The Annals of Statistics
1975, Vol. 3, No. 1, 84-97

THE LOG LIKELIHOOD RATIO IN SEGMENTED REGRESSION[1]

By Paul I. Feder

General Electric Company

This paper deals with the asymptotic distribution of the log likelihood ratio statistic in regression models which have different analytical forms in different regions of the domain of the independent variable.

It is shown that under suitable identifiability conditions, the asymptotic chi square results of Wilks and Chernoff are applicable. It is shown by example that if there are actually fewer segments than the number assumed in the model, then the least squares estimates are not asymptotically normal and the log likelihood ratio statistic is not asymptotically χ^2. The asymptotic behavior is then more complicated, and depends on the configuration of the observation points of the independent variable.

1. Introduction. In many regression situations it is necessary to consider models which are composed of several segments. The segments form a continuous function but may give rise to discontinuities in slope at the transition points between segments. An important special case of these model is that of *broken line regression*. The least squares fitting of such models is complicated by the fact that the transition points are unknown and must be estimated from the data. Hudson [11] has developed an algorithm to obtain the fit iteratively.

The analytical properties of segmented polynomial functions have been studied extensively by mathematicians working in the theory of approximation. They call these segmented polynomial functions *spline functions*. The transition points between the segments are called *knots*. If the locations of the knots are estimated from the data, the splines are said to have *free knots*. An introduction to the literature on spline functions can be had by consulting Greville [8] and Schoenberg [14]. A number of papers on statistical aspects of fitting spline functions to data have appeared recently. See for example Kimeldorf and Wahba [12] and Studden [16].

The main difference in emphasis between spline approximation theory and segmented regression theory is that the segments are considered in spline theory solely for mathematical convenience whereas the motivation for segmented regression models is typically a change in the underlying physical mechanism. However, this difference does not affect the analytical considerations.

A number of papers have studied the asymptotic distribution theory of the least squares estimators in segmented models. Feder [7] considered the relatively

Received April 1971; revised February 1974.

[1] This research was supported at Princeton University by Contract DA-31-124-ARO-(D)-215 and at General Electric Corporate Research and Development. This is a generalization of a part of the author's doctoral dissertation written at Stanford University under the support of National Science Foundation Grant GP-5705.

AMS 1970 *subject classifications.* 62E20, 62J05.

Key words and phrases. Regression, segmented, likelihood ratio testing, asymptotic theory.

general situation where the model consists of r segments, each of which is reasonably arbitrary. See Sections 1 and 2 of [7] for a discussion of the literature relating to segmented regression and for an illustration of some of the technical difficulties that preclude the use of "standard" asymptotic maximum likelihood and regression arguments to study the asymptotic theory. None of the papers in the literature discuss the case in which there are really fewer segments than the number assumed. An important special case is the behavior of the parameter estimates from a two-segment model when there is really just one segment.

A related problem is the determination of the asymptotic distributions of statistics suitable for testing various hypotheses about the parameters. In particular it is of interest to examine the test of the hypothesis that there are fewer segments than appear in the model. The regression parameters then are not all well defined, since spurious parameters are estimated. In particular the estimates of the fictitious changeover points are poorly behaved. Farley and Hinich [4], Brown and Durbin [2], Bacon and Watts [1], and Hinkley [10] discuss procedures for testing the hypothesis that a two-phase regression function in fact consists of just one phase. Hinkley [10] reports on empirical grounds that with a two-phase broken line regression model and with normal observation errors, the null distribution of the likelihood ratio statistic for testing the hypothesis of a one-segment regression is "very close to the χ^2 distribution with 3 degrees of freedom." Quandt [13] reports for the same problem "The distribution of the relevant likelihood ratio λ is analyzed on the basis of the empirical distribution resulting from some sampling experiments. The hypothesis that $-2\log\lambda$ has the χ^2 distribution with the appropriate degrees of freedom is rejected and an empirical table of percentage points is obtained."

In this paper we show that if the true regression is *identified* (Definition 2.1) under the null hypothesis, then the asymptotic theory of Wilks [19] and Chernoff [3] is applicable (Section 3, especially Theorem 3.2).

In order to get a feeling for what happens in the unidentified case, a simple yet instructive example is considered (Section 4) in which the model specifies a two-segment regression but the true regression consists of just one segment. It is shown (for the example) that the parameter estimates are not asymptotically normal. The null distribution of $-2\log\lambda$ is shown to be that of the maximum of a number of correlated χ_1^2 and χ_2^2 random variables and to vary with the configuration of the observation points of the independent variable. It is then shown that this behavior is typical of that in the general two-phase broken line fit.

Insight into the null distribution theory of the likelihood ratio statistic is obtained by looking at the problem geometrically. This approach is also undoubtedly applicable to the study of regression models other than those discussed in this paper.

2. Definition of the model and background discussion. The material in this section is taken from Feder [7]. The notation used is the same as there.

Consider an r phase, segmented regression function of the form

$$\mu(\boldsymbol{\xi}; t) = f_i(\boldsymbol{\theta}_i; t) \qquad \text{for} \quad t \in [\tau_{i-1}, \tau_i] \qquad (i = 1, \cdots, r),$$

where $A \equiv \tau_0 \leqq \cdots \leqq \tau_r \equiv B$. The model's assumptions and restraints are discussed in Section 2 of [7].

For given n, assume that n observations, $X_{n1}, \cdots, X_{nn}$ are taken where

$$X_{ni} = \mu(\boldsymbol{\xi}; t_{ni}) + e_{ni}.$$

Let $s(\boldsymbol{\xi})$ denote $\sum (X_{ni} - \mu(\boldsymbol{\xi}; t_{ni}))^2$.

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_r)'$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_{r-1})'$, and $\boldsymbol{\xi} = (\boldsymbol{\theta}, \boldsymbol{\tau})'$. Let $\boldsymbol{\xi}_0 \equiv (\boldsymbol{\theta}^{(0)}, \boldsymbol{\tau}^{(0)})'$ denote the true state of nature. The segments $f_j(\boldsymbol{\theta}_j^{(0)}; t)$ are abbreviated in the sequal as $f_j^{(0)}(t)$ or simply $f_j^{(0)}$, for notational convenience. Let Θ denote the set of $\boldsymbol{\theta}$'s which lead to functions $\mu(\boldsymbol{\xi}; t)$ satisfying the model's continuity restraints for at least one vector, $\boldsymbol{\tau}$, of transition points. The vector $\boldsymbol{\tau}$ may or may not be uniquely determined. Examples of cases in which $\boldsymbol{\tau}$ is and is not uniquely determined are discussed in Section 3 of [7]. Let Ξ denote the set of corresponding $\boldsymbol{\xi}$'s and let $U = \{\mu(\boldsymbol{\xi}; t); \boldsymbol{\xi} \in \Xi\}$.

