THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

THE THEORY OF NONLINEAR REGRESSION AS IT RELATES TO

SEGMENTED POLYNOMIAL REGRESSIONS WITH ESTIMATED JOIN POINTS

by

A. Ronald Gallant

ABSTRACT

This report is a summary of the topics in nonlinear regression which are particularly relevant in applications of segmented polynomial regressions with unknown join points. These are nonlinear regression models whose distinguishing characteristic is that the partial derivatives of the response function with respect to the parameters to be estimated may not exist.

The regularity conditions usually assumed in nonlinear regression theory to obtain convergence of the modified Gauss-Newton method of computing least squares parameter estimators, the asymptotic normality of these estimators, and the asymptotic distribution of Likelihood Ratio test statistics for hypotheses of parameter location are reviewed. The extent to which the derivative conditions can be weakened to accommodate segmented polynomial regressions constrained to be once continuously differentiable with respect to the input variable is discussed.

Hartley's method of testing hypotheses of the location of subsets of the parameters which enter a response function nonlinearly is reviewed. This method does not require the imposition of regularity conditions involving derivatives and its use in segmented polynomial regressions which violate the derivative conditions is discussed.

The topics presented are illustrated with examples taken from applications.

THE THEORY OF NONLINEAR REGRESSION AS IT RELATES TO

SEGMENTED POLYNOMIAL REGRESSIONS WITH ESTIMATED JOIN POINTS

by

A. Ronald Gallant[*]

[*] Assistant Professor of Statistics and Economics, Institute of Statistics,
North Carolina State University, Raleigh, North Carolina 27607

1. INTRODUCTION

This report summarizes the theory of nonlinear regression which is particularly relevant in statistical applications of segmented polynomial approximating functions where the join points are unknown and must be estimated from the data. A formal statement of the theoretical results cited is included as an Appendix.

The nonlinear parameters appearing in such models are, of course, the unknown abscissae where one polynomial submodel stops and the next polynomial submodel starts; the remaining coefficients of the model enter in a linear fashion. In many data analysis situations the join points are known a priori or are of no particular interest in the analysis (Fuller, 1969). In the former case the model is linear subject known restrictions on the parameters (referred to as B_0, B_1, ... in the later sections) and in the latter case the same is true provided one is content to treat his visual estimates of the join points as a priori knowledge. In these situations, Section 2 may be of interest as a means to put the model in proper form for standard multiple regression computing programs. On the other hand, one can avoid reparameterization in these situations by employing the numerical methods given in Gerig and Gallant (1974). This algorithm is available as a SAS procedure upon request from the author.

2. REPARAMETERIZATION OF SEGMENTED POLYNOMIAL REGRESSIONS

It will be convenient in the later sections to be able to express or represent a segmented polynomial regression according to the specification commonly employed in nonlinear regression theory: A set of responses y_t to inputs $\underset{\sim}{x}_{y_t}$ are generated according to the regression equations

$$y_t = f(\underset{\sim}{x}_t, \underset{\sim}{\theta}^*) + e_t \qquad (t = 1, 2, \ldots, n)$$

where the errors e_t are independently and normally distributed with mean zero and unknown variance σ^2. The response function $f(\underset{\sim}{x}, \underset{\sim}{\theta})$ has a known functional form depending on k-dimensional input vectors $\underset{\sim}{x}$ contained in a known set $\underset{\sim}{X}$ and a p-dimensional parameter $\underset{\sim}{\theta}$ contained in a known set $\underset{\sim}{\Omega}$. The true but unknown value of the parameter is denoted by $\underset{\sim}{\theta}^*$.

Segmented polynomial regressions are usually written as

$$y_t = g(x_t) + e_t$$

where $g(x)$ is the sequence of grafted polynomial submodels

$$g(x) = \begin{cases} g_1(x, \underset{\sim}{\beta}_1) & \alpha_0 \leq x \leq \alpha_1 \\ g_2(x, \underset{\sim}{\beta}_2) & \alpha_1 < x \leq \alpha_2 \\ \quad \vdots & \quad \vdots \\ g_r(x, \underset{\sim}{\beta}_r) & \alpha_{r-1} < x \leq \alpha_r \end{cases}$$

corresponding to the partitioning of the interval $[\alpha_0, \alpha_r]$ into

$$\alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{r-1} < \alpha_r \quad .$$

The endpoints α_0 and α_r are known but the intermediate points of join

$$\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{r-1})'$$

are unknown and must be estimated from the data. The submodels $g_j(x, \underset{\sim}{\beta}_j)$ $(j = 1, \ldots, r)$ consist of the polynomials

$$g_1(x, \underset{\sim}{\beta}_1) = \beta_{01} + \beta_{11}x + \beta_{21}x^2 + \ldots + \beta_{q_1 1} x^{q_1} \quad ,$$

$$g_2(x, \underset{\sim}{\beta}_2) = \beta_{02} + \beta_{12}x + \beta_{22}x^2 + \ldots + \beta_{q_2 2} x^{q_2} \quad ,$$

$$\vdots$$

$$g_r(x, \underset{\sim}{\beta}_r) = \beta_{0r} + \beta_{1r}x + \beta_{2r}x^2 + \ldots + \beta_{q_r r} x^{q_r} \quad .$$

As pointed out by Fuller (1969), it is often desirable to impose a continuity restriction on the model

$$B_0: \quad g_j(\alpha_j, \underset{\sim}{\beta}_j) = g_{j+1}(\alpha_j, \underset{\sim}{\beta}_{j+1}) \quad j = 1, 2, \ldots, r-1$$

and, in addition, require that the response functions $g(x)$ be once continuously differentiable by imposing

$$B_1: \quad \frac{\partial}{\partial x} g_j(\alpha_j, \underset{\sim}{\beta}_j) = \frac{\partial}{\partial x} g_{j+1}(\alpha_j, \underset{\sim}{\beta}_{j+1}) \quad j = 1, 2, \ldots, r-1 \ .$$

As mentioned earlier, it is helpful to be able to express such a regression model in the single equation form

$$y_t = f(\underset{\sim}{x}, \underset{\sim}{\theta}) + e_t$$

in order to apply the ideas of nonlinear regression theory. This may be done along the lines (Gallant and Fuller, 1973) summarized here.

The basic idea is to rewrite the grafted polynomial model in terms of the basis functions

$$1, \ x, \ x^2, \ \ldots, \ x^q$$

$$T_0(\alpha_i - x), \ T_1(\alpha_i - x), \ \ldots, \ T_q(\alpha_i - x) \qquad i = 1, 2, \ldots, r-1$$

where $q = \max\{q_i\}$ and

$$
T_k(z) \ = \
\begin{cases}
z^k & z \geq 0 \\
\\
0 & z < 0
\end{cases}
\ .
$$

Note that $T_0(z)$ is discontinuous, $T_1(z)$ is a continuous function, $T_2(z)$ is once continuously differentiable, $T_3(z)$ is twice continuously differentiable, and so on. Consequently, the imposition of the continuity restriction B_0 requires that the functions $T_0(\alpha_i - x)$ be deleted from the basis, imposition of the once continuously differentiable restriction B_1 requires that the functions $T_1(\alpha_i - x)$ be deleted from the basis, and so on.

As an example, the quadratic-quadratic model

$$g_1(x, \ \underset{\sim}{\beta}_1) = \beta_{01} + \beta_{11}x + \beta_{21}x^2 \qquad \alpha_0 \leq x \leq \alpha_1$$

$$g_2(x, \ \underset{\sim}{\beta}_2) = \beta_{02} + \beta_{12}x + \beta_{22}x^2 \qquad \alpha_1 < x \leq \alpha_2$$

subject to the continuity restriction

$$B_0: \quad \beta_{01} + \beta_{11}\alpha_1 + \beta_{21}\alpha_1^2 = \beta_{02} + \beta_{12}\alpha_1 + \beta_{22}\alpha_1^2$$

is written in terms of the basis functions as

$$f_1(x, \underset{\sim}{\theta}) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 T_1(\theta_6 - x) + \theta_5 T_2(\theta_6 - x)$$

where the correspondences between parameters in the two representations are

$$\theta_1 = \beta_{02}, \quad \theta_2 = \beta_{12}, \quad \theta_3 = \beta_{22}, \quad \theta_4 = \beta_{12} - \beta_{11} + 2(\beta_{22} - \beta_{21})\alpha_1, \quad \theta_5 = \beta_{22} - \beta_{21},$$

$\theta_6 = \alpha_1$. The data given in Table 1 can be represented by this model as seen from the least squares fit in Figure 1.

