EXPLICIT ESTIMATORS OF PARAMETRIC
FUNCTIONS IN NONLINEAR REGRESSION

by

A. Ronald Gallant

Institute of Statistics
Mimeo Series No. 1108
February, 1977

EXPLICIT ESTIMATORS OF PARAMETRIC
FUNCTIONS IN NONLINEAR REGRESSION

by

A. Ronald Gallant^{*}

ABSTRACT

The possibility of employing explicitly defined functions of the observations as estimators of parametric functions in nonlinear regression analysis is explored. A general theory of best average mean square error estimation leading to explicit estimators is set forth. Such estimators are given a Bayesian interpretation as Fourier expansions of the estimator which minimizes expected posterior square error loss. In an example a linear function of the observations performs better than the maximum likelihood estimator of the parametric function of interest and nearly as well as the Bayes estimator according to the criterion of average square error loss.

1. Introduction

The article explores the possibility of employing explicitly defined functions of the observations as estimators of parametric functions in nonlinear regression analysis. There are two motivations for the use of explicit estimators as opposed to the customary least squares or maximum likelihood approach. The first is to avoid the need for iterative methods in the computations and attendant problems of noneconvergence and convergence to erroneous local optima. The second is that the explicit functional form of the estimator may permit the discovery of the sampling distribution of the estimator and thus avoid reliance on an asymptotic theory.

Reasonable success along these lines is achieved, assuming normally distributed errors. A general theory of best average mean square error estimation leading to explicit estimators is set forth in the next two sections. Such estimators are given a Bayesian interpretation as Fourier expansions of the estimator which minimizes expected posterior square error loss in Section 4. In the example of Section 5, a linear function of the observations performs better than the maximum likelihood estimator and nearly as well as the Bayes' estimator according to the criterion of average mean square error. A procedure for finding an exact confidence interval for the parametric function which uses the linear estimator is given in Section 6. The article concludes with some comments.

2. The Structure of the Problem

The observed data y_t, x_t ($t = 1, 2, \dots, n$) is assumed to have been generated according to the statistical model

$$y_t = f(x_t, \theta) + e_t \quad (t = 1, 2, \dots, n)$$

The form of the response function $f(x, \theta)$ is known, the unknown parameter θ is a p -vector known to be contained in the parameter space Θ , the inputs x_t are known k -vectors, the univariate responses y_t are observed, and the unobservable errors e_t are assumed normally distributed with mean zero and finite variance σ^2 . The parametric function to be estimated is

$$g^*(\gamma)$$

where $\gamma = (\theta', \sigma')'$.

The estimators considered are of a class obtained as follows. The observed y_t , denoted as the n -vector

$$y = (y_1, y_2, \dots, y_n)'$$

are transformed according to

$$z = Z(y)$$

The function Z may be any vector valued mapping of n -space to m -space but the convenient choice in applications is to take a basis for the polynomials in y of, say, second degree as the components of $Z(y)$. The class of estimators $\hat{g}$ of the form

$$\hat{g} = a'z$$

where a is an m -vector restricted to a linear space G are considered.

Estimators of this form are explicit functions of the observations,

viz. $\hat{g}(y) = a'Z(y)$

The estimator $\hat{g}$ to be chosen from this class is that which minimizes average mean square error with respect to a weighting measure ρ defined on $\Gamma = \Theta \times (0, \infty)$. That is, one seeks to find $\hat{g}$ of the form $\hat{g} = a'z$ minimizing

$$AMSE(\hat{g}) = \int_{\Gamma} \mathcal{E}_{\gamma} [\hat{g} - g^*(\gamma)]^2 d\rho(\gamma)$$

where a is restricted to G .

It shall be assumed throughout that the components of $Z(y)$ are measurable and square integrable with respect to the n -dimensional normal distribution with mean vector

$$f(\theta) = [f(x_1, \theta), f(x_2, \theta), \dots, f(x_n, \theta)]'$$

and variance-covariance matrix $\sigma^2 I$; denote the corresponding density by $n[y|f(\theta), \sigma^2]$; and, denote expectation with respect to $n[y|f(\theta), \sigma^2]$ by $E_Y(\cdot)$. The functions $g^*(\gamma)$, σ^2 , $E_Y[Z_i(y)]$ and the like are assumed measurable and square integrable with respect to ρ .

3. A Sampling Theoretic Derivation of the Estimator

Let $b_1, b_2, \dots, b_r$ be m -vectors constituting a basis for G and arrange them as column vectors in the m by r matrix

$$B = [b_1 : b_2 : \dots : b_r] .$$

The following result is derived in this section: in the class of estimators of the form $a'z$ with a in G , that choice of a which minimizes the average mean square error with respect to ρ is

$$\hat{a} = B \{ B' \left[\int_{\Gamma} E_Y(z z') d\rho(\gamma) \right] B \}^{-1} B' \left[\int_{\Gamma} g^*(\gamma) E_Y(z) d\rho(\gamma) \right] .$$

The notation $\int_{\Gamma} E_Y(z z') d\rho(\gamma)$ denotes the m by m matrix with typical element

$$\int_{\Gamma} E_Y(z_i z_j) d\rho(\gamma) .$$

Similarly $\int_{\Gamma} g^*(\gamma) E_Y(z) d\rho(\gamma)$ denotes the m -vector with typical element

$$\int_{\Gamma} g^*(\gamma) E_Y(z_i) d\rho(\gamma) .$$

The notation A^- denotes any generalized inverse of A . If there are no restrictions on the choice of a then B may be taken as the identity matrix and the formula simplifies to

$$\hat{a} = \left[\int_{\Gamma} \mathcal{E}_{\gamma}(zz') d\rho(\gamma) \right]^{-1} \int_{\Gamma} g^*(\gamma) \mathcal{E}_{\gamma}(z) d\rho(\gamma) .$$

