THE POWER OF THE LIKELIHOOD RATIO TEST OF LOCATION IN NONLINEAR REGRESSION MODELS

by

A. R. GALLANT

Institute of Statistics
Mimeograph Series No. 889
Raleigh - September 1973

The Power of the Likelihood Ratio Test Of Location in Nonlinear Regression Models

by

A. R. Gallant*

ABSTRACT

The Likelihood Ratio Test statistic T for the hypothesis $H: \theta = \theta_o$ against $A: \theta \neq \theta_o$ is considered when the data are generated according to the nonlinear model $y = f(x,\theta) + e$ with variance unknown. A random variable X is obtained such that $n\cdot(T-X)$ converges in probability to zero; the distribution function of X is derived assuming normal errors.

The power of the Likelihood Ratio test is tabulated for selected sample sizes and selected departures from the null hypothesis by using the distribution function of X to approximate the distribution function of T. Monte-Carlo power estimates for an exponential model are compared to power points calculated using this approximation to gain a feel for the adequacy of the approximation in applications.

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

1. INTRODUCTION

This paper considers the hypothesis of location:

$$H: \theta = \theta_o \quad \text{against} \quad A: \theta \neq \theta_o$$

at the α level of significance when the data are responses y_t to inputs x_t generated according to the nonlinear regression model

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

The unknown parameter θ is known to be contained in the parameter space Ω which is a subset of the p-dimensional reals. The inputs x_t are contained in X which is a subset of the k-dimensional reals. The errors e_t are assumed independent and normally distributed with mean zero and unknown variance σ^2.

The Likelihood Ratio test and the large sample distribution of the test statistic are obtained in Section 3. The power function obtained from this distribution is tabulated at selected departures from the null hypothesis and selected sample sizes in Section 4. Monte-Carlo estimates of power are compared with the large sample values for an exponential model in Section 5 in order to gain a feel for the adequacy of the large sample approximation in smaller samples. Section 6 contains summary and concluding remarks.

The results presented in this paper are for the case σ^2 unknown. If σ^2 is known, the large sample distribution of the Likelihood Ratio test statistic is obtained, but not tabulated, in [2;3].

2. NOTATION AND ASSUMPTIONS

The following notation will be useful in the remainder of the paper.

<u>Notation:</u> Given the regression model

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

where $\theta \in \Omega \subset R^p$, the observations

$$(y_t, x_t) \qquad (t = 1, 2, \ldots, n),$$

and the hypothesis of location

$$H: \ \theta = \theta_o \quad \text{against} \quad A: \ \theta \neq \theta_o$$

we define:

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

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

$$e = (e_1, e_2, \ldots, e_n)' \qquad (n \times 1),$$

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

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

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

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

$$\delta = f(\theta) - f(\theta_o) \qquad (n \times 1),$$

$$\lambda_1 = \delta' P \delta / (2\sigma^2),$$

$$\lambda_2 = \delta' P^{\perp} \delta / (2\sigma^2),$$

$$\hat{\sigma}^2(y) = \inf_\Omega \frac{1}{n} \Sigma_{t=1}^{n}\{y_t - f(x_t,\theta)\}^2 ,$$

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

$g(t;v,\lambda)$ = the non-central chi-squared density function with v degrees freedom and non-centrality λ [4, p. 74],

$$G(x;v,\lambda) = \int_0^x g(t;v,\lambda)dt ,$$

$n(t;\mu,\sigma^2)$ = the normal density function with mean μ and variance σ^2,

$$N(x;\mu,\sigma^2) = \int_{-\infty}^{x} n(t;\mu,\sigma^2)\, dt ,$$

$p(i,\lambda)$ = the Poisson density function with mean λ.

In order to obtain asymptotic results, it is necessary to specify the behavior of the inputs x_t as n becomes large. A general way of specifying the limiting behavior of nonlinear regression inputs is due to Malinvaud [6]. Malinvaud's definitions are repeated below for the readers convenience; a more complete discussion and examples are contained in his paper.