If the quadratic-quadratic model is restricted to be continuously differentiable by imposing

$$B_1: \quad \beta_{11} + 2\beta_{21}\alpha_1 = \beta_{12} + 2\beta_{22}\alpha_1$$

one obtains

$$f_2(x, \underset{\sim}{\theta}) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 T_2(\theta_5 - x)$$

where the parameter correspondences are

$$\theta_1 = \beta_{02}, \quad \theta_2 = \beta_{12}, \quad \theta_3 = \beta_{22}, \quad \theta_4 = \beta_{22} - \beta_{21}, \quad \theta_5 = \alpha_1 \quad .$$

The data given in Table 2 can be represented by this model as seen in Figure 2.

FIGURE 1. Quadratic-Quadratic Model Subject to B_0 and B_1 Fitted to the Data of Table 1
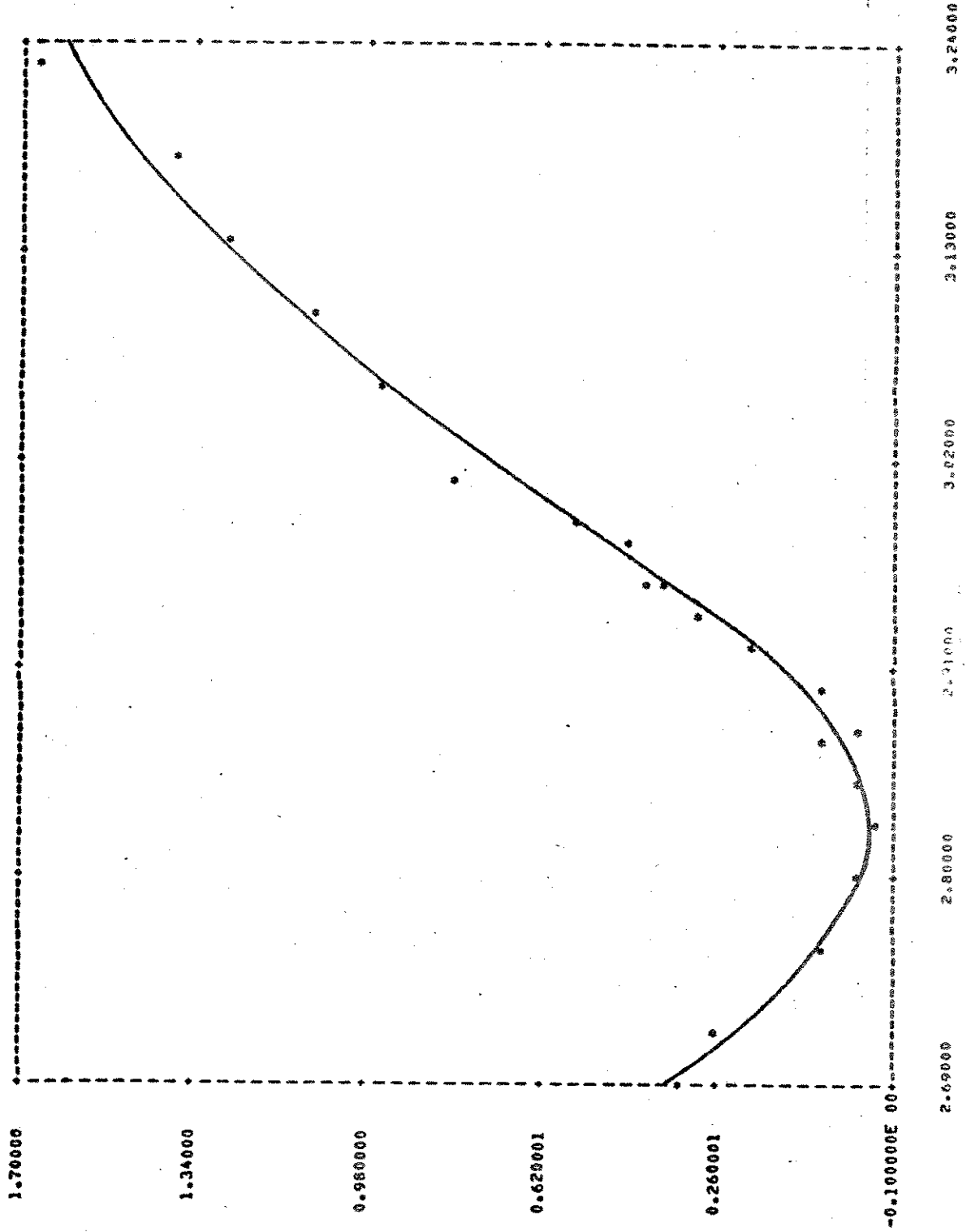
FIGURE 2. Quadratic-Quadratic Model Subject to B_0
Fitted to the Data of Table 2

TABLE 1. Specific retention volume of cycloheptene in polyethylene terephthalate

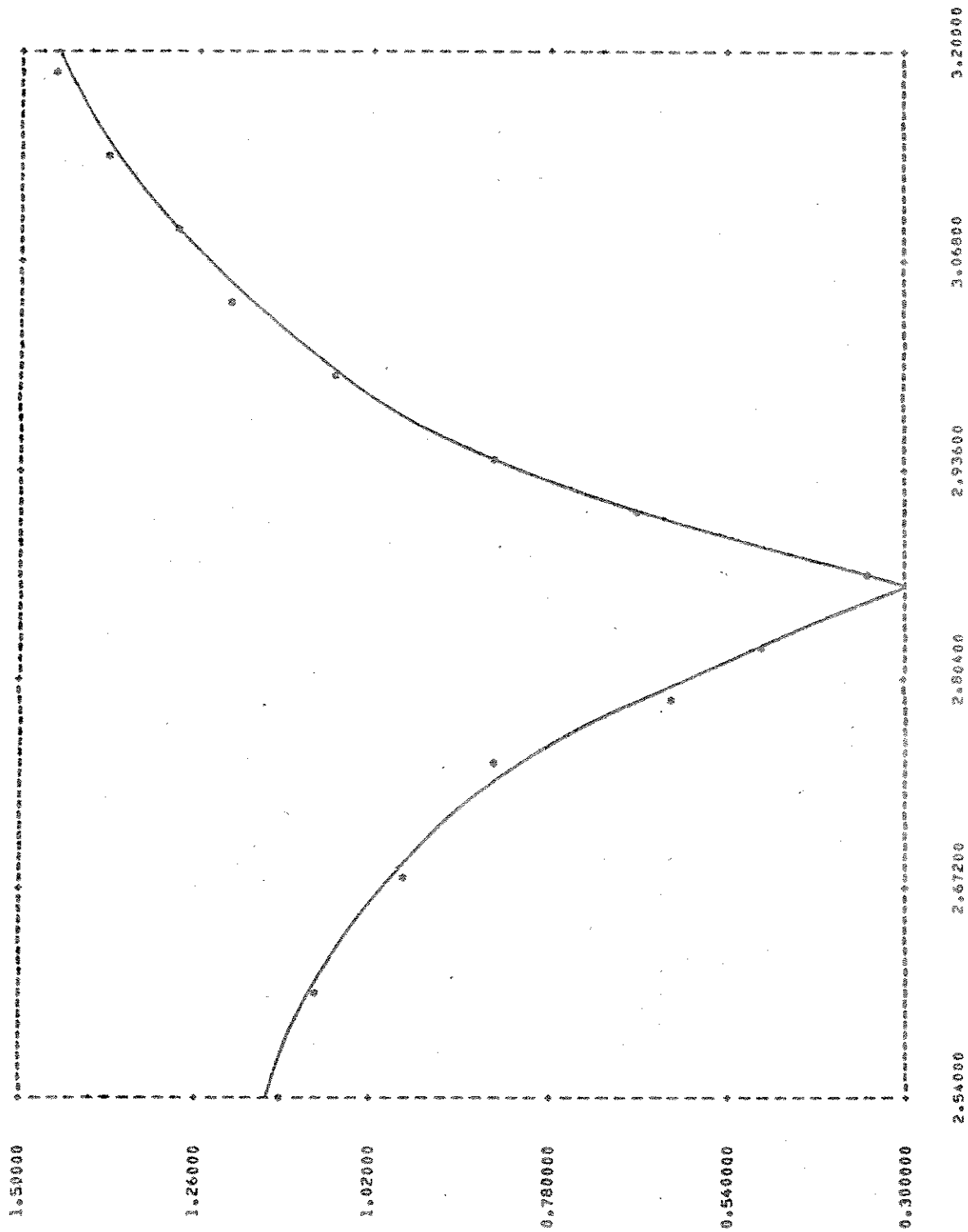| Reciprocal $\times 10^3$ of temperature in degrees Kelvin	Natural logarithm of specific volume in CC. per gm.
2.69323	0.35680
2.72182	0.27624
2.76395	0.064185
2.80269	0.00116999
2.82885	-0.038708
2.85388	-0.011101
2.87686	0.068929
2.87852	-0.013547
2.90191	0.068643
2.92568	0.20731
2.93772	0.29910
2.95420	0.38649
2.95945	0.43128
2.97885	0.44319
2.99132	0.56516
3.01386	0.80240
3.05997	0.95394
3.10173	1.11987
3.13971	1.27006
3.18471	1.41183
3.18471	1.40225
3.23310	1.66507