Note that for computing $\hat{a}$ the most convenient choice for the functions $Z_i(y)$ are the polynomials in y . With this choice the elements of $\mathcal{E}_{\gamma}(z)$ and $\mathcal{E}_{\gamma}(zz')$ may be obtained as the moments about zero of a spherical normal distribution centered at $f(\theta)$; these will be polynomials in $f(x_i, \theta)$ and σ . A natural ordering of the functions $Z_i(y)$ is such that if $m = 1 + n$ the components of $Z(y)$ are a basis for the polynomials in y of first degree, if $m = 1 + n + n(n+1)/2$ the components of $Z(y)$ are a basis for the polynomials in y of second degree, and so on. The degree of the polynomial may be varied by varying m or by choice of B ; also, a sufficiency reduction may be accomplished by choice of columns for B . The use of B for these purposes may permit some flexibility in writing code to implement the estimator.

The verification of these facts follows; the remainder of this section may be skipped at first reading without affecting continuity.

Define

$$g_{b_i}(\gamma) = \mathcal{E}_{\gamma}(b'_i z) \quad (i = 1, 2, \dots, r) .$$

Let H be the $r \times r$ matrix with typical element

$$h_{ij} = \int_{\Gamma} g_{b_i}(\gamma) g_{b_j}(\gamma) d\rho(\gamma) \quad (i, j = 1, 2, \dots, r)$$

and let h be the r -vector with typical element

$$h_i = \int_{\Gamma} g_{b_i}(\gamma) g^*(\gamma) d\rho(\gamma) \quad (i = 1, 2, \dots, r) .$$

The matrix H is positive semi-definite since, for an arbitrary r -vector α ,

$$\alpha' H \alpha = \int_{\Gamma} \left[\sum_{i=1}^r \alpha_i g_{b_i}(\gamma) \right]^2 d\rho(\gamma) \geq 0 .$$

Moreover, the equations $H\alpha = h$ are consistent with a solution

$$\bar{\alpha} = H^{-}h$$

where H^{-} denotes any generalized inverse of H . Set

$$\bar{g}(\gamma) = \sum_{i=1}^r \bar{\alpha}_i g_{b_i}(\gamma) .$$

Lemma 1. Let α be an arbitrary r -vector. Then

$$\begin{aligned} \int_{\Gamma} [g^*(\gamma) - \sum_{i=1}^r \alpha_i g_{b_i}(\gamma)]^2 d\rho(\gamma) \\ = (\alpha - \bar{\alpha})' H (\alpha - \bar{\alpha}) + \int_{\Gamma} [g^*(\gamma) - \bar{g}(\gamma)]^2 d\rho(\gamma) . \end{aligned}$$

Proof: Add and subtract $\bar{g}$ and expand the integrand to obtain

$$\begin{aligned} \int (g^* - \sum_{i=1}^r \alpha_i g_{b_i})^2 d\rho \\ = \int (g^* - \bar{g})^2 d\rho + \int (\bar{g} - \sum_{i=1}^r \alpha_i g_{b_i})^2 d\rho \\ + 2 \int (g^* - \bar{g}) (\bar{g} - \sum_{i=1}^r \alpha_i g_{b_i}) d\rho . \end{aligned}$$

By substituting $\bar{g} = \sum_{i=1}^r \bar{\alpha}_i g_{b_i}$ in the last two terms one obtains

$$\begin{aligned} = \int (g^* - \bar{g})^2 d\rho + \int [\sum_{i=1}^r (\bar{\alpha}_i - \alpha_i) g_{b_i}]^2 d\rho \\ + 2 \int \sum_{i=1}^r (\bar{\alpha}_i - \alpha_i) g_{b_i} (g^* - \sum_{j=1}^r \bar{\alpha}_j g_{b_j}) d\rho \end{aligned}$$

Substitution, using the equations defining H and h , yields

$$= \int (g^* - \bar{g})^2 d\rho + (\bar{\alpha} - \alpha)' H (\bar{\alpha} - \alpha) + 2(\bar{\alpha} - \alpha) (h - H\bar{\alpha}) .$$

The cross product term is zero because $\bar{\alpha}$ is a solution of $H\bar{\alpha} = h$. $\square$

Let $V(\gamma)$ be the variance-covariance matrix of z with typical element

$v_{ij}(\gamma)$. Let $\bar{V}$ be the $m \times m$ matrix with typical element

$$\bar{v}_{ij} = \int_{\Gamma} v_{ij}(\gamma) d\rho(\gamma) .$$

Note that $\bar{V}$ is positive semi-definite since, for an arbitrary m -vector a ,

$$a' \bar{V} a = \int_{\Gamma} a' V(\gamma) a d\rho(\gamma) \geq 0$$

Lemma 2. Let a be an arbitrary choice of an m -vector from G and let the r -vector α satisfy $a = B\alpha$. Then

$$AMSE(a'z) = a' \bar{V} a + (\alpha - \bar{\alpha})' H (\alpha - \bar{\alpha}) + \int_{\Gamma} [g^*(\gamma) - \bar{g}(\gamma)]^2 d\rho(\gamma) .$$

