requires that the vector α_0 be identifiable in $\{g_t(\alpha_0): t = 1, 2, \dots\}$ and this implicitly imposes conditions on the components of the model which make up the systematic component $g_t(\alpha_0)$. For instance, in the constrained linear model $y_t = A(\alpha_0)z_t + u_t$ where the elements of the matrix A are subject to a number of restrictions and can therefore be regarded as functions of a smaller set of parameters comprising the vector α and z_t is a vector of exogenous variables, Assumption 2(ii) implies that if the matrix tail product of z is non-singular then α_0 is identifiable in $A(\alpha_0)$. Similar assumptions have been made by other writers [3, 4, and 6].

For α in ϕ and some positive definite matrix S_T we define

$$Q_T(\alpha) = \text{tr} \{S_T(g(\alpha) - g(\alpha_0), g(\alpha) - g(\alpha_0))_T\}.$$

The asymptotic behavior of $Q_T(\alpha)$ is given by the following result whose proof is straightforward and is omitted.

LEMMA 1: *If Assumption 2 is satisfied and S_T converges almost surely to a positive definite matrix S , then $Q_T(\alpha)$ converges almost surely to*

$$Q(\alpha) = \text{tr} \{S(g(\alpha) - g(\alpha_0), g(\alpha) - g(\alpha_0))\}$$

and Q is continuous on ϕ . Moreover, $Q(\alpha)$ is positive for all α in ϕ not equal to α_0 .

We now have the following theorem:

THEOREM 1: *If Assumptions 1 and 2 are satisfied and if S is an arbitrary positive definite matrix, then the MDE $\alpha_T(S)$ converges almost surely to α_0 and $(y - g(\alpha_T(S)), y - g(\alpha_T(S)))_T$ converges almost surely to Ω .*

This theorem is a simple extension of Jennrich's Theorem 6 [3] and can be established in essentially the same way. Using the strong law of large numbers² we know that $R_T(\alpha) \rightarrow Q(\alpha) + \text{tr}(S\Omega)$ almost surely and uniformly for α in ϕ . Hence, if α^* is a limit point³ of the set of points in the sequence $\{\alpha_T(S)\}$, it follows from Lemma 1 and the inequality $R_T(\alpha_T(S)) \leq R_T(\alpha^0)$ that $Q(\alpha^*) + \text{tr}(S\Omega) \leq \text{tr}(S\Omega)$. Since Q attains its minimum only at α_0 , we have $\alpha^* = \alpha_0$ and $\alpha_T(S) \rightarrow \alpha_0$ almost surely. The fact that $(y - g(\alpha_T(S)), y - g(\alpha_T(S)))_T \rightarrow \Omega$ almost surely now follows because $Q(\alpha_0) = 0$.

THEOREM 2: *If Assumptions 1 and 2 are satisfied then the QML estimator $\hat{\alpha}_T \rightarrow \alpha_0$ almost surely.*

PROOF: Writing $D_T(\alpha) \doteq \log \det (y - g(\alpha), y - g(\alpha))_T$ we know that

$$(2) \quad D_T(\hat{\alpha}_T) = \inf_{\alpha \in \phi} D_T(\alpha)$$

and the existence of $\hat{\alpha}_T$ is assured by the continuity of the elements of g and the compactness of ϕ . By the strong law of large numbers and Jennrich's Theorem 4 [3] it follows that

$$(y - g(\alpha), y - g(\alpha))_T \rightarrow \Omega + (g(\alpha_0) - g(\alpha), g(\alpha_0) - g(\alpha)) = \bar{D}(\alpha)$$

almost surely and uniformly in ϕ . We denote a limit point of the set of points in the sequence $\{\hat{\alpha}_T\}$ by α^* and, setting $D(\alpha) = \log \det \{\bar{D}(\alpha)\}$, we have the inequality $\bar{D}(\alpha^*) \leq \bar{D}(\alpha_0)$, in view of (2), so that by Assumption 2(ii) $\alpha^* = \alpha_0$. Hence, all subsequences of $\{\hat{\alpha}_T\}$ converge to α_0 and the theorem is proved.

² More precisely, the strong law of large numbers ensures that $(u, u)_T \rightarrow \Omega$ almost surely and $(g(\alpha) - g(\alpha^0), u)_T \rightarrow 0$ almost surely and uniformly in ϕ by Jennrich's Theorem 4 [3].