Source: Hsiung (1974)

TABLE 2. Specific retention volume of methylene chloride in polyethylene
terephthalate

Reciprocal $\times 10^3$ of temperature in degrees Kelvin	Natural logarithm of specific volume in CC. per gm.
2.54323	1.16323
2.60960	1.10458
2.67952	0.98832
2.75330	0.87471
2.79173	0.62060
2.82965	0.51175
2.87026	0.35371
2.91120	0.66954
2,94637	0.85555
3.00030	1.07086
3.04228	1.22272
3,09214	1.29113
3.13971	1.38480
3.19081	1.46728

Source: Hsiung (1974)

3. NONLINEAR REGRESSION THEORY

The topics in nonlinear theory which are of primary importance in applications are:

i)   the modified Gauss-Newton method for computing least squares estimators (Hartley, 1961);

ii) the consistency and asymptotic normality of the least squares estimator (Jennrich, 1969; Malinvaud, 1970; Malinvaud, 1966); and,

iii) the asymptotic null and non-null distribution of Likelihood Ratio tests for the location of parameter estimates (Gallant, 1973b; Gallant, 1974).

The results in these references are obtained subject to the requirement that the second order partial derivatives of $f(x, \theta)$ with respect to the parameters θ exist and be continuous. As we have seen from the previous discussion of grafted polynomial models, this requires the deletion of the functions $T_0(\alpha_i - x)$, $T_1(\alpha_i - x)$ and $T_2(\alpha_i - x)$ from the set of basis functions for grafted polynomial models or, equivalently, the imposition of the conditions B_0, B_1, and B_2 on the response function $g(x)$.

In Gallant (1973a), a set of assumptions are listed which are sufficient to obtain these results and allow the restoration of the functions $T_2(\alpha_i - x)$ to the set of basis functions or, equivalently, the elimination of the need to impose the restriction B_2 on the response function $g(x)$. Subject to choosing one's inputs as "constant in repeated samples" (Theil, 1971, p. 364), it is verified in Gallant (1971) that the quadratic-quadratic model subject to B_0 and B_1,

$$f_2(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 T_2(\theta_5 - x) ,$$

satisfies this set of assumptions and the additional assumptions required to obtain the convergence of the modified Gauss-Newton method. The methods of proof employed are not unique to the quadratic-quadratic model and, working by analogy, one should be able to verify that these assumptions are satisfied by any segmented polynomial model subject to B_0 and B_1 when the inputs are "constant in repeated samples."

A listing of Gallant's (1973a) assumptions and a formal statement of the theoretical results are given in the Appendix to this report. A summary of these topics as applied to segmented polynomial regression models with unknown join points and subject to the conditions B_0 and B_1 is presented below.

Assuming that the methods of Section 1 have been employed to write the model in terms of the basis functions as

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

we define;

$$y = (y_1, y_2, \ldots, y_n)' \qquad (n \times 1),$$

$$f(\theta) = (f(x_1, \theta), \ f(x_2, \theta), \ \ldots, f(x_n, \theta))' \qquad (n \times 1),$$

$$\nabla f(x, \theta) = \text{the } p \times 1 \text{ vector whose } j^{th} \text{ element is } \frac{\partial}{\partial \theta_j} f(x, \theta),$$

$$F(\theta) = \text{the } n \times p \text{ matrix whose } t^{th} \text{ row is } \nabla' f(x_t, \theta),$$

$$SSE(\theta) = \Sigma_{t=1}^n (y_t - f(x_t, \theta))^2 = (y - f(\theta))'(y - f(\theta)),$$

$$\hat{\theta} = \text{the } p \times 1 \text{ vector minimizing } SSE(\theta) \text{ over } \Omega,$$

$$\hat{\sigma}^2(y) = SSE(\hat{\theta})/n \ .$$

The starting value θ_0 for the modified Gauss-Newton method is obtained by plotting the data, choosing join points by visual inspection, and estimating the remaining parameters by ordinary multiple regression methods. As mentioned earlier, a grafted polynomial model is linear in those parameters which are not join points and they may be estimated by ordinary least squares. The combined visual estimates of the join points and corresponding least squares estimates of the remaining parameters provide the starting value θ_0.

The modified Gauss-Newton algorithm (Hartley, 1961) proceeds as follows:

0) From the starting estimate θ_0 compute

$$D_0 = [F'(\theta_0)F(\theta_0)]^{-1}F'(\theta_0)[y - f(\theta_0)] .$$

Find a λ_0 between 0 and 1 such that

$$SSE(\theta_0 + \lambda_0 D_0) \leq SSE(\theta_0) .$$

1) Set $\theta_1 = \theta_0 + \lambda_0 D_0$. Compute

$$D_1 = [F'(\theta_1)F(\theta_1)]^{-1}F'(\theta_1)[y - f(\theta_1)] .$$

Find a λ_1 between 0 and 1 such that

$$SSE(\theta_1 + \lambda_1 D_1) \leq SSE(\theta_1) .$$

2) Set $\theta_2 = \theta_1 + \lambda_1 D_1$.

$$\vdots$$

These iterations are continued until terminated according to some stopping rule such as

$$\|\theta_i - \theta_{i+1}\| \leq \epsilon \|\theta_i\| \qquad (i = 1, 2, \ldots, p)$$

and simultaneously

$$|SSE(\theta_i) - SSE(\theta_{i+1})| \leq \epsilon |SSE(\theta_i)|$$

where ϵ is some tolerance. (The use of the modified Gauss-Newton method to fit grafted polynomial models is explained in more detail in Gallant and Fuller (1973).)

Let $\hat{\theta}$ be the value of the parameter to which the iterations have converged and let

$$\hat{C} = (SSE(\hat{\theta})/n)[\frac{1}{n} F'(\hat{\theta})F(\hat{\theta})]^{-1} .$$

Assuming that $\hat{\theta}$ is, in fact, the least squares estimator, it is strongly consistent for θ^* and $\sqrt{n}(\hat{\theta} - \theta^*)$ converges in distribution to a p-dimensional multivariate normal with mean zero and a variance-covariance matrix estimated consistently by $\hat{C}$. Consequently, one may expect

$$\hat{\theta}_i \pm (1.96) \sqrt{c_{ii}/n}$$

to be a reasonably accurate 95% confidence interval for θ_i in applications.