<u>Definition</u>. Let $\mathcal{Q}$ be the Borel subsets of X and $\{x_t\}_{t=1}^{\infty}$ be the sequence of inputs chosen from X. Let $I_A(x)$ be the indicator function of a subset A of X. The measure μ_n on $(X,\mathcal{Q})$ is defined by

$$\mu_n(A) = n^{-1}\Sigma_{t=1}^{n} I_A(x_t)$$

for each $A \in \mathcal{Q}$.

<u>Definition</u>. A sequence of measures $\{\mu_n\}$ on $(X,\mathcal{Q})$ is said to converge weakly to a measure μ on $(X,\mathcal{Q})$ if for every real valued, bounded, continuous

function g with domain X

$$\int g(x)d\mu_n(x) \rightarrow \int g(x)d\mu(x)$$

as $n \rightarrow \infty$.

The assumptions below are used to obtain the large sample distribution of the Likelihood Ratio test statistic.

Assumptions. The parameter space Ω and the set X are compact subsets of the p-dimensional and k-dimensional reals, respectively. The response function $f(x,\theta)$ and the partial derivatives $\frac{\partial}{\partial\theta_i} f(x,\theta)$ and $\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta)$ are continuous on $X \times \Omega$. The sequence of inputs $\{x_t\}_{t=1}^{\infty}$ are chosen such that the sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to a measure μ defined over $(X,\mathcal{C})$. The true value of θ, denoted by θ^o, is contained in an open set which, in turn, is contained in Ω. If $f(x,\theta) = f(x,\theta^o)$ except on a set of μ measure zero, it is assumed that $\theta = \theta^o$. The $p \times p$ matrix

$$\text{\textbf{\$}} = [\int \frac{\partial}{\partial\theta_i} f(x,\theta^o) \frac{\partial}{\partial\theta_j} f(x,\theta^o)]$$

is non-singular. As mentioned earlier, the errors $\{e_t\}$ are independent with density $n\{x;o,\sigma^2\}$ where σ^2 is non-zero, finite, and unknown.

These assumptions are patterned after those used by Malinvaud [6] to show that the Maximum Likelihood (least squares) estimator is consistent. In addition, it can be shown [2;3] under these assumptions that a measurable function $\hat{\theta}(y)$ minimizing $(y - f(\theta))'(y - f(\theta))$ over Ω exists and that $\sqrt{n}\,(\hat{\theta}(y) - \theta^0)'$ is asymptotically normally distributed with mean zero and variance-covariance matrix $\sigma^2\,\text{\textbf{\$}}^{-1}$.

The following theorem is proved in [2;3].

Theorem 1. Under the assumptions listed above, the estimator $\hat{\sigma}^2(y)$ is consistent for σ^2 and is characterized by

$$\hat{\sigma}^2(y) = e'P^{\perp}e/n + a_n$$

where $n \cdot a_n$ converges in probability to zero. (The matrix P is evaluated at $\theta = \theta^o$, the true parameter value. There is an N such that $F'(\theta^o)F(\theta^o)$ is non-singular for all $n > N$.)

Assumptions which allow Ω to be an unbounded set and do not require that the second partial derivatives of $f(x,\theta)$ exist yet are sufficient for the conclusion of Theorem 1 are given in [2].

3. LARGE SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO TEST STATISTIC

The Likelihood of the sample y is

$$L(y;\theta,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\tfrac{1}{2}\sigma^{-2}(y - f(\theta))'(y - f(\theta))\} .$$

The Maximum Likelihood estimators under H are $\tilde{\theta} = \theta_o$ and $\tilde{\sigma}^2(y) = n^{-1}(y - f(\theta_o))'(y - f(\theta_o))$; over the entire parameter space they are $\hat{\theta}(y)$ minimizing $(y - f(\theta))'(y - f(\theta))$ over Ω and

$$\hat{\sigma}^2(y) = n^{-1}(y - f(\hat{\theta}(y)))'(y - f(\hat{\theta}(y))) = \inf_{\Omega} n^{-1}(y - f(\theta))'(y - f(\theta)) .$$