³ Such a point always exists by the Bolzano-Weierstrass theorem when the set of points in the sequence $\{\alpha_T(S)\}$ is infinite. If this set were finite, then there would be no limit points (in the set of points of $\{\alpha_T(S)\}$) but we can use α^* to denote a subsequential limit of the sequence. The argument then follows in the same way. Of course, α^* is also a subsequential limit when the set of points in the sequence is infinite.

Theorem 1 still holds if we replace S by S_T and require that S_T converge almost surely to a positive definite matrix. Therefore, the sequence of estimators $\{\alpha_T^{(n)}: n = 1, 2, \dots\}$ defined by the operations

$$(3) \quad \alpha_T^{(n)} = \alpha_T(\{S_T^{(n-1)}\}^{-1})$$

and

$$(4) \quad S_T^{(n)} = (y - g(\alpha_T^{(n)}), y - g(\alpha_T^{(n)}))_T$$

are all strongly consistent, given $S_T^{(0)} = S$, an arbitrary positive definite matrix. The fact that the estimators $\alpha_T^{(n)}$ converge almost surely to the same value α_0 as $T \rightarrow \infty$ suggests that the iterative procedure involved in (3) and (4) should be numerically stable for large enough T . We consider this question in the next section.

3. NUMERICAL STABILITY OF THE ITERATED MDE

First of all, we introduce a number of additional assumptions that will be needed in this section.

ASSUMPTION 3: *The derivatives of $g_i(\alpha)$ up to the second order exist and are continuous on ϕ . The matrix tail products involving $g_i(\alpha)$, $\partial g_{it}(\alpha)/\partial \alpha$, and $\partial^2 g_{it}(\alpha)/\partial \alpha \partial \alpha'$ exist for all $i = 1, \dots, m$.*

We denote by $W_t(\alpha)$ the matrix whose (i, j) th element is $w_{ijt}(\alpha) = \partial g_{jt}(\alpha)/\partial \alpha_i$ and define for every positive definite matrix S of order n the matrix

$$M_T(S, \alpha) = T^{-1} \sum_{t=1}^T W_t(\alpha)' S W_t(\alpha).$$

ASSUMPTION 4: *For any positive definite matrix S of order m , $M_T(S, \alpha_0)$ has a positive definite limit as $T \rightarrow \infty$.*

Note that Assumption 4 implies that $M_T(S, \alpha_0)$ is positive definite for large enough T .

ASSUMPTION 5: *The vector α_0 lies in the interior of ϕ .*

In view of Assumption 3, each estimator $\alpha_T^{(n)}$ in the sequence defined by (3) and (4) satisfies the necessary conditions

$$(5) \quad T^{-1} \sum_{t=1}^T W_t(\alpha_T^{(n)})' \{S_T^{(n-1)}\}^{-1} (y_t - g_t(\alpha_T^{(n)})) = 0$$

where $S_T^{(n-1)} = (y - g(\alpha_T^{(n-1)}), y - g(\alpha_T^{(n-1)}))_T$. For (5) to be well defined we need $S_T^{(n-1)}$ to be nonsingular. Later in the paper (in the proof of Theorem 3) we will show that this is so for large enough T . Since $S_T^{(n-1)}$ depends on the previous

member $\alpha_T^{(n-1)}$ of the sequence $\{\alpha_T^{(n)}\}$ we can write (5) as the implicit function iteration

$$(6) \quad H_T(\alpha_T^{(n)}, \alpha_T^{(n-1)}) = 0 \quad (n = 2, 3, \dots).$$

To ensure that at each stage $\alpha_T^{(n)}$ is a minimum distance estimator it is usual to require that $\alpha_T^{(n)}$ minimizes

$$q_T^{(n)}(\alpha) = \text{tr} [\{S_T^{(n-1)}\}^{-1}(y - g(\alpha), y - g(\alpha))_T]$$

in ϕ . This brings us back to the iteration defined by (3) and (4). An alternative which we adopt in the following is to combine (6) with the requirement that both $\alpha_T^{(n)}$ and $\alpha_T^{(n-1)}$ are restricted to an open sphere in ϕ that contains α_0 , has center $\bar{\alpha}_T$ where $H_T(\bar{\alpha}_T, \bar{\alpha}_T) = 0$, and fixed radius independent of T . Such a sphere is constructed below and it is shown that, for large enough T , $\alpha_T^{(n)}$ is then uniquely defined by (6) in this region and is, moreover, the MDE for which $\alpha_T^{(n)}(\alpha)$ attains its global minimum in ϕ . The initial condition on $\alpha_T^{(1)}$ can be met by requiring $\alpha_T^{(1)}$ to be a MDE (with, for example, $S^{(0)} = I_m$) so that, for large enough T , $\alpha_T^{(1)}$ lies in a suitable neighborhood of α_0 . For large T , there will, therefore, be no real difference between this formulation of the iteration and that based on (3) and (4).