Proof:

$$\begin{aligned} AMSE(a'z) &= \int Var(a'z) + Bias(a'z) d\rho \\ &= \int_{\Gamma} a' V(\gamma) a + [g^*(\gamma) - \alpha' \mathcal{E}_{\gamma}(B'z)]^2 d\rho(\gamma) \end{aligned}$$

By integrating the first term and applying Lemma 1 to the second, one obtains

$$= a' \bar{V} a + (\alpha - \bar{\alpha})' H (\alpha - \bar{\alpha}) + \int (g^* - \bar{g})^2 d\rho . \quad \square$$

Theorem 1. In the class of estimators of the form $a'z$ with a in G , that choice of a which minimizes

$$AMSE(a'z) = \int_{\Gamma} [a'z - g^*(\gamma)]^2 d\rho(\gamma)$$

is

$$\hat{a} = B(B' \bar{V} B + H)^{-1} h .$$

An alternative form, for computations, is

$$\hat{a} = B \left[B' \int_{\Gamma} \mathcal{E}_{\gamma}(zz') d\rho(\gamma) B \right]^{-1} B' \int_{\Gamma} g^*(\gamma) \mathcal{E}_{\gamma}(z) d\rho(\gamma)$$

Proof: Rearranging the terms of the expression in Lemma 2,

$$AMSE(\alpha' B' z) = \alpha' (B' \bar{V} B + H) \alpha - 2\alpha' H \bar{\alpha} + \text{const.}$$

This is a quadratic equation in α with stationary point

$$\hat{\alpha} = (B' \bar{V} B + H)^{-1} H \bar{\alpha}$$

and positive semi-definite Hessian matrix $B' \bar{V} B + H$. Note that $H \bar{\alpha} = h$ to obtain the first form given in the theorem. The second form is obtained by verifying these equalities algebraically:

$$H = B' \int \mathcal{E}(z) \mathcal{E}'(z) d\rho B, \quad ,$$

$$B' \bar{V} B = B' \int \mathcal{E}(zz') - \mathcal{E}(z) \mathcal{E}'(z) d\rho B, \quad ,$$

and

$$h = B' \int g^* \mathcal{E}(z) d\rho. \quad \square$$

4. Bayesian Interpretation of the Estimator

The problem of estimating $g^*(\gamma)$ may be recast in a Bayesian setting by regarding ρ as the prior distribution on Γ and $n(y|f(\theta), \sigma^2)$ as the conditional distribution of y given γ . Take for the Bayes' rule the mean of $g^*(\gamma)$ with respect to the posterior density, denoted by $\tilde{g}(y)$. This choice minimizes expected posterior loss for a square error loss function.

For a variety of purposes it may be convenient to approximate the Bayes' rule by truncating a series expansion of $\tilde{g}(y)$. Consider a Fourier series expansion in terms of a set of basis functions $\{Z_i(y)\}_{i=1}^{\infty}$. The most natural choice of a weight function, in the present context, is the marginal distribution of the observations

$$p(y) = \int_{\Gamma} n[y|f(\theta), \sigma^2] d\rho(\gamma).$$

Let $\{\phi_i(y)\}_{i=1}^{\infty}$ be a sequence of orthonormal basis functions with respect to $p(y)$ generated from the sequence $\{Z_i(y)\}_{i=1}^{\infty}$ by the Gram-Schmidt process (with

linear dependencies in the Z_i accommodated by deletion). The Fourier series expansion of the Bayes rule is

$$\tilde{g}(y) = \sum_{i=1}^{\infty} c_i \phi_i(y)$$

with Fourier coefficients

$$c_i = \int_{\mathcal{Y}} \tilde{g}(y) \phi_i(y) p(y) dy .$$

The main result of this section is that if this expansion is truncated at m then ^{1/}

$$\hat{a}'Z(y) = \sum_{i=1}^m c_i \phi_i(y)$$

provided $G = R^m$. This result is a corollary of the more general result, Theorem 2, that $\hat{a}'Z(y)$ is the best approximation of $\tilde{g}(y)$ with respect to the norm

$$\|g\| = \left[\int_{\mathcal{Y}} g^2(y) p(y) dy \right]^{\frac{1}{2}}$$

in the class of functions of the form $a'Z(y)$ with a in G ; see Hewlitt and Stromberg (1965, Theorem 16.16) .

This representation of the estimator as a Fourier series expansion of the Bayes' rule is helpful because the statistical behavior $\hat{a}'z$ corresponding to a choice of weighting measure p may be anticipated from that of the Bayes' rule.

The verification of these facts follows; the remainder of this section may be omitted at first reading without affecting continuity. The reference for notation and terminology is (Hewlitt and Stromberg 1965, Ch. IV, Sec. 16).

Define the measure μ on the Lebesgue subsets of $\mathcal{Y} \times \Gamma$ by

$$\mu(A) = \int_{\Gamma} \int_{\mathcal{Y}} I_A(y, \gamma) n[y|f(\theta), \sigma^2] dy d\phi(\gamma)$$

where $\mathcal{Y}$ is n -dimensional Euclidean space and $I_A(y, \gamma)$ denotes the indicator function of the set A . Consider the collection of measurable, square integrable functions with respect to this (probability) measure denoted by $\mathcal{L}_2(\mathcal{Y} \times \Gamma, \mu)$.