It is shown in [7] (Section 3) that under suitable identifiabilityconditions (which imply that no two adjacent f_j's are identical), $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ and $(\hat{\tau}_j - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$, where $\hat{\boldsymbol{\theta}}$ and $\hat{\tau}_j$ are the least squares estimators (l.s.e.'s) of $\boldsymbol{\theta}$ and τ_j, and m_j is the lowest order t-derivative in which $f_j^{(0)}$ and $f_{j+1}^{(0)}$ differ at $t = \tau_j^{(0)}$. Furthermore, if ω is a subset of Ξ, $\bar{\omega}$ is its closure and $\boldsymbol{\xi}^{(0)} \in \bar{\omega}$, then the above assertion applies equally well to $\hat{\boldsymbol{\xi}}_\omega$, the l.s.e. among all $\boldsymbol{\xi} \in \bar{\omega}$.

A *pseudo problem* is formed by deleting all of the observations in intervals $L_j(n)$, $j = 1, \cdots, r-1$ of length $d_j(n)$ about each of the $\tau_j^{(0)}$. The intervals $L_j(n)$ are chosen so that $d_j(n) \to 0$ but $(n/\log\log n)^{(1/2m_j)} d_j(n) \to \infty$. See [7], Section 4, for a detailed definition and description of the pseudo problem. Let $\hat{\boldsymbol{\xi}}^* \equiv (\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\tau}}^*)$ denote the l.s.e. in the pseudo problem (abbreviated p.l.s.e.). It is shown in Section 4 of [7] that under suitable identifiability assumptions, $\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$, $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = o_p(n^{-\frac{1}{2}})$, $(\hat{\tau}_j^* - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}})$, $(\hat{\tau}_j^* - \tau_j^{(0)})^{m_j} - (\hat{\tau}_j - \tau_j^{(0)})^{m_j} = o_p(n^{-\frac{1}{2}})$. This implies that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$, $(\hat{\tau}_j - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}})$, and that $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}^*$ have the same asymptotic distribution. If $\boldsymbol{\theta}^{(0)}$ is an interior point of Θ then the asymptotic distribution is normal. In particular, this is the case if $m_1 = \cdots = m_{r-1} = 1$.

We now present several definitions that will be used in the next section.

DEFINITION 2.1. (Definition 3.1 of [7]). The parameter $\boldsymbol{\theta}$ is *identified* at $\boldsymbol{\mu}^{(0)}$ by the vector $\mathbf{t} = (t_1, t_2, \cdots, t_k)'$ if the system of k simultaneous equations $\boldsymbol{\mu}(\boldsymbol{\xi}; \mathbf{t}) = \boldsymbol{\mu}^{(0)}$ uniquely determines $\boldsymbol{\theta}^{(0)}$.

Let $F_n(s_2) - F_n(s_1) = n^{-1}\{$number of observations in $(s_1, s_2]\}$. Assume that the t_{ni} are selected to satisfy the

Hypothesis. $F_n(s) \to F(s)$ in distribution, where $F(s)$ is a distribution function with $F(0) = 0$, $F(1) = 1$.

DEFINITION 2.2. (Definition 3.3 of [7]). A *center of observations* is a point of increase of F.

Let $\Delta_j^{(0)}$ denote the set of τ's such that $f_j(\boldsymbol{\theta}_j^{(0)}; \tau) = f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; \tau)$ and which lie to the right of those centers of observation that uniquely determine $\boldsymbol{\theta}_j$ and to the left of those centers of observation that uniquely determine $\boldsymbol{\theta}_{j+1}$. (For brevity, one can describe $\Delta_j^{(0)}$ as the set of τ's which are *compatible* with the centers of observation.)

DEFINITION 2.3. (Definition 3.9 of [7]). The parameter $\boldsymbol{\theta}$ is *well identified* at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ if (i) $\boldsymbol{\theta}$ is identified at $\mu^{(0)}$ by $\mathbf{t}$, (ii) for each j, $1 \leqq j \leqq r-1$, $\Delta_j^{(0)}$ is the one point set, $\{\tau_j^{(0)}\}$.

3. Asymptotic distribution of $-2\log\lambda$. We consider the problem of determining the asymptotic distribution of statistics suitable for testing hypotheses about $\boldsymbol{\xi}$. More precisely, suppose it is desired to test $H_1\colon \boldsymbol{\xi} \in \omega_1$ vs. $H_2\colon \xi \in \omega_2$, where ω_1 and ω_2 are disjoint subsets of Ξ. In analogy with normal theory likelihood ratio testing, define $\hat{\boldsymbol{\xi}}_1 \equiv \hat{\boldsymbol{\xi}}_{\omega_1}$, $\hat{\boldsymbol{\xi}}_2 \equiv \hat{\boldsymbol{\xi}}_{\omega_2}$, $\lambda = [s(\hat{\boldsymbol{\xi}}_2)/s(\hat{\boldsymbol{\xi}}_1)]^{n/2}$. Similarly, define $\hat{\boldsymbol{\xi}}_1^*$, $\hat{\boldsymbol{\xi}}_2^*$, λ^* for the pseudo problem. Let n, n^* be the sample sizes in the original and pseudo problems respectively and let $n^{**} = n - n^*$. Let $\sum$, $\sum^*$ represent summation in the original and pseudo problems respectively and let $\sum^{**} = \sum - \sum^*$. Suppose $\boldsymbol{\xi}^{(0)} \in \bar{\omega}_1 \cap \bar{\omega}_2$.

It is shown in Theorem 3.1 below that $-2\log\lambda$ and $-2\log\lambda^*$ have the same asymptotic distribution. For ease of notation in the discussion below, define $\nu_{ni} = \mu(\boldsymbol{\xi}; t_{ni}) - \mu(\boldsymbol{\xi}^{(0)}; t_{ni})$.

THEOREM 3.1. *If $\boldsymbol{\xi}$ is well identified at $\boldsymbol{\mu}_0^{(0)}$ by $\mathbf{t}$ and the components of $\mathbf{t}$ are centers of observations then* $\log\lambda = \log\lambda^* + o_p(1)$.

PROOF. The hypotheses of the theorem imply that the conditions for Corollaries 3.17 and 3.20 of [7] are satisfied. Corollaries 3.17 and 3.20 of [7] imply that for $i = 1, 2$ $\hat{\boldsymbol{\theta}}_{\omega_i} - \boldsymbol{\theta}^{(0)}$, $\hat{\boldsymbol{\theta}}_{\omega_i}^* - \boldsymbol{\theta}^{(0)}$, $(\hat{\tau}_{\omega_i,j} - \tau_j^{(0)})^{m_j}$, and $(\hat{\tau}_{\omega_i,j}^* - \tau_j^{(0)})^{m_j}$ are all $O_p(n^{-\frac{1}{2}}(\log\log \mathrm{n})^{\frac{1}{2}})$. Select a sequence a_n such that $a_n(\log\log n)^{-\frac{1}{2}} \to \infty$ and $a_n = o((n/n^{**})^{\frac{1}{2}})$. This is possible since $n^{**} = o(n/\log\log n)$ ([7], Section 4). Let $V_n = \{\boldsymbol{\xi} \in \Xi\colon |\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}| < a_n n^{-\frac{1}{2}}, \tau_j \in L_j(n), j = 1, \cdots, r-1\}$. Then $\hat{\boldsymbol{\xi}}_i$, $\hat{\boldsymbol{\xi}}_i^*$ for $i = 1, 2$ lie in V_n with large probability as $n \to \infty$. From the definitions of $s(\boldsymbol{\xi})$, X_{ni}, and ν_{ni} it follows that $s(\boldsymbol{\xi}) = \sum (e_{ni} - \nu_{ni})^2$, $s^*(\xi) = \sum^* (e_{ni} - \nu_{ni})^2$. Thus