THEOREM 3: *If Assumptions 1 through 5 are satisfied then there exists a neighborhood N of α_0 in ϕ such that for almost every $y = \{y_i: 1, 2, \dots\}$ there is an integer $T(y)$ for which the iteration (6) will converge from any starting value $\alpha_T^{(1)}$ in N for all $T \geq T(y)$. The point of attraction of this iteration is $\hat{\alpha}_T$, the QML estimator of α_0 .*

STRUCTURE OF THE PROOF: Since the proof of the theorem is lengthy it may be useful to comment on the main steps in the development of the argument:

(i) We first note that the QML estimator $\hat{\alpha}_T$ is one point of attraction of the iteration (6). Later on in the proof we verify that this point of attraction is unique in a fixed neighborhood of α_0 for large enough T .

(ii) We then investigate the properties of the derivatives of the implicit function H_T and use these to establish that (a) for starting values of the iteration in a suitable fixed neighborhood of α_0 (which we construct in the proof) $\alpha_T^{(n)}$ can be expressed uniquely in the explicit form $\alpha_T^{(n)} = f_T(\alpha_T^{(n-1)})$, and that (b) this explicit iteration is numerically stable and converges to the same point of attraction for all starting values in this neighborhood. The verification of (a) which occupies the main body of the proof requires more than local uniqueness in a neighborhood of $\bar{\alpha}_T$ where $H_T(\bar{\alpha}_T, \bar{\alpha}_T) = 0$. Instead, we need to establish global uniqueness for all points in a neighborhood of α_0 , which remains fixed for all values of T considered. This is done in the theorem by appealing to one of the Gale-Nikaido univalence theorems [1].⁴

PROOF: By the statement "almost every y " we mean almost every y with respect to the probability measure induced by the stochastic properties of the u_t on the space of all realizations of the y_t process.

⁴ Use of the implicit function theorem alone ensures only a unique solution in a neighborhood of $\bar{\alpha}_T$. This neighborhood cannot be guaranteed to include a fixed neighborhood of α_0 for all large T as is required by the theorem.

To prove the numerical stability of (6) we must first verify that there is a point of attraction $\bar{\alpha}_T$ in ϕ for which

$$(7) \quad H_T(\bar{\alpha}_T, \bar{\alpha}_T) = 0.$$

The equation system (7) provides the necessary conditions for $D_T(\alpha)$ to have a stationary point at $\bar{\alpha}_T$ in the interior of ϕ . According to Theorem 2 and Assumption 5, such a point is given by the estimator $\hat{\alpha}_T$ when T is sufficiently large. More precisely, for almost every y , there is a $T_1(y)$ such that $\hat{\alpha}_T$ lies in the interior of ϕ for all $T \geq T_1(y)$. Hence, the existence of at least one point of attraction is assured and in the argument that follows we treat $\bar{\alpha}_T = \hat{\alpha}_T$. We will show later in the proof that this point of attraction is unique in a fixed neighborhood of α_0 for large enough T . But, first, we turn to investigate the derivatives of H_T .

From (5) and (6) we obtain⁵

$$(8) \quad \frac{\partial H_T(\alpha^{(n)}, \alpha^{(n-1)})}{\partial \alpha^{(n)}} = A_T(\alpha^{(n)}, \alpha^{(n-1)}) + U_T(\alpha^{(n)}, \alpha^{(n-1)})$$

where

$$A_T(\alpha^{(n)}, \alpha^{(n-1)}) = -T^{-1} \sum_{t=1}^T W_t(\alpha^{(n)})' \{S_T^{(n-1)}\}^{-1} W_t(\alpha^{(n)})$$

and U_T is the matrix whose (i, j) th element is

$$u_{ijT}(\alpha^{(n)}, \alpha^{(n-1)}) = T^{-1} \sum_{t=1}^T \frac{\partial^2 g_t(\alpha^{(n)})'}{\partial \alpha_i \partial \alpha_j} \{S_T^{(n-1)}\}^{-1} (y_t - g_t(\alpha^{(n)})).$$