Consider the hypothesis of location

$$H: \tau = \tau_0 \qquad \text{against} \qquad A: \tau \neq \tau_0$$

where θ has been partitioned according to

$$\theta' = (\rho', \tau')$$

with ϱ being an $r \times 1$ vector of nuisance parameters. Define $\tilde{\varrho}$ to be the $r \times 1$ vector minimizing $\Sigma_{t=1}^{n}(y_t - f(x_t, (\varrho, \tau_0))$ over $(\varrho, \tau_0) \in \Omega$. The value of $\tilde{\varrho}$ is obtained using the modified Gauss–Newton method as above by treating $f(x_t, (\varrho, \tau_0))$ as a response function with parameter ϱ. Let

$$\tilde{\sigma}^2(y) = \frac{1}{n} \Sigma_{t=1}^{n}(y_t - f(x_t, (\varrho, \tau_0)))^2 .$$

The Likelihood Ratio Test for H against A is to reject H when

$$T(y) = \tilde{\sigma}^2(y)/\hat{\sigma}^2(y)$$

is larger than c where $P[T(y) > c|H] = \alpha$. The Likelihood Ratio test statistic $T(y)$ may be characterized as

$$T(y) = X + c_n$$

where c_n converges almost surely to zero. The point c^* such that the approximating random variable X satisfies

$$P[X > c^*|H] = \alpha$$

is

$$c^* = 1 + (p-r)F_\alpha/(n-p)$$

where F_α denotes the upper $\alpha \cdot 100$ percentage point of an F-distribution with $p-r$ numerator degrees freedom and $n-p$ denominator degrees freedom. (See the Appendix for the non-null behavior of X.)

In applications, one would test the hypothesis $H: \tau = \tau_0$ against $A: \tau \neq \tau_0$ by computing $\hat{\sigma}^2(y)$ and $\tilde{\sigma}^2(y)$, forming the ratio $T(y) = \tilde{\sigma}^2(y)/\hat{\sigma}^2(y)$ and rejecting when $T(y)$ exceeds c^*. Also, one may set a confidence interval

I about a coordinate τ of $\underset{\sim}{\theta}$ by putting all τ in I for which the hypothesis H: $\tau = \tau_0$ is accepted; that is, for which $T(\underset{\sim}{y}) < c^*$.

As an example, we may apply these methods to the data given in Table 1 using the quadratic-quadratic model subject to B_0 and B_1

$$y_t = \theta_1 + \theta_2 x_t + \theta_3 x_t^2 + \theta_4 T_2(\theta_5 - x_t) + e_t .$$

The results from a modified Gauss-Newton fit are displayed graphically in Figure 1 and the numerical results are given in Table 3. The modified Gauss-Newton method did not perform well for these data necessitating a choice of a start value very near the correct answer as determined from the entries in Table 4. The cause for this poor performance seems to be a very poorly conditioned $F'(\underset{\sim}{\theta})F(\underset{\sim}{\theta})$ matrix; see Table 3. Examples where the performance of the Gauss-Newton method is much better are given in Gallant and Fuller (1973).

The asymptotic normality of $\hat{\underset{\sim}{\theta}}$, the Likelihood Ratio test, and Hartley's test (discussed in the next section) may each be used to set a 95% confidence interval on the point of join θ_5 . Using the numerical results in Tables 3 and 4 the confidence intervals thus obtained are:

1) Asymptotic normality of the least squares estimator

$$I = \{\theta_5: \quad \hat{\theta}_5 - (1.96)(.39) \leq \theta_5 \leq \hat{\theta}_5 + (1.96)(.39)\}$$

$$= [2.178, 3.705]$$

2) Inversion of the Likelihood Ratio test

$$I = \{\theta_5: \quad T(\underset{\sim}{y}) \leq 1.262\} = [2.915, 2.975]$$

3) Inversion of Hartley's test

$$I = \{\theta_5: \quad F(\underset{\sim}{y}) \leq 3.62\} = [2.875, 2.965] .$$

The lengths of these intervals are 1.53, .06, and .09 respectively.

TABLE 3. Modified Gauss-Newton Results

```
*************************************************************************
*                                                                     *
*                                                                     *
*               MODIFIED GAUSS-NEWTON ITERATIONS                      *
*                                                                     *
*                                                                     *
*************************************************************************
```

STEP	ERROR SUM OF SQUARES	PARAMETER NUMBER	INTERMEDIATE ESTIMATE
0	0.3414313D-01	1	-0.55508545D 02
		2	0.32164322D 02
		3	-0.44838362D 01
		4	0.29673080D 02
		5	0.29399824D 01
1	0.3414115D-01	1	-0.55512069D 02
		2	0.32166603D 02
		3	-0.44842048D 01
		4	0.29672975D 02
		5	0.29401019D 01

---- UNABLE TO IMPROVE ON THIS ESTIMATE

```
*************************************************************************
*                                                                     *
*                                                                     *
*        CHECK THESE ITERATIONS TO SEE IF CONVERGENCE HAS             *
*        BEEN ACHIEVED BEFORE USING THE FOLLOWING RESULTS             *
*                                                                     *
*                                                                     *
*************************************************************************
```

TABLE 2. (continued)

```
+------------------------------------------------------------------+
|                                                                  |
|        MODIFIED GAUSS-NEWTON PARAMETER ESTIMATES                 |
|                                                                  |
+------------------------------------------------------------------+

   PARAMETER
    NUMBER          ESTIMATE        STANDARD ERROR      Z-STATISTIC

       1            -55.5121           14.5208            -3.8229
       2             32.1666            9.45202            3.4031
       3             -4.48420           1.53661           -2.9182
       4             29.6730            2.03210           14.6021
       5              2.94010           0.390501           7.5271


          ESTIMATED VARIANCE          0.155187D-02

       RESIDUAL SUM OF SQUARES        0.3414115D-01
       NUMBER OF OBSERVATIONS                  22
       NUMBER OF PARAMETERS                     5
```

TABLE 3. (continued)

```
****************************************************************************
*                                                                          *
*                                                                          *
*            CORRELATION MATRIX OF THE PARAMETER ESTIMATES                  *
*                                                                          *
*                                                                          *
****************************************************************************
```

1.000000	-0.999935	0.999739	-0.426821	-0.794536
-0.999935	1.000000	-0.999934	0.430923	0.789949
0.999739	-0.999934	1.000000	-0.434792	-0.785379
-0.426821	0.430923	-0.434792	1.000000	-0.168183
-0.794536	0.789949	-0.785379	-0.168183	1.000000

TABLE 4. Tests for the point of join using a quadratic-quadratic model subject to B_0 and B_1 for the data of Table 1

Hypothesized Join Point	Corresponding Least Squares Estimates Of the Linear Parameters					Hartley's Test	Likelihood Ratio Test
θ_5^0	θ_1	θ_2	θ_3	θ_4	SSE	$F(y)$	$T(y)$
2.85999	-0.12112	-3.87760	1.37240	37.6367	0.097515	7.50994	2.85608
2.86999	-5.56382	-0.31805	0.79124	35.7040	0.085059	3.95173	2.49126
2.87999	-11.3023	3.43134	0.17964	34.0864	0.073391	2.05554	2.14952
2.88999	-17.4076	7.41454	-0.46917	32.7759	0.063047	1.41018	1.84656
2.89999	-23.9942	11.7057	-1.16719	31.7538	0.053955	1.35767	1.58027
2.90998	-31.0819	16.3171	-1.91637	30.9664	0.046236	1.55706	1.35419
2.91998	-38.6981	21.2595	-2.71821	30.3796	0.040295	1.83090	1.18018
2.92998	-46.8439	26.5508	-3.57560	29.9583	0.036153	2.12069	1.05887
2.93998	-55.5085	32.1643	-4.48384	29.6731	0.034143	2.35441	1.00000
2.94998	-64.6956	38.1059	-5.44356	29.5143	0.034271	2.68254	1.00375
2.95998	-74.4353	44.3938	-6.45749	29.4709	0.036548	3.26427	1.07044
2.96998	-84.7119	51.0141	-7.52281	29.5398	0.040612	4.27518	1.18947
2.97998	-95.6625	58.0540	-8.65336	29.7347	0.046188	5.90773	1.35278
2.98998	-107.171	65.4347	-9.83591	30.0233	0.053705	7.76941	1.57294
2.99998	-119.070	73.0443	-11.0518	30.3753	0.063389	9.18396	1.85658
3.00998	-131.308	80.8461	-12.2945	30.7831	0.075120	9.76091	2.20016
3.01998	-143.985	88.9031	-13.5741	31.2578	0.088674	9.66892	2.59714

The .05 critical point for Hartley's test is 3.63.

The .05 critical point for the Likelihood Ratio test is 1.262.