$$
\begin{aligned}
(3.1)\qquad s(\boldsymbol{\xi}) &= s^*(\boldsymbol{\xi}) + \textstyle\sum^{**} (e_{ni} - \nu_{ni})^2 \\
&= s^*(\boldsymbol{\xi}) + \textstyle\sum^{**} e_{ni}^2 - 2\sum^{**} e_{ni}\nu_{ni} + \sum^{**} \nu_{ni}^2 .
\end{aligned}
$$

It follows from the definition of V_n and the continuity of the functions $f_{jk}(t)$ that

$$
\sup_{\boldsymbol{\xi}\in V_n} \max_t |\nu(\boldsymbol{\xi}; t)| = O(a_n n^{-\frac{1}{2}}) .
$$

Therefore $\sup_{\boldsymbol{\xi}\in V_n} \sum^{**} \nu_{ni}^2 = O(n^{**}a_n^{\,2}n^{-1}) = o(1)$. Lemma 4.11 of [7] implies $\sup_{\boldsymbol{\xi}\in V_n} |\sum^{**} e_{ni}\nu_{ni}| \leqq \sup_{\boldsymbol{\xi}\in V_n} \{\max_i |\nu_{ni}|\} O_p(n^{**\frac{1}{2}}) = O(a_n n^{-\frac{1}{2}}) O_p(n^{**\frac{1}{2}}) = o_p(1)$.

We thus have, for $\boldsymbol{\xi} \in V_n$

$$s(\boldsymbol{\xi}) = s^*(\boldsymbol{\xi}) + \textstyle\sum^{**} e_{ni}^2 + o_p(1) \tag{3.2}$$

where $o_p(1)$ is uniformly small for $\boldsymbol{\xi} \in V_n$.

By definition of λ

$$\log \lambda = \frac{n}{2} \log \frac{s(\hat{\boldsymbol{\xi}}_2)}{s(\hat{\boldsymbol{\xi}}_1)} \equiv \frac{n}{2} \log \frac{s_2}{s_1} ,$$

and similarly for λ^*, so that

$$\log \lambda - \log \lambda^* = \frac{n^*}{2}\left[\log \frac{s_2}{s_1} - \log \frac{s_2{}^*}{s_1{}^*}\right] + \frac{n^{**}}{2} \log \frac{s_2}{s_1} . \tag{3.3}$$

Since $\hat{\boldsymbol{\xi}}_k$, $\hat{\boldsymbol{\xi}}_k{}^*$ for $k = 1, 2$ lie in V_n with large probability as $n \to \infty$, it follows that for $k = 1, 2$

$$\begin{aligned} s_k &\equiv \textstyle\sum (e_{ni} - \hat{\nu}(\boldsymbol{\xi}_k; t_{ni}))^2 = \sum e_{ni}^2 - 2 \sum e_{ni} \hat{\nu}_{kni} + \sum \hat{\nu}_{kni}^2 \\ &= \textstyle\sum e_{ni}^2 + O_p(a_n) + O_p(a_n{}^2) = \sum e_{ni}^2 + O_p(a_n{}^2) . \end{aligned} \tag{3.4a}$$

This implies

$$\begin{aligned} n^{**} \log s_2/s_1 &= n^{**} \log [(\textstyle\sum e_{ni}^2 + O_p(a_n{}^2))/(\sum e_{ni}^2 + O_p(a_n{}^2))] \\ &= n^{**} \log [1 + O_p(n^{-1}a_n{}^2)] = o_p(1) . \end{aligned}$$

Since $\hat{\boldsymbol{\xi}}_i$ and $\hat{\boldsymbol{\xi}}_i{}^*$, $i = 1, 2$ are restricted l.s.e.'s (and p.l.s.e.'s), $s(\hat{\boldsymbol{\xi}}_i) \leqq s(\hat{\boldsymbol{\xi}}_i{}^*)$ and $s^*(\hat{\boldsymbol{\xi}}_i{}^*) \leqq s^*(\hat{\boldsymbol{\xi}}_i)$. Thus, from (3.2)

$$\begin{aligned} 0 &\leqq s(\hat{\boldsymbol{\xi}}_i{}^*) - s(\hat{\boldsymbol{\xi}}_i) \\ &= s^*(\hat{\boldsymbol{\xi}}_i{}^*) + \textstyle\sum^{**} e_{ni}^2 + o_p(1) - s^*(\hat{\boldsymbol{\xi}}_i) - \sum^{**} e_{ni}^2 + o_p(1) \leqq o_p(1) . \end{aligned}$$

Therefore

$$s(\hat{\boldsymbol{\xi}}_i{}^*) = s(\hat{\boldsymbol{\xi}}_i) + o_p(1) \tag{3.4b}$$

$$s^*(\hat{\boldsymbol{\xi}}_i{}^*) = s^*(\hat{\boldsymbol{\xi}}_i) + o_p(1) . \tag{3.4c}$$

Equations (3.2), (3.3), and (3.4c) imply that

$$\begin{aligned} \log \lambda - \log \lambda^* &= \frac{n^*}{2}\left\{\log \frac{s_2}{s_1} - \log \frac{s_2 - \sum^{**} e_{ni}^2 + o_p(1)}{s_1 - \sum^{**} e_{ni}^2 + o_p(1)}\right\} + o_p(1) \\ &= -\frac{n^*}{2} \log\left[\frac{1 - s_2{}^{-1} \sum^{**} e_{ni}^2 + o_p(n^{-1})}{1 - s_1{}^{-1} \sum^{**} e_{ni}^2 + o_p(n^{-1})}\right] + o_p(1) . \end{aligned}$$

This, with equation (3.4a) implies that

$$\begin{aligned} \log \lambda - \log \lambda^* &= -\frac{n^*}{2} \log\left[\frac{(1 - \sum^{**} e_{ni}^2/\sum e_{ni}^2) + O_p(n^{**}n^{-2}a_n{}^2) + o_p(n^{-1})}{(1 - \sum^{**} e_{ni}^2/\sum e_{ni}^2) + O_p(n^{**}n^{-2}a_n{}^2) + o_p(n^{-1})}\right] \\ &\quad + o_p(1) \\ &= -\frac{n^*}{2} \log [1 + O_p(n^{**}n^{-2}a_n{}^2) + o_p(n^{-1})] + o_p(1) \\ &= O_p(n^{**}n^{-1}a_n{}^2) + o_p(1) = o_p(1) . \qquad \square \end{aligned}$$

Since $-2 \log \lambda$ has the same asymptotic distribution as $-2 \log \lambda^*$, it suffices to consider the latter. From now on the asterisks will be dropped for notational convenience, keeping in mind that the discussion refers to the pseudo problem.

For the sake of simplicity we restrict attention here to the case when the assumptions of Theorem 3.1 are satisfied, $\boldsymbol{\theta}^{(0)}$ is interior to Θ, ω_1 is locally a hyperplane, and ω_2 is its complement. The generalizations to ω_1, ω_2 being cones, as in Chernoff [3], or to $\boldsymbol{\theta}^{(0)}$ a boundary point of Θ, are reasonably straightforward and lead to similar and related results. From Theorem 4.13 of [7], $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{G}^{-1}$, where $\mathbf{G}$ is the positive definite information matrix with

$$G_{jk} = \int_0^1 \frac{\partial\mu(\boldsymbol{\xi}^{(0)}; t)}{\partial\theta_j} \frac{\partial\mu(\boldsymbol{\xi}^{(0)}; t)}{\partial\theta_k} dF(t) .$$

Note that if $m_1 = \cdots = m_{r-1} = 1$, as is the case in broken line regression, then $\boldsymbol{\theta}^{(0)}$ is interior to Θ.

Since the changeover points $\tau_1, \cdots, \tau_{r-1}$ are functions of $\boldsymbol{\theta}$, the hypothesis spaces can be equivalently expressed as $H_1\colon \boldsymbol{\theta} \in \rho_1$ vs. $H_2\colon \boldsymbol{\theta} \in \rho_2$, where ρ_1, ρ_2 are the projections of ω_1, ω_2 onto the $\boldsymbol{\theta}$-space.

THEOREM 3.2. *Suppose* (i) $\boldsymbol{\xi}$ *is well identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$ *and the components of* $\mathbf{t}$ *are centers of observation.*

(ii) $\boldsymbol{\theta}^{(0)}$ *is interior to* Θ *which is locally a* q*-dimensional Euclidean space.*

(iii) ρ_1 *is locally an* m*-dimensional linear subspace of* Θ *and* ρ_2 *its complement. Then* $-2 \log \lambda$ *converges in distribution to* χ^2_{q-m} *as* $n \to \infty$.

PROOF. Since $\hat{\boldsymbol{\theta}}_{\omega_1} - \boldsymbol{\theta}^{(0)}$ and $\hat{\boldsymbol{\theta}}_{\omega_2} - \boldsymbol{\theta}^{(0)}$ are $O_p(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}})$, the restricted p.l.s.e.'s $\hat{\tau}_{j,\omega_1}$, $\hat{\tau}_{j,\omega_2}$ fall within $L_j(n)$ with large probability for n large, $j = 1, \cdots, r-1$. In this region $s(\boldsymbol{\xi})$ is a quadratic function of $\boldsymbol{\theta}$ and does not depend on $\boldsymbol{\tau}$. For the remainder of the proof denote $s(\boldsymbol{\xi})$ by $s(\boldsymbol{\theta})$. Assumption (iii) implies that under the null hypothesis $\boldsymbol{\theta}^{(0)} \in \bar{\rho}_1 \cap \bar{\rho}_2$. Thus $s(\hat{\boldsymbol{\theta}}_{\omega_i}) \leqq s(\boldsymbol{\theta}^{(0)})$, $i = 1, 2$ and so we obtain for $i = 1, 2$

$$\begin{aligned} (3.5)\qquad 0 &\geqq s(\hat{\boldsymbol{\theta}}_{\omega_i}) - s(\boldsymbol{\theta}^{(0)}) \\ &= (\hat{\boldsymbol{\theta}}_{\omega_i} - \boldsymbol{\theta}^{(0)})' \frac{\partial s(\boldsymbol{\theta}^{(0)})}{\partial \boldsymbol{\theta}} + \tfrac{1}{2}(\hat{\boldsymbol{\theta}}_{\omega_i} - \boldsymbol{\theta}^{(0)})' \frac{\partial^2 s(\boldsymbol{\theta}^{(0)})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}} (\hat{\boldsymbol{\theta}}_{\omega_i} - \boldsymbol{\theta}^{(0)}) . \end{aligned}$$

The matrix $(2n)^{-1}\partial^2 s(\boldsymbol{\theta}^{(0)})/\partial\boldsymbol{\theta}\, \partial\boldsymbol{\theta}$ converges to the $q \times q$ information matrix $\mathbf{G}$ where $G_{jk} = \int_0^1 [\partial\mu(\boldsymbol{\xi}^{(0)}; t)]/\partial\theta_j \cdot [\partial\mu(\boldsymbol{\xi}^{(0)}; t)]/\partial\theta_k\, dF(t)$. The proof of Lemma 3.15 of [7] and continuity imply that $\mathbf{G}$ is positive definite. The vector $n^{-\frac{1}{2}}\partial s(\boldsymbol{\theta}^{(0)})/\partial\boldsymbol{\theta}$ has mean $\mathbf{0}$ and uniformly bounded covariance matrix as $n \to \infty$. Thus $\partial^2 s(\boldsymbol{\theta}^{(0)})/\partial\boldsymbol{\theta}\, \partial\boldsymbol{\theta} \approx 2n\mathbf{G}$ and $\partial s(\boldsymbol{\theta}^{(0)})/\partial\boldsymbol{\theta} = O_p(n^{\frac{1}{2}})$ as $n \to \infty$.

Since the right-hand part of equation (3.5) is non-positive it necessarily follows from the above order calculations that $\boldsymbol{\theta}_{\omega_i} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$, $i = 1, 2$.

For all $\boldsymbol{\theta}$ with $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ sufficiently close to 0

$$(3.6)\qquad s(\boldsymbol{\theta}) = s(\hat{\boldsymbol{\theta}}) + \frac{n}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \left\{n^{-1} \frac{\partial s^2(\boldsymbol{\theta}^{(0)})}{\partial\boldsymbol{\theta}\, \partial\boldsymbol{\theta}}\right\} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) .$$

The second derivative matrix $n^{-1}\partial^2 s(\boldsymbol{\theta}^{(0)})/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}$ is $2G + o(1)$ and is thus positive definite for n sufficiently large.

Equation (3.6) implies that

$$s(\hat{\boldsymbol{\theta}}_{\omega_i}) = s(\hat{\boldsymbol{\theta}}) + n \inf_{\boldsymbol{\theta}\in\rho_i} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + o_p(1) \qquad i = 1, 2 . \tag{3.7}$$

Define $R_i = n \inf_{\boldsymbol{\theta}\in\rho_i} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$, $i = 1, 2$. Since ρ_2 is dense in Θ, $R_2 = 0$. Since $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$ converges in distribution to $N(\mathbf{0}, \sigma^2\mathbf{G}^{-1})$ and ρ_1 is locally an m-dimensional hyperplane, it follows that R_1 converges in distribution to $\sigma^2\chi^2_{q-m}$. Furthermore, $n^{-1}s(\hat{\boldsymbol{\theta}}) = n^{-1}\sum e_{ni}^2 + O_p(n^{-1}) = \sigma^2 + o_p(1)$.

Equation (3.7) implies

$$\begin{aligned} -2\log\lambda \equiv n \log \frac{s(\hat{\boldsymbol{\theta}}_{\omega_1})}{s(\hat{\boldsymbol{\theta}}_{\omega_2})} &= n \log\left[\frac{1 + R_1/s(\hat{\boldsymbol{\theta}}) + o_p(n^{-1})}{1 + R_2/s(\hat{\boldsymbol{\theta}}) + o_p(n^{-1})}\right] \\ &= n[(R_1/s(\hat{\boldsymbol{\theta}}) + o_p(n^{-1})] \end{aligned} \tag{3.8}$$

since $R_2 = 0$. Equation (3.8) and the paragraph immediately above it imply the statement of the theorem.

REMARK 1. Equation (3.8) suggests that $-2\log\lambda$ may be better approximated by an $F_{q-m,n-q}$ distribution.

REMARK 2. The local power of the likelihood ratio test is obtained from the noncentral χ^2 distribution as discussed in Wald [18] or Feder [6].

We illustrate Theorem 3.2 with a simple application to broken line regression. Suppose that $\mu(\boldsymbol{\theta}; t) = \theta_{11} + \theta_{12}t$ for $0 \leqq t \leqq \tau$ and $\theta_{21} + \theta_{22}t$ for $\tau \leqq t \leqq 1$. Here $q = 4$. The continuity restraint is $\theta_{11} + \theta_{12}\tau = \theta_{21} + \theta_{22}\tau$. Obviously $\partial\mu(\boldsymbol{\theta}^{(0)}; t)/\partial\theta_{11} = 1 \cdot (0 \leqq t \leqq \tau^{(0)})$, $\partial\mu(\boldsymbol{\theta}^{(0)}; t)/\partial\theta_{12} = t \cdot (0 \leqq t \leqq \tau^{(0)})$, $\partial\mu(\boldsymbol{\theta}^{(0)}; t)/\partial\theta_{21} = 1 \cdot (\tau^{(0)} \leqq t \leqq 1)$, $\partial\mu(\boldsymbol{\theta}^{(0)}; t)/\partial\theta_{22} = t \cdot (\tau^{(0)} \leqq t \leqq 1)$, where $(\cdots)$ represents indicator function notation. This implies that $\mathbf{G}$ is the 4×4 matrix whose (i, j)th element is $\int_0^{\tau^{(0)}} t^{i+j-2}\,dF(t)$ if $i \leqq 2$ and $j \leqq 2$, $\int_{\tau^{(0)}}^1 t^{i+j-6}\,dF(t)$ if $i \geqq 3$ and $j \geqq 3$, and 0 otherwise.

Suppose that it is desired to test the hypothesis $H: \boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}$. Then $\rho_1 = \{\boldsymbol{\theta}^{(0)}\}$, $\rho_2 = \Theta - \{\boldsymbol{\theta}^{(0)}\}$ and $\mathscr{L}\{-2\log\lambda\} \to \chi_4^2$. If it is desired to test the hypothesis $H: \tau = \tau^{(0)}$, then $\rho_1 = \{\boldsymbol{\theta}: \theta_{11} + \theta_{12}\tau^{(0)} = \theta_{21} + \theta_{22}\tau^{(0)}\}$. This is locally a three-dimensional hyperplane passing through $\boldsymbol{\theta}^{(0)}$. Thus $L\{-2\log\lambda\} \to \chi_1^2$.

4. The null case. We first discuss an example of a two-segment model where the underlying regression consists of but one segment. The example is simple but instructive. It turns out that the parameter estimates are *not* asymptotically normal and $-2\log\lambda$ is *not* asymptotically χ^2 with the "appropriate" number of degrees of freedom. Instead, the distribution of $-2\log\lambda$ is that of the maximum of a number of correlated chi square variates. It is then shown that this type of behavior applies in general when there is really just one segment in a two-segment model.

Hartigan [9] and Shorack [15] discuss other situations where the distribution of $-2\log\lambda$ is not asymptotically χ^2 with the "appropriate" number of degrees

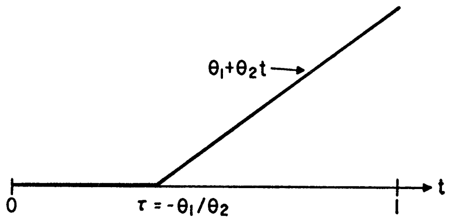

FIG. 4.1.

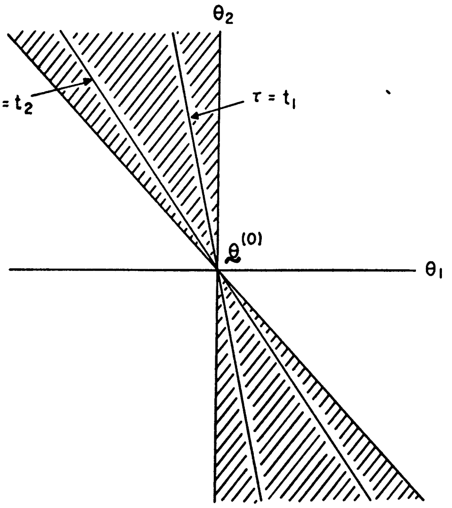

FIG. 4.2.

of freedom. They are concerned with models in which the parameters are constrained by inequality relationships. The principal difference between their models and the one discussed here is that their regression spaces are convex sets whereas those that arise in this section are not. This lack of convexity complicates the distribution theory somewhat, as will be seen in the example below.

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ and $\mu(\boldsymbol{\theta}; t) = 0$ if $0 \leqq t \leqq \tau$, $\mu(\boldsymbol{\theta}, t) = \theta_1 + \theta_2 t$ if $\tau \leqq t \leqq 1$. Θ consists of the subset of Euclidean 2-space contained between the lines $\theta_1 = 0$ and $\theta_2 = -\theta_1$. The parameter values on the two-sided ray $\theta_1/\theta_2 = -\tau$ correspond to those regression functions with given τ. Note that all points along the line $\theta_2/\theta_1 = -1$ correspond to the regression fnnction $\mu(\boldsymbol{\theta}; t) \equiv 0$. Suppose $\theta_1^{(0)} = \theta_2^{(0)} = 0$ and it is desired to test the hypothesis $H_0\colon \theta_1 = \theta_2 = 0$.

Suppose $Y_{ij} = \mu(\boldsymbol{\theta}^{(0)}; t_i) + e_{ij}$ where $e_{ij} \sim \mathcal{N}(0, 1)$ and suppose that γn, $\bar{\gamma} n$ independent observations are taken at t_1, t_2 respectively where $0 < \gamma < 1$ and $\bar{\gamma} = 1 - \gamma$. Let $X_1 = (\gamma n)^{-1} \sum Y_{1j}$, $X_2 = (\bar{\gamma} n)^{-1} \sum Y_{2j}$. Then the residual sum of squares is

$$\begin{aligned} s(\theta_1, \theta_2) &= S + \gamma n(X_1 - \theta_1 - \theta_2 t_1)^2 + \bar{\gamma} n(X_2 - \theta_1 - \theta_2 t_2)^2 & 0 \leqq \tau < t_1 \\ &= S + \gamma n X_1^2 + \bar{\gamma} n(X_2 - \theta_1 - \theta_2 t_2)^2 & t_1 \leqq \tau < t_2 \\ &= S + \gamma n X_1^2 + \bar{\gamma} n X_2^2 & \tau \geqq t_2 \end{aligned}$$

where $S = \sum (Y_{1j} - X_1)^2 + \sum (Y_{2j} - X_2)^2 = n[1 + O_p(n^{-\frac{1}{2}})]$. We can eliminate the third case from consideration, since if $\theta_1 + \theta_2 t_2 = X_2$ then $s(\theta_1, \theta_2) = S + \gamma n X_1^2$. Thus $\hat{\tau} \leqq t_2$ always.

Three types of fits to the data can occur, corresponding to (1) $0 < \hat{\tau} < t_1$, (2) $\hat{\tau} = 0$, or (3) $t_1 < \hat{\tau} < t_2$. These three cases occur if

(1) $X_2/X_1 > t_2/t_1$
(2) $-(t_2/t_1)[(1 + (\gamma/\bar{\gamma})(t_1/t_2)^2)^{\frac{1}{2}} - 1] < X_2/X_1 \leqq t_2/t_1$
(3) $X_2/X_1 \leqq -(t_2/t_1)[(1 + (\gamma/\bar{\gamma})(t_1/t_2)^2)^{\frac{1}{2}} - 1]$.

TABLE 4.1

Case 1	Case 2	Case 3
$\hat{\theta}_1 = \dfrac{X_1 t_2 - X_2 t_1}{t_2 - t_1}$	$\hat{\theta}_1 = 0$	$\hat{\theta}_1 = -\dfrac{X_2 \hat{\tau}}{t_2 - \hat{\tau}}$
$\hat{\theta}_2 = \dfrac{X_2 - X_1}{t_2 - t_1}$	$\hat{\theta}_2 = \dfrac{t_1 \gamma X_1 + t_2 \bar{\gamma} X_2}{t_1^2 \gamma + t_2^2 \bar{\gamma}}$	$\hat{\theta}_2 = \dfrac{X_2}{t_2 - \hat{\tau}}$
$\hat{\tau} = \dfrac{X_2 t_1 - X_1 t_2}{X_2 - X_1}$	$\hat{\tau} = 0$	$t_1 \leqq \hat{\tau} < t_2$
$s(\hat{\theta}_1, \hat{\theta}_2) = S$	$s(\hat{\theta}_1, \hat{\theta}_2) = S + \dfrac{n\gamma\bar{\gamma}(t_2 X_1 - t_1 X_2)^2}{\gamma t_1^2 + \bar{\gamma} t_2^2}$	$s(\hat{\theta}_1, \hat{\theta}_2) = S + n\gamma X_1^2$

In Case 3 we might put $\hat{\tau} = t_1$, for definiteness.

Let $Z_1 = X_1(\gamma n)^{\frac{1}{2}}$, $Z_2 = X_2(\bar{\gamma} n)^{\frac{1}{2}}$. Then $(Z_1, Z_2) \sim N(\mathbf{0}, \mathbf{I})$. By direct calculation,

$$\begin{aligned}
-2 \log \lambda &= n \log \frac{s(\hat{\boldsymbol{\theta}}_{\omega_1})}{s(\hat{\boldsymbol{\theta}}_{\omega_2})} = n \log\left[\frac{S + (Z_1^2 + Z_2^2)}{S + (s(\hat{\theta}_1, \hat{\theta}_2) - S)}\right] \\
&= n\left[\frac{Z_1^2 + Z_2^2}{S} - \frac{s(\hat{\theta}_1, \hat{\theta}_2) - S}{S} + o_p(n^{-1})\right] \\
&= Z_1^2 + Z_2^2 - [s(\hat{\theta}_1, \hat{\theta}_2) - S] + o_p(1)\,.
\end{aligned}$$

Thus from Table 4.1

$$\begin{aligned}
-2 \log \lambda &= Z_1^2 + Z_2^2 + o_p(1) \qquad \frac{Z_2}{Z_1} > \frac{t_2 \bar{\gamma}^{\frac{1}{2}}}{t_1 \gamma^{\frac{1}{2}}} \\
&= \frac{(t_1 Z_1 \gamma^{\frac{1}{2}} + t_2 Z_2 \bar{\gamma}^{\frac{1}{2}})^2}{\gamma t_1^2 + \bar{\gamma} t_2^2} + o_p(1) \quad -\frac{t_2 \bar{\gamma}^{\frac{1}{2}}}{t_1 \gamma^{\frac{1}{2}}}\left[\left(1 + \frac{t_1^2 \gamma}{t_2^2 \bar{\gamma}}\right)^{\frac{1}{2}} - 1\right] < \frac{Z_2}{Z_1} \leqq \frac{t_2 \bar{\gamma}^{\frac{1}{2}}}{t_1 \gamma^{\frac{1}{2}}} \\
&= Z_2^2 + o_p(1) \qquad \frac{Z_2}{Z_1} \leqq -\frac{t_2 \bar{\gamma}^{\frac{1}{2}}}{t_1 \gamma^{\frac{1}{2}}}\left[\left(1 + \frac{t_1^2 \gamma}{t_2^2 \bar{\gamma}}\right)^{\frac{1}{2}} - 1\right].
\end{aligned}$$

We see from Table 4.1 and the above expressions that $\hat{\boldsymbol{\theta}} = O_p(n^{-\frac{1}{2}})$, $\hat{\tau}$ behaves wildly in the sense that in Case 1 it is the ratio of two random variables with mean 0 and in Case 3 it is not even uniquely determined, $-2 \log \lambda = O_p(1)$, $\hat{\boldsymbol{\theta}}$ is *not* normally distributed, and the distribution of $-2 \log \lambda$ is a mixture of a χ_2^2 random variable and the maximum of two correlated χ_1^2 random variables. This is somewhat smaller than the χ_2^2 distribution that might be naively expected on the basis of the "standard" theory of likelihood ratio testing. This means that a test of H_0 based on the χ_2^2 distribution would be overly conservative.

It is useful to consider the same example from a slightly different viewpoint. This will allow more ready generalization. We again summarize the observations by (Z_1, Z_2) where $L\{(Z_1, Z_2)\} = N(\mathbf{0}, \mathbf{I})$. The problem can be regarded as a least squares regression problem where the regression space is a two-dimensional subset of two-dimensional Euclidean space. Let (η_1, η_2) denote the mean of (Z_1, Z_2), to be estimated by least squares.

$$\begin{array}{lll} \text{If} & 0 \leqq \tau < t_1 & (\eta_1, \eta_2) = \theta_1(\gamma^{\frac{1}{2}}, \bar{\gamma}^{\frac{1}{2}}) + \theta_2(t_1\gamma^{\frac{1}{2}}, t_2\bar{\gamma}^{\frac{1}{2}}) \\ \text{If} & t_1 \leqq \tau < t_2 & (\eta_1, \eta_2) = \bar{\gamma}^{\frac{1}{2}}(\theta_1 + \theta_2 t_2)(0, 1) \\ \text{If} & t_2 \leqq \tau & (\eta_1, \eta_2) = (0, 0)\,. \end{array}$$

The regression space is pictured in Figure 4.3. If $\tau > t_2$ the mean vector is at the origin. If $t_1 \leqq \tau \leqq t_2$ the mean vector lies along the η_2 axis. If $0 \leqq \tau < t_1$ the mean vector is contained in one of the wedge-shaped regions indicated in Figure 4.3. The changeover point, τ, is constant on two-sided rays through the origin in the shaded area.

Since $\theta_1^{(0)} = \theta_2^{(0)} = 0$, the true mean is at the origin. The residual sum of squares is

$$s(\theta_1, \theta_2) = S + (Z_1 - \eta_1)^2 + (Z_2 - \eta_2)^2$$

where $S = \sum (Y_{ij} - X_1)^2 + \sum (Y_{2j} - X_2)^2$. In other words, $s(\theta_1, \theta_2) - S$ is the squared distance from the vector $\mathbf{Z} \equiv (Z_1, Z_2)$ to the regression space.

The symbols ρ, φ, and ω are used below in a context unrelated to their usage earlier in the paper. Recall that $-2 \log \lambda = Z_1^2 + Z_2^2 - (s(\theta_1, \theta_2) - S) + o_p(1)$.

Let (ρ, φ) be the representation of $\mathbf{Z}$ in polar coordinates. Then ρ, φ are independent and $L(\rho^2) = \chi_2^2$, $L(\varphi) = U(0, 2\pi)$.

Let $2\omega = \tan^{-1}(t_2\bar{\gamma}^{\frac{1}{2}}/t_1\gamma^{\frac{1}{2}})$. It can be seen from Figure 4.3 that $\omega - \pi/4 < \varphi < 2\omega$ or $\omega + 3\pi/4 < \varphi < 2\omega + \pi$ implies that $-2 \log \lambda$ is asymptotically the

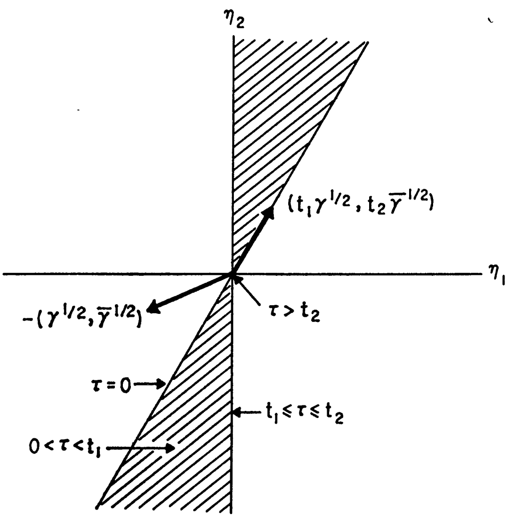

FIG. 4.3.

squared length of the projection of $\mathbf{Z}$ onto the one-dimensional hyperplane (line) $\tau = 0$, $2\omega < \varphi < \pi/2$ or $2\omega + \pi < \varphi < 3\pi/2$ implies that $-2 \log \lambda$ is asymptotically the squared length of $\mathbf{Z}$ itself, and $\pi/2 < \varphi < \omega + 3\pi/4$ or $3\pi/2 < \varphi < \omega + 7\pi/4$ implies that $-2 \log \lambda$ is the squared length of the projection of $\mathbf{Z}$ onto the one-dimensional hyperplane $\tau = t_1$. The second case results in an asymptotic χ_2^2 distribution for $-2 \log \lambda$ and the first and third cases result in asymptotic distributions somewhere between χ_1^2 and χ_2^2. The first and third cases are characterized by $\hat{\tau} = 0$ and $\hat{\tau} = t_1$ respectively.

The information regarding which of the three cases occurs is itself relevant to testing the hypothesis, since Cases 1 and 3 are more likely to occur under H_0 than under the alternative hypothesis. The principle of conditionality asserts that the marginal distribution of $-2 \log \lambda$ as the edge of fit varies should be considered, rather than the conditional distribution given the edge of fit.

If t_2 is bounded away from 0 this marginal distribution is bounded from below by a chi square distribution with one degree of freedom and from above by a chi square distribution with two degrees of freedom. The lower bound is sharp but the upper bound can be improved somewhat.

The extreme cases correspond to $t_1 \gamma^{\frac{1}{2}}/t_2 \bar{\gamma}^{\frac{1}{2}}$ being small or large. Consider t_1, t_2 to be fixed so that the extremes correspond to extreme values of $\gamma/\bar{\gamma}$. As $\gamma/\bar{\gamma} \to 0$ the statistic $-2 \log \lambda$ approaches Z_1^2, which has a chi square distribution with 1 degree of freedom. This occurs when the great majority of the observations are at t_2. The fit is then quite sensitive to X_2, which is very close to 0. The residual sum of squares is then approximately $n\gamma X_1^2 \equiv Z_1^2$. As $\gamma/\bar{\gamma} \to \infty$ the statistic $-2 \log \lambda$ approaches $Z_1^2 + Z_2^2$ when $Z_1 Z_2 > 0$ and $\max(Z_1^2, Z_2^2)$ when $Z_1 Z_2 < 0$. Thus an upper bound on the distribution of $-2 \log \lambda$ is the distribution which is χ_2^2 with probability $\frac{1}{2}$ and the maximum of two independent χ_1^2's with probability $\frac{1}{2}$. This occurs when the great majority of the observations are at t_1. The fit is then quite sensitive to X_1, which is very close to 0. The residual sum of squares is usually 0 when $X_1 X_2 > 0$ and is essentially $\min(Z_1^2, Z_2^2)$ when $X_1 X_2 < 0$.

An interesting feature of this example is that in contrast to the standard likelihood ratio testing situation, the asymptotic distribution of $-2 \log \lambda$ varies with the configuration of the observation points of the independent variable.

The same geometry holds when n distinct observations are taken. Suppose that $\eta(t) = \theta(t - \tau)^+$ and n observations are taken at $t_1, t_2, \cdots, t_n$. The regression set is a two-dimensional subset of n-space, composed of $n - 1$ contiguous segments of two-dimensional hyperplanes such as those in Figure 4.3. Each segment is composed of two-sided rays through the origin. The ith segment consists of all $\eta(t)$ such that $t_{i-1} < \tau < t_i$. (See Figure 4.4.)

The segments are joined by rays through the origin corresponding to $\tau = t_i$, $i = 0, 1, \cdots, n - 1$. The projection of $\mathbf{Z}$ onto the regression space lies either in a segment or on a line, corresponding to whether τ lies between two observations or exactly at an observation. It is seen from Figure 4.4 that the chances

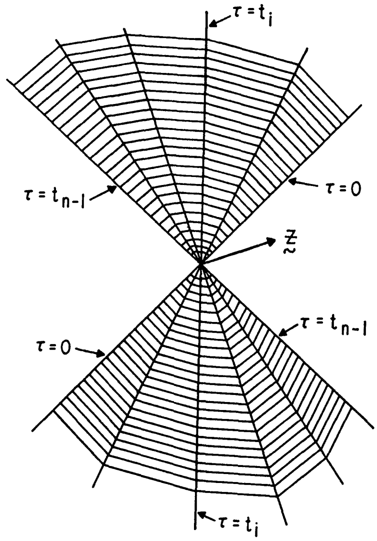

FIG. 4.4.

are greater under the null hypothesis than under the alternative that the projection lies on a line $\tau = t_i$. Thus as with $n = 2$ it is necessary to consider the marginal distribution of the projection length as the edge of fit varies. The observation vector **Z** has an n-dimensional unit normal distribution centered at the origin.

Let $\bar{t}^{(k)}$ denote $(n - k)^{-1} \sum (i \geqq k + 1)t_i$. It is easily shown that the orthogonal projection of **Z** onto the two-dimensional hyperplane which contains the segment $t_k < \tau < t_{k+1}$ is parameterized by $\theta = \sum (i \geqq k + 1)Z_i(t_i - \bar{t}^{(k)})/\sum (i \geqq k + 1) \times (t_i - \bar{t}^{(k)})^2$ and $\tau = [\theta \sum (i \geqq k + 1)t_i - \sum (i \geqq k + 1)Z_i]/(n - k)\theta$. This implies that for any constant c, $\theta(c\mathbf{Z}) = c\theta(\mathbf{Z})$ and $\tau(c\mathbf{Z}) = \tau(\mathbf{Z})$. Similarly for the orthogonal projection of **Z** onto the one-dimensional hyperplane determined by $\tau = t_i$. These results imply that $\hat{\tau}$ is constant for **Z** varying on rays through the origin. Thus, the n-dimensional observation space can be partitioned into cones through the origin in such a way that if **Z** lies in certain cones then $-2 \log \lambda$ is asymptotically the squared norm of the projection of **Z** onto a particular two-dimensional hyperplane and if **Z** lies in the other cones than $-2 \log \lambda$ is asymptotically the squared norm of the projection of **Z** onto one of the one-dimensional subspaces $\tau = t_i$. The location of $\hat{\tau}$ depends only on the direction of **Z**. The conditions that ensure that the projection of **Z** onto the regression space lies in a particular segment or line are:

(1) the projection of **Z** onto the hyperplane containing the segment or line lies in the regression space;

(2) the length of the projection is the maximum over all segments or lines for which (1) is satisfied.

Note that since the regression space is not convex, the Kuhn–Tucker conditions, as discussed in Hartigan [9], are not sufficient. Thus the distribution theory is more complex than simply a mixture of chi squares, as described by Hartigan [9], Section 3 or Shorack [15], Section 2.

The asymptotic distribution of $-2 \log \lambda$ is the distribution of the maximum of a large number of correlated χ_1^2 and χ_2^2 random variables. Hinkley [10] reports on empirical grounds that the distribution resembles a χ_3^2. As $n \to \infty$ the number of random variables over which the maximum is taken increases, but the correlations between pairs of random variables approach 1. Thus some sort of balance is maintained. The precise correlation structure depends on the spacings of the observations and presumably different limiting distributions would result from different spacings of the independent variable.

If τ is arbitrarily set equal to a τ^* (e.g. $\frac{1}{2}$) and the regression is fitted subject to this additional constraint, then $L\{-2 \log\} \to \chi_1^2$. This implies that the asymptotic distribution of $-2 \log \lambda$ is bounded from below by χ_1^2.

Essentially the same results hold for the general two-segment model, $\eta(t) = a + bt + c(t - \tau)^+$. Suppose it is desired to test the hypothesis $H\colon c = 0$ when $c^{(0)} = 0$. In this case $-2 \log \lambda = s(\hat{a}, \hat{b}) - s(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{\tau})$, the difference between the residual sums of squares from the two and four parameter fits respectively. Let $\hat{Z}_i = Z_i - \hat{a} - \hat{b}t_i$. The previous discussion is valid with $\mathbf{Z}$ replaced by $\hat{\mathbf{Z}}$. Thus, $-2 \log \lambda$ is the squared norm of the projection of $\hat{\mathbf{Z}}$ onto a one- or two-dimensional subspace, depending on the direction of $\hat{\mathbf{Z}}$. Again, $L\{-2 \log \lambda\} \geqq \chi_1^2$.

REFERENCES

[1] Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58** 525–534.

[2] Brown, R. L. and Durbin, J. (1968). Methods of investigating whether a regression relationship is constant over time. *European Meeting Amsterdam, Selected Statistical Papers, I.* 37–45.

[3] Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578.

[4] Farley, J. U. and Hinich, M. J. (1970). A test for a shifting slope coefficient in a linear model. *J. Amer. Statist. Assoc.* **65** 1320–1329.

[5] Feder, P. I. (1967). On the likelihood ratio statistic with applications to broken line regression. Technical Report No. 23, Department of Statistics, Stanford Univ.

[6] Feder, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is 'near' the boundaries of the hypothesis regions. *Ann. Math. Statist.* **39** 2044–2055.

[7] Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems—identified case. *Ann. Statist.* **3** 49–83.

[8] Greville, T. N. C., ed. (1969). *Theory and Applications of Spline Functions.* Academic Press, New York.

[9] Hartigan, J. A. (1967). Distribution of the residual sum of squares in fitting inequalities. *Biometrika* **54** 69–83.

[10] Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56** 495–504.

[11] HUDSON, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *J. Amer. Statist. Assoc.* **61** 1097–1129.

[12] KIMELDORF, G. S. and WAHBA, G. A. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

[13] QUANDT, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.* **55** 324–330.

[14] SCHOENBERG, I. J., ed. (1969). *Approximations with Special Emphasis on Spline Functions.* Academic Press, New York.

[15] SHORACK, G. R. (1967). Testing against ordered alternatives in model I analysis of variance; normal theory and nonparametric. *Ann. Math. Statist.* **38** 1740–1752.

[16] STUDDEN, W. J. and VAN ARMAN, D. J. (1969). Admissible designs for polynomial spline regression. *Ann. Math. Statist.* **40** 1557–1569.

[17] SYLWESTER, D. L. (1965). On maximum likelihood estimation for two-phase linear regression. Technical Report No. 11, Department of Statistics, Stanford Univ.

[18] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

[19] WILKS, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.

GENERAL ELECTRIC COMPANY
CORPORATE RESEARCH AND DEVELOPMENT
BLDG. 37, ROOM 578
SCHENECTADY, NEW YORK 12301