We let β and γ be two vectors in ϕ and consider $\partial H_T(\beta, \gamma) / \partial \alpha^{(n)}$. The (i, j) th element of this matrix is

$$(9) \quad -T^{-1} \sum_i \sum_{kl} w_{iki}(\beta) s_T^{kl}(\gamma) w_{jil}(\beta) + T^{-1} \sum_i \frac{\partial^2 g_i(\beta)'}{\partial \alpha_i \partial \alpha_j} S_T(\gamma)^{-1} (y_i - g_i(\beta))$$

where $S_T(\gamma) = (y - g(\gamma), y - g(\gamma))_T$ and $S_T(\gamma)^{-1} = (s_T^{ij}(\gamma))$. Writing $y - g(\gamma) = y - g(\alpha_0) + g(\alpha_0) - g(\gamma)$, it is clear from the strong law of large numbers and Jennrich's Theorem 4 [3] that $S_T(\gamma)$ converges almost surely and uniformly in γ to

$$S(\gamma) = \Omega + (g(\alpha_0) - g(\gamma), g(\alpha_0) - g(\gamma))$$

as $T \rightarrow \infty$. $S(\gamma)$ is positive definite for all γ and it follows from the uniform convergence of $S_T(\gamma)$ that $S_T(\gamma)$ is nonsingular for T sufficiently large. Hence, (9) is well defined for large enough T .

By Assumption 3, (9) converges almost surely as $T \rightarrow \infty$ to

$$(10) \quad -\text{tr} \{S(\gamma)^{-1} (\partial g(\beta) / \partial \alpha_i, \partial g(\beta) / \partial \alpha_j)\} \\ + \text{tr} \{S(\gamma)^{-1} (g(\alpha_0) - g(\beta), \partial^2 g(\beta) / \partial \alpha_i \partial \alpha_j)\}$$

⁵ In much of what follows we omit the subscript T on $\alpha_T^{(n)}$ to simplify the notation. This is unlikely to cause confusion.

uniformly for (β, γ) in $\phi \times \phi$. Hence, $\partial H_T(\beta, \gamma)/\partial \alpha^{(n)}$ converges uniformly in (β, γ) as $T \rightarrow \infty$ to

$$H^1(\beta, \gamma) = A(\beta, \gamma) + U(\beta, \gamma)$$

where the (i, j) th elements of A and U are the first and second terms, respectively, of (10). The elements of $U(\alpha_0, \alpha_0)$ are clearly zero and, since $S(\alpha_0) = \Omega$, it follows from Assumption 4 that H^1 is nonsingular at $(\beta, \gamma) = (\alpha_0, \alpha_0)$. But the elements of H^1 are continuous functions of β and γ , so there exists a closed sphere S^* with center α_0 in ϕ such that H^1 is nonsingular for all (β, γ) in $S^* \times S^*$. Moreover, $\partial H_T(\beta, \gamma)/\partial \alpha^{(n)} \rightarrow H^1(\beta, \gamma)$ uniformly in (β, γ) and, therefore, there exists an integer $T_2(y)$ such that $\partial H_T(\beta, \gamma)/\partial \alpha^{(n)}$ is nonsingular in $S^* \times S^*$ for all $T \geq T_2(y)$. We select the radius of S^* in such a way that the boundary of S^* lies in the interior of ϕ . Then, from the strong convergence of $\bar{\alpha}_T$ to α_0 , we know that there exists an integer $T_3(y) \geq T_1(y)$ for which $\bar{\alpha}_T(y)$ is an interior point of S^* when $T \geq T_3(y)$. Hence, $\partial H_T(\bar{\alpha}_T, \bar{\alpha}_T)/\partial \alpha^{(n)}$ is nonsingular for all $T \geq T_4(y) = \max \{T_2(y), T_3(y)\}$.

We now consider $\partial H_T(\beta, \gamma)/\partial \alpha^{(n-1)}$. The i 'th row (which, for convenience, we write as a column vector) of this matrix is

$$(11) \quad -T^{-1} \sum_{\tau} W_{\tau}(\beta) S_T(\gamma)^{-1} \frac{\partial S_T(\gamma)}{\partial \alpha_i} S_T(\gamma)^{-1} (y_{\tau} - g_{\tau}(\beta))$$

for T large enough to ensure that $S_T(\gamma)$ is nonsingular. Since

$$\partial S_T(\gamma)/\partial \alpha_i = -(\partial g(\gamma)/\partial \alpha_i, y - g(\gamma))_T - (y - g(\gamma), \partial g(\gamma)/\partial \alpha_i)_T$$

it follows from Assumption 3 that $\partial S_T(\gamma)/\partial \alpha_i$ converges almost surely and uniformly in γ to

$$(12) \quad S^i(\gamma) = -(\partial g(\gamma)/\partial \alpha_i, y - g(\gamma)) - (y - g(\gamma), \partial g(\gamma)/\partial \alpha_i).$$