4. HARTLEY'S METHOD

The theory presented in the previous section required the deletion of the functions $T_0(\alpha_i - x)$ and $T_1(\alpha_i - x)$ from the set of basis functions or, equivalently, the imposition of the constraints B_0 and B_1 on the response function $g(x)$. In applications where it is not appropriate to impose these conditions (see Figure 1) it is still possible to test some hypotheses and set exact confidence intervals on some parameters using a method due to Hartley (1964). In those applications where the imposition of B_0 and B_1 is appropriate, Hartley's method is not recommended for setting confidence intervals. Usually such intervals will be wider than those constructed using the methods of Section 2 and sometimes the confidence interval will not contain the least squares estimate of the parameter.

Consider the hypothesis

$$H: \underset{\sim}{\tau} = \underset{\sim}{\tau}_0 \qquad \text{against} \qquad A: \underset{\sim}{\tau} \neq \underset{\sim}{\tau}_0$$

for the regression model

$$y_t = f(\underset{\sim}{x}_t, \underset{\sim}{\theta}) + e_t$$

where $\underset{\sim}{\theta}$ has been partitioned according to

$$\underset{\sim}{\theta}' = (\underset{\sim}{\rho}', \underset{\sim}{\tau}') \quad .$$

If the regression model with $\underset{\sim}{\tau}$ specified to be $\underset{\sim}{\tau}_0$,

$$y_t = f(\underset{\sim}{x}_t, (\underset{\sim}{\rho}, \underset{\sim}{\tau}_0)) + e_t \quad ,$$

is a linear model in the remaining parameters $\underset{\sim}{\rho}$ then Hartley's method is

applicable. For example, the quadratic-quadratic model subject to B_0 with the join point θ_6 specified as θ_6^0 is

$$y_t = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 T_1(\theta_6^0 - x) + \theta_5 T_2(\theta_6^0 - x) + e_t$$

which is linear in the parameters $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$. If the hypothesis $H: \underset{\sim}{\tau} = \underset{\sim}{\tau}_0$ is true and one fits the linear model

$$y_t = f(\underset{\sim}{x}_t, (\varrho, \underset{\sim}{\tau}_0)) + z_t' \underset{\sim}{\delta} + e_t$$

where $\underset{\sim}{z}_t$ denotes arbitrarily chosen regressors then the least squares estimate $\hat{\underset{\sim}{\delta}}$ is normally distributed with the zero vector as its expectation. Consequently, the standard F-test for the hypothesis $H': \underset{\sim}{\delta} = \underset{\sim}{0}$ in the linear model

$$y_t = f(\underset{\sim}{x}_t, (\varrho, \underset{\sim}{\tau}_0)) + z_t' \underset{\sim}{\delta} + e_t$$

is test for $H: \underset{\sim}{\tau} = \underset{\sim}{\tau}_0$ in the nonlinear model

$$y_t = f(\underset{\sim}{x}_t, (\varrho, \underset{\sim}{\tau}_0)) + e_t \quad .$$

The regressors $\underset{\sim}{z}_t$ must be chosen to obtain good power for the test under the alternative $A: \underset{\sim}{\tau} \neq \underset{\sim}{\tau}_0$. A method of choosing the extra regressors which works well for grafted polynomial models is the following. Plot the residuals from fitting

$$y_t = f(x_t, (\varrho, \underset{\sim}{\tau}_0)) + e_t$$

against x_t for various choices of $\underset{\sim}{\tau}_0$. From a visual inspection of these plots, try to choose regressors which are functions of the x_t and would give a "good" least squares fit to the plotted residuals. For example, the choice

$$z_t' = (x_t^3, x_t^4)$$

yields adequate results for the hypothesis $H: \theta_6 = \theta_6^0$ of the quadratic-quadratic model subject to B_0 .

A confidence interval may be set on the point of join for grafted polynomial models with a single join point using this method. Letting τ correspond to this join point, one includes in the confidence interval I those points τ_0 for which the hypothesis $H': \underset{\sim}{\delta} = \underset{\sim}{0}$ corresponding to $H: \tau = \tau_0$ is accepted. In models with multiple joins

$$\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{r-1})$$

this same process can be used to construct joint confidence regions.

One might note that Hartley's method depends only on the assumption of normal errors e_t so that probability statements based on the method are exact and not approximate as with the methods of Section 2. However, in this author's opinion, the deficiencies of the method noted earlier outweigh the advantages of exact probability statements in situations where either approach is applicable.

As an example, we may apply these methods to the data given in Table 2 using the quadratic-quadratic model subject to B_0

$$y_t = \theta_1 + \theta_2 x_t + \theta_3 x_t^2 + \theta_4 T_1(\theta_6 - x_t) + \theta_5 T_2(\theta_6 - x_t) + e_t \ .$$

The parameter estimate

$$\hat{\underset{\sim}{\theta}} = (-56, \ 45, \ -8.8, \ 12, \ -.93, \ 2.86)'$$

is obtained by minimizing $SSE(\underset{\sim}{\theta})$ with respect to $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ for specified θ_6 by ordinary least squares and then varying θ_6 as seen in Table 5. A 95% confidence region for θ_6 obtained by inverting Hartley's test is

TABLE 5.   Tests for the point of join using quadratic-quadratic subject to B_0 for the data of Table 2.

Hypothesized Join Point	Corresponding Least Squares Estimates of the Linear Parameters					SSE	Hartley's Test
θ_6^0	θ_1	θ_2	θ_3	θ_4	θ_5		$F(y)$
2.78999	-110.722	86.4351	-16.6933	8.61481	18.5916	0.18597	8.23017
2.79998	-114.040	88.8261	-17.1214	9.73849	17.4259	0.15144	6.78867
2.80998	-111.403	86.7330	-16.7065	10.6115	15.3369	0.11787	5.10833
2.81998	-99.2890	77.5055	-14.9505	10.8299	12.2575	0.093665	3.30890
2.82998	-81.9816	64.3856	-12.4663	10.5483	8.82243	0.078989	1.93258
2.83998	-80.3509	63.1001	-12.2132	11.5916	6.17212	0.047203	1.74485
2.84998	-71.4793	56.3848	-10.9436	12.0699	2.78332	0.024865	2.23804
2.85998	-56.3863	45.0162	-8.80521	11.8987	-0.92789	0.014930	3.34372
2.86998	-37.7755	31.0290	-6.18039	11.1772	-4.48722	0.016594	3.34716
2.87998	-27.3095	23.1908	-4.71490	11.2467	-7.85562	0.032228	5.06854
2.88998	-13.0604	12.5410	-2.72793	10.7713	-11.1180	0.058569	4.67633
2.89998	3.90675	-0.12182	-0.36903	9.72309	-13.8512	0.092839	4.32828
2.90998	21.3302	-13.1081	2.04683	8.27283	-15.7642	0.12986	4.58415
2.91998	29.0508	-18.8532	3.11376	7.91674	-17.9723	0.14798	5.48984
2.92998	37.5027	-25.1315	4.27759	7.31554	-19.8466	0.16822	6.58497
2.93997	47.3566	-32.4412	5.63068	6.26935	-20.9070	0.19132	7.79108
2.94997	54.4744	-37.7147	6.60558	5.43235	-21.7840	0.20767	8.70193

The .05 critical point for Hartley's test is 4.74.

$$I = \{\theta_6: \quad F(\underset{\sim}{y}) \leq 4.74\}$$

$$= [2.815, \ 2.875] \cup [2.885, \ 2.915] \ .$$

The disconnectedness of this confidence region can, of course, be eliminated by including the omitted interval (2.875, 2.885). This would not alter the truth of the statement

$$\text{"}P[I \quad \text{contains} \quad \theta_6^*] \geq .95.\text{"}$$

REFERENCES

Billingsly, P. and Topsoe, F. (1967) "Uniformity in weak convergence," Z. Wahrscheinlichkeitstheorie und Verw. Geb., 7, pp. 1-16.

Fuller, W. A. (1969) "Grafted polynomials as approximating functions," The Australian Journal of Agricultural Economics, 13, pp. 35-46.

Gallant, A. R. (1971) Statistical inference for nonlinear regression models. Ph.D. dissertation, Iowa State University.

