The Annals of Statistics
1975, Vol. 3, No. 1, 49-83

ON ASYMPTOTIC DISTRIBUTION THEORY IN SEGMENTED REGRESSION PROBLEMS—IDENTIFIED CASE[1]

BY PAUL I. FEDER

General Electric Company

This paper deals with the asymptotic distribution theory of least squares estimators in regression models having different analytical forms in different regions of the domain of the independent variable. An important special case is that of broken line regression, in which each segment of the regression function is a different straight line. The residual sum of squares function has many corners, and so classical least squares techniques cannot be directly applied. It is shown, however, that the problem can be transformed into a new problem in which the sum of squares function is locally smooth enough to apply the classical techniques. Asymptotic distribution theory is discussed for the new problem and it is shown that the results are also valid for the original problem. Results related to the usual normal theory are derived.

1. Introduction. Frequently in regression problems, a model is assumed which supposes that the regression function is of a single parametric form throughout the entire domain of interest. However, in many problems it is necessary to consider regressions which have different analytical forms in different regions of the domain. An important special case is that of broken line regression, in which each segment is a different straight line. Dunicz [5] provides an example in which such a model naturally arises in a chemical process. Sprent [23] gives examples from agriculture and biology where such regression models are appropriate.

One class of segmented models consists of functions where each segment is in the form of a linear model. Robison [21] gives procedures for obtaining confidence intervals when the regression function is one polynomial, $\mu_1(t)$, for $t \leqq \tau$ and a second polynomial, $\mu_2(t)$, for $t > \tau$, with $\mu_1(\tau) = \mu_2(\tau)$. However he assume that it is known between which observation points τ lies, and furthermore does not restrain his estimate of τ to lie between these appropriate observation points. Quandt [19, 20] discusses methods of estimating the coefficients of segmented regression functions and heuristically obtains from sampling experiments a sampling distribution of the likelihood ratio statistic for the test of no

Received July 1970; revised February 1974.

[1] This research was supported at Yale University by the Army, Navy, Air Force and NASA under a contract administered by the Office of Naval Research, at Princeton University under Army Research Office 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, least squares estimation, asymptotic theory, splines.

switch in the form of the regression. Farley, Hinich, and McGuire [6, 7] propose a routine screening procedure to detect parameter instability in time series regression models.

In recent years mathematicians interested in approximation theory have devoted much attention to the theory of *spline functions*. Spline functions consist polynomial segments. They are continuous and usually continuously differentiable; however, they have discontinuous higher order derivatives at the change-over points between the polynomial segments. The class of spline functions is a considerable extension of the class of polynomials. Splines have been found very useful for approximation and interpolation. See Greville [10] and Schoenberg [22] for details. In the theory of spline approximation, the change-over points between segments (or *knots*) are chosen merely for analytical convenience, whereas in segmented regression theory, the change-over points usually have intrinsic physical meaning in that they correspond to structural changes in the underlying model. However, the technical problems are the same in both situations. Poirer [17] relates spline theory and segmented regression theory. He develops tests to detect structural changes in the model and to decide whether certain of the model coefficients vanish. However, he makes the simplifying and restrictive assumption that the locations of the change-over points between segments are known.

The principal difficulties in the estimation problem occur when it is not known between which consecutive observations of the independent variable the change-over points lie. If for each k it is known that τ_k, the kth change-over point, lies between the successive observations $t_{j(k)}$ and $t_{j(k)+1}$, then the fitting problem is relatively simple. For each admissible set of change-over points $\tau_1, \cdots, \tau_{r-1}$, one obtains separate least squares fits (functions of $\tau_1, \cdots, \tau_{r-1}$) within each segment, subject only to the restraint of continuity at the change-over points. This can be readily accomplished by the use of Lagrange multipliers. One then chooses that set of admissible τ_k's for which the best fit is obtained. This is not too difficult a job computationally if the set of admissible τ's is relatively small. On the other hand, if the restraining region is large, the problem is very likely to reduce to that of fitting separate regressions in each segment, without imposing any of the constraints. The asymptotic distribution theory then depends on the magnitudes of $t_{j(k)+1} - t_{j(k)}$ as compared with $n^{-\frac{1}{2}}$.

In 1966, Hudson [14] considered the problem of obtaining computational procedures for the least squares fit of a continuous, segmented, linear regression function when no prior knowledge is assumed regarding the location of the change-over points. Bellman and Roth [1] applied dynamic programming methods to this problem. In the present paper, asymptotic distribution theory of the least squares estimates is discussed for models such as those considered by Hudson.

In 1965 and 1972 Sylwester [24, 25] considered the case of two straight line segments and one unknown change-over point. In 1967 Feder [8] treated the case where all the segments are dth degree polynomials differing only in their linear term. The present paper updates these results to treat a considerably larger class of models. Feder [9] considered the problem of likelihood ratio testing in segmented regression models. He showed by example that the asymptotic null distribution of the likelihood ratio test that a two-segment model in fact consist of just one segment is not unique but depends on the spacing of the independent variables. Hinkley [11] considered the case of two straight line segments and reports, on empirical grounds, that the asymptotic normality of the estimates of $(\tau_1, \cdots, \tau_{r-1})$ derived in this paper may not be an adequate approximation for moderate sample sizes. He presents an informal argument to derive alternative approximations. Hinkley [13] considered the problem of estimating and making inferences about the point of change of distribution in a sequence of random variables, which is related to the problem of the present paper.

For reasons of simplicity and to avoid peripheral issues, this paper confines attention to the case in which all segments of the regression function are in the form of linear models. However the techniques employed should suffice, by use of appropriate Taylor expansions, to handle many cases in which the segments are nonlinear.

2. Definition of the model and summary of results. Consider an r phase, segmented regression function of the form

$$
\begin{aligned}
\mu(\boldsymbol{\xi}; t) &= f_1(\boldsymbol{\theta}_1; t) && \text{for } A \equiv \tau_0 \leqq t \leqq \tau_1 \\
&= f_2(\boldsymbol{\theta}_2; t) && \text{for } \tau_1 \leqq t \leqq \tau_2 \\
&= f_r(\boldsymbol{\theta}_r; t) && \text{for } \tau_{r-1} \leqq t \leqq B \equiv \tau_r \,.
\end{aligned} \tag{2.1}
$$

This can be compactly represented as

$$\mu(\boldsymbol{\xi}; t) = \textstyle\sum_{j=1}^{r} f_j(\boldsymbol{\theta}_j; t) I_j(t) \tag{2.2}$$

where $I_j(t)$ is the indicator function of the interval $[\tau_{j-1}, \tau_j)$. It is assumed that $\mu(\boldsymbol{\xi}; t)$ is continuous at $t = \tau_j$, $j = 1, \cdots, r-1$. In this model, A and B are known constants (assumed 0, 1 without loss of generality) and $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_r$, $\tau_1, \cdots, \tau_{r-1}$ are unknown parameters. Boldface symbols will represent vectors and matrices. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_r)$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_{r-1})$, and $\boldsymbol{\xi} = (\boldsymbol{\theta}, \boldsymbol{\tau})$. Assume that for each j, $\boldsymbol{\theta}_j \equiv (\theta_{j1}, \cdots, \theta_{jK(j)})$ is a $K(j)$ dimensional vector and that there exist known functions $f_{j1}(t), \cdots, f_{jK(j)}(t)$ such that

$$f_j(\boldsymbol{\theta}_j; t) = \textstyle\sum_{k=1}^{K(j)} \theta_{jk} f_{jk}(t) \tag{2.3}$$

where $\{f_{jk}(t)\}$ are linearly independent functions on the interval $[\tau_{j-1}, \tau_j]$.

In addition, it will be assumed that there exists an $s < \infty$ such that any linear combination of the functions $\{f_{jk}(t)\}$ has at most s sign changes in derivative on the interval $[0, 1]$. This condition is satisfied by most functions usually encountered, such as polynomials, sines and cosines, and exponentials.

Let Θ denote the set of "admissible" vectors $\boldsymbol{\theta}$. That is, Θ is the collection of $\boldsymbol{\theta}$'s which lead to functions $\mu(\boldsymbol{\xi}; t)$ satisfying the continuity restraints. For each $\boldsymbol{\theta} \in \Theta$ consider the set of $\boldsymbol{\tau}$'s (depending on $\boldsymbol{\theta}$) which lead to functions $\mu(\boldsymbol{\xi}; t)$ satisfying the continuity restraints. Form the vectors $\boldsymbol{\xi} = (\boldsymbol{\theta}, \boldsymbol{\tau}(\boldsymbol{\theta})) \equiv (\boldsymbol{\theta}, \boldsymbol{\tau})$. Let Ξ denote the set of these $\boldsymbol{\xi}$'s and let $U = \{\mu(\boldsymbol{\xi}; t) : \boldsymbol{\xi} \in \Xi\}$. Throughout the discussion attention will be confined to $\boldsymbol{\theta}$'s in Θ and to $\boldsymbol{\xi}$'s in Ξ.

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

$$X_{ni} = \mu(\boldsymbol{\xi}; t_{ni}) + e_{ni} \,. \tag{2.4}$$

Assume that the observation errors, e_{ni}, are independently and identically distributed with $E(e_{ni}) = 0$, Var $(e_{ni}) = \sigma^2$, unknown, and $E|e_{ni}|^{2(1+\delta)} < \infty$ for some $\delta > 0$. Let $\boldsymbol{\varphi} \equiv (\boldsymbol{\xi}, \sigma^2)$ and let $\boldsymbol{\varphi}^{(0)} \equiv (\boldsymbol{\theta}_1^{(0)}, \cdots, \boldsymbol{\theta}_r^{(0)}, \tau_1^{(0)}, \cdots, \tau_{r-1}^{(0)}, \sigma_0^2)$ denote the true state of nature.

Asymptotic properties of the least squares estimators of $\boldsymbol{\varphi}$ will be examined. Since the change-over points $\tau_1^{(0)}, \cdots, \tau_{r-1}^{(0)}$ are unknown, the derivation of asymptotic properties is considerably more involved than one might at first expect. For each fixed $\boldsymbol{\tau}$, the estimate $\hat{\boldsymbol{\theta}}$ (a function of $\boldsymbol{\tau}$) is chosen to minimize

$$s(\boldsymbol{\xi}) = \frac{1}{n} \textstyle\sum_{i=1}^{n} (X_{ni} - \mu(\boldsymbol{\xi}; t_{ni}))^2 \tag{2.5}$$

subject to the continuity restraints. The residual mean square then can be expressed as $\bar{s}(\boldsymbol{\tau})$. It is necessary to obtain the minimum of this function.

There is no guarantee that $s(\boldsymbol{\xi})$ possesses asymptotically even one continuous derivative in the neighborhood of $\boldsymbol{\tau}^{(0)}$. For instance, if $f_j(\boldsymbol{\theta}_j; t) = a_j + b_2 t$, $b_j \neq b_{j+1}$, then $s(\boldsymbol{\xi})$ and $\bar{s}(\boldsymbol{\tau})$ possess discontinuities in the τ derivatives along $\tau_j = t_{ni}$ for each i, j.

Consider for example the simple 2 phase model where $\mu(\boldsymbol{\xi}; t) = c$ for $0 \leqq t \leqq \tau$ and t for $\tau \leqq t \leqq 1$. (See Fig. 2.1.)
Here, $\boldsymbol{\theta}$ is the scalar c and $\tau = c$. From equation (2.5)

$$s(\boldsymbol{\xi}) \equiv s(c, c) = \frac{1}{n} \textstyle\sum_{t_{ni} \leqq c} (X_{ni} - c)^2 + \frac{1}{n} \textstyle\sum_{t_{ni} > c} (X_{ni} - t_{ni})^2 \,.$$

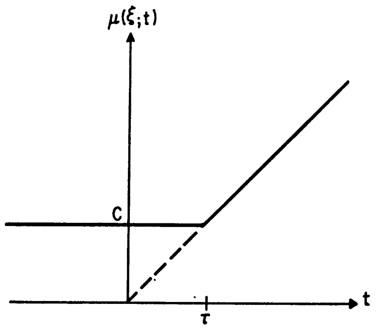

FIG. 2.1.

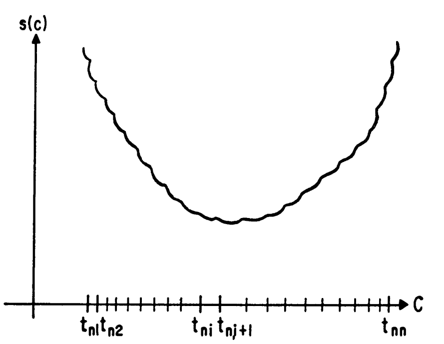

FIG. 2.2.