Writing $y_{\tau} - g_{\tau}(\beta) = y_{\tau} - g_{\tau}(\alpha_0) + g_{\tau}(\alpha_0) - g_{\tau}(\beta)$ in (11), we see that the (i, j) th element of $\partial H_T(\beta, \gamma)/\partial \alpha^{(n-1)}$ converges almost surely and uniformly in (β, γ) to

$$(13) \quad -\text{tr} \{S(\gamma)^{-1} S^i(\gamma) S(\gamma)^{-1} (y - g(\beta), \partial g(\beta)/\partial \alpha_j)\}.$$

The square matrix of order p whose (i, j) th element is given by (13) we denote by $H^2(\beta, \gamma)$. It then follows from (12) and (13) that the matrix H^2 is zero when $\beta = \gamma = \alpha_0$.

From the continuity of (13) and the uniform convergence of $\partial H_T(\beta, \gamma)/\partial \alpha^{(n-1)}$, we can find for any $\varepsilon > 0$ an integer $T_5(y)$ and an open sphere R^* with center α_0 and radius r such that

$$\|\partial H_T(\beta, \gamma)/\partial \alpha^{(n-1)}\| < \varepsilon$$

for all $T \geq T_5(y)$ and for all β and γ in R^* . The double vertical bars are used to denote the Euclidean norm. We select

$$\varepsilon = \frac{3/4}{\sup_{T \geq T_2(y)} \{ \sup_{\beta, \gamma \in S^*} \|\{\partial H_T(\beta, \gamma)/\partial \alpha^{(n)}\}^{-1}\| \}}$$

so that, taking R^* to be within S^* , we have

$$(14) \quad \begin{aligned} & \| \{ \partial H_T(\beta, \gamma) / \partial \alpha^{(n)} \}^{-1} \{ \partial H_T(\beta, \gamma) / \partial \alpha^{(n-1)} \} \| \\ & \leq \| \{ \partial H_T(\beta, \gamma) / \partial \alpha^{(n)} \}^{-1} \| \| \partial H_T(\beta, \gamma) / \partial \alpha^{(n-1)} \| < 3/4 \end{aligned}$$

for all $T \geq T_6(y) = \max \{ T_4(y), T_5(y) \}$ and all $(\beta, \gamma) \in R^* \times R^*$.

To prove the theorem we now need a neighborhood of N of α_0 in which the iteration is stable for all T sufficiently large. We must show first that there is a sequence of spherical neighborhoods $\{S_T\}$ with center $\bar{\alpha}_T$ and fixed radius and a sequence of unique continuously differentiable functions $\{f_T\}$ such that, for large enough T , (i) $S_T \subset R^*$, (ii) $f_T(\bar{\alpha}_T) = \bar{\alpha}_T$, and (iii) $H_T(f_T(\gamma), \gamma) = 0$ for all $\gamma \in S_T$. For, if conditions (i), (ii), and (iii) are satisfied we can write the iteration (6) as

$$(15) \quad \alpha_T^{(n)} = f_T(\alpha_T^{(n-1)})$$

when the starting value $\alpha_T^{(1)} \in S_T$. Then, for some α^* on the line segment joining $\alpha_T^{(n-1)}$ and $\bar{\alpha}_T$ we have

$$\alpha_T^{(n)} - \bar{\alpha}_T = \frac{\partial f_T(\alpha^*)}{\partial \alpha_T^{(n-1)}} (\alpha_T^{(n-1)} - \bar{\alpha}_T),$$

and since

$$\frac{\partial f_T}{\partial \alpha_T^{(n-1)}} = - \left(\frac{\partial H_T}{\partial \alpha_T^{(n)}} \right)^{-1} \frac{\partial H_T}{\partial \alpha_T^{(n-1)}}$$

it follows from (14) that the iteration (15) is numerically stable (with the point of attraction $\bar{\alpha}_T$) for all starting values in S_T . Finally, if we can find a fixed neighborhood N of α_0 with the property that $N \subset S_T$ when T is sufficiently large, the iteration (15) will converge for all starting values $\alpha_T^{(1)} \in N$ and the theorem will be proved.