The Likelihood Ratio is, therefore,

$$\frac{\max\{L(y,\theta,\sigma^2):\ \theta = \theta_o,\ 0 < \sigma^2 < \infty\}}{\max\{L(y,\theta,\sigma^2):\ \theta \in \Omega\ ,\ 0 < \sigma^2 < \infty\}} = \left[\frac{\tilde{\sigma}^2(y)}{\hat{\sigma}^2(y)}\right]^{-\frac{n}{2}} .$$

Thus, the Likelihood Ratio test has the form: reject the null hypothesis H when

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

is larger than c where $P[T(y) > c \mid \theta = \theta_o] = \alpha$.

The following lemma is needed to prove the main result of this section.

<u>Lemma 1</u>. Under the Assumptions listed in Section 2

$$1/\hat{\sigma}^2(y) = n/e'P^{\perp}e + b_n$$

where $n \cdot b_n$ converges in probability to zero.

<u>Proof</u>. Choose τ such that $0 < \tau < \sigma^2$ and let a_n be as in Theorem 1. Let $\delta > 0$ and $\epsilon > 0$ be given. By Theorem 1, there is an N such that $n > N$ implies $P(K_n) \geq 1-\delta$ where

$$K_n = [\tau < \hat{\sigma}^2(y)] \cap [\tau < e'P^{\perp}e/n] \cap [(1/\tau^2)n|a_n| + (1/\tau^3)n|a_n|^2 < \epsilon]$$

since $\hat{\sigma}^2(y)$ and $e'P^{\perp}e/n$ converge in probability to σ^2 and $n \cdot a_n$ converges in probability to zero. By Taylor's theorem, for e in K_n

$$n(1/\hat{\sigma}^2(y) - n/e'P^{\perp}e) = -(e'P^{\perp}e/n)^{-2}(n \cdot a_n) + (e'P^{\perp}e/n + \lambda\, a_n)^{-3}(n \cdot a_n^2)$$

for some λ between 0 and 1. Thus, $e \in K_n$ implies

$$n|1/\hat{\sigma}^2(y) - n/e'P^{\perp}e| \leq (1/\tau^2)n|a_n| + (1/\tau^3)n|a_n|^2 < \epsilon$$

whence $e \in K_n$ and $n > N$ imply

$$1-\delta \leq P(K_n) \leq P[n|1/\hat{\sigma}^2(y) - n/e'P^{\perp}e| < \epsilon] . \quad \blacksquare$$

<u>Theorem 2</u>. Under the Assumptions listed in Section 2 the Likelihood Ratio test statistic may be characterized by

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

where $n \cdot c_n$ converges in probability to zero and the distribution function of X is:

$$
\begin{cases}
0, & x \leq 1,\ \lambda_2 = 0, \\
\int_0^\infty G(t/[x-1] + 2x\lambda_2/[x-1]^2;\ n-p,\ \lambda_2/[x-1]^2)g(t;p,\lambda_1)dt, & x < 1,\ \lambda_2 > 0, \\
\int_0^\infty N(-t;\ 2\lambda_2,\ 8\lambda_2)\ g(t;p,\lambda_1)dt & x = 1,\ \lambda_2 > 0, \\
1 - \int_0^\infty G(t/[x-1] + 2x\lambda_2/[x-1]^2;\ n-p,\ \lambda_2/[x-1]^2)g(t;p,\lambda_1)dt, & x > 1\ .
\end{cases}
$$

<u>Proof</u>. By the preceeding Lemma

$$
\begin{aligned}
T(y) &= (y - f(\theta_o))'(y - f(\theta_o))/e'P^\perp e + b_n(y - f(\theta_o))'(y - f(\theta_o))/n \\
&= (e+\delta)'(e+\delta)/e'P^\perp e + b_n(e+\delta)'(e+\delta)/n \\
&= X + c_n
\end{aligned}
$$

where δ is evaluated at θ^o. Now $n \cdot c_n = n \cdot b_n(e'e/n + 2\,\delta'e/n + \delta'\delta/n)$ and $n \cdot b_n$ converges in probability to zero. The term $e'e/n$ converges in probability to σ^2 by the Strong Law of Large Numbers. The term $2\,\delta'e/n$ has mean zero and variance $4\,\sigma^2\delta'\delta/n^2$. Since $\{f(x,\theta^o) - f(x,\theta_o)\}^2$ is a continuous function of x we have

$$
\delta'\delta/n = \int \{f(x,\theta^o) - f(x,\theta_o)\}^2 d\mu_n(x) \rightarrow \int \{f(x,\theta^o) - f(x,\theta_o)\}^2 d\mu(x)
$$

as $n \rightarrow \infty$ by the weak convergence of the measures μ_n. Thus, $\mathrm{Var}(2\,\delta'e/n) \rightarrow 0$ and $2\,\delta'e/n$ converges in probability to zero by Chebysheff's inequality. The last term $\delta'\delta/n$ converges to a finite constant as shown above. Thus, $n \cdot c_n$ converges in probability to zero.

Set $z = \frac{1}{\sigma} e$ and $\gamma = \frac{1}{\sigma} \delta$. The random variables $(z_1, z_2, \ldots, z_n)$ are independent with density $n\{t;0,1\}$. For an arbitrary constant a, the random variable $(z + a\gamma)'P(z + a\gamma)$ is a non-central Chi-squared with p degrees freedom and non-centrality $a^2\gamma'P\gamma/2$ since P is idempotent with rank p. Similarly, $(z + b\gamma)'P^{\perp}(z + b\gamma)$ is a non-central Chi-squared with $n-p$ degrees freedom and non-centrality $b^2\gamma'P^{\perp}\gamma/2$. These two random variables are independent because $PP^{\perp} = 0$; see Graybill [4, p. 74ff].

Let $a > 0$.

$$
\begin{aligned}
P[X > a + 1] &= P[(z+\gamma)'(z+\gamma) > (a+1)z'P^{\perp}z] \\
&= P[(z+\gamma)'P(z+\gamma) > az'P^{\perp}z - 2\gamma'P^{\perp}z - \gamma'P^{\perp}\gamma] \\
&= P[(z+\gamma)'P(z+\gamma) > a(z-a^{-1}\gamma)'P^{\perp}(z-a^{-1}\gamma) - (1+a^{-1})\gamma'P^{\perp}\gamma] \\
&= \int_0^{\infty} P[t > a(z-a^{-1}\gamma)'P^{\perp}(z-a^{-1}\gamma) - (1+a^{-1})\gamma'P^{\perp}\gamma] \\
&\quad \times g(t;p,\gamma'P\gamma/2)\,dt \\
&= \int_0^{\infty} P[(z-a^{-1}\gamma)'P^{\perp}(z-a^{-1}\gamma) < (t + (a^{-1}+1)\gamma'P^{\perp}\gamma)/a] \\
&\quad \times g(t;p,\gamma'P\gamma/2)\,dt \\
&= \int_0^{\infty} G(t/a + (a+1)\gamma'P^{\perp}\gamma/a^2;\ n-p,\ \gamma'P^{\perp}\gamma/[2a^2]) \\
&\quad \times g(t;p,\gamma'P\gamma/2)\,dt
\end{aligned}
$$

By substituting $x = a+1$, $\lambda_1 = \gamma'P\gamma/2$, and $\lambda_2 = \gamma'P^{\perp}\gamma/2$ one obtains the form of the distribution function for $x > 1$.

The derivations for the remaining cases are analogous and are omitted. ▮

The large sample approximation of the critical point c will be denoted by c^* and defined by $P[X > c^* | \delta=0] = \alpha$ where X is as in Theorem 2. The point c^* can be obtained from a table of F as follows. When $\delta=0$

$$P[X > c^*] = P[e'Pe/e'P^{\perp}e > c^*-1] = P[F > (n-p)(c^*-1)/p] .$$

Let F_α denote the $\alpha \cdot 100$ percentage point of an F random variable with p numerator degrees freedom and n-p denominator degrees freedom; c^* is given by

$$c^* = 1 + pF_\alpha/(n-p) .$$

Note that is $c^* \leq 1$ then $P[X > c^*] = 1$ when $\delta=0$. It is assumed that $0 < \alpha < 1$ and hence that $c^* > 1$ throughout the rest of the paper.

4. PARTIAL TABULATION OF THE POWER FUNCTION

The conclusion of Theorem 2 states that $n \cdot (T-X)$ converges in probability to zero as n tends to infinity. This is a relatively rapid approach of the difference T-X to zero which leads one to expect that the probability $P[X > t]$ would be a good approximation to $P[T > t]$ even in small samples.

The probability $P[X > c^*]$ with c^* chosen for $\alpha = .05$ is tabulated in Tables 1 through 9 for values of p, n, λ_1 and λ_2 thought to be representative of those occurring most frequently in applications.

The details of the numerical evaluation of $P[X > c^*]$ are as follows. The density $g(t;v,\lambda)$ may be put in the form [4, p. 76]

$$g(t;v,\lambda) = \Sigma_{i=0}^{\infty} p(i;\lambda)\ g(t;v+2i,0) .$$

1. Power: $p=2$, $n=30$, $c^* = 1.23860$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.124	.206	.378	.536	.667	.769	.844	.934	.974	.990
.0001	.050	.124	.206	.378	.536	.667	.769	.844	.934	.974	.990
.001	.050	.124	.207	.378	.536	.667	.769	.844	.934	.974	.990
.01	.051	.125	.208	.380	.538	.669	.770	.845	.934	.974	.990
.1	.060	.139	.224	.397	.553	.681	.779	.851	.936	.975	.990

2. Power: $p=3$, $n=30$, $c^* = 1.32893$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.106	.171	.314	.456	.585	.692	.778	.892	.951	.979
.0001	.050	.106	.171	.314	.456	.585	.692	.778	.892	.951	.979
.001	.050	.106	.171	.314	.457	.585	.692	.778	.892	.951	.979
.01	.051	.107	.173	.316	.458	.586	.693	.779	.892	.951	.979
.1	.058	.118	.185	.330	.472	.598	.703	.786	.896	.953	.980

3. Power: $p=5$, $n=30$, $c^* = 1.52060$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.089	.134	.239	.353	.465	.569	.661	.802	.892	.944
.0001	.050	.089	.134	.239	.353	.465	.569	.661	.802	.892	.944
.001	.050	.089	.134	.239	.353	.465	.569	.661	.802	.892	.944
.01	.051	.089	.135	.240	.354	.466	.570	.662	.803	.892	.944
.1	.056	.097	.144	.251	.365	.477	.580	.670	.808	.895	.946

4. **Power:** $p=2$, $n=60$, $c^* = 1.10897$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.129	.218	.399	.563	.695	.795	.866	.947	.980	.993
.0001	.050	.129	.218	.399	.563	.695	.795	.866	.947	.980	.993
.001	.050	.129	.218	.399	.563	.695	.795	.866	.947	.980	.993
.01	.051	.130	.219	.401	.564	.696	.796	.866	.947	.980	.993
.1	.060	.144	.235	.417	.578	.707	.803	.871	.948	.981	.993

5. Power: $p=3$, $n=60$, $c^* = 1.14532$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.111	.182	.337	.488	.622	.730	.813	.916	.966	.987
.0001	.050	.111	.182	.337	.488	.622	.730	.813	.916	.966	.987
.001	.050	.111	.182	.337	.489	.622	.730	.813	.916	.966	.987
.01	.051	.112	.183	.338	.490	.623	.731	.813	.917	.966	.987
.1	.059	.124	.198	.355	.505	.636	.740	.820	.920	.967	.987

6. **Power:** $p=5$, $n=60$, $c^* = 1.21664$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.094	.145	.265	.393	.517	.627	.720	.853	.928	.967
.0001	.050	.094	.145	.265	.393	.517	.627	.720	.853	.928	.967
.001	.050	.094	.145	.265	.393	.517	.627	.720	.853	.928	.967
.01	.051	.094	.146	.267	.395	.518	.628	.720	.853	.929	.967
.1	.056	.103	.156	.279	.407	.530	.638	.729	.858	.931	.968

7. Power: $p=2$, $n=120$, $c^* = 1.05337$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.131	.222	.409	.574	.707	.805	.874	.951	.982	.994
.0001	.050	.131	.222	.409	.574	.707	.805	.874	.951	.982	.994
.001	.050	.131	.222	.408	.574	.707	.805	.874	.951	.982	.994
.01	.050	.131	.222	.408	.574	.706	.804	.874	.951	.982	.994
.1	.057	.140	.232	.417	.580	.710	.807	.875	.951	.982	.994

8. Power: $p=3$, $n=120$, $c^* = 1.06762$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.113	.186	.347	.504	.639	.748	.829	.928	.972	.990
.0001	.050	.113	.186	.347	.504	.639	.748	.829	.928	.972	.990
.001	.050	.113	.187	.348	.504	.640	.748	.829	.928	.972	.990
.01	.051	.115	.189	.350	.507	.642	.749	.930	.928	.972	.990
.1	.062	.131	.209	.373	.528	.659	.762	.839	.932	.973	.990

9. Power: $p=5$, $n=120$, $c^* = 1.09925$

λ_2	λ_1										
	0	.5	1	2	3	4	5	6	8	10	12
0.0	.050	.096	.150	.277	.412	.540	.652	.745	.873	.942	.976
.0001	.050	.096	.150	.277	.412	.540	.652	.745	.873	.942	.976
.001	.050	.096	.150	.277	.412	.540	.653	.745	.874	.942	.976
.01	.051	.097	.152	.279	.414	.542	.654	.746	.874	.943	.976
.1	.058	.107	.164	.294	.430	.556	.666	.756	.879	.945	.976

Using this expression and rearranging terms

$$P[X > c^*] = \Sigma_{i=0}^{\infty} \Sigma_{j=0}^{\infty} p(i;\lambda_2/[c^*-1]^2)p(j;\lambda_1)$$

$$\times \int_0^{\infty} G(t/[c^*-1] + 2c^*\lambda_2/[c^*-1]^2;\ n-p + 2i,\ 0)\ g(t;\ p + 2j,\ 0)dt.$$

This expression was evaluated on an IBM 370/165 using the IBM Scientific Subroutine Package [5] subroutines DLGAM, CDTR, and DQL16. A listing of the authors program is available on request.

5. A MONTE-CARLO COMPARISON

In an effort to gain a feel for the adequacy of the large sample approximation

$$P[T > t] \doteq P[X > t]$$

in samples of moderate size, Monte-Carlo estimates of $P[T > c^*]$ were obtained for the model

$$y_t = \theta_1 e^{\theta_2 x_t} + e_t .$$

Thirty inputs were chosen from $X = [0,1]$ by replicating the points $0(.1).7$ three times and the points $.8(.1)1$ twice. The parameter space was taken as $\Omega = [0,1] \times [0,1]$ and the null hypothesis as H: $\theta_0 = (1/2, 1/2)$. For the null hypothesis and selected departures from the null hypothesis, five thousand random samples were generated according to the model with σ^2 taken as .04. The point estimate $\hat{p}$ of $P[T > c^*]$ is, of course, the ratio of the number of times T exceeded c^* to five thousand. The variance of $\hat{p}$ was estimated by

$$Var(\hat{p}) = P[X > c^*] \cdot P[X \leq c^*]/5000 .$$

The results are presented in Table 10.

Certain points should be mentioned about the choice of values of $\theta \neq \theta_0$ in the Monte-Carlo study. The ratio λ_2/λ_1 is minimized (=0) for $\theta \neq (\frac{1}{2}, \frac{1}{2})$ of the form $(\theta_1, \frac{1}{2})$ and, based on a numerical evaluation of λ_1 and λ_2 over Ω, maximized for θ of the form $(\frac{1}{2}, \frac{1}{2}) \pm r(\cos(\frac{5}{8}\pi), \sin(\frac{5}{8}\pi))$. Three points were chosen to be of the first form, and two of the latter form. Further, two sets of points were paired with respect to λ_1. This was done to evaluate the variation in power when λ_2 changes while λ_1 is held fixed.

6. REMARKS

Considering the standard errors of the Monte-Carlo estimates of $P[T > c^*]$, the estimates support the use of $P[X > c^*]$ to approximate power in this instance. Generalizations beyond this statement carry the usual risks of generalizing from Monte-Carlo studies.

In most applications, λ_2 will be quite small relative to λ_1 as in the Monte-Carlo study. This being the case, a value of $P[X > c^*]$ computed with $\lambda_2 = 0$ would be adequate. Note that if $\lambda_2 = 0$, then

$$P[X > c^*] = P[F' > F_\alpha]$$

where F' denotes a non-central F with p numerator degrees freedom, n-p denominator degrees freedom, and non-centrality λ_1[4, p. 77-78]. In other words, the first row of Tables 1 through 9 are a tabulation of the power of the

10. Monte-Carlo Power Estimates for an Exponential Model

Parameters		Non-centralities		Power		
					Monte-Carlo	
θ_1	θ_2	λ_1	λ_2	$P[X > c^*]$	$\hat{p}$	$SE(\hat{p})$
.5	.5	0	0	.050	.0532	.00308
.5398	.5	.9854	0	.204	.2058	.00570
.4237	.6849	.9853	.00034	.204	.2114	.00570
.5856	.5	4.556	0	.727	.7140	.00630
.3473	.8697	4.556	.00537	.728	.7312	.00629
.62	.5	8.958	0	.957	.9530	.00287

F-test. Thus, in most applications, an adequate indication of the power of the Likelihood Ratio test can be obtained from charts of the power of the F-test such as [1] and [7].

One last point might be mentioned. To reject H when $T(y)$ exceeds $c^* = 1 + p\,F_\alpha/(n-p)$ is equivalent to rejecting H when $S(y)$ exceeds F_α where

$$S(y) = \frac{[\tilde{\sigma}^2(y) - \hat{\sigma}^2(y)]/p}{\hat{\sigma}^2(y)/(n-p)} .$$

This form of the Likelihood Ratio test is analagous to the F-test used in linear regression and can be compared directly with tabled F-test critical points. As stated above, the behavior of $S(y)$ under the alternative $A: \theta \neq \theta^o$ will be adequately approximated in most applications by a non-central F with p numerator degrees freedom, $n-p$ denominator degrees freedom, and non-centrality λ_1. The size of λ_2 will give an indication of the adequacy of the approximation.

REFERENCES

[1] Fox, M., "Charts of the Power of the F-test," The Annals of Mathematical Statistics, 27 (June, 1956), 484-97.

[2] Gallant, A. R., "Inference for Nonlinear Models," Institute of Statistics Mimeo Series No. 875, Raleigh, 1973.

[3] Gallant, A. R., "Statistical Inference for Nonlinear Regression Models," Ph.D. dissertation, Iowa State University, 1971.

[4] Graybill, F. A., An Introduction to Linear Statistical Models, Vol. I, New York: McGraw-Hill, 1961.

[5] IBM Corporation, System 1360 Scientific Subroutine Package, Version III, White Plains, New York: IBM Corporation, Technical Publications Department, 1968.

[6] Malinvaud, E., "The Consistency of Nonlinear Regressions," The Annals of Mathematical Statistics, 41 (June, 1970), 956-69.

[7] Pearson, E. S. and Hartley, H. O., "Charts of the Power Function of the Analysis of Variance Tests, Derived from the Non-central F-distribution," Biometrika, 38(1951), 112-30.