Gallant, A. R. (1973a) "Inference for nonlinear models," Institute of Statistics Mimeograph Series, No. 875. Raleigh: North Carolina State University.

Gallant, A. R. (1973b) "The power of the likelihood ratio test of location in nonlinear regression models," Institute of Statistics Mimeograph Series, No. 889. Raleigh: North Carolina State University.

Gallant, A. R. (1974) "Testing for the location of a subset of the parameters of a nonlinear regression model," Unpublished.

Gallant, A. R. and Fuller, W. A. (1973) "Fitting segmented polynomial regression models whose join points have to be estimated," Journal of the American Statistical Association, 68, pp. 144-147.

Gerig, T. M. and Gallant, A. R. (1974) "Computing methods for linear models subject to linear parametric constraints," Journal of Statistical Computation and Simulation. In press.

Graybill, F. A. (1961) An introduction to linear statistical models, Vol. I. New York: McGraw-Hill.

Hartley, H. O. (1961) "The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares," Technometrics, 3, pp. 269-280.

Hartley, H. O. (1964) "Exact confidence regions for the parameters in non-linear regression laws," Biometrika, 51, pp. 347-353.

Hsiung, P. O. (1974) Study of the interaction of organic liquids with polymers by means of gas chromatography. Ph.D. dissertation, North Carolina State University.

Jennrich, R. I. (1969) "Asymptotic properties of non-linear least squares estimators," Annals of Mathematical Statistics, 40, pp. 633-645.

Malinvaud, E. (1966) Statistical methods of Econometrics. Chicago: Rand McNally and Company.

Malinvaud, E. (1970) "The consistency of nonlinear regression," Annals of Mathematical Statistics, 41, pp. 956-969.

Theil, H. (1971) Principles of econometrics. New York: John Wiley and Sons, Inc.

A P P E N D I X

5.   FORMAL STATEMENT OF THEORETICAL RESULTS

Notation. For the regression model

$$y_t = f(x_t, \theta^*) + e_t \qquad (t = 1, 2, \ldots, n)$$

where $f(x, \theta)$ is a known function defined on $\mathcal{X} \times \Omega$ and θ^* denotes the true but unknown value of the parameter θ define:

$$y = (y_1, y_2, \ldots, y_n)' \qquad (n \times 1) \;,$$

$$f(\theta) = (f(x_1, \theta),\; f(x_2, \theta),\; \ldots,\; f(x_n, \theta))' \qquad (n \times 1) \;,$$

$$\nabla f(x, \theta) = \text{the } p \times 1 \text{ vector whose } j^{th} \text{ element is } \frac{\partial}{\partial \theta_j} f(x, \theta) \;,$$

$$\nabla^2 f(x, \theta) = \text{the } p \times p \text{ matrix whose } i,j^{th} \text{ element is } \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x, \theta) \;,$$

$$F(\theta) = \text{the } n \times p \text{ matrix whose } t^{th} \text{ row is } \nabla' f(x_t, \theta) \;,$$

$$SSE(\theta) = (y - f(\theta))'(y - f(\theta)) \;,$$

$$\hat{\theta}(y) = \text{any Borel measurable function of } y \text{ with } SSE(\hat{\theta}) = \inf_{\Omega} SSE(\theta) \;,$$

$$\hat{\sigma}^2(y) = n^{-1} SSE(\hat{\theta}) \;;$$

with respect to the hypothesis

$$H: \tau = \tau_0 \qquad \text{against} \quad A: \tau \neq \tau_0$$

where θ has been partitioned according to

$$\theta = (\rho, \tau)$$

define:

$$\underset{\sim}{P} = \underset{\sim}{F}(\underset{\sim}{\theta}^*)[\underset{\sim}{F}'(\underset{\sim}{\theta}^*)\underset{\sim}{F}(\underset{\sim}{\theta}^*)]^{-1}\underset{\sim}{F}'(\underset{\sim}{\theta}^*) \qquad (n \times n)$$

$$\underset{\sim}{P}^{\perp} = \underset{\sim}{I} - \underset{\sim}{P} \qquad (n \times n) \ ,$$

$$\underset{\sim}{f}_0(\underset{\sim}{\rho}) = \underset{\sim}{f}((\underset{\sim}{\rho}, \underset{\sim}{\tau}_0)) \qquad (n \times 1) \ ,$$

$$\nabla_{\underset{\sim}{\rho}}f(\underset{\sim}{x}, \underset{\sim}{\theta}) = \text{the } r \times 1 \text{ vector whose } j^{th} \text{ element is } \frac{\partial}{\partial \rho_j} f(\underset{\sim}{x}, (\underset{\sim}{\rho}, \underset{\sim}{\tau})) \ ,$$

$$\underset{\sim}{F}_0(\underset{\sim}{\rho}) = \text{the } n \times r \text{ matrix whose } t^{th} \text{ row is } \nabla'_{\underset{\sim}{\rho}}f(\underset{\sim}{x}_t, (\underset{\sim}{\rho}, \underset{\sim}{\tau}_0)) \ ,$$

$$\underset{\sim}{P}_0(\underset{\sim}{\rho}) = \underset{\sim}{F}_0(\underset{\sim}{\rho})[\underset{\sim}{F}'_0(\underset{\sim}{\rho})\underset{\sim}{F}_0(\underset{\sim}{\rho})]^{-1}\underset{\sim}{F}'_0(\underset{\sim}{\rho}) \qquad (n \times n) \ ,$$

$$\underset{\sim}{P}_0^{\perp}(\underset{\sim}{\rho}) = \underset{\sim}{I} - \underset{\sim}{P}_0(\underset{\sim}{\rho}) \qquad (n \times n) \ ,$$

$$\underset{\sim}{\delta}_0(\underset{\sim}{\rho}) = \underset{\sim}{f}(\underset{\sim}{\theta}^*) - \underset{\sim}{f}_0(\underset{\sim}{\rho}) \qquad (n \times 1) \ ,$$

$$\widetilde{\sigma}^2(\underset{\sim}{y}) = \inf_{(\underset{\sim}{\rho}, \underset{\sim}{\tau}_0) \, \epsilon \, \Omega} n^{-1}SSE((\underset{\sim}{\rho}, \underset{\sim}{\tau}_0)) \ ,$$

$$T(\underset{\sim}{y}) = \widetilde{\sigma}^2(\underset{\sim}{y})/\hat{\sigma}^2(\underset{\sim}{y})$$

Densities and Distributions. Let $g(t; v, \lambda)$ denote the non-central chi-squared density function with v degrees freedom and non-centrality λ (Graybill, 1961, p. 74) and let $G(t; v, \lambda)$ denote the corresponding distribution function. Let $n(t; \mu, \sigma^2)$ denote the normal density function with mean μ and variance σ^2 and let $N(t; \mu, \sigma^2)$ denote the corresponding distribution function. Define $H(x; v_1, v_2, \lambda_1, \lambda_2)$ to be the distribution function given by

$$0, \qquad\qquad\qquad\qquad\qquad\qquad x \leq 1, \lambda_2 = 0 ,$$

$$\int_0^\infty G(t/[x-1] + 2\,x\lambda_2/[x-1]^2; v_2, \lambda_2/[x-1]^2)g(t; v_1, \lambda_1)dt, \qquad x < 1, \lambda_2 > 0 ,$$

$$\int_0^\infty N(-t; 2\lambda_2, 8\lambda_2)g(t; v_1, \lambda_1)dt, \qquad\qquad\qquad x = 1, \lambda_2 > 0 ,$$

$$1 - \int_0^\infty G(t/[x-1] + 2x\lambda_2/[x-1]^2; v_2, \lambda_2/[x-1]^2)g(t; v_1, \lambda_1)dt, \qquad x > 1 .$$

Remark. The distribution function $H(x; v_1, v_2, \lambda_1, \lambda_2)$ is partially tabulated in Gallant (1973b) and can be approximated by the non-central F-distribution with argument $v_2(x-1)/v_1$, v_1 numerator degrees freedom, v_2 denominator degrees freedom, and non-centrality λ_1 when the parameter λ_2 is small.