We construct S_T as the open sphere with center $\bar{\alpha}_T$ and radius $r/2$ and let E_T be the smallest closed cube containing S_T . E_T may always be constructed in such a way that its sides are parallel to the axes in R^p , so that the cube can be regarded as a closed rectangular region. (This will be needed later.) Since $\bar{\alpha}_T \rightarrow \alpha_0$ almost surely, there exists an integer $T_7(y)$ such that $E_T \subset R^*$ for all $T \geq T_7(y)$. Thus, we have $S_T \subset E_T \subset R^* \subset S^*$ for all $T \geq T_7(y)$. We now let N be the open sphere with center α_0 and radius $r/8$. We can readily show that, for $T \geq T_7(y)$, $N \subset S_T$.

We can now return to the remark made at the beginning of the proof that the point of attraction $\bar{\alpha}_T = \hat{\alpha}_T$ is unique in a neighborhood of α_0 for large T . We take the neighborhood N and suppose $\beta \in N$ is another point of attraction. Both β and $\bar{\alpha}_T$ are fixed points of the function f_T in (15) and expanding $f_T(\beta)$ in a Taylor series about $\bar{\alpha}_T$ we have

$$f_T(\beta) = f_T(\bar{\alpha}_T) + \frac{\partial f_T(\beta^*)}{\partial \alpha^{(n-1)}} (\beta - \bar{\alpha}_T)$$

so that

$$(16) \quad \bar{\beta} - \bar{\alpha}_T = \frac{\partial f_T(\beta^*)}{\partial \alpha(n-1)}(\bar{\beta} - \bar{\alpha}_T)$$

where β^* lies on the line segment joining $\bar{\beta}$ and $\bar{\alpha}_T$. But, in view of (14) and the fact that $N \subset R^*$, we have

$$\|\partial f_T(\beta^*)/\partial \alpha^{(n-1)}\| < 3/4,$$

for large T , so that (16) implies that $\bar{\beta} = \bar{\alpha}_T$.

Having constructed the neighborhood N of α_0 and the spheres S_T such that $N \subset S_T \subset R^*$, it remains to prove that conditions (ii) and (iii) are satisfied. The i 'th element of the vector $H_T(\beta, \gamma)$ is

$$T^{-1} \sum_t (\partial g_t(\beta)/\partial \alpha_i) S_T(\gamma)^{-1} (y_t - g_t(\beta))$$

which, according to earlier arguments about $S_T(\gamma)$ in this proof and by Assumption 3, converges almost surely as $T \rightarrow \infty$ to

$$(17) \quad \text{tr} \{S(\gamma)^{-1}(y - g(\beta)), \partial g(\beta)/\partial \alpha_i\}$$

uniformly in β and γ . We denote by $H(\beta, \gamma)$ the vector whose i 'th element is (17). Then, $H(\alpha_0, \alpha_0) = 0$ and $\partial H(\beta, \gamma)/\partial \beta = H^1(\beta, \gamma)$ by the uniform convergence of $\partial H_T(\beta, \gamma)/\partial \beta$. Thus, $\partial H(\beta, \gamma)/\partial \beta$ and $\partial H_T(\beta, \gamma)/\partial \beta$ are nonsingular for all β and γ in S^* and $T \geq T_2(y)$.

We now define the mapping $F: S^* \times S^* \rightarrow R^{2p}$ and the sequence of mappings $\{F_T: S^* \times S^* \rightarrow R^{2p}\}$ by

$$z = H(\beta, \gamma), w = \gamma; (\beta, \gamma) \in S^* \times S^*$$

and

$$\{z = H_T(\beta, \gamma), w = \gamma; (\beta, \gamma) \in S^* \times S^*\},$$

respectively. It follows that

$$(18) \quad \frac{\partial F(\beta, \gamma)}{\partial (\beta, \gamma)} = \begin{bmatrix} \partial H/\partial \beta & 0 \\ \partial H/\partial \gamma & I \end{bmatrix}$$

is nonsingular for all $(\beta, \gamma) \in S^* \times S^*$ and

$$\frac{\partial F_T(\beta, \gamma)}{\partial (\beta, \gamma)} = \begin{bmatrix} \partial H_T/\partial \beta & 0 \\ \partial H_T/\partial \gamma & I \end{bmatrix}$$

is nonsingular for all $(\beta, \gamma) \in S^* \times S^*$ and all $T \geq T_2(y)$. Moreover, from the continuity of the elements of $\partial H(\beta, \gamma)/\partial \beta$ and the fact that $\partial H(\alpha_0, \alpha_0)/\partial \beta$ is positive definite, it follows that $\partial H(\beta, \gamma)/\partial \beta$ is a positive quasi-definite matrix on $S^* \times S^*$ (i.e., the symmetric part $\{\partial H/\partial \beta + (\partial H/\partial \beta)'\}/2$ is positive definite). Thus, $\partial H(\beta, \gamma)/\partial \beta$ is a P matrix (see [1]) on $S^* \times S^*$ and, from the form of (18), it is clear that