The inner product between g_1, g_2 in $\mathcal{L}_2(\mathcal{Y} \times \Gamma, \mu)$ is

$$\langle g_1, g_2 \rangle = \int_{\Gamma} \int_{\mathcal{Y}} g_1(y, \gamma) g_2(y, \gamma) d\mu(y, \gamma)$$

and the corresponding norm is

$$\|g\| = [\langle g, g \rangle]^{\frac{1}{2}} = \left[\int_{\Gamma} \int_{\mathcal{Y}} g^2(y, \gamma) d\mu(y, \gamma) \right]^{\frac{1}{2}} .$$

These definitions comprise a Hilbert space on the collection of measurable, square integrable functions with argument (y, γ) (Hewlitt and Stromberg 1965, Eg. 16.8).

An estimator, $\hat{g}(y)$ is a measurable, square integrable function of y only, that is a member of $\mathcal{L}_2(\mathcal{Y} \times \Gamma, \mu)$ which is constant with respect to variation in γ . The average mean square error of an estimator is

$$\text{AMSE} [\hat{g}(y)] = \|\hat{g} - g^*\|^2 ,$$

the square of the distance between the estimator $\hat{g}(y)$ and the parametric function $g^*(\gamma)$. The mean of $g^*(\gamma)$ with respect to the posterior density

$$\tilde{g}(y) = \int_{\Gamma} g^*(\gamma) n[y|\mathbf{f}(\theta), \sigma^2] d\gamma / p(y) ,$$

the Bayes' rule, is the orthogonal projection of the parametric function $g^*(\gamma)$ into the collection of estimators with respect to the inner product $\langle, \rangle$.

That is, the Bayes' rule satisfies the Pythagorean identity

$$\|\hat{g} - g^*\|^2 = \|\hat{g} - \tilde{g}\|^2 + \|\tilde{g} - g^*\|^2 .$$

Applying this identity to functions of the form $a'Z(y)$ with a in G the equation

$$\text{AMSE} [a'Z(y)] = \|a'z - \tilde{g}\|^2 + \|\tilde{g} - g^*\|^2$$

is obtained. Since the second term does not vary with a it follows that an

attempt to minimize $AMSE [a'Z(y)]$ by varying a represents an attempt to minimize $\|a'z - \tilde{g}\|$.

Note that neither $a'Z(y)$ nor $\tilde{g}(y)$ vary with γ . The dimension of the latter minimization problem may, therefore, be reduced and attention may be restricted to measurable functions on $\mathcal{Y}$ which are square integrable with respect to the measure ν defined on the Lebesgue subsets of $\mathcal{Y}$ by

$$\nu(A) = \int_{\Gamma} \int_{\mathcal{Y}} I_A(y) d\mu(y, \gamma)$$

That is, to minimize $\|a'z - g^*\|$ with respect to the norm on $\mathcal{L}_2(\mathcal{Y} \times \Gamma, \mu)$ it suffices to find a in G which minimizes $\|a'z - \tilde{g}\|$ with respect to the norm on $\mathcal{L}_2(\mathcal{Y}, \nu)$. We have, by Theorem 1, that the choice

$$\hat{a} = B[B' \int_{\Gamma} \mathcal{E}_{\gamma}(zz') d\rho(\gamma) B]^{-1} B' \int_{\Gamma} g^*(\gamma) \mathcal{E}_{\gamma}(z) d\rho(\gamma)$$

minimizes $AMSE [a'Z(y)]$ whence this choice is the solution of both of these minimization problems. For computations, we have by Fubini's theorem that for integrable $g(y)$

$$\begin{aligned} \int_{\mathcal{Y}} g(y) d\nu(y) &= \int_{\Gamma} \int_{\mathcal{Y}} g(y) n(y|f(\theta), \sigma^2) dy d\rho(\gamma) \\ &= \int_{\mathcal{Y}} g(y) \int_{\Gamma} n(y|f(\theta), \sigma^2) d\rho(\gamma) dy \\ &= \int_{\mathcal{Y}} g(y) p(y) dy . \end{aligned}$$

The foregoing discussion is summarized as Theorem 2.

Theorem 2. Let $\tilde{g}(y)$ be the Bayes' rule which minimizes expected posterior square error loss when estimating $g^*(\gamma)$ with prior measure ρ on Γ . Let $p(y)$ be the marginal density of the observations,

$$p(y) = \int_{\Gamma} n[y|f(\theta), \sigma^2] d\rho(\gamma) .$$

The estimator $\hat{g}(y)$ of Theorem 1 is the best approximation of the Bayes' rule in the class of estimators of the form $a'Z(y)$ with a in G in the sense of minimizing the distance

$$\|a'z - \tilde{g}\| = \left[\int_{\mathcal{Y}} [a'Z(y) - \tilde{g}(y)]^2 p(y) dy \right]^{\frac{1}{2}} .$$

5. Example

Consider the statistical model

$$y_t = f(x_t, \theta) + e_t$$

with response function

$$f(x, \theta) = \begin{cases} \theta_1(e^{-x\theta_2} - e^{-x\theta_1})/(\theta_1 - \theta_2) & \theta_1 \neq \theta_2 , \\ \theta_1 x e^{-x\theta_1} & \theta_1 = \theta_2 , \end{cases}$$

parameter space

$$\Theta = \{(\theta_1, \theta_2): \theta_1 \geq \theta_2\} ,$$

and design

$$\{x_t\} = \{.25, .5, 1, 1.5, 2, 4, .25, .5, 1, 1.5, 2, 4\} .$$

This model has appeared several times in the nonlinear regression literature (Box and Lucas 1959; Guttman and Meeter 1964; Gallant 1976) and its use here will facilitate comparison with standard methods of nonlinear regression analysis.