Definition. (Malinvaud, 1970) Let $\underset{\sim}{G}$ be the Borel subsets of $\underset{\sim}{X}$ and let $\{\underset{\sim}{x}_t\}_{t=1}^\infty$ be a sequence of inputs chosen from $\underset{\sim}{X}$. Let $I_A(\underset{\sim}{x})$ be the indicator function of a subset A of $\underset{\sim}{X}$. The measure μ_n on $(\underset{\sim}{X}, \underset{\sim}{G})$ is defined by

$$\mu_n(A) = n^{-1}\Sigma_{t=1}^n I_A(\underset{\sim}{x}_t)$$

for each $A \in \underset{\sim}{G}$.

Definition. (Billingsly and Topsoe, 1967) A sequence of measures $\{\mu_n\}$ on $(\underset{\sim}{X}, \underset{\sim}{G})$ is said to converge weakly to a measure μ on $(\underset{\sim}{X}, \underset{\sim}{G})$ if for every real valued, bounded, continuous function g with domain $\underset{\sim}{X}$

$$\int g(\underset{\sim}{x})d\mu_n(\underset{\sim}{x}) \to \int g(\underset{\sim}{x})d\mu(\underset{\sim}{x})$$

as $n \to \infty$.

For each assumption below it is implicitly assumed that any lower numbered assumption necessary for existence of terms is satisfied. For example, the statement of Assumption 7 requires Assumption 6 for definition of μ and Assumption 2 for the $\underset{\sim}{G}$ measurability of $\{\underset{\sim}{x}: f(\underset{\sim}{x}, \underset{\sim}{\theta}) \neq f(\underset{\sim}{x}, \underset{\sim}{\theta}^*)\}$.

Assumption 1. $\underset{\sim}{\Omega}$ is a closed subset of R^p.

Assumption 2. $f(\underset{\sim}{x}, \underset{\sim}{\theta})$ is continuous on $\underset{\sim}{\mathfrak{X}} \times \underset{\sim}{\Omega}$.

Assumption 3. For given n and almost every $\underset{\sim}{x}$ there is a $\underset{\sim}{\theta}$ in $\underset{\sim}{\Omega}$ minimizing $SSE(\underset{\sim}{\theta})$.

Assumption 4. The errors $\{e_t\}$ are independent and identically distributed with mean zero and finite variance $\sigma^2 > 0$.

Assumption 5. $\underset{\sim}{\mathfrak{X}}$ is a compact subset of R^k.

Assumption 6. The sequence of measures $\{\mu_n\}$ determined by $\{\underset{\sim}{x}_t\}$ converges weakly to a measure μ on $(\underset{\sim}{\mathfrak{X}}, \underset{\sim}{G})$.

Assumption 7. If $\underset{\sim}{\theta} \neq \underset{\sim}{\theta}^*$ and $\underset{\sim}{\theta} \in \underset{\sim}{\Omega}$ then $\mu\{\underset{\sim}{x}: f(\underset{\sim}{x}, \underset{\sim}{\theta}) \neq f(\underset{\sim}{x}, \underset{\sim}{\theta}^*)\} > 0$.

Assumption 8. Given $M > 0$ there is an N and a K such that for all $n > N$ and all $\underset{\sim}{\theta} \in \underset{\sim}{\Omega}$ if $n^{-1} \Sigma_{t=1}^n f^2(\underset{\sim}{x}_t, \underset{\sim}{\theta}) < M$ then $\|\underset{\sim}{\theta}\| < K$.

Assumption 9. There is a bounded open sphere $\underset{\sim}{\Omega}^o$ containing $\underset{\sim}{\theta}^*$ whose closure $\overline{\underset{\sim}{\Omega}^o}$ is a subset of $\underset{\sim}{\Omega}$.

Assumption 10. $\underset{\sim}{\nabla} f(\underset{\sim}{x}, \underset{\sim}{\theta})$ exists and is continuous on $\underset{\sim}{\mathfrak{X}} \times \underset{\sim}{\Omega}$.

Assumption 11. The matrix

$$\mathbf{\not{L}} = [\int \frac{\partial}{\partial \theta_i} \; f(\underset{\sim}{x}, \underset{\sim}{\theta}^*) \; \frac{\partial}{\partial \theta_j} \; f(\underset{\sim}{x}, \underset{\sim}{\theta}^*) d\mu(\underset{\sim}{x})]_{p \times p}$$

is non-singular.

Assumption 12. There is a function $\varphi(\underset{\sim}{x}, \underset{\sim}{\theta})$ which is uniformly bounded for $(\underset{\sim}{x}, \underset{\sim}{\theta}) \; \epsilon \; \underset{\sim}{\mathfrak{X}} \times \overline{\underset{\sim}{\Omega}}^{\circ}$ such that

$$f(\underset{\sim}{x}, \underset{\sim}{\theta}) = f(\underset{\sim}{x}, \underset{\sim}{\theta}^*) + \underset{\sim}{\nabla}' f(\underset{\sim}{x}, \underset{\sim}{\theta}^*)(\underset{\sim}{\theta} - \underset{\sim}{\theta}^*) + \varphi(\underset{\sim}{x}, \underset{\sim}{\theta}) \| \underset{\sim}{\theta} - \underset{\sim}{\theta}^* \|^2 \; .$$

Assumption 13. The response function f, inputs $\{\underset{\sim}{x}_t\}$ and errors $\{e_t\}$ are such that given a sequence of random variables $\{\widetilde{\underset{\sim}{\theta}}_n\}$ with $\widetilde{\underset{\sim}{\theta}}_n \xrightarrow{\text{a.s.}} \underset{\sim}{\theta}^*$ as $n \rightarrow \infty$ it follows that

$$n^{-\frac{1}{2}} \Sigma_{t=1}^n \{\frac{\partial}{\partial \theta_i} \; f(\underset{\sim}{x}_t, \widetilde{\underset{\sim}{\theta}}_n) - \frac{\partial}{\partial \theta_i} \; f(\underset{\sim}{x}_t, \underset{\sim}{\theta}^*)\} \; e_t \xrightarrow{\;P\;} 0$$

as $n \rightarrow \infty$ for $i = 1, 2, \ldots, p$.

The verification of Assumption 8 is unnecessary when those which preceed it as satisfied and

Assumption 14. $\underset{\sim}{\Omega}$ is bounded.

Also, the verification of Assumptions 12, 13 is unnecessary when those which preceed them and

Assumption 15. The partial derivatives $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \; f(\underset{\sim}{x}, \underset{\sim}{\theta})$ exist and are continuous over $\underset{\sim}{\mathfrak{X}} \times \underset{\sim}{\Omega}$.

Additional assumptions which are required to derive the asymptotic properties of the Likelihood Ratio test of $H: \underset{\sim}{\tau} = \underset{\sim}{\tau}_0$ against $A: \underset{\sim}{\tau} \neq \underset{\sim}{\tau}_0$ when $\underset{\sim}{\theta}$ has been partitioned according to $\underset{\sim}{\theta}' = (\underset{\sim}{\rho}', \underset{\sim}{\tau}')$ are:

Assumption 16. There is a unique point ρ_0 which minimizes

$$\int [f(x, \theta^*) - f(x, (\rho, \tau_0))]^2 d\mu(x) \quad \text{over} \quad R = \{\rho: (\rho, \tau_0) \in \Omega\}$$

Assumption 17. There is a bounded open sphere R^0 containing ρ_0 whose closure $\bar{R}^0$ is a subset of R .

Assumption 18. The matrix

$$\mathcal{L}_0 = [\int \frac{\partial}{\partial \rho_i} f(x, (\rho_0, \tau_0)) \frac{\partial}{\partial \theta_j} f(x, (\rho_0, \tau_0)) d\mu(x)]_{r \times r}$$

is non-singular.

Theorem. If a regression model satisfies Assumptions 1 through 8 then $\hat{\theta}$ converges almost surely to θ^* and $\hat{\sigma}^2$ converges almost surely to σ^2 .

Proof. Gallant (1973a).

Theorem. If a regression model satisfies Assumptions 1 through 13 then $\sqrt{n}(\hat{\theta} - \theta^*)$ converges in distribution to a multivariate normal with mean 0 and variance-covariance matrix $\mathcal{L}^{-1}$. Moreover, $n^{-1} F'(\hat{\theta}) F(\hat{\theta})$ converges almost surely to $\mathcal{L}$.

Proof. Gallant (1973a).