$\partial F(\beta, \gamma)/\partial(\beta, \gamma)$ is a P matrix on the same set. Since the eigenvalues of $\partial H_T(\beta, \gamma)/\partial\beta$ converge uniformly to the eigenvalues of $\partial H(\beta, \gamma)/\partial\beta$, there exists an integer $T_8(y) \geq T_2(y)$ for which the eigenvalues of $\partial H_T(\beta, \gamma)/\partial\beta$ are positive whenever $T \geq T_8(y)$ and $(\beta, \gamma) \in S^* \times S^*$. Thus, $\partial H_T(\beta, \gamma)/\partial\beta$ and, therefore, $\partial F_T(\beta, \gamma)/\partial(\beta, \gamma)$ are P matrices on $S^* \times S^*$ when $T \geq T_8(y)$.

By the Gale-Nikaido univalence theorem [1, Theorem 4], the sequence of functions $\{F_T: T \geq T_8(y)\}$ are 1:1 mappings of the corresponding sequence of closed rectangles $\{E_T \times E_T: T \geq T_8(y)\}$ onto $\{F_T(E_T \times E_T): T \geq T_8(y)\}$. We can write the sequence of inverse functions as

$$\{\beta = G_T(z, w), \gamma = w; (z, w) \in F_T(E_T \times E_T)\},$$

and the elements of G_T are continuously differentiable by the inverse function theorem of advanced calculus. Since $S_T \subset E_T \subset R^*$ when $T \geq T_7(y)$, we have

$$(19) \quad \{\beta = G_T(z, w), \gamma = w; (z, w) \in F_T(S_T \times S_T)\}$$

for $T \geq T^*(y) = \max \{T_7(y), T_8(y)\}$. S_T contains $\bar{\alpha}_T$ and $\bar{\alpha}_T$ satisfies $H_T(\bar{\alpha}_T, \bar{\alpha}_T) = 0$, so that $(0, w) \in F_T(S_T \times S_T)$ for all $w \in S_T$. Setting $z = 0$ in (19) we have, for all $T \geq T^*(y)$, $\bar{\alpha}_T = G_T(0, \bar{\alpha}_T)$ and $H_T(G_T(0, \gamma), \gamma) = 0$ for all $\gamma \in S_T$. Thus, conditions (ii) and (iii) given earlier in the proof are satisfied.

The above argument is valid for almost all y and the theorem is proved with the spherical neighborhood N of α_0 and the integer $T(y) = \max \{T_6(y), T^*(y)\}$.

Q.E.D.

If the starting point of the iteration (6) is $\alpha_T^{(1)} = \alpha_T(S)$ for some positive definite matrix S , then Theorems 1 and 3 have the following corollary:

COROLLARY: *If Assumptions 1–5 are satisfied and $\alpha_T^{(1)} = \alpha_T(S)$ for some positive definite matrix S , then for almost all y there is an integer $T(y, S)$ such that the iteration (6) converges to $\hat{\alpha}_T$ whenever $T \geq T(y, S)$.*

4. CONCLUSION

Theorem 3 tells us that, for any starting point $\alpha_T^{(1)}$ in a sufficiently small neighborhood of the true value α_0 , the iterated MDE converges to the QML estimator with probability approaching one as $T \rightarrow \infty$. Given that $\alpha_T^{(1)}$ in the implicit function iteration (6) is itself a MDE, the condition that $\alpha_T^{(1)}$ be sufficiently close to α_0 can be translated into a further restriction on the sample size, as indicated by the Corollary of Theorem 3. Intuition suggests that a poor choice of the arbitrary matrix S in $\alpha_T^{(1)} = \alpha_T(S)$ may necessitate a large sample size before convergence to the QML estimator is assured. Nevertheless, given two different starting points $\alpha_T(S_1)$ and $\alpha_T(S_2)$ to the iteration, where $S_1 \neq S_2$ are two arbitrary positive definite matrices, we can be sure that the iterations converge on the same point ($\hat{\alpha}_T$) provided $T \geq T(y) = \max \{T(y, S_1), T(y, S_2)\}$. In this sense, the iterated MDE is independent of the choice of the positive definite matrix S used at the start of the iteration.