The response function $f(x, \theta)$ describes the relative yield at time x of an intermediate substance in a chemical reaction; consider estimation of the time at which the maximum yield occurs:

$$g^*(\theta) = (\theta_1 - \theta_2)^{-1} \ln(\theta_1/\theta_2) .$$

Some additional notation required is: let $F(\theta)$ be the n by p matrix with typical element $(\partial/\partial\theta_j) f(x_t, \theta)$ where t is the row index and put

$$C = [F'(\theta)F(\theta)]^{-1} .$$

To obtain the explicit estimator of the time of maximum yield, the weighting measure ρ is taken as the product measure comprised of a uniform distribution on the ellipsoid $(\theta - \tau)'C^{-1}(\theta - \tau) \leq r^2$ with C evaluated at τ and the inverted gamma distribution (Zellner 1971, p. 371) with parameters v and s^2 . Appropriate transformations yield an expression which is convenient for use with quadrature formulae:

$$\int g(\theta, \sigma) d\rho(\gamma) \\ = [\pi r^2 \Gamma(\alpha)]^{-1} \int_0^\infty \mu^{\alpha-1} \left\{ \int_{y' y \leq r^2} g[R'y + \tau, 1/\sqrt{(\gamma\mu)}] dy \right\} e^{-\mu} d\mu .$$

where $\alpha = v/2$, $\gamma = 2/(vs^2)$, and R is obtained by factoring C as $C = R'R$.

The integral within braces may be evaluated by means of a 12-point quadrature formula (Stroud 1971, p. 281)^{2/} which integrates polynomials in y up to degree seven exactly; the outer integral may be evaluated using a 4-point quadrature formula (IBM 1968, DQL4) which integrates polynomials in μ up to degree seven exactly. In terms of the arguments of $g(\theta, \sigma)$, if g is a polynomial of degree less than or equal to seven in θ and degree of degree less than or equal to $v-2$ in σ it will be integrated exactly.

The parametric choices employed here are $\tau = (1.4, .4)$, $s = .025$, $v = 10$, and $r^2 = 8 s^2$. These choices were motivated by the study of the sampling distribution of the least squares estimator reported in (Gallant 1976). With these choices, the numerical integration procedure above is equivalent to the substitution of a discrete prior measure $\hat{\rho}$,

$$\int g(\theta, \sigma) d\hat{\rho}(y) = \sum_i p_i g(\theta_i, \sigma_i) ,$$

for the measure ρ . This discrete measure is displayed in Table 1. In fact, it may be best to regard $\hat{\rho}$ as the prior in the sequel.

Table 1 here

Two explicit estimators are considered, a linear function^{3/} in y

$$\hat{g}_1(y) = a_0 + \sum_{i=1}^{12} a_i y_i$$

and a quadratic function in y

$$\hat{g}_2(y) = a_0 + \sum_{i=1}^{12} a_i y_i + \sum_{i \leq j}^{12} a_{ij} y_i y_j .$$

The performance of these estimators relative to the Bayes' estimator and the maximum likelihood estimator is shown in Table 2. The table gives the value of the parametric function, the bias and the root mean square error for each estimator at each parameter value of the quadratic formula together with the root average mean square error for each estimator.

Table 2 here

These conclusions obtain from Table 2 according to the criterion of average mean square error. The slight improvement in the performance of the explicit quadratic estimator $\hat{g}_2$ relative to the explicit linear estimator $\hat{g}_1$ is not worth the added complexity entailed. The explicit linear estimator $\hat{g}_1$ is a serious competitor of the Bayes' estimator in view of its simplicity. The explicit linear estimator is improvement over the maximum likelihood estimator.

The details of the computations are as follows. The moments of the explicit estimators were obtained exactly using the usual formulae for the moments of linear and quadratic functions of normally distributed random variables. The moments of the Bayes' estimator were obtained from 2000 Monte-Carlo trials using the control variate method of variance reduction (Hammersly and Handscomb 1964, p. 59-60). The random number generator employed was GGNOF (IMSL 1975); the explicit linear estimator was used as the control variate. The estimated standard error of $\sqrt{\text{AMSE}}$ for the Bayes' estimator was .000034695. The moments of the maximum likelihood estimator were approximated using the moments of the asymptotic distribution. As a partial check on accuracy the moments of $g(\hat{\theta})$ were estimated for $\gamma = (1.4, .4, .025)$ from an existing set of 4000 simulated values of the least squares estimator $\hat{\theta}$ (Gallant 1976); the results were $\text{Est}\{\mathcal{E}[g(\hat{\theta})]\} = 1.25283$ and $\text{Est}\{\text{Var}[g(\hat{\theta})]\} = .00135907$ yielding $\text{Est}\{\sqrt{\text{MSE}}[g(\hat{\theta})]\} = .03687$ which compares favorably with .03670, the value reported in Table 2. All code was written in double precision, save the use of GGNOF, using an IBM 370/165.

6. Confidence Intervals

Any procedure for finding an exact confidence region for γ may be used to find an exact confidence region for $g^*(\gamma)$. Let $\hat{\Gamma}$ be a random set function depending on y such that

$$P_{\gamma}(\{y: \gamma \text{ in } \hat{\Gamma}\}) \geq 1 - \alpha$$

Let $\hat{G} = \{g^*(\gamma): \gamma \text{ in } \hat{\Gamma}\}$; since

$$\{y: g^*(\gamma) \text{ in } \hat{G}\} \supset \{y: \gamma \text{ in } \hat{\Gamma}\}$$

it follows that

$$P_Y(\{y: g^*(y) \text{ in } \hat{G}\}) \geq 1 - \alpha \quad .$$

Thus $\hat{G}$ is an exact confidence interval for $g^*(y)$ - exact in the sense that the probability statement is correct, not approximate.