Theorem. We are given a regression model

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

and the data pairs (y_t, x_t) $(t = 1, 2, \ldots, n)$.

Conditions. There is a convex, bounded subset S of R^p and a θ_0 interior to S such that:

1) $\nabla f(x_t, \theta)$ exists and is continuous over $\bar{S}$ for $t = 1, 2, \ldots, n$.

2) $\theta \in S$ implies the rank of $F(\theta)$ is p .

3) $SSE(\theta_0) < \inf\{SSE(\theta);\ \theta$ a boundary point of $S\}$.

4) There do not exist θ', θ'' in S such that

$$\nabla\ SSE(\theta') = \nabla\ SSE(\theta'') = 0 \qquad \text{and} \qquad SSE(\theta') = SSE(\theta'') .$$

Construction. Construct the sequence $\{\theta_k\}_{k=1}^{\infty}$ as follows:

0) Compute $D_0 = [F'(\theta_0)F(\theta_0)]^{-1}F'(\theta_0)[y - f(\theta_0)]$.

Find λ_0 which minimizes $SSE(\theta_0 + \lambda D_0)$ over $\Lambda_0 = \{\lambda:\ 0 \leq \lambda \leq 1,$ $\theta_0 + \lambda D_0 \in \bar{S}\}$.

1) Set $\theta_1 = \theta_0 + \lambda_0 D_0$.

Compute $D_1 = [F'(\theta_1)F(\theta_1)]^{-1}F'(\theta_1)[y - f(\theta_1)]$.

Find λ_1 which minimizes $SSE(\theta_1 + \lambda D_1)$ over $\Lambda_1 = \{\lambda:\ 0 \leq \lambda \leq 1,$ $\theta_1 + \lambda D_1 \in \bar{S}\}$.

2) Set $\theta_2 = \theta_1 + \lambda_1 D_1$.

$\vdots$

Conclusions. The for the sequence $\{\theta_k\}_{k=1}^{\infty}$ it follows that:

1) θ_k is an interior point of S for $k = 1, 2, \ldots$.

2) The sequence $\{\theta_k\}$ converges to a limit θ^* which is interior to S .

3) $\nabla\ SSE(\theta^*) = 0$.

Proof: Gallant (1971).

In the following theorem, f_0, δ_0, P_0, and $P_0^\perp$ denote the corresponding functions evaluated at the point $\rho = \rho_0$ given by Assumption 16.

Theorem. For a regression model with normally distributed errors which satisfies Assumptions 1 through 13 and Assumptions 16 through 18 the Likelihood Ratio test statistic for H: $\tau = \tau_0$ against A: $\tau \neq \tau^0$ may be characterized by

$$T(y) = X + c_n$$

where c_n converges in probability to zero and

$$X = (e + \delta_0)' P_0^\perp (e + \delta_0)/e' P^\perp e \ .$$

If the condition $P_0 P = P_0$ is satisfied, the random variable X has the distribution function $H(x; p-r, n-p, \lambda_1, \lambda_2)$ where $\lambda_1 = \delta_0'(P - P_0)\delta_0/(2\sigma^2)$ and $\lambda_2 = \delta_0' P^\perp \delta_0/(2\sigma^2)$.

When the null hypothesis H: $\tau = \tau_0$ is true the condition $P_0 P = P_0$ is satisfied and X is distributed as $H(x; p-r, n-p, 0, 0)$.

Proof. Gallant (1974).

INSTITUTE OF STATISTICS
MIMEOGRAPH SERIES
(These are available at 1¢ per page)

867. SEN, P. K. On unbiased estimation for randomized response model. 21 pp.
868. HUNTER, JEFFREY J. Renewal theory in two dimensions: asymptotic results. 29 pp.
869. CHAKRAVARTI, I. M. On random search using binary systems derived from finite projective planes. 10 pp.
870. CHAKRAVORTI, S. R. On some tests of growth curves models under Behren-Fisher's situation. 27 pp.
871. JOHNSON, N. L. Return to repetitions. 11 pp.
872. CASTILLO-MORALES, ALBERTO. Drawing an optimal tree from a distance matrix. Thesis. 100 pp.
873. JOHNSON, N. L. and S. KOTZ. A vector valued multivariate hazard rate: I.
874. GUALTIEROTTI, A. and S. CAMBANIS. Some remarks on the equivalence of Gaussian processes. 11 pp.
875. GALLANT, A. R. Inference for non-linear models. 30 pp.
876. JEFFCOAT, COLIN E. Some related queuing models with dependent service and inter-arrival time. 94 pp.
877. FELDMAN, JOSEPH GERALD. Evaluating the effect of group therapy in the prognosis of coronary patients. Thesis 94 pp.
878. MUSTAFA, AHMED F.M. Fecundability differentials among acceptors and nonacceptors of family planning; A simulation experiment. Thesis. 90 pp.
879. NOUR, EL SAYED. A stochastic model for the study of human fertility. 78 pp.
880. SYMONS, M. Bayes' modification of some clustering criteria. 22 pp.
881. MAHMOUD, MAHMOUD RIAD. Sequential decision procedures for testing hypotheses concerning general estimable parameters. Thesis. 33 pp.
882. KOONG, LING JUNG and H. L. LUCAS. A mathematical model for the joint metabolism of nitrogen and energy. Thesis. 116 pp.
883. JOHNSON, N. L. and S. KOTZ. A vector-valued multivariate hazard rate: II. 14 pp.
884. DAVIES, H. I. and EDWARD J. WEGMAN. Sequential nonparametric density estimation. 19 pp.
885. O'FALLON, JUDITH. Discriminate analysis under truncation. Thesis.
886. LEADBETTER, M. R. On extreme values in stationary sequences. 35 pp.
887. LEADBETTER, M. R. Locating the maximum of a stationary normal process. 8 pp.
888. SIMONS, GORDON D. Generalized cumulative distribution functions II: The σ-lower finite case.
889. GALLANT, A. R. The power of the likelihood ratio test of location in nonlinear regression models. 13 pp.
890. GALLANT, A. R. Non-linear Regression Models. 26 pp.
891. FERGUSON, R. A. and J. G. FRYER. Jackknifing maximum likelihood estimates in the case of a single parameter. 23 pp.
892. FERGUSON, R. A. and J. G. FRYER. Jackknifing maximum likelihood estimates for a special class of distributions. 10 pp.
893. CHAKRAVORTI, S. R. Maximum likelihood estimation for the growth curve model with unequal dispersion matrices. 14 pp.
894. NOUR, EL SAYED and EDWARD WEGMAN. A stochastic model for human fertility: I.
895. KALLENBERG, OLAV. On symmetrically distributed random measures. 21 pp.
896. CHAKRAVORTI, S. R. On asymptotic properties of the maximum likelihood estimates of the general growth curve model. 12 pp.
897. LEE, ALAN J. Some results in the theory of stochastic processes.
898. CURRAN, THOMAS COOKE. Stochastic models of oxygen transport in respiring tissues. 166 pp.
899. GALLANT, A RONALD and Thomas M. GERIG. Constrained linear models: A SAS implementation. 22 pp.
900. GALLANT, A. RONALD and THOMAS M. GERIG. Seemingly unrelated nonlinear regressions.
901. SMEACH, STEPHEN CHARLES Comparison of stochastic and deterministic models for the kinetic behavior of certain structured enzyme systems. 256 pp.
902. HOCHBERG, JOSEPH. An extension of the T method to unbalanced linear models of fixed effects.
903. WOOLSON, ROBERT. Some alternative measures of asymptotic reltive efficiency for the multiparameter testing problem with application to the growth curve problem.
904. SIMONS, GORDON. Optimality within the class of sequential probability ratio tests.
905. SEN, P. K. Nonparametric tests for interchangeability under competing risks.
906. SEN, P. K. An almost sure invariance principle for the extrema of certain sample functions.
907. HASMINSKII, R. V. and I. A. IBRAGIMOV. An example of singular statistical experiments admitting local exponential approximation.
908. KALLENBERG, OLAV. A limit theorem for thinning of point processes.
909. KALLENBERG, OLAV. On the existence and path properties of stochastic integrals.
910. CHAKRAVORTI, S. R. A note on step-down procedures in the MANOVA problem with unequal dispersion matrices.