Another implication of the proof of Theorem 3 is that, if we start with an MDE $\alpha_T^{(1)} = \alpha_T(S)$, the unique solution $\alpha_T^{(n)}$ of (6) is indeed the MDE which minimizes $q_T^{(n)}(\alpha)$, provided T is sufficiently large. To show this we consider $\alpha_T^{(2)}$ which satisfies $H_T(\alpha_T^{(2)}, \alpha_T^{(1)}) = 0$ and thus the necessary conditions for $q_T^{(2)}(\alpha)$ to have a local minimum at $\alpha_T^{(2)}$. But $S_T^{(1)} \rightarrow \Omega$ almost surely and

$$\frac{\partial^2 q_T^{(2)}(\alpha)}{\partial \alpha \partial \alpha'} \rightarrow \frac{\partial^2 q(\alpha)}{\partial \alpha \partial \alpha'}$$

almost surely and uniformly in α , where

$$q(\alpha) = \text{tr} \{ \Omega^{-1} (g(\alpha_0) - g(\alpha), g(\alpha_0) - g(\alpha)) \} + n.$$

Moreover, from the proof of Theorem 3 we know that

$$\frac{\partial^2 q(\alpha)}{\partial \alpha \partial \alpha'} = -A(\alpha, \alpha_0) - U(\alpha, \alpha_0)$$

and this matrix is positive definite for all $\alpha \in S^*$. Hence, $\partial^2 q_T^{(2)}(\alpha) / \partial \alpha \partial \alpha'$ is positive definite for $\alpha \in S^*$ and large T . In particular, it is positive definite for $\alpha = \alpha_T^{(2)}$ which lies in $S_T \subset S^*$ whenever $\alpha_T^{(1)}$ lies in S_T . Thus, $q_T^{(2)}$ is a local minimum of $q_T^{(2)}(\alpha)$. Since $\alpha_T^{(2)}$ is the only turning point of $q_T^{(2)}(\alpha)$ in S_T , and since the global minimum of $q_T^{(2)}(\alpha)$ converges almost surely to α_0 , it follows that $\alpha_T^{(2)}$ is the global minimum of $q_T^{(2)}(\alpha)$ for large enough T . The proof for $n > 2$ follows in the same way.

Before concluding this paper, two comments are in order. First, our discussion has been rather abstract so that we have not yet mentioned that $\alpha_T^{(1)}$ itself must, in general, be found by an iterative procedure, and a starting value for this primary iteration will be needed. However, provided $\alpha_T^{(1)} = \alpha_T(S)$ does minimize $\text{tr} \{ S^{-1} \times (y - g(\alpha), y - g(\alpha)) \}$, this fact does not affect our results. Finally, it may be worth emphasizing that the equivalence of the iterated MDE and the QML estimator for sufficiently large T has been established under conditions which, in themselves, ensure that the QML estimator is strongly consistent. This fact is of some importance (c.f. [5, pp. 338–340]). For, although the estimator $\hat{\alpha}_T$ can be formally regarded as the MDE $\alpha_T(S_T^{-1})$ where $S_T = (y - g(\hat{\alpha}_T), y - g(\hat{\alpha}_T))_T$, the fact that $\hat{\alpha}_T \rightarrow \alpha_0$ almost surely does not follow from Theorem 1 because, in the absence of Theorem 2, we cannot assert that S_T converges almost surely as $T \rightarrow \infty$ to a positive definite matrix.

University of Essex

Manuscript received August, 1974.

REFERENCES

[1] GALE, D., AND H. NIKAIIDO: "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159 (1965), 81–93. (Reprinted in *Readings in Mathematical Economics*, Vol. 1, ed. by P. Newman, Baltimore: Johns Hopkins Press, 1968.)
 [2] HANNAN, E. J.: "Non-Linear Time Series Regression," *Journal of Applied Probability*, 8 (1971), 767–780.
 [3] JENNRICH, R. I.: "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633–643.

- [4] MALINVAUD, E.: "The Consistency of Nonlinear Regressions," *Annals of Mathematical Statistics*, 41 (1970), 956-969.
- [5] ———: *Statistical Methods of Econometrics*, second edition. Amsterdam: North Holland, 1970.
- [6] ROBINSON, P. M.: "Non-Linear Regression for Multiple Time-Series," *Journal of Applied Probability*, 9 (1972), 758-768.
- [7] SARGAN, J.: "The Identification and Estimation of Sets of Simultaneous Stochastic Equations," mimeograph. London: London School of Economics, 1972.