When g^* does not depend on σ^2 , as in the example, consider

$$U = a_0 + A'y$$

where a_0 is a q -vector of known constants and A is an n by q matrix of known constants. Then, following Hartley (1964),

$$\hat{\Theta} = \{\theta : [U - a_0 - A'f(\theta)](A'A)^{-1}[U - a_0 - A'f(\theta)]/[q s^2(\theta)] \leq F_\alpha\}$$

where

$$s^2(\theta) = [y - f(\theta)]'[I - A(A'A)^{-1}A'] [y - f(\theta)]/(n-q)$$

is an exact confidence region for θ ; F_α denotes the upper α percentage point of the F distribution with q degrees freedom for the numerator and $n-q$ degrees freedom for the denominator. An exact confidence interval for $g^*(\theta)$ is

$$\hat{G} = \{g^*(\theta): \theta \text{ in } \hat{\Theta}\} \quad .$$

To illustrate with the example, consider the sample

$$y = \begin{bmatrix} 0.31753 \\ 0.42208 \\ 0.62973 \\ 0.56630 \\ 0.54830 \\ 0.26603 \\ 0.29474 \\ 0.49830 \\ 0.58632 \\ 0.63670 \\ 0.56983 \\ 0.26299 \end{bmatrix}$$

generated according to the model with $\theta = (1.4, .4)$ and $\sigma = .025$ (corresponding

to $i = 49$ in Tables 1 and 2). For a_0 and A_1 take:^{4/}

$$a_0 = \begin{bmatrix} 1.44126 \\ 0.19376 \end{bmatrix} ,$$

$$A = \begin{bmatrix} -0.27126 & 0.45343 \\ -0.33201 & 0.58293 \\ -0.18570 & 0.42717 \\ 0.035484 & 0.14643 \\ 0.21061 & -0.077778 \\ 0.38468 & -0.30496 \\ -0.27126 & 0.45343 \\ -0.33201 & 0.58293 \\ -0.18570 & 0.42717 \\ 0.035484 & 0.14653 \\ 0.21061 & -0.077778 \\ 0.38468 & -0.30496 \end{bmatrix}$$

With these choices, the first element of $U = a_0 + A'y$ is the best linear average mean square error estimator of $g^*(\theta)$ and the second element of U is the best linear average mean square error of θ_1 .

The estimate of $g^*(\theta)$ is the value 1.2255 observed for U_1 and the exact confidence interval is

$$\hat{G} = (1.0730, 1.3138) .$$

By way of comparison, the maximum likelihood estimate of $g^*(\theta)$ is 1.2077 and a confidence interval obtained from the asymptotic theory of the least squares estimator is

$$(1.1206, 1.2948)$$

which is approximate, not exact. The true value of $g^*(\theta)$ is 1.2528 .

This procedure for finding a confidence interval is entirely analagous to Scheffe's method for finding multiple confidence intervals which are simultaneously valid. The parametric confidence statement " θ in $\hat{\Theta}$ " may be used for several different functions of θ yielding a system of confidence intervals which are simultaneously valid. As with Scheffe's method, the procedure is conservative.

In this instance, the interval was 38% longer than that found using the asymptotic theory of the maximum likelihood estimator of $g^*(\theta)$.

The first attempt to find an exact confidence interval for this sample used the statistic $[U_1 - a_{01} - A_1' f(\theta)]'(A_1' A_1)^{-1}[U_1 - a_{01} - A_1' f(\theta)]/s_1^2(\theta)$ to find the preliminary region $\hat{\Theta}$. However, the region was constrained on the left only by the parametric constraint $\theta_2 \leq \theta_1$ and appeared to continue indefinitely to the right. The addition of the best linear average mean square error estimator of θ_1 as a second component of U remedied the problem and produced a small elliptically shaped region for $\hat{\Theta}$.

7. Discussion

It is seen that linear estimators for use in nonlinear regression analysis may be found by truncating a Fourier series expansion of a Bayes' rule. And that the linearity of these estimators may be exploited to find exact confidence intervals for a parametric function of interest. If the example is not misleading, the estimator is competitive with both the Bayes' estimator and the maximum likelihood estimator according to the criterion of mean square error.

It is by chance, not by design, should the example prove misleading. The example itself was taken directly from Guttman and Meeter (1964) so as to facilitate comparison. The parameters of the prior were not chosen iteratively so as to influence the final results; their values were set for the reasons stated and, once chosen, were not altered. The sample used to illustrate the confidence interval procedure was the only one considered.

The most serious deterrent to use of the explicit estimator in applications is the requirement that the user produce a prior. However, in nonlinear regression analysis one must usually have some knowledge of the situation in order to find starting values for computing the least squares estimator. But the fact that

different investigators will obtain different results depending on the subjective choice of a prior when all else remains the same is disturbing . On the other hand, one can envisage an industrial application involving repetitive estimations for the purpose of calibration where the simplicity of a linear estimator would be an overwhelming advantage. The estimator may find use in such contexts.

It is possible to develop a formal theory of unbiased estimation by borrowing the ideas of estimability from linear regression and proceeding along lines roughly analogous to those used here. In examples, there was the problem that unbiased estimators of the form $a'z$ usually did not exist and another that, when they did, they depended on only one or two of the n observations. This did not appear to be a fruitful line of inquiry and was abandoned.