Thus for all c's between $t_{n,j}$ and $t_{n,j+1}$

$$\frac{\partial s(c)}{\partial c} = -\frac{2}{n} \sum_{i=1}^{j} (X_{ni} - c) , \qquad \frac{\partial^2 s(c)}{\partial c^2} = \frac{2j}{n} .$$

It is seen (Fig. 2.2) that there is a discontinuity in derivative at each observation point, t_{ni}.

The classical derivations of asymptotic normality of maximum likelihood estimators assume that the log likelihood function (or equivalently $s(\boldsymbol{\xi})$) asymptotically behaves like a paraboloid in some neighborhood of $\boldsymbol{\xi}^{(0)}$. (See Cramér [3], page 501, for instance.) Hence the classical arguments are not directly applicable here. The method of approach and results of this paper are outlined below.

1. It is shown that $\hat{\boldsymbol{\varphi}}$, the unrestricted least squares estimator (l.s.e.) of $\boldsymbol{\varphi}$, is consistent under suitable identifiability assumptions, which tacitly assume that no two consecutive $f_j(\boldsymbol{\theta}_j^{(0)}; t)$ are identical. (See Theorem 3.6.)

2. If f_j and f_{j+1} are identical, the parameter space is overspecified since the regression function does not depend on τ_j, which therefore cannot be estimated consistently. However, it is shown that under certain circumstances, with large probability as $n \to \infty$, the fitted regression function will reflect this situation. (See Corollary 3.22.) One segment can then be deleted from the model and the regression function recomputed in the reduced model. For the remainder of the summary it is assumed that no two adjacent f_j's are identical.

3. Under suitable identifiability conditions, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}})$ 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 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 $t = \tau_j^{(0)}$. (See Theorems 3.16 and 3.18.)

4. If ω is a subset of Ξ and $\boldsymbol{\xi}^{(0)} \in \bar{\omega}$, the closure of ω, then statements 1, 2, and 3 apply equally well to $\boldsymbol{\varphi}_\omega \equiv (\boldsymbol{\xi}_\omega, \sigma_\omega^2)$, the l.s.e. among all $\boldsymbol{\xi} \in \omega$.

5. A *pseudo problem* is formed by deleting $o(n/\log \log n)$ strategically placed observations near the true change-over points. It is observed that 1, 2, 3, and 4 are still valid in the pseudo problem. (See beginning of Section 4, in particular Theorem 4.1.)

6. Let $\hat{\boldsymbol{\xi}}^* \equiv (\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\tau}}^*)$ denote the l.s.e. in the pseudo problem. It is shown that under 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. (See Lemmas 4.3, 4.12, and 4.16.)

7. The asymptotic distribution of $\hat{\boldsymbol{\theta}}^*$, $\hat{\boldsymbol{\tau}}^*$ is obtained by "classical" methods. (See Lemmas 4.4, 4.8, and Theorem 4.15.)

8. Several examples are presented that illustrate the results.

9. Several unresolved problems are mentioned.

Much notation is used in the later sections. Some of the notation frequently used is set out below for ease of reference.

$q = K(1) + K(2) + \cdots + K(r)$

$\mathbf{t} = (t_1, t_2, \cdots, t_k)$ or $\mathbf{t} = (t_1, t_2, \cdots, t_q)$

$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\xi}; \mathbf{t}) \equiv (\mu(\boldsymbol{\xi}; t_1), \mu(\boldsymbol{\xi}; t_2), \cdots, \mu(\boldsymbol{\xi}; t_k))$

$\boldsymbol{\mu}^{(0)} = \boldsymbol{\mu}(\boldsymbol{\xi}^{(0)}; \mathbf{t})$; $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\hat{\boldsymbol{\xi}}; \mathbf{t})$

$\nu \equiv \nu(\boldsymbol{\xi}; t) \equiv \mu(\boldsymbol{\xi}; t) - \mu(\boldsymbol{\xi}^{(0)}; t) \equiv \mu(t) - \mu_0(t) \equiv \mu - \mu_0$

$\nu_{ni} = \nu(\boldsymbol{\xi}; t_{ni})$, $\hat{\nu}_{ni} = \nu(\hat{\boldsymbol{\xi}}; t_{ni})$

$||\mu|| = \max_{0 \leq t \leq 1} |\mu(t)|$ or $||\boldsymbol{\mu}(\boldsymbol{\xi}; \mathbf{t})|| = (\sum \mu^2(\boldsymbol{\xi}; t_{ni}))^{\frac{1}{2}}$

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_r)$; $\boldsymbol{\theta}_j = (\theta_{j1}, \cdots, \theta_{jK(j)})$

$\Theta = \{\text{admissible } \boldsymbol{\theta}\text{'s}\}$; $\Xi = \{(\boldsymbol{\theta}, \boldsymbol{\tau}(\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta\}$; $U = \{\mu(\boldsymbol{\xi}; t); \boldsymbol{\xi} \in \Xi\}$

$\boldsymbol{\tau} = (\tau_1, \cdots, \tau_{r-1})$; $\boldsymbol{\xi} = (\boldsymbol{\theta}, \boldsymbol{\tau})$; $\boldsymbol{\varphi} = (\boldsymbol{\xi}, \sigma^2)$

$\boldsymbol{\theta}^{(0)}, \boldsymbol{\tau}^{(0)} =$ "true" states of nature; $P_0 = P$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$, $\tau = \tau^{(0)}$

$\boldsymbol{\xi}^{(0)} = (\boldsymbol{\theta}^{(0)}, \boldsymbol{\tau}^{(0)})$; $\boldsymbol{\varphi}^{(0)} = (\boldsymbol{\xi}^{(0)}, \sigma_0^2)$

$\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}})$ = least squares estimator (l.s.e.) of $\boldsymbol{\xi}^{(0)}$

$\hat{\boldsymbol{\xi}}_\omega = (\hat{\boldsymbol{\theta}}_\omega, \hat{\boldsymbol{\tau}}_\omega)$ = l.s.e. restricted to $\omega \subset \Xi$

$\hat{\boldsymbol{\xi}}^* \equiv (\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\tau}}^*) \equiv$ least squares estimator in the pseudo problem (see Section 4)

$\hat{\boldsymbol{\xi}}_\omega^* = (\hat{\boldsymbol{\theta}}_\omega^*, \hat{\boldsymbol{\tau}}_\omega^*)$ = restricted l.s.e. in the pseudo problem

$f_j(\boldsymbol{\theta}_j; t) \equiv \sum_{k=1}^{K(j)} \theta_{jk} f_{jk}(t)$; $f_{jk}(t) \equiv f_{jk}$; $f_j(\boldsymbol{\theta}_j^{(0)}; t) \equiv f_j^{(0)}(t) \equiv f_j^{(0)}$

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

$Ee_{ni} = 0$; $\operatorname{Var}(e_{ni}) = \sigma^2$; $E|e_{ni}|^{2(1+\delta)} < \infty$ for some $\delta > 0$

$H_n(s)$ = distribution of $\{t_{ni}\}$; $H_n(s) \rightarrow_d H(s)$ as $n \rightarrow \infty$;

$H_n(A) = \int_A dH_n(s)$

$D^{\pm}(h, j, k) \equiv D^{\pm}$ = the kth left and right t-derivatives respectively of $f_h(\boldsymbol{\theta}_h^{(0)}; t)$ at $t = \tau_j^{(0)}$. If $D^+ = D^-$, their common value is denoted as D.

n^* = sample size in the pseudo problem; $n^{**} = n - n^*$

$\sum^*$ = summation over the n^* terms of the pseudo problem; $\sum^{**} = \sum_{i=1}^n - \sum^*$

The calculus of O_p and o_p (Definition 2.1 below and discussed rigorously by Pratt [18]) is used throughout the paper without any explanation. Loosely speaking, one can operate with O_p and o_p in asymptotic calculations as with O and o.

DEFINITION 2.1. (O_p and o_p). A sequence of random variables $\{Y_n\}$ is said to be

(a) $O_p(1)$ if for every $\varepsilon > 0$ there exist constants $D(\varepsilon)$ and $N(\varepsilon)$ such that $n > N(\varepsilon)$ implies $P[|Y_n| < D(\varepsilon)] \geqq 1 - \varepsilon$;

(b) $o_p(1)$ if for every $\varepsilon > 0$, $\delta > 0$ there exists a constant $N(\varepsilon, \delta)$ such that $n > N(\varepsilon, \delta)$ implies $P[|Y_n| < \delta] \geqq 1 - \varepsilon$.

(c) A sequence of random variables $\{Y_n\}$ is said to be $O_p(r_n)(o_p(r_n))$ if the sequence $\{Y_n/r_n\}$ is $O_p(1)(o_p(1))$.

3. Consistency and rate of convergence of $\hat{\boldsymbol{\xi}}$. The questions of consistency and rate of convergence of $\hat{\boldsymbol{\xi}}$ to $\boldsymbol{\xi}^{(0)}$ are considered in this section.

At the outset, the nonstatistical notion of *identifiability* of the regression function immediately arises. That is, assuming no observation errors, at which t values must $\mu(\boldsymbol{\xi}; t)$ be observed in order to uniquely determine it over the entire interval $[0, 1]$? It will be shown that under suitable identifiability assumptions $\boldsymbol{\theta}$ converges to $\boldsymbol{\theta}^{(0)}$ at the rate $O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ and τ_j converges to $\tau_j^{(0)}$ at a rate determined by the number of t-derivatives in which $f_j(\boldsymbol{\theta}_j^{(0)}; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; t)$ agree at $t = \tau_j^{(0)}$.

It will be assumed throughout that $f_j(\boldsymbol{\theta}_j^{(0)}; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; t)$ agree in $m_j - 1$ t-derivatives at $t = \tau_j^{(0)}$ but differ in the m_jth. Further, it will be assumed that f_j and f_{j+1} each have continuous left and right m_jth t-derivatives at $t = \tau_j^{(0)}$, $j = 1, 2, \cdots, r-1$.

DEFINITION 3.1. The parameter $\boldsymbol{\theta}$ is *identified* at $\mu^{(0)}$ 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 $\boldsymbol{\theta}^{(0)}$.

LEMMA 3.2. *If $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ then there exist neighborhoods N, T where N is a (k-dimensional) neighborhood of $\boldsymbol{\mu}^{(0)}$ and T is a (k-dimensional) neighborhood of $\mathbf{t}$ such that*

(a) *for all (k-dimensional) vectors $\boldsymbol{\mu} \in N$ and $\mathbf{t}' \in T$ such that $\boldsymbol{\mu}$ can be represented as $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\xi}; \mathbf{t}')$ for some $\xi \in \Xi$, $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}$ by $\mathbf{t}'$*

(b) *there exists a constant, C, such that the transformation $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\mu}; \mathbf{t}')$ satisfies the Lipschitz condition $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \leqq C\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|$ whenever $\mathbf{t}' \in T$ and $\boldsymbol{\mu}_1 \equiv \boldsymbol{\mu}(\boldsymbol{\xi}_1; \mathbf{t}')$, $\boldsymbol{\mu}_2 \equiv \boldsymbol{\mu}(\boldsymbol{\xi}_2; \mathbf{t}')$ are both in N.*

PROOF. Since $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$, it follows that for any possible choice of parameters $\tau_1, \cdots, \tau_{r-1}$ (and consequent segments $\{[\tau_{j-1}, \tau_j), j = 1, \cdots, r\}$) consistent with $\boldsymbol{\theta}^{(0)}$, for each j there must exist $K(j)$ components $t_{j1}, \cdots, t_{jK(j)}$ within the segment $(\tau_{j-1}, \tau_j) \cap (\tau_{j-1}^{(0)}, \tau_j^{(0)})$ such that the $K(j)$ by $K(j)$ matrix $\mathbf{A}_j(t_{j1}, \cdots, t_{jK(j)})$, with (i, k)th element $f_{j,k}(t_{ji})$, is nonsingular. By continuity, the t_{ji}'s may be perturbed slightly without disturbing the nonsingularity of $\mathbf{A}_j$. Assertions (a) and (b) follow directly from the properties of nonsingular linear transformations

REMARK 1. Nothing has yet been mentioned about the determination of $\tau_1, \cdots, \tau_{r-1}$. The τ's may or may not be uniquely determined once $\boldsymbol{\theta}$ is known. This will be discussed at length later on.

REMARK 2. The proof of Lemma 3.2 shows that if $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t} \equiv (t_1, t_2, \cdots, t_k)$, then $k \geqq q$ and there exists a q-dimensional subvector $\tilde{\mathbf{t}}$ of $\mathbf{t}$, such that $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\tilde{\mathbf{t}}$.

REMARK 3. In order that $\boldsymbol{\theta}$ be identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ it is necessary that no two adjacent segments of $\mu(\boldsymbol{\xi}^{(0)}; t)$ be identically the same.

REMARK 4. Since $\mu(\boldsymbol{\xi}; t)$ effectively depends on q parameters (actually $q + r - 1$ parameters related by $r - 1$ continuity restraints), it must be observed at a minimum of q points in order to be identified. It is clear from the linear independence of the f_{jk}'s that the placement of $K(j)$ distinct observations between each pair of consecutive change-over points $\tau_j^{(0)}$, $\tau_{j+1}^{(0)}$ is necessary and sufficient to identify $\boldsymbol{\theta}^{(0)}$. In particular, if the n observations are equally spaced and no two adjacent segments of $\mu(\boldsymbol{\xi}^{(0)}; t)$ are identical, then for n sufficiently large a $\mathbf{t}$ exists such that $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$.

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

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

DEFINITION 3.3. A *center of observations* is a point of increase of H.

The principal result of this section is that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ if there is a vector $\mathbf{t}$ whose components are centers of observations and which identifies $\boldsymbol{\theta}$ at $\boldsymbol{\mu}^{(0)}$.

Lemma 3.5 below implies that $\mu(\boldsymbol{\xi}; t)$ must be near $\mu(\boldsymbol{\xi}^{(0)}; t)$ for at least one value of t close to each center of observations. The consistency of $\hat{\boldsymbol{\theta}}$ is a consequence of this.

Condition (*) of Lemma 3.4 guarantees that the least squares estimator $\hat{\boldsymbol{\theta}}$ is contained in a sphere with center $\boldsymbol{\theta}^{(0)}$ and with radius d^*.

LEMMA 3.4. *Suppose that there exists an* $\varepsilon > 0$ *such that for every* $K > 0$ *there exist* $d(K)$, $n(K)$ *such that* $d > d(K)$, $n > n(K)$ *imply*

$$(*) \qquad \inf_{\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)}\|>d\}} H_n\{t: |\mu(\boldsymbol{\xi}; t) - \mu(\boldsymbol{\xi}^{(0)}; t)| > K\} > \varepsilon .$$

Then there exists a d^* *such that*

$$\lim_{n\to\infty} P_0\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}\| \leqq d^*\} = 1 .$$

PROOF. Take $K = 3\sigma_0/\varepsilon^{\frac{1}{2}}$. For $n > n(K)$, $d > d(K)$, and $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\| > d$,

$$\inf_{\boldsymbol{\theta}} \sum_{i=1}^{n} (X_{ni} - \mu(\boldsymbol{\xi}; t_{ni}))^2 = \inf \sum_{i=1}^{n} (e_{ni} - \nu_{ni})^2 .$$

The triangle inequality implies

$$\begin{aligned}\sum (e_{ni} - \nu_{ni})^2 &\geqq [(\sum \nu_{ni}^2)^{\frac{1}{2}} - (\sum e_{ni}^2)^{\frac{1}{2}}]^2 \\ &= (\sum \nu_{ni}^2)^{\frac{1}{2}}[(\sum \nu_{ni}^2)^{\frac{1}{2}} - 2(\sum e_{ni}^2)^{\frac{1}{2}}] + \sum e_{ni}^2 \\ &\geqq (\sum \nu_{ni}^2)^{\frac{1}{2}}[3\sigma_0 n^{\frac{1}{2}} - 2\sigma_0 n^{\frac{1}{2}} + o_p(n^{\frac{1}{2}})] + \sum e_{ni}^2 .\end{aligned}$$

The last inequality is a consequence of condition (*). Note that the $o_p(n^{\frac{1}{2}})$ term does not depend on $\boldsymbol{\theta}$. Thus, if $d > d(K)$, $n > n(K)$,

$$\sum (e_{ni} - \upsilon_{ni})^2 \geqq n^{\frac{1}{2}}(\sum \upsilon_{ni}^2)^{\frac{1}{2}}(\sigma_0 + o_p(1)) + \sum e_{ni}^2 \qquad \text{as } n \to \infty$$

where $o_p(1)$ is independent of $\boldsymbol{\theta}$. This implies that with probability approaching 1 as $n \to \infty$, $\sum (e_{ni} - \upsilon_{ni})^2 > \sum e_{ni}^2$ *uniformly* for all $\boldsymbol{\theta}$ such that $||\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}|| > d(K)$. In other words, (with the inf restricted to such $\boldsymbol{\theta}$),

$$\lim P_0\{\inf_{\boldsymbol{\theta}} \sum (X_{ni} - \mu(\boldsymbol{\xi}, t_{ni}))^2 > \sum e_{ni}^2\} = 1 .$$

Since the least squares estimator $\hat{\boldsymbol{\theta}}$ minimizes the residual sum of squares function it follows that with probability approaching 1 as $n \to \infty$ $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}|| \leqq d(K)$. □

LEMMA 3.5. *If t_0 is a center of observations, $\delta > 0$, $\eta > 0$, and condition* (*) *of Lemma* 3.4 *holds, then*

$$P_0\{|\mu(\hat{\boldsymbol{\xi}}; t) - \mu(\boldsymbol{\xi}^{(0)}; t)| \geqq \eta \text{ for all } t \text{ such that } |t - t_0| \leqq \delta\} \to 0 .$$

PROOF. Let S denote $\{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}|| \leqq d^*\}$. Let $\tilde{\boldsymbol{\xi}}$ denote the least squares estimator of $\boldsymbol{\xi}$ with $\boldsymbol{\theta}$ restricted to S. Lemma 3.4 implies that with large probability as $n \to \infty$, $\hat{\boldsymbol{\xi}}$, the unrestricted least squares estimator, is equal to $\tilde{\boldsymbol{\xi}}$, the restricted least squares estimator. We utilize Theorems 1 and 4 of Jennrich [15] to discuss the behavior of $\tilde{\boldsymbol{\xi}}$.

Since t_0 is a center of observations, the number of observations in the interval $[t_0 - \delta, t_0 + \delta]$ is $\lambda n + o(n)$ for some $\lambda > 0$. Let $F = \{\nu : |\nu(\boldsymbol{\xi}; t)| \geqq \eta$ for all t with $|t - t_0| \leqq \delta$ and $||\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}|| \leqq d^*\}$. Now

$$\begin{aligned} \sum (X_{ni} - \mu(\boldsymbol{\xi}; t_{ni}))^2 &= \sum (e_{ni} - \nu_{ni})^2 \\ &= \sum e_{ni}^2 + \sum \nu_{ni}^2 - 2 \sum e_{ni}\nu_{ni} . \end{aligned}$$

If $\upsilon \in F$, then $\sum \upsilon_{ni}^2 \geqq \lambda\eta^2 n[1 + o(1)]$ where $o(1)$ does not depend on $\boldsymbol{\theta}$. Furthermore, the arguments of Theorems 1 and 4 of Jennrich [15] imply that $n^{-1} \sum e_{ni}\upsilon_{ni}$ converges to 0 in probability *uniformly* for $\boldsymbol{\theta} \in S$, since S is a compact set. Thus $\inf_{\nu \in F} \sum (e_{ni} - \nu_{ni})^2 \geqq \sum e_{ni}^2 + \lambda\eta^2 n + o_p(n)$. Since the least squares estimate minimizes the residual sum of squares, it follows that

$$\sum (X_{ni} - \mu(\tilde{\boldsymbol{\xi}}; t_{ni}))^2 \leqq \sum (X_{ni} - \mu(\boldsymbol{\xi}^{(0)}; t_{ni}))^2 = \sum e_{ni}^2 .$$

Thus, with large probability as $n \to \infty$, $\upsilon(\tilde{\boldsymbol{\xi}}; t)$ is not in F. Since $\hat{\boldsymbol{\xi}} = \tilde{\boldsymbol{\xi}}$ with large probability as $n \to \infty$, it follows that with large probability as $n \to \infty$, $\upsilon(\hat{\boldsymbol{\xi}}; t)$ is not in F. That is, $|\upsilon(\hat{\boldsymbol{\xi}}; t)| \leqq \eta$ for some t with $|t - t_0| \leqq \delta$. □

THEOREM 3.6. (*Consistency*). *If* (i) *Condition* (*) *of Lemma* 3.4 *holds*, (ii) $\boldsymbol{\theta}$ *is identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$, (iii) *the components of t are centers of observations, then*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = o_p(1) \tag{3.1}$$

$$\hat{\sigma}^2 - \sigma_0^2 = o_p(1) . \tag{3.2}$$

PROOF. Let N, T be (k-dimensional) neighborhoods of $\boldsymbol{\mu}^{(0)}$ and $\mathbf{t}$ within which the assertions of Lemma 3.2 hold. It follows from Lemma 3.5 that given $\varepsilon > 0$, with probability approaching one as $n \to \infty$ there exists a $\mathbf{t}' \in T$ such that $\boldsymbol{\mu}(\hat{\boldsymbol{\xi}}; \mathbf{t}') \in N$ and $||\boldsymbol{\nu}(\hat{\boldsymbol{\xi}}; \mathbf{t})|| \leqq \varepsilon$.

From Lemma 3.2, $\hat{\boldsymbol{\theta}} \equiv \boldsymbol{\theta}(\hat{\boldsymbol{\mu}}; \mathbf{t}')$ is uniquely determined and

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}\| \leqq C\|\boldsymbol{\mu}(\hat{\boldsymbol{\xi}}; \mathbf{t}') - \boldsymbol{\mu}(\boldsymbol{\xi}^{(0)}; \mathbf{t}')\| \leqq C\varepsilon \,. \tag{3.3}$$

Since ε is arbitrary, equation (3.3) implies (3.1).

Equation (3.2) follows directly since

$$\hat{\sigma}^2 = \frac{1}{n} \sum (e_{ni} - \hat{\nu}_{ni})^2 \leqq \frac{1}{n} \sum e_{ni}^2 = \sigma_0^2 + o_p(1) \,.$$

On the other hand, $\hat{\sigma}^2 = 1/n \sum e_{ni}^2 + 1/n \sum \hat{\nu}_{ni}^2 - 1/n \sum e_{ni}\hat{\nu}_{ni}$. Lemma 3.4 and the uniform convergence in probability to 0 of $\sum e_{ni}\nu_{ni}$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\| \leqq d^*$ implies that $\hat{\sigma}^2 \geqq \sigma_0^2 + o_p(1)$. Thus $\hat{\sigma}^2 = \sigma_0^2 + o_p(1)$. □

REMARK. One might expect that any function which satisfies the constraints of the model and which fits the data better than $\mu_0(t)$ has to be close to $\mu_0(t)$ somewhere in the neighborhood of any point t around which $r(n)$ observations are taken, as long as $r(n) \to \infty$. Lemma 3.5 implies that this is the case if $r(n) = \lambda n$. However, the following example indicates that this is not true generally.

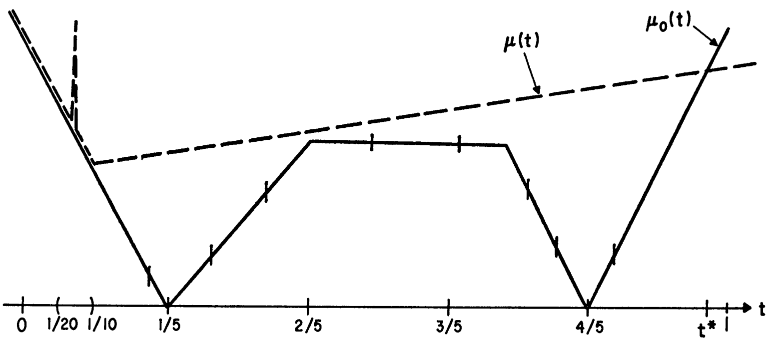

FIG. 3.1.

Suppose the model is a five segment broken line and $\mu(\boldsymbol{\xi}^{(0)}, t)$ is as shown in Figure 3.1. Suppose that $\log n$ observations are taken at each of the eight t-values indicated with hash marks. Further, suppose n observations are made at t_* and n observations are uniformly spread over the subinterval $I \equiv (\frac{1}{20}, \frac{1}{10})$. Let T_{nM} be the t-value in I at which the maximum disturbance, e_{nM}, occurs. Define $\tilde{\mu}(t)$ as:

$$\begin{aligned} \tilde{\mu}(t_{ni}) &= \mu_0(t_{ni}) & t_{ni} &< t_{nM} \\ &= X_{nM} & t_{ni} &= t_{nM} \\ &= \mu_0(t_{ni}) & t_{nM} &< t_{ni} < \tfrac{1}{10} \\ &= \mu_0(t_*) & t_{ni} &= t_* \,. \end{aligned}$$

Define $\tilde{\mu}(t)$ elsewhere by the condition that it conforms to the five segment model. It may be verified that asymptotically $\tilde{\mu}(t)$ will fit the data better than $\mu_0(t)$, $\mu_0(t)$ is identified, but $\tilde{\mu}(t)$ does not at all resemble $\mu_0(t)$.

Suppose $\boldsymbol{\xi}^{(0)} \in \bar{\omega}$, the closure of the set $\omega \subset \Xi$. The proofs of Lemma 3.5 and

Theorem 3.6 directly imply the consistency of $\hat{\boldsymbol{\theta}}_\omega$, the restricted l.s.e. This is formally stated as

COROLLARY 3.7. *If* $\boldsymbol{\xi}^{(0)} \in \bar{\omega} \subset \Xi$, *then under the hypotheses of Theorem* 3.6 $\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}^{(0)} = o_p(1)$, $\hat{\sigma}_\omega^2 - \sigma_0^2 = o_p(1)$, *where* $\hat{\boldsymbol{\theta}}_\omega$ *is the restricted* l.s.e. *of* $\boldsymbol{\theta}$, *and* $\hat{\sigma}_\omega^2$ *is the restricted* l.s.e. *of* σ^2.

Thus far, no mention has been made of the behavior of $\hat{\boldsymbol{\tau}}$. It turns out that under suitable conditions $\hat{\boldsymbol{\tau}}$ converges to $\boldsymbol{\tau}^{(0)}$ in probability and the rate of convergence depends on the number of t-derivatives that adjacent segments have in common at the change-over points $\tau_1^{(0)}, \cdots, \tau_{r-1}^{(0)}$.

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 all centers of observation involved in the identification of $\boldsymbol{\theta}_j$ and to the left of all centers of observation involved in the identification of $\boldsymbol{\theta}_{j+1}$. (For brevity, one can describe $\Delta_j^{(0)}$ as the set of τ's which are *compatible*, with respect to $\boldsymbol{\theta}_j^{(0)}$, $\boldsymbol{\theta}_{j+1}^{(0)}$, with the centers of observation.) It will be shown that $\hat{\tau}_j$ lies near an element of $\Delta_j^{(0)}$ with large probability as $n \to \infty$.

It is shown in Lemma 3.8 that if $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_{j+1}$ are close to $\boldsymbol{\theta}_j^{(0)}$ and $\boldsymbol{\theta}_{j+1}^{(0)}$ respectively, then each element of Δ_j (the set of τ_j's compatible, with respect to $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$, with the centers of observation) lies close to an element of $\Delta_j^{(0)}$.

LEMMA 3.8. *Let* $\mathscr{N}_j$ *be any collection of neighborhoods which covers the set* $\Delta_j^{(0)}$. *There exist neighborhoods* N_j *and* N_{j+1} *about* $\boldsymbol{\theta}_j^{(0)}$ *and* $\boldsymbol{\theta}_{j+1}^{(0)}$ *respectively, such that if* $\boldsymbol{\theta} \in \Theta$ *and* $\boldsymbol{\theta}_j \in N_j$, $\boldsymbol{\theta}_{j+1} \in N_{j+1}$ *respectively, then if* τ_j *is compatible with respect to* $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$ *with the centers of observation,* τ_j *must be contained in an element of* $\mathscr{N}_j$.

PROOF. Suppose to the contrary that there exists a collection of neighborhoods $\mathscr{N}_j$ such that for all N_j, N_{j+1} as described in the statement, there exists $\boldsymbol{\theta} \in \Theta$ with $\boldsymbol{\theta}_j \in N_j$, $\boldsymbol{\theta}_{j+1} \in N_{j+1}$ and such that τ_j's compatible with respect to $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$ exist outside of the elements of $\mathscr{N}_j$. Then there exists a sequence $\boldsymbol{\theta}^{(n)} \in \Theta$ such that $\boldsymbol{\theta}_{j,n}$ and $\boldsymbol{\theta}_{j+1,n}$ converge to $\boldsymbol{\theta}_j^{(0)}$ and $\boldsymbol{\theta}_{j+1}^{(0)}$ respectively and a sequence $\{\tau_{j,n}\}$ such that for each n, $\tau_{j,n}$ belongs to no element of $\mathscr{N}_j$, is compatible, with respect to $\boldsymbol{\theta}_{jn}$, $\boldsymbol{\theta}_{j+1,n}$, with the centers of observation, and is such that $f_j(\boldsymbol{\theta}_{j,n}; \tau_{j,n}) = f_{j+1}(\boldsymbol{\theta}_{j+1,n}; \tau_{j,n})$. Thus there exists a subsequence of $\tau_{j,n}$'s which converges to $\tau_* \notin \Delta_j^{(0)}$ and which is compatible with respect to $\boldsymbol{\theta}_j^{(0)}$, $\boldsymbol{\theta}_{j+1}^{(0)}$ with the centers. By continuity, $f_j(\boldsymbol{\theta}_j^{(0)}; \tau_*) = f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; \tau_*)$, which contradicts the definition of $\Delta_j^{(0)}$. $\square$

An important special case of Lemma 3.8 occurs when $\Delta_j^{(0)}$ is the one point set, $\{\tau_j^{(0)}\}$. This suggests

DEFINITION 3.9. The parameter $\boldsymbol{\theta}$ is *well-identified* at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ if

(i) $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$

(ii) for each j, $1 \leqq j \leqq r - 1$, $\Delta_j^{(0)}$ is the one point set, $\{\tau_j^{(0)}\}$.

The consistency of $\hat{\boldsymbol{\tau}}$ is then an immediate consequence of Theorem 3.6, Lemma 3.8, and Definition 3.9. This is stated in Theorem 3.10.

THEOREM 3.10. *If* (i) *Condition* (*) *of Lemma* 3.4 *holds,* (ii) $\boldsymbol{\theta}$ *is well-identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$, (iii) *the components of* $\mathbf{t}$ *are centers of observations, then*

$$\hat{\varphi} - \varphi^{(0)} = o_p(1) \,. \tag{3.4}$$

Several examples may help to distinguish among the notions of identified, well-identified, and unidentified. The two phase broken line is well-identified by the four points pictured in example 1 of Figure 3.2. The two phase parabola-straight line function is identified by the five points pictured in example 2 of Figure 3.2. However, it is not well-identified since the change-over point may occur in one of two places. The two phase parabola-straight line function is *not* identified by the five points pictured in example 3 of Figure 3.2. This is because it is possible to partition the points in two different ways, each in accordance with different $\boldsymbol{\theta}$'s; namely 1, 2 and 3, 4, 5 or alternatively 1, 2, 3 and 4, 5. Both of these possibilities are pictured in example 3.

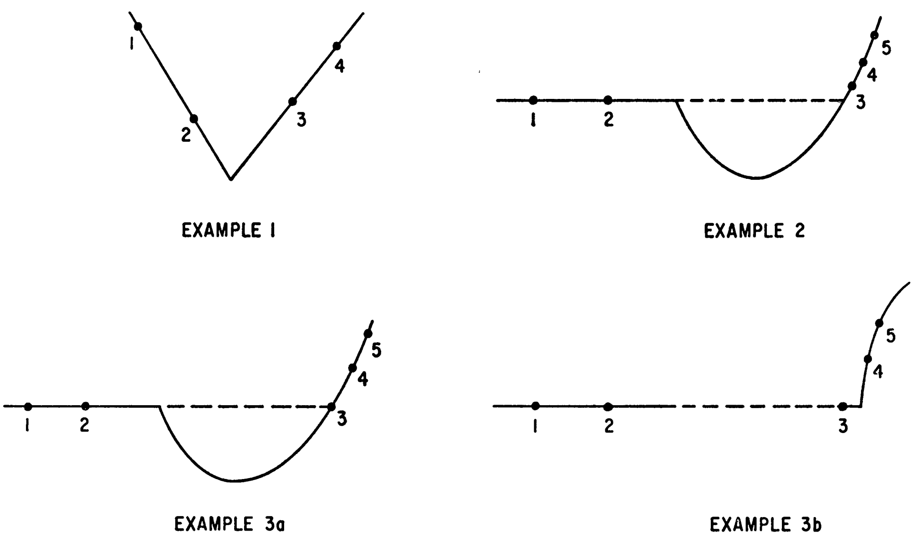

FIG. 3.2.

As a fourth example, if the model specifies a two-phase broken line but the segments are in reality colinear, then the regression function is *not* identified, since a nonunique segment can be adjoined in such a way that the resulting function conforms to the model. However if the additional assumption is made that there are at least two centers of observation within each segment, then the function is identified, but obviously not well-identified.

It will now be shown that under the identifiability assumptions of Theorem 3.6, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}})$. It will be shown in the next section that under a mild additional assumption, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$.

We first prove three preliminary lemmas which enter into the rate of convergence argument. Note that notation may differ from that in the applications.

Let $\mathscr{V}$ be an inner product space and $\mathscr{X}$, $\mathscr{Y}$ subspaces of $\mathscr{V}$. Suppose $x \in \mathscr{X}$, $y \in \mathscr{Y}$, $z = x + y$, and x^*, y^* are the orthogonal projections of z onto $\mathscr{X}$, $\mathscr{Y}$, respectively.

LEMMA 3.11. *Suppose there exists an $\alpha < 1$ such that $x \in \mathscr{X}$, $y \in \mathscr{Y}$ implies $|(x, y)| \leqq \alpha ||x|| \, ||y||$. Then*

$$||x + y|| \leqq (||x^*|| + ||y^*||)/(1 - \alpha) . \tag{3.5}$$

PROOF. It follows from the definition of x^*, y^* that $(x + y - y^*, y) = 0$, $(x + y - x^*, x) = 0$. Thus $||y||^2 = (y, y^*) - (x, y) \leqq ||y|| \, ||y^*|| + \alpha ||x|| \, ||y||$ and $||x||^2 = (x, x^*) - (x, y) \leqq ||x|| \, ||x^*|| + \alpha ||x|| \, ||y||$. The triangle inequality and the above two relations imply $||x + y|| \leqq ||x|| + ||y|| \leqq (||x^*|| + \alpha ||y||) + (||y^*|| + \alpha ||x||)$, from which (3.5) follows. □

Lemma 3.12 is an obvious multivariate generalization of Kolmogorov's inequality.

LEMMA 3.12. *Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are independent $p \times 1$ random vectors, with $E(\mathbf{X}_i) = \mathbf{0}$, $\mathrm{Cov}\,(\mathbf{X}_i) = \mathbf{\Sigma}_i$. Let $\mathbf{S}_k = \sum_{i=1}^k \mathbf{X}_i$, $\mathbf{\Sigma}^{(k)} = \sum_{i=1}^k \mathbf{\Sigma}_i$, and let $\mathbf{M}$ denote a positive definite matrix. Then*

$$P\{\max_{1 \leq k \leq n} \mathbf{S}_k' \mathbf{M} \mathbf{S}_k \geqq \varepsilon^2\} \leqq \varepsilon^{-2} \operatorname{tr} (\mathbf{\Sigma}^{(n)} \mathbf{M}) . \tag{3.6}$$

PROOF. The quadratic form $\mathbf{S}_k' \mathbf{M} \mathbf{S}_k$ can be written as $||\mathbf{M}^{\frac{1}{2}} \mathbf{S}_k||^2 \equiv \sum_{j=1}^p [(\mathbf{M}^{\frac{1}{2}} \mathbf{S}_k)_j]^2$ (where $\mathbf{M}^{\frac{1}{2}}$ denotes the symmetric square root of $\mathbf{M}$). The derivation of the usual Kolmogorov inequality is directly applicable to this sum. (See Doob [4], bottom of page 315.) □

Lemma 3.13 will be used to show that the length of the projection of the disturbance vector $\mathbf{e}$ onto a certain random one-dimensional subspace is not too large. This lemma is used in the proof of Theorem 3.14, the principal rate of convergence result.

The conditions and assumptions of the lemma are rather opaque and were of course motivated by the needs of Theorem 3.14; it may be preferable to read it after the proof of Theorem 3.14.

Assume $0 < p(n) \leqq n$ and $p(n) \to \infty$ as $n \to \infty$. Assume further that N_n is a sequence of random variables such that $1 \leqq N_n \leqq n$ and $N_n = O_p(p(n))$. For brevity denote N_n by N. Let ζ_i, $1 \leqq i \leqq \infty$ be a fixed set of constants and let $\alpha \geqq 0$. Let $\zeta_{n1}, \cdots, \zeta_{nn}$ be constants such that $\sup_n \max_{1 \leq i \leq n} |\zeta_{ni}| < \infty$ and such that $n^\alpha \zeta_{ni} = \zeta_i(1 + \rho_{ni})$ where $m_n^{-\frac{1}{2}} \equiv \max_{1 \leq i \leq n} |\rho_{ni}| = o(p^{-\frac{1}{2}}(n))$ as $n \to \infty$. Let $h(i) = o(1)$ as $i \to \infty$.

LEMMA 3.13. *Suppose for every $K > 0$ such that $Kp(n) < n$ for n sufficiently large,*

$$\lim_{n \to \infty} \sum_{i=1}^{Kp(n)} \left(\frac{\zeta_i^2 \log \log \sum_{j=1}^i \zeta_j^2}{h(i) \sum_{j=1}^i \zeta_j^2} \right)^{1+\delta} < \infty . \tag{3.7}$$

Then

$$T_N \equiv \sum_{i=1}^N \zeta_{ni} e_{ni} / (\sum_{i=1}^N \zeta_{ni}^2)^{\frac{1}{2}} = O_p((\log \log n)^{\frac{1}{2}}) \quad \text{as } n \to \infty . \tag{3.8}$$

REMARK 1. Define $\log\log \sum_{j=1}^{i} \zeta_j^2$ to be 0 if it is otherwise undefined or negative.

REMARK 2. This lemma is used in the proof of Theorem 3.14 to account for the influence on the least squares estimates of the observations near the change-over points.

REMARK 3. The δ in the exponent in equation (3.7) is the δ referred to in the paragraph immediately below equation (2.4), where it is assumed that $E|e_{ni}|^{2(1+\delta)} < \infty$.

PROOF. Multiply each of the ζ_{ni} by n^α. Thus

$$T_N = \sum_{i=1}^{N} \zeta_i(1 + \rho_{ni})e_{ni}/(\sum_{i=1}^{N} \zeta_i^2(1 + \rho_{ni})^2)^{\frac{1}{2}} .$$

Since $\max_{1\leq i\leq n} |\rho_{ni}| = o(p^{-\frac{1}{2}}(n))$, we have that for every $\varepsilon > 0$, $(\sum_{i=1}^{N} \zeta_i^2(1 + \rho_{ni})^2)^{\frac{1}{2}} > (1-\varepsilon)(\sum_{i=1}^{N} \zeta_i^2)^{\frac{1}{2}}$ for n sufficiently large. Therefore, for n sufficiently large,

$$\begin{aligned}(1-\varepsilon)|\textstyle\sum_{i=1}^{N} \zeta_i \rho_{ni} e_{ni}|/(\sum_{i=1}^{N} \zeta_i^2(1+\rho_{ni})^2)^{\frac{1}{2}} \\ < |\textstyle\sum_{i=1}^{N} \zeta_i \rho_{ni} e_{ni}|/(\sum_{i=1}^{N} \zeta_i^2)^{\frac{1}{2}} \leqq m_n^{-\frac{1}{2}} \sum_{i=1}^{N} |\zeta_i|\,|e_{ni}|/(\sum_{i=1}^{N} \zeta_i^2)^{\frac{1}{2}} \\ \leqq m_n^{-\frac{1}{2}}(\textstyle\sum_{i=1}^{N} e_{ni}^2)^{\frac{1}{2}} \leqq m_n^{-\frac{1}{2}}(\sum_{i=1}^{cm_n} e_{ni}^2)^{\frac{1}{2}}[1 + o_p(1)] = O_p(1) .\end{aligned}$$

Now consider $V_N \equiv \sum_{i=1}^{N} \zeta_i e_{ni}/(\sum_{i=1}^{N} \zeta_i^2)^{\frac{1}{2}}$. For every $\varepsilon > 0$ there exists a $K = K(\varepsilon)$ such that $N < Kp(n) \leqq n$ with probability greater than $1-\varepsilon$ as $n \to \infty$. Thus, with probability greater than $1-\varepsilon$ as $n \to \infty$, $|V_N| \leqq \max_{1\leq k\leq Kp(n)} |V_k|$. If $\sum_{i=1}^{Kp(n)} \zeta_i^2$ is bounded as $n \to \infty$, then Kolmogorov's inequality implies that that the numerator is $O_p(1)$. The denominator is obviously bounded away from 0. If $\sum_{i=1}^{Kp(n)} \zeta_i^2 \to \infty$ as $n \to \infty$, define e_{ni}^T as e_{ni} if $e_{ni}^2 < h(i) \sum_{j=1}^{i} \zeta_j^2/(\zeta_i^2 \log\log \sum_{j=1}^{i} \zeta_j^2)$ and as 0 otherwise.

Imagine the finite sequence $\{\sum_{i=1}^{k} \zeta_i e_{ni}^T/(\sum_{i=1}^{k} \zeta_i^2)^{\frac{1}{2}}, k = 1, \cdots, Kp(n)\}$, to be the beginning of an infinite sequence. The proof of the first part of the law of the iterated logarithm on pages 261–262 of Loève [16] holds true for this infinite sequence. This implies that for every $\beta > 0$

$$P\{|\textstyle\sum_{i=1}^{k} \zeta_i e_{ni}^T| > (1+\beta)[\sum_{i=1}^{k} \zeta_i^2]^{\frac{1}{2}}[\log\log \sum_{i=1}^{k} \zeta_i^2]^{\frac{1}{2}} \text{ infinitely often}\} = 0 .$$

This in turn implies that

$$\lim_{M\to\infty} P\{\sup_{k\geq M} |[\textstyle\sum_{i=1}^{k} \zeta_i e_{ni}^T][\sum_{i=1}^{k} \zeta_i^2]^{-\frac{1}{2}}[\log\log \sum_{i=1}^{k} \zeta_i^2]^{-\frac{1}{2}}| > 1 + \beta\} = 0$$

or equivalently that for every $\varepsilon > 0$ there exists an $M(\varepsilon)$ such that for all n for which $Kp(n) \geqq M(\varepsilon)$, $P\{\max_{M(\varepsilon)\leq k\leq Kp(n)} |\text{same}| > 1 + \beta\} < \varepsilon$. This implies that

$$\max_{1\leq k\leq Kp(n)} \textstyle\sum_{i=1}^{k} \zeta_i e_{ni}^T/(\sum_{i=1}^{k} \zeta_i^2)^{\frac{1}{2}} = O_p([\log\log \sum_{i=1}^{Kp(n)} \zeta_i^2]^{\frac{1}{2}}) .$$

Since $\zeta_i < Cn^\alpha(1 + m_n^{-\frac{1}{2}})$, $1 \leqq i \leqq n$, it follows that the above expression is $O_p([\log\log n]^{\frac{1}{2}})$ as $n \to \infty$.

It now remains to show that e_{ni} can be substituted for e_{ni}^T in the above order relation. This is done using the first part of the Borel–Canlelli Lemma (see

Loève [16], page 228): namely, if $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n$ infinitely often$) = 0$. Now,

$$\sum_{i=1}^{Kp(n)} P\{e_{ni}^T \neq e_{ni}\} = \sum_{i=1}^{Kp(n)} P\left\{|e_{ni}| \geqq \left[\frac{h(i)\sum_{j=1}^{i}\zeta_j^2}{\zeta_i^2 \log\log \sum_{j=1}^{i}\zeta_j^2}\right]^{\frac{1}{2}}\right\}$$
$$\leqq E|e_{ni}|^{2(1+\delta)} \sum_{i=1}^{Kp(n)} \left(\frac{\zeta_i^2 \log\log \sum_{j=1}^{i}\zeta_j^2}{h(i)\sum_{j=1}^{i}\zeta_j^2}\right)^{1+\delta}$$

which, by hypothesis, remains bounded as $n \to \infty$.

If we pretend that $\{e_{ni}, 1 \leqq i \leqq Kp(n)\}$ is the beginning of an infinite sequence, then the Borel-Cantelli Lemma implies that $P\{e_{ni}^T \neq e_{ni}$ infinitely often$\} = 0$. Equivalently, for every $\varepsilon > 0$ there exists an $M(\varepsilon)$ such that $P\{e_{ni} - e_{ni}^T = 0$ for all $i \geqq M(\varepsilon)\} \geqq 1 - \varepsilon$. This implies that $\max_{1 \leq k \leq Kp(n)} |\sum_{i=1}^{k} \zeta_1(e_{ni} - e_{ni}^T)|/(\sum_{i=1}^{k}\zeta_i^2)^{\frac{1}{2}} = o_p(1)$ as $n \to \infty$. □

Let n_{0j}, N_j denote the indices of the observations which occur immediately preceding $\tau_j^{(0)}$, $\hat{\tau}_j$, $j = 1, \cdots, r-1$. Assume that the spacing of the observations around the change-over points satisfies the following condition:

There exist functions $p_j(n)$, $j = 1, \cdots, r-1$ such that for each j the assumptions of Lemma 3.13 are satisfied with $N = |N_j - n_{0j}|$, $p(n) = p_j(n)$, and $\zeta_{ni} = f_j^{(0)}(t_{n,n_{0j}\pm i}) - f_{j+1}^{(0)}(t_{n,n_{0j}\pm i})$.

Remark. If $\mu(t)$ is a broken line and the t_{ni} are equally spaced, then equation (3.7) holds true with $h(i) = i^{-\delta/2}$, $p(n) = n$, and N, ζ_{ni} defined as in the preceding paragraph. To see this, express $f_j^{(0)}(t) = a_j^{(0)} + b_j^{(0)}(t - \tau_j^{(0)})$, $f_{j+1}^{(0)}(t) = a_j^{(0)} + b_{j+1}^{(0)}(t - \tau_j^{(0)})$. Then ζ_{ni} is $(b_j^{(0)} - b_{j+1}^{(0)})i/n(1 + o(1))$ and so $n\zeta_{ni} \to (b_j^{(0)} - b_{j+1}^{(0)})i \equiv \zeta_i$. The assertion in (3.7) is now easily verified.

Next, a key theorem in the rate of convergence argument is stated and proved. The theorem guarantees that within any subset of t values which contains a "substantial" portion of the observations, the estimated regression function must be "quite close" to the true regression function, at least at one point. An assumption is required to the effect that "enough" observations are taken within each (true) segment of the regression function.

Suppose S_j is a subset of $(\tau_{j-1}^{(0)}, \tau_j^{(0)})$, $j = 1, \cdots, r$ and that with probability approaching one as $n \to \infty$, $S_j \in (\hat{\tau}_{j-1}, \hat{\tau}_j)$, $j = 1, \cdots, r$. Let $\mathbf{M}_j$ denote the $K(j) \times K(j)$ matrix $(1/n \sum_{S_j} f_{jh}(t_{ni}) f_{jk}(t_{ni}))$, $h, k = 1, 2, \cdots, K(j)$. Assume that the minimum eigenvalue of $\mathbf{M}_j$ is $\lambda_j + o(1)$ as $n \to \infty$ where $\lambda_j > 0$, $j = 1, \cdots, r$. In other words, the subset S_j contains a proportion of the information which is bounded away from 0 as $n \to \infty$.

Theorem 3.14. *Suppose W is a subset of* $[0, 1]$ *such that* $H(W) > 0$. *Then*

$$\min_{t_{ni} \in W} |\hat{\nu}(t_{ni})| = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) . \tag{3.9}$$

Proof. $(r = 2)$. The proof is given only for the case $r = 2$, to simplify the notation, with no substantial loss in generality. Let $\mathbf{t}$ denote the n-vector

$(t_{n1}, \cdots, t_{nn})$ and let $\boldsymbol{\mu}_0$ denote the n-vector $(\mu(\boldsymbol{\xi}^{(0)}; t_{n1}), \cdots, \mu(\boldsymbol{\xi}^{(0)}; t_{nn}))$ within (and only within) this proof.

There are $n + 1$ ways in which the t_{ni} may be divided among two segments. Consider the kth of these partitions: $(t_{n1}, \cdots, t_{n,k-1})$, $(t_{nk}, \cdots, t_{nn})$. Let $\mathscr{F}_k$ be the linear space spanned by the q $(\equiv K(1) + K(2))$ n-vectors where the ith component of the jth vector is

$$\begin{array}{lll} f_{1j}(t_{ni}) & \text{if } i \leqq k-1 & \text{and } j \leqq K(1) \\ 0 & \text{if } i \geqq k & \text{and } j \leqq K(1) \\ 0 & \text{if } i \leqq k-1 & \text{and } j \equiv K(1) + h > K(1) \\ f_{2h}(t_{ni}) & \text{if } i \geqq k & \text{and } j \equiv K(1) + h > K(1) \end{array}$$

and let $\mathscr{F}_k^+ = \mathscr{F}_k \oplus [\boldsymbol{\mu}_0(\mathbf{t})]$ denote the direct sum of the two vector spaces. Let Q_k^+ denote the orthogonal projection onto $\mathscr{F}_k^+$, Q_k the orthogonal projection onto $\mathscr{F}_k$.

Let $\mathbf{X} \equiv (X_{n1}, \cdots, X_{nn})'$ and let $\boldsymbol{\mu}_k^*$ be the orthogonal projection of $\mathbf{X}$ onto $\mathscr{F}_k^+$, $\tilde{\boldsymbol{\mu}}_k$ the closest point to $\mathbf{X}$ in $\mathscr{F}_k^+$, subject to the underlying continuity restrictions. This is displayed schematically in Figure 3.3.

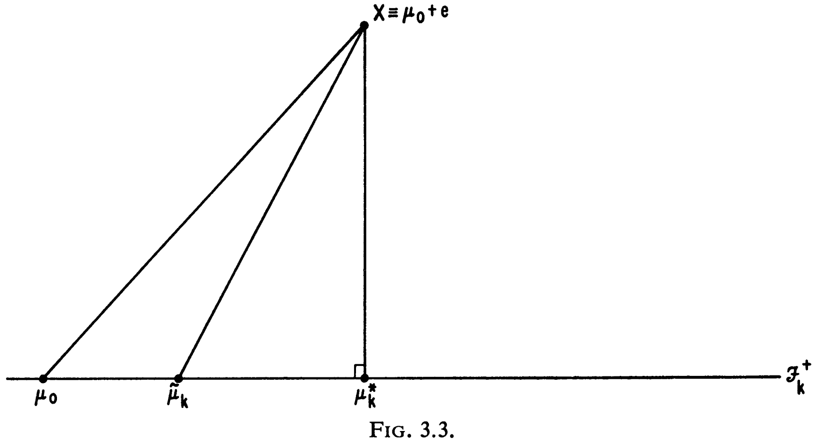

FIG. 3.3.

Then

$$||\mathbf{X} - \boldsymbol{\mu}_k^*||^2 + ||\boldsymbol{\mu}_k^* - \tilde{\boldsymbol{\mu}}_k||^2 = ||\mathbf{X} - \tilde{\boldsymbol{\mu}}_k||^2 \leqq ||\mathbf{X} - \boldsymbol{\mu}_0||^2$$

which implies

$$||\mathbf{X} - \boldsymbol{\mu}_0||^2 - ||\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0||^2 + ||\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0||^2 \\ - 2(\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0, \tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0) + ||\tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0||^2 \leqq ||\mathbf{X} - \boldsymbol{\mu}_0||^2 .$$

Thus

$$||\tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0||^2 \leqq 2(\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0, \tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0) \leqq 2||\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0||\,||\tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0||$$

and so

$$||\tilde{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_0|| \leqq 2||\boldsymbol{\mu}_k^* - \boldsymbol{\mu}_0|| = 2||Q_k^+\mathbf{e}|| . \tag{3.10}$$

This computation is an important step in showing that $||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0|| = O_p((\log\log n)^{\frac{1}{2}})$.

The estimated regression belongs to the random linear space, $\hat{\mathscr{F}}$, which is spanned by q n-vectors whose components are almost the same as those of the q n-vectors that span $\mathscr{F}_k$, except that the condition $i \leqq k - 1$ is replaced by $i \leqq N_1$, so that there are now a random number of "0" components in the vectors. This implies that $\hat{\boldsymbol{\nu}}(\mathbf{t})$ belongs to $\hat{\mathscr{F}}^+ \equiv \hat{\mathscr{F}} \oplus [\boldsymbol{\mu}_0(\mathbf{t})]$, the direct sum of the two vector spaces. The vector space $\hat{\mathscr{F}}^+$ is also generated by the direct sum of $\hat{\mathscr{F}}$ and the vector $\boldsymbol{\zeta}$, where $\boldsymbol{\zeta}' = (0, \cdots, 0, f_1^{(0)}(t_{n,N_1+1}) - f_2^{(0)}(t_{n,N_1+1}), \cdots, f_1^{(0)}(t_{n,n_{01}}) - f_2^{(0)}(t_{n,n_{01}}), 0, \cdots, 0)$. Let $\hat{Q}^+$, $\hat{Q}$ denote the orthogonal projections onto $\hat{\mathscr{F}}^+$, $\hat{\mathscr{F}}$ respectively.

If it can be shown that $||\hat{Q}^+\mathbf{e}|| = O_p((\log\log n)^{\frac{1}{2}})$, then equation (3.10) implies $||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0|| = O_p((\log\log n)^{\frac{1}{2}})$. Thus any subset, W, of the t_{ni}'s which contains a proportion of the observations asymptotically bounded away from 0, must contain a t_{ni} for which $\hat{\nu}(t_{ni}) = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$. This implies (3.9).

It now remains to show that $||\hat{Q}^+e|| = O_p((\log\log n)^{\frac{1}{2}})$. Recall that it was assumed that with large probability as $n \to \infty$, $S_j \in (\hat{\tau}_{j-1}, \hat{\tau}_j) \cap (\tau_{j-1}^{(0)}, \tau_j^{(0)})$ $j = 1, 2$. Since the minimum eigenvalues, λ_1, λ_2 of the information matrices $\mathbf{M}_1, \mathbf{M}_2$ are asymptotically bounded away from 0 as $n \to \infty$, it follows that the proportions of observations in the subsets S_1, S_2 must each be asymptotically bounded away from 0 as $n \to \infty$. With large probability as $n \to \infty$ the sets S_1, S_2 do not intersect the interval $(\hat{\tau}_1, \tau_1^{(0)})$ and so the components of the vector $\boldsymbol{\zeta}$ are identically 0 for $t_{ni} \in S_1$, $t_{ni} \in S_2$. Thus, intuitively, one would expect the vector $\boldsymbol{\zeta}$ to be at a substantial angle to $\hat{\mathscr{F}}$ (in fact almost orthogonal) as $n \to \infty$. This intuition can be quantified by demonstrating the existence of an $\alpha < 1$ such that with probability approaching one as $n \to \infty$, $|(\boldsymbol{\zeta}, \mathbf{g})| \leqq \alpha||\boldsymbol{\zeta}||\,||\mathbf{g}||$ for all $\mathbf{g} \in \hat{\mathscr{F}}$. The calculation of such an α is not difficult but is omitted to avoid digressions. It thus follows from Lemma 3.11 that

$$||\hat{Q}^+e|| \leqq \{1/(1 - \alpha)(||\hat{Q}\mathbf{e}|| + |(\boldsymbol{\zeta}, \mathbf{e})|/||\boldsymbol{\zeta}||)\}(1 + o_p(1)) .$$

Therefore, if it can be shown that $||\hat{Q}\mathbf{e}|| = O_p(1)$, $(\boldsymbol{\zeta}, \mathbf{e})/||\zeta|| = O_p((\log\log n)^{\frac{1}{2}})$, then

$$||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0|| \leqq 2||\hat{Q}^+e|| = O_p((\log\log n)^{\frac{1}{2}}) .$$

That $||\hat{Q}\mathbf{e}|| = O_p(1)$ is a consequence of Lemma 3.12 and the assumptions regarding S_j, $\mathbf{M}_j$, $j = 1, 2$. From standard least squares results

$$||\hat{Q}\mathbf{e}||^2 \leqq \textstyle\sum_{j=1}^2 \mathbf{v}_j'\mathbf{M}_j^{-1}[1 + o_p(1)]\mathbf{v}_j = \{\sum_{j=1}^2 \mathbf{v}_j'\mathbf{M}_j^{-1}\mathbf{v}_j\}(1 + o_p(1))$$

where $n^{\frac{1}{2}}\mathbf{v}_1' = (\sum_{i=0}^{N_1} f_{11}(t_{ni})e_{ni}, \cdots, \sum_{i=0}^{N_1} f_{1K(1)}(t_{ni})e_{ni})$ and where $n^{\frac{1}{2}}\mathbf{v}_2' = (\sum_{i=N_1+1}^{n} f_{21}(t_{ni})e_{ni}, \cdots, \sum_{i=N_1+1}^{n} f_{2K(2)}(t_{ni})e_{ni})$. Lemma 3.12 implies that $P\{\mathbf{v}_j'\mathbf{M}_j^{-1}\mathbf{v}_j \geqq \lambda^2\} \leqq \lambda^{-2} \operatorname{tr}\{\mathbf{M}_j^*\mathbf{M}_j^{-1}\}$ where $\mathbf{M}_j^*$ is the $K(j) \times K(j)$ matrix $(n^{-1}\sum_{i=1}^n f_{jh}(t_{ni})f_{jk}(t_{ni}))$, $h, k = 1, 2, \cdots, K(j)$. Since $\operatorname{tr}\{\mathbf{M}_j^*\mathbf{M}_j^{-1}\} = O(1)$, it follows that $\mathbf{v}_j'\mathbf{M}_j^{-1}\mathbf{v}_j = O_p(1)$ and so $||\hat{Q}\mathbf{e}|| = O_p(1)$.

It now remains to show that $(\boldsymbol{\zeta}, \mathbf{e})/||\boldsymbol{\zeta}|| = O_p((\log\log n)^{\frac{1}{2}})$. But this follows as a direct consequence of Lemma 3.13. Thus the proof of the theorem is complete.

LEMMA 3.15. *If $\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by the q-vector $\mathbf{t} \equiv (t_1, \cdots, t_q)$ and the components of* $\mathbf{t}$ *are centers of observations, then the assumptions of Theorem* 3.14 *relating to S_j, $\mathbf{M}_j$, $j = 1, \cdots, r$ are satisfied.*

PROOF. It suffices to demonstrate S_j, $\mathbf{M}_j$, $j = 1, \cdots, r$. Let S_j consist of the union of small neighborhoods about each of the components of $\mathbf{t}$ that lie within $(\tau_{j-1}^{(0)}, \tau_j^{(0)})$, and are not limit points of $\Delta_j^{(0)}$. Define $\mathbf{M}_j$ as previously. Since the elements of S_j are not limit points of $\Delta_j^{(0)}$ and Lemma 3.8 implies that $\hat{\tau}_j$ is arbitrarily close to $\Delta_j^{(0)}$ with large probability as $n \to \infty$, it follows that $S_j \in (\hat{\tau}_{j-1}, \hat{\tau}_j)$, $j = 1, \cdots, r$, with probability approaching one as $n \to \infty$. Let $\mathbf{G}_n$ be the $q \times q$ block diagonal matrix consisting of r diagonal blocks and $r(r-1)$ off diagonal blocks. The jth diagonal block is the $K(j) \times K(j)$ information matrix $\mathbf{M}_j$. The (i, j)th off diagonal block is the $K(i) \times K(j)$ matrix consisting entirely of zeroes. If it can be shown that the smallest eigenvalue of $\mathbf{G}_n$ is bounded away from 0, then the same must be true for each of the $\mathbf{M}_j$. Let $\mathbf{G}^* \equiv (\mathbf{f}(\mathbf{t}))$ denote the $q \times q$ matrix with (i, j)th element

$$\begin{aligned}
G_{ij}^* &= f_{1p}(t_i) && i \leqq K(1) \quad \text{and} \quad j = p \leqq K(1) \\
&= 0 && i > K(1) \quad \text{and} \quad j = p \leqq K(1) \\
&= f_{gp}(t_i) && i = K(1) + \cdots + K(g-1) + h \leqq K(1) + \cdots + K(g) \\
& && \text{and} \quad j = K(1) + \cdots + K(g-1) + p \leqq K(1) + \cdots + K(g) \\
&= 0 && i < K(1) + \cdots + K(g-1) \quad \text{or} \quad i > K(1) + \cdots + K(g) \\
& && \text{and} \quad j = K(1) + \cdots + K(g-1) + p \leqq K(1) + \cdots + K(g)\,.
\end{aligned}$$

It follows from Lemma 3.2 that $\mathbf{G}^*$ is nonsingular, and the smallest eigenvalue of $\mathbf{G}^{*\prime}\mathbf{G}^*$ is greater than some $C > 0$. By continuity, there exist intervals, I_j, about each of the components of $\mathbf{t}$ such that if $\mathbf{t}^*$ is any q-dimensional vector with components $t_j^* \in I_j$, $j = 1, \cdots, q$, then the smallest eigenvalue of $(\mathbf{f}(\mathbf{t}^*))'(\mathbf{f}(\mathbf{t}^*))$ is greater than $C/2$. Since the components of $\mathbf{t}$ are centers of observations, there exists a $\gamma > 0$ such that each I_j contains at least $2\gamma n + o(n)$ of the t_{ni}. Form $[\gamma n]$ vectors $\mathbf{t}_{nh} \equiv (t_{nh1}, \cdots, t_{nhq})$, $h = 1, \cdots, [\gamma n]$, each having one component in I_j, $j = 1, \cdots, q$. Thus,

$$\mathbf{G}_n \geqq \frac{1}{n} \sum_{h=1}^{[\gamma n]} (\mathbf{f}(\mathbf{t}_{nh}))'(\mathbf{f}(\mathbf{t}_{nh}))\,.$$

($\mathbf{A} \geqq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite.) Therefore, the smallest eigenvalue of $\mathbf{G}_n$ is greater than $\gamma C/2 + o(1)$. This implies that $\mathbf{G}_n$ is strictly positive definite for n sufficiently large. □

Theorem 3.14 and Lemma 3.15 together imply the principal rate of convergence result, namely:

THEOREM 3.16. (*Rate of convergence*). *If* (i) *$\boldsymbol{\theta}$ is identified at $\boldsymbol{\mu}^{(0)}$ by* $\mathbf{t}$ *and the components of* $\mathbf{t}$ *are centers of observations*, (ii) *the spacing of the observations* *around the each of the breakpoints is such that the assumptions of Lemma* 3.13 *are satisfied, then*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) . \tag{3.11}$$

PROOF. Theorem 3.14 implies that within any small neighborhood of a center of observations there exists a t_{ni} such that

$$\hat{\mu}(t_{ni}) - \mu_0(t_{ni}) = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) .$$

Lemma 3.2 implies the conclusion of the theorem.

COROLLARY 3.17. *If* $\xi^{(0)} \in \bar{\omega} \subset \Xi$, *then under the hypotheses of Theorem* 3.16

$$\hat{\boldsymbol{\theta}}_\omega - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) . \tag{3.12}$$

For simplicity, it will be assumed in the sequel that $\boldsymbol{\theta}$ is *well-identified* at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$.

The rate of convergence of $\hat{\boldsymbol{\tau}}$ to $\boldsymbol{\tau}^{(0)}$ will now be considered. Suppose that $f_j(\boldsymbol{\theta}_j^{(0)}; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; t)$ have $m_j - 1$ t-derivatives in common at $\tau_j^{(0)}$, $j = 1, \cdots, r-1$. Further, suppose that f_j and f_{j+1} have continuous left and right m_jth t-derivatives at $t = \tau_j^{(0)}$ and differ in both of these derivatives. For brevity, denote these assumptions by *conditions* (τ). Let $D^+(h, j, k)$, $D^-(h, j, k)$ denote the kth right and left t-derivatives respectively of $f_h(\boldsymbol{\theta}_h^{(0)}; t)$ at $t = \tau_j^{(0)}$. If they coincide, denote their common value by $D(h, j, k)$.

Expand $f_j(\boldsymbol{\theta}_j; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}; t)$ in Taylor series about $\boldsymbol{\theta}_j^{(0)}$, $\tau_j^{(0)}$ and $\boldsymbol{\theta}_{j+1}^{(0)}$, $\tau_j^{(0)}$ respectively. Recall that $f_j(\boldsymbol{\theta}_j^{(0)}; \tau_j^{(0)}) = f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; \tau_j^{(0)})$, $D(j, j, k) = D(j+1, j, k)$, $k = 1, 2, \cdots, m_j - 1$.

For $\boldsymbol{\theta} \in \Theta$ in the neighborhood of $\boldsymbol{\theta}^{(0)}$, the intersection point, τ_j, of the two segments $f_j(\boldsymbol{\theta}_j, t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}; t)$, is obtained by solving the equation $f_{j+1}(\boldsymbol{\theta}_{j+1}; \tau_j) - f_j(\boldsymbol{\theta}_j; \tau_j) = 0$. For $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$, τ_j near $\boldsymbol{\theta}_j^{(0)}$, $\boldsymbol{\theta}_{j+1}^{(0)}$, $\tau_j^{(0)}$,

$$\begin{aligned} 0 &= f_{j+1}(\boldsymbol{\theta}_{j+1}; \tau_j) - f_j(\boldsymbol{\theta}_j; \tau_j) \\ &= \left[\frac{\partial f_{j+1}^{(0)}}{\partial \boldsymbol{\theta}_{j+1}} + o(1)\right]' (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_{j+1}^{(0)}) - \left[\frac{\partial f_j^{(0)}}{\partial \boldsymbol{\theta}_j} + o(1)\right]' (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^{(0)}) \\ &\quad + \frac{1}{m_j!}[D^\pm(j+1, j, m_j) - D^\pm(j, j, m_j) + o(1)](\tau_j - \tau_j^{(0)})^{m_j} \end{aligned}$$

where $D^\pm$ is D^+ if $\tau_j > \tau_j^{(0)}$ and D^- if $\tau_j < \tau_j^{(0)}$. Thus

$$\begin{aligned} &\frac{1}{m_j!}[D^\pm(j+1, j, m_j) - D^\pm(j, j, m_j) + o(1)](\tau_j - \tau_j^{(0)})^{m_j} \\ &\quad = \left[\frac{\partial f_j^{(0)}}{\partial \boldsymbol{\theta}_j} + o(1)\right]' (\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^{(0)}) - \left[\frac{\partial f_{j+1}^{(0)}}{\partial \boldsymbol{\theta}_{j+1}} + o(1)\right]' (\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_{j+1}^{(0)}) . \end{aligned} \tag{3.13}$$

Equation (3.13) and Theorem 3.16 imply

$$(\hat{\tau}_j - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) . \tag{3.14}$$

This is stated formally in the theorem below.

THEOREM 3.18. *If* (i) $\boldsymbol{\theta}$ *is well-identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$ *and the components of* $\mathbf{t}$ *are centers of observations*, (ii) *conditions* (τ) *above are satisfied, and* (iii) *the observations are spaced around the breakpoints in a manner so that the assumptions of Lemma* 3.13 *are satisfied, then*

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}} \qquad \textit{and}$$
$$(\hat{\tau}_j - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})\,, \qquad j = 1, \cdots, r-1\,.$$

An important special case of this theorem is stated as

COROLLARY 3.19. *If the hypotheses of Theorem* 3.18 *are satisfied and in addition* $m_1 = \cdots = m_{r-1} = 1$, *then*

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})\,. \tag{3.15}$$

Just as with Corollary 3.17

COROLLARY 3.20. *If* $\boldsymbol{\xi}^{(0)} \in \bar{\omega} \subset \Xi$ *and the hypotheses of Theorem* 3.18 *are satisfied, then* $(\hat{\tau}_{\omega,j} - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$, $j = 1, \cdots, r-1$.

COROLLARY 3.21. *If* (i) *the conditions of Theorem* 3.18 *are satisfied, and* (ii) $E(e^4) < \infty$, *then* $\hat{\sigma}^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$. *If in addition* $\boldsymbol{\xi}^{(0)} \in \bar{\omega}$, *then* $\hat{\sigma}_\omega^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$.

It was previously remarked that a necessary condition for identification of the underlying regression function is that no two adjacent segments are identically the same. However, if one imposes the additional assumption that a sufficient number of previously specified centers of observation lie within each segment, then $\boldsymbol{\theta}^{(0)}$ is identified even though two adjacent segments may be the same. However, $\boldsymbol{\theta}^{(0)}$ is obviously not well-identified. Thus Theorem 3.16 can be used to decide whether or not two adjacent segments are distinct.

More precisely, if the specified centers of observations lie within the appropriate segments and estimated coefficients from two adjacent segments (having the same functional form) differ by less than $\log n/n^{\frac{1}{2}}$, or if an estimated segment does not contain appropriate disjoint subintervals, each with at least $n/\log n$ observations, then it can be inferred that two adjacent segments are identical.

This is stated formally as

COROLLARY 3.22. *Suppose the assumptions of Theorem* 3.16 *hold and it is known a priori that certain specified centers of observation lie strictly within each segment and are sufficient to identify* $\boldsymbol{\theta}^{(0)}$. *If* (*and only if*) $\mu(\boldsymbol{\xi}^{(0)}; t)$ *contains two identical adjacent segments, then with probability approaching one as* $n \to \infty$, *either*

(i) $|\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j+1}| \leqq \log n/n^{\frac{1}{2}}$

or

(ii) *there exists an estimated segment* $[\hat{\tau}_j, \hat{\tau}_{j+1}]$ *which does not contain appropriate disjoint subintervals, each with at least* $n/\log n$ *observations.*

PROOF. For the sake of simplicity assume that $r = 2$. For all least squares solutions with the a priori specified centers contained in the appropriate segments, the argument leading up to the proof of Theorem 3.16 goes through directly. Thus the only way for (i) to be violated is for (ii) to hold.

4. Asymptotic distribution of $\hat{\xi}$. It will be assumed, unless specifically mentioned to the contrary, that $\boldsymbol{\theta}$ is well identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ and the components of $\mathbf{t}$ are centers of observations. It was shown in Theorem 3.18 that under these conditions and a mild additional assumption $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ and $(\hat{\tau}_j - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ where m_j is the lowest order t-derivative in which the segments $f_j^{(0)}$ and $f_{j+1}^{(0)}$ disagree at $t = \tau_j^{(0)}$. This enables one to discuss the asymptotic distribution of $\hat{\xi}$.

The principal idea (due to Sylwester [24]) is to form a *pseudo problem* 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 $(n/\log\log n)^{(1/2m_j)} d_j(n) \to \infty$. The term pseudo problem is used because in practice the statistician does not know $(\tau_j^{(0)})$ and thus cannot delete such observations.

Assume that the t_{ni} are distributed in a manner that implies only $o(n/\log\log n)$ observations are deleted by this process. (If $H(t)$ is continuous and has finite slope at each $\tau_j^{(0)}$, then this will be the case.) Intuitively, deleting $o(n/\log\log n)$ observations eliminates a percentage of the information which approaches zero as $n \to \infty$. Thus, from an asymptotic point of view, the deletion of these observations should not affect any of the distribution theory. It will be shown that this is in fact the case.

More precisely, it will be shown that

$$\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}), \qquad (\hat{\tau}_j^* - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}})$$
$$\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = o_p(n^{-\frac{1}{2}}), \qquad (\hat{\tau}_j^* - \tau_j^{(0)})^{m_j} - (\hat{\tau}_j - \tau_j^{(0)})^{m_j} = o_p(n^{-\frac{1}{2}})$$

where $\hat{\xi}^* \equiv (\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\tau}}^*)$ is the l.s.e. in the pseudo problem (p.l.s.e.). This is a great simplification since the p.l.s.e. behaves asymptotically as if it were known between which two consecutive observations each of the $\tau_j^{(0)}$ are located. Thus it is possible to use classical techniques to discuss the asymptotic behavior of $\hat{\xi}^*$. Notation:

(i) n^*: sample size in the pseudo problem, $n^{**} = n - n^* = o(n/\log\log n)$.
(ii) $\hat{\boldsymbol{\theta}}^*$: unrestricted l.s.e. in the pseudo problem (p.l.s.e.).
(iii) $\hat{\boldsymbol{\theta}}_\omega^*$: restricted l.s.e. in the pseudo problem.
(iv) $\sum^*$: the summation over the n^* terms of the pseudo problem.
(v) $\sum^{**}$: $\sum_{i=1}^n - \sum^*$.

Generally, a single asterisk refers to the pseudo problem.

Theorems 3.10 and 3.18 carry over directly in the pseudo problem. Thus

THEOREM 4.1. *If* $\boldsymbol{\xi}^{(0)} \in \bar{\omega} \subset \Xi$, $\boldsymbol{\theta}$ *is well-identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$ *whose components are centers of observations (in the pseudo problem), the conditions of Lemma* 3.13

are satisfied (in the pseudo problem), and $f_j(\boldsymbol{\theta}_j^{(0)}; t)$, $f_{j+1}(\boldsymbol{\theta}_{j+1}^{(0)}; t)$ *have at most* $m_j - 1$ *t-derivatives in common at* $t = \tau_j^{(0)}$, *then*

$$\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}), \quad (\tau_j^* - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \tag{4.1}$$

$$\hat{\boldsymbol{\theta}}_\omega^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}), \quad (\hat{\tau}_{\omega,j}^* - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}). \tag{4.2}$$

The asymptotic behavior of $\hat{\boldsymbol{\xi}}$ will now be considered by first treating $\hat{\boldsymbol{\xi}}^*$. Recall that Θ was defined to be the collection of $\boldsymbol{\theta}$'s which lead to functions $\mu(\boldsymbol{\xi}; t)$ satisfying the continuity restraints, and Ξ the set of corresponding $\boldsymbol{\xi}$'s. Thus, Θ is a subset of q-dimensional space, and different asymptotic behavior occurs, depending on whether $\boldsymbol{\theta}^{(0)}$ is a boundary point or an interior point of Θ. It will be shown later that $\boldsymbol{\theta}^{(0)}$ interior to Θ implies $\hat{\boldsymbol{\theta}}^*$ has an asymptotic normal distribution.

LEMMA 4.2. *Suppose* $D^+(j, j, m_j) = D^-(j, j, j, m_j)$, $D^+(j+1, j, m_j) = D^-(j+1, j, m_j)$, $j = 1, \cdots, r-1$. *If* $m_1, \cdots, m_{r-1}$ *are all odd,* $\boldsymbol{\theta}^{(0)}$ *is an interior point of* Θ. *If any of the* m_j *are even, then* $\boldsymbol{\theta}^{(0)}$ *is a boundary point of* Θ.

PROOF. Recall equation (3.13): If τ_j is the intersection point of $f_j(\boldsymbol{\theta}_j; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}; t)$

$$\frac{1}{m_j!}[D^\pm(j+1, j, m_j) - D^\pm(j, j, m_j)](\tau_j - \tau_j^{(0)})^{m_j} = \left[\frac{\partial f_j^{(0)}}{\partial \boldsymbol{\theta}_j} + o(1)\right]'(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^{(0)}) - \left[\frac{\partial f_{j+1}^{(0)}}{\partial \boldsymbol{\theta}_{j+1}} + o(1)\right]'(\boldsymbol{\theta}_{j+1} - \boldsymbol{\theta}_{j+1}^{(0)}) \tag{3.13}$$

for all $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$ sufficiently close to $\boldsymbol{\theta}_j^{(0)}$, $\boldsymbol{\theta}_{j+1}^{(0)}$. If m_j is odd and $D^+ = D^-$ then for all $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$, equation (3.13) can be solved for τ_j. However, if m_j is even, then (3.13) cannot hold both for $\boldsymbol{\theta}_j^{(0)} + \delta_j$, $\boldsymbol{\theta}_{j+1}^{(0)} + \boldsymbol{\delta}_{j+1}$ and $\boldsymbol{\theta}_j^{(0)} - \boldsymbol{\delta}_j$, $\boldsymbol{\theta}_{j+1}^{(0)} - \boldsymbol{\delta}_{j+1}$. Thus any neighborhood of $(\boldsymbol{\theta}_j^{(0)}, \boldsymbol{\theta}_{j+1}^{(0)})$ must contain points which are not in Θ and so $\boldsymbol{\theta}^{(0)}$ is a boundary point of Θ. □

An important special case of Lemma 4.2 occurs when each segment is a straight line. Then $m_1 = m_2 = \cdots = m_{r-1} = 1$.

The next lemma shows that the rate of convergence of $\hat{\boldsymbol{\theta}}^*$ to $\boldsymbol{\theta}^{(0)}$ is $n^{-\frac{1}{2}}$ rather than $n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}$.

LEMMA 4.3. $\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$.

PROOF. In analogy with equation (2.5) define

$$s^*(\boldsymbol{\xi}) = n^{-1} \sum{}^* (X_{ni} - \mu(\boldsymbol{\xi}; t_{ni}))^2 .$$

Theorem 4.1 implies that $\hat{\tau}_j^* \in L_j(n)$, $j = 1, \cdots, r-1$ with large probability as $n \to \infty$. Since there are no observations (of the pseudo problem) within the intervals $L_j(n)$ it follows that $s^*(\boldsymbol{\xi})$ within this region does not depend on τ and is a paraboloid in $\boldsymbol{\theta}$. In particular, it is twice differentiable in $\boldsymbol{\theta}$. For the remainder of the proof denote $s^*(\boldsymbol{\xi})$ by $s^*(\boldsymbol{\theta})$. Thus, with large probability as $n \to \infty$

$$\begin{aligned} 0 &\geqq s^*(\hat{\boldsymbol{\theta}}^*) - s^*(\boldsymbol{\theta}^{(0)}) \\ &= n^{-\frac{1}{2}} \frac{\partial s^*(\boldsymbol{\theta}^{(0)})'}{\partial \boldsymbol{\theta}} [n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})] \\ &\quad + \tfrac{1}{2}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})]' \left[\frac{1}{n} \frac{\partial^2 s^*(\boldsymbol{\theta}^{(0)})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}}\right] [n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})]\,. \end{aligned} \tag{4.3}$$

The proof of Lemma 3.15 implies that the smallest eigenvalue of $n^{-1}\, \partial^2 s^*(\boldsymbol{\theta}^{(0)})/\partial \boldsymbol{\theta}\, \boldsymbol{\theta}\partial$ is bounded away from zero as $n \to \infty$, and the matrix is positive definite and converges to a limit as $n \to \infty$. The vector $n^{-\frac{1}{2}}[\partial s^*(\boldsymbol{\theta}^{(0)})/\partial \boldsymbol{\theta}]$ has mean $\mathbf{0}$ and uniformly bounded variance as $n \to \infty$, so that it is $O_p(1)$. Thus equation (4.3) implies that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)}) = O_p(1)$. □

LEMMA 4.4. *If $\boldsymbol{\theta}^{(0)}$ is an interior point of Θ then*

$$\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})] \to \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1}) \tag{4.4}$$

where $\mathscr{L}(V)$ represents the distribution (law) of V and

$$\mathbf{G}_{jk} = \int_0^1 \frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t)}{\partial \theta_j} \frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t)}{\partial \theta_k}\, dH(t) \tag{4.5}$$

where $\mathbf{G}$ is the $q \times q$ information matrix and is strictly positive definite.

PROOF. Define $s^*(\boldsymbol{\xi})$ as in the proof of Lemma 4.3. Since $\boldsymbol{\theta}^{(0)}$ is interior to Θ, any $\boldsymbol{\theta}$ within a neighborhood of $\boldsymbol{\theta}^{(0)}$ is "admissible" as a possibility for $\hat{\boldsymbol{\theta}}^*$. Thus, with large probability as $n \to \infty$, $\hat{\boldsymbol{\theta}}^*$ is the point at which the *unrestrained* minimum of $s^*(\boldsymbol{\xi})$ occurs. Thus $\hat{\boldsymbol{\theta}}^*$ may be obtained by setting the derivative of $s^*(\boldsymbol{\xi})$ equal to 0. This implies

$$\begin{aligned} &\left\{\frac{1}{n} \sum{}^* \left(\frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t_{ni})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t_{ni})'}{\partial \boldsymbol{\theta}}\right)\right\} (\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)}) \\ &\qquad = \frac{1}{n} \sum{}^* \frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t_{ni})}{\partial \boldsymbol{\theta}} e_{ni}\,. \end{aligned} \tag{4.6}$$

The proof of Lemma 3.15 implies that the smallest eigenvalue of the matrix on the left-hand side of equation (4.6) is bounded away from zero. Since this matrix converges to $\mathbf{G}$, by continuity $\mathbf{G}$ must be positive definite. Thus

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)}) = \{\mathbf{G}^{-1} + o(1)\} n^{-\frac{1}{2}} \sum{}^* \frac{\partial \mu(\boldsymbol{\xi}^{(0)}; t_{ni})}{\partial \boldsymbol{\theta}} e_{ni}\,.$$

The Lindeberg–Feller central limit theorem for double sequences (see [16], page 295) implies the assertion of the lemma.

COROLLARY 4.5. *In the case of broken line regression (i.e., when each segment is a straight line), $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$ has an asymptotic normal distribution with mean $\mathbf{0}$ and covariance $\mathbf{G}^{-1}$.*

Before discussing the case when $\boldsymbol{\theta}^{(0)}$ is a boundary point of Θ, it is necessary to introduce two definitions. These appear in Chernoff [2].

DEFINITION 4.6. A set C is *positively homogeneous* if $\gamma \in C$ implies $c\gamma \in C$ for $c > 0$.

DEFINITION 4.7. The set ρ is *approximated by the positively homogeneous set* C if $\inf_{\gamma \in C} |\gamma - \eta| = o(|\eta|)$ for $\eta \in \rho$, $\eta \to 0$ and $\inf_{\eta \in \rho} |\gamma - \eta| = o(|\gamma|)$ for $\gamma \in C$, $\gamma \to 0$.

Let $\Theta - \boldsymbol{\theta}^{(0)}$ denote the translate of Θ by $\boldsymbol{\theta}^{(0)}$; i.e., $\Theta - \boldsymbol{\theta}^{(0)} = \{\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}; \boldsymbol{\theta} \in \Theta\}$. This amounts to translating the origin of the parameter space to $\boldsymbol{\theta}^{(0)}$.

LEMMA 4.8. *If* (i) $\boldsymbol{\theta}^{(0)}$ *is a boundary point of* Θ, (ii) $\Theta - \boldsymbol{\theta}^{(0)}$ *is approximated by the convex, positively homogeneous set* C, *then the asymptotic distribution of* $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})$ *is that of the closest (in the metric determined by the asymptotic information matrix,* $\mathbf{G}$) *point in* C *to* Z, *where* $\mathscr{L}(Z) = \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$.

PROOF. The p.l.s.e. $\hat{\boldsymbol{\theta}}^*$ is that element of Θ which minimizes $s^*(\boldsymbol{\xi})$. Let $\bar{\boldsymbol{\theta}}^*$ denote the parameter which minimizes $s^*(\boldsymbol{\xi})$ without regard for the continuity restraints at the change-over points. Then $\mathscr{L}[n^{\frac{1}{2}}(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})] \to \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$, where $\mathbf{G}$ is specified in the statement of Lemma 4.4.

Recall that within the region of interest of the parameter space, $s^*(\boldsymbol{\xi})$ does not depend on $\boldsymbol{\tau}$ and is a paraboloid in $\boldsymbol{\theta}$. It will thus be convenient for the remainder of the proof to denote $s^*(\boldsymbol{\xi})$ by $s^*(\boldsymbol{\theta})$. The change-over points between segments will be understood to lie in $L_j(n)$, $j = 1, \cdots, r-1$ regardless of whether or not $\boldsymbol{\theta} \in \Theta$.

For all $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}^{(0)}$

$$s^*(\boldsymbol{\theta}) = s^*(\bar{\boldsymbol{\theta}}^*) + (\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta})'\mathbf{G}_n(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \tag{4.7}$$

where $\mathbf{G}_n$ is the matrix on the left-hand side of equation (4.6). $\mathbf{G}_n$ converges to $\mathbf{G}$, where $\mathbf{G}$ is given in equation (4.5). With large probability $\hat{\boldsymbol{\theta}}^*$ is that element of Θ which minimizes $(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta})'\mathbf{G}_n(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta})$. It may be shown by an argument that utilizes the convexity of C that $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\gamma}}^* + o_p(n^{-\frac{1}{2}})$ where $\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\theta}^{(0)}$ minimizes $(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{G}(\bar{\boldsymbol{\theta}}^* - \boldsymbol{\theta})$ over all $\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)} \in C$. (The argument is a bit lengthy and so has been omitted. It is available from the author upon request.) This implies that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})$ and $n^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\theta}^{(0)})$ have the same asymptotic distribution, which may be shown by a continuity argument to be that which is asserted in the statement of the lemma. This completes