Footnotes

- 1/ If there are deletions then the expansion is truncated at m less the number of deletions.
- 2/ Read $B_2 = [(551 + 41 \sqrt{29})/6264]V$ to correct an error in their formula.
- 3/ See the next section for the values of a_i .
- 4/ Note that u is a function of the sufficient statistic

$$[y_1^+ y_7, y_2^+ y_8, y_3^+ y_9, y_4^+ y_{10}, y_5^+ y_{11}, y_6^+ y_{12}, \sum_{i=1}^6 (y_1 - y_{i+6})^2]$$

1. Quadrature Formula

i	Parameter Value			Weight
	θ_{1i}	θ_{2i}	σ_i	P_i
1	1.52972	0.40788	0.01824	0.01297
2	1.27028	0.39212	0.01824	0.01297
3	1.40000	0.43339	0.01824	0.01297
4	1.40000	0.36661	0.01824	0.01297
5	1.44837	0.41539	0.01824	0.02157
6	1.35163	0.40951	0.01824	0.02157
7	1.44837	0.39049	0.01824	0.02157
8	1.35163	0.38461	0.01824	0.02157
9	1.49649	0.43070	0.01824	0.00923
10	1.30351	0.41897	0.01824	0.00923
11	1.49649	0.38103	0.01824	0.00923
12	1.30351	0.36930	0.01824	0.00923
13	1.52972	0.40788	0.02625	0.05084
14	1.27028	0.39212	0.02625	0.05084
15	1.40000	0.43339	0.02625	0.05084
16	1.40000	0.36661	0.02625	0.05084
17	1.44837	0.41539	0.02625	0.08456
18	1.35163	0.40951	0.02625	0.08456
19	1.44837	0.39049	0.02625	0.08456
20	1.35163	0.38461	0.02625	0.08456
21	1.49649	0.43070	0.02625	0.03618
22	1.30351	0.41897	0.02625	0.03618
23	1.49649	0.38103	0.02625	0.03618
24	1.30351	0.36930	0.02625	0.03618
25	1.52972	0.40788	0.04231	0.01025
26	1.27028	0.39212	0.04231	0.01025
27	1.40000	0.43339	0.04231	0.01025
28	1.40000	0.36661	0.04231	0.01025
29	1.44837	0.41539	0.04231	0.01704
30	1.35163	0.40951	0.04231	0.01704
31	1.44837	0.39049	0.04231	0.01704
32	1.35163	0.38461	0.04231	0.01704
33	1.49649	0.43070	0.04231	0.00729
34	1.30351	0.41897	0.04231	0.00729
35	1.49649	0.38103	0.04231	0.00729
36	1.30351	0.36930	0.04231	0.00729
37	1.52972	0.40788	0.09843	0.00002
38	1.27028	0.39212	0.09843	0.00002
39	1.40000	0.43339	0.09843	0.00002
40	1.40000	0.36661	0.09843	0.00002
41	1.44837	0.41539	0.09843	0.00003
42	1.35163	0.40951	0.09843	0.00003
43	1.44837	0.39049	0.09843	0.00003
44	1.35163	0.38461	0.09843	0.00003
45	1.49649	0.43070	0.09843	0.00001
46	1.30351	0.41897	0.09843	0.00001
47	1.49649	0.38103	0.09843	0.00001
48	1.30351	0.36930	0.09843	0.00001
49	1.40000	0.40000	0.02500	0.0

2. Bias and Root Mean Square Error of the Explicit Estimator, Bayes' Estimator, and the Maximum Likelihood Estimator of the Time of Maximum Yield

i	$g^*(\theta_i)$	Explicit Estimators							
		Linear		Quadratic		Bayes' Estimator		MLE	
		Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$	Bias	$\sqrt{MSE}$
1	1.17830	0.02820	0.03270	0.02903	0.03300	0.02040	0.02496	0.0	0.02480
2	1.33851	-0.03340	0.03727	-0.03247	0.03666	-0.02239	0.02801	0.0	0.02916
3	1.21310	0.01704	0.02376	0.01656	0.02329	0.01151	0.02161	0.0	0.02613
4	1.29663	-0.01514	0.02243	-0.01559	0.02255	-0.01149	0.02163	0.0	0.02753
5	1.20910	0.01785	0.02434	0.01749	0.02380	0.00704	0.02027	0.0	0.02575
6	1.26747	-0.00395	0.01702	-0.00448	0.01738	-0.00194	0.02289	0.0	0.02735
7	1.23907	0.00678	0.01789	0.00632	0.01744	0.00263	0.02127	0.0	0.02626
8	1.29970	-0.01686	0.02363	-0.01729	0.02419	-0.00726	0.02199	0.0	0.02789
9	1.16859	0.03204	0.03607	0.03250	0.03614	0.02449	0.02741	0.0	0.02481
10	1.28317	-0.01000	0.01934	-0.00994	0.01977	-0.00710	0.02002	0.0	0.02798
11	1.22640	0.01153	0.02017	0.01166	0.01981	0.00898	0.01908	0.0	0.02579
12	1.35002	-0.03789	0.04135	-0.03735	0.04118	-0.02753	0.03112	0.0	0.02909
13	1.17830	0.02820	0.03692	0.02911	0.03684	0.02494	0.03309	0.0	0.03569
14	1.33851	-0.03340	0.04102	-0.03239	0.04099	-0.03007	0.03695	0.0	0.04196
15	1.21310	0.01704	0.02929	0.01664	0.02885	0.01520	0.02845	0.0	0.03761
16	1.29663	-0.01514	0.02822	-0.01552	0.02861	-0.01471	0.02802	0.0	0.03962
17	1.20910	0.01785	0.02977	0.01756	0.02913	0.01259	0.02806	0.0	0.03706
18	1.26747	-0.00395	0.02415	-0.00440	0.02457	-0.00321	0.02926	0.0	0.03936
19	1.23907	0.00678	0.02477	0.00640	0.02426	0.00469	0.02790	0.0	0.03779
20	1.29970	-0.01686	0.02918	-0.01721	0.02982	-0.01211	0.02969	0.0	0.04013
21	1.16859	0.03204	0.03993	0.03258	0.03973	0.03121	0.03571	0.0	0.03571
22	1.28317	-0.01000	0.02584	-0.00986	0.02649	-0.00914	0.02649	0.0	0.04027
23	1.22640	0.01153	0.02647	0.01174	0.02586	0.01165	0.02563	0.0	0.03712
24	1.35002	-0.03789	0.04476	-0.03727	0.04486	-0.03476	0.04018	0.0	0.04187
25	1.17830	0.02820	0.04765	0.02934	0.04677	0.04050	0.04787	0.0	0.05754
26	1.33851	-0.03340	0.05089	-0.03216	0.05172	-0.04443	0.05237	0.0	0.06754
27	1.21310	0.01704	0.04201	0.01687	0.04159	0.02285	0.03675	0.0	0.06062
28	1.29663	-0.01514	0.04128	-0.01528	0.04166	-0.02189	0.03600	0.0	0.06388
29	1.20910	0.01785	0.04235	0.01780	0.04148	0.02273	0.03812	0.0	0.05975
30	1.26747	-0.00395	0.03860	-0.00417	0.03920	-0.00561	0.03260	0.0	0.06345
31	1.23907	0.00678	0.03899	0.00663	0.03831	0.00851	0.03273	0.0	0.06092
32	1.29970	-0.01686	0.04194	-0.01697	0.04278	-0.02180	0.03849	0.0	0.06469
33	1.16859	0.03204	0.05001	0.03281	0.04921	0.04696	0.05255	0.0	0.05756
34	1.28317	-0.01000	0.03968	-0.00963	0.04080	-0.01437	0.03233	0.0	0.04491
35	1.22640	0.01153	0.04010	0.01197	0.03904	0.01640	0.03301	0.0	0.05984
36	1.35002	-0.03789	0.05395	-0.03704	0.05470	-0.05235	0.05754	0.0	0.06749
37	1.17830	0.02820	0.09368	0.03103	0.09044	0.06311	0.06893	0.0	0.13386
38	1.33851	-0.03340	0.09538	-0.03047	0.09924	-0.06961	0.07975	0.0	0.15736
39	1.21310	0.01704	0.09095	0.01856	0.09057	0.03653	0.04546	0.0	0.14104
40	1.29663	-0.01514	0.09061	-0.01360	0.09138	-0.03638	0.04126	0.0	0.14860
41	1.20910	0.01785	0.09110	0.01948	0.08953	0.03933	0.04596	0.0	0.13900
42	1.26747	-0.00395	0.08943	-0.00248	0.09092	-0.00660	0.02791	0.0	0.14761
43	1.23907	0.00678	0.08959	0.00831	0.08838	0.01261	0.02662	0.0	0.14172
44	1.29970	-0.01686	0.09091	-0.01529	0.09292	-0.03781	0.04393	0.0	0.15051
45	1.16859	0.03204	0.09491	0.03449	0.09222	0.07476	0.07675	0.0	0.13392
46	1.28317	-0.01000	0.08990	-0.00794	0.09278	-0.02217	0.02464	0.0	0.15101
47	1.22640	0.01153	0.09008	0.01365	0.08773	0.02529	0.03212	0.0	0.13921
48	1.35002	-0.03789	0.09704	-0.03535	0.10029	-0.08423	0.08419	0.0	0.15702
49	1.25276	0.00163	0.02275	0.00100	0.02267	0.00072	0.02915	0.0	0.03670
$\sqrt{AMSE}$:		0.03237		0.03234		0.03095		0.04122	

References

- Box, G. E. P. and Lucas, H. L. (1959) "Design of Experiments in Non-Linear Situations", Biometrika, 46, 77-90 .
- Gallant, A. Ronald (1976). "Confidence Regions for the Parameters of a Nonlinear Regression Model", Institute of Statistics Mimeograph Series, No. 1077, Raleigh: North Carolina State University
- Guttman, Irwin and Meeter, Duane A. (1964) "On Beale's Measures of Non-Linearity", Technometrics, 7, 623-637.
- Hammersly, J. M. and Handscomb, D. C. (1964), Monte-Carlo Methods, New York: John Wiley and Sons, Inc.
- Hartley, H. O. (1964) "Exact Confidence Regions for the Parameters in Non-Linear Regression Laws", Biometrika, 51, 347-353.
- Hewitt, Edwin and Stromberg, Karl (1965), Real and Abstract Analysis, New York: Springer-Verlag Inc.
- I. B. M. Corporation (1968), System 360 Scientific Subroutine Package, Version III, White Plains, New York: Technical Publications Department.
- I.M.S.L. (1975), IMSL Library 1, Fifth Ed., Houston, Texas: International Mathematical and Statistical Libraries, Inc.
- Stroud, A. H. (1971), Approximate Calculation of Multiple Integrals, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Zellner, Arnold (1971), An Introduction to Bayesian Inference in Econometrics, New York: John Wiley and Sons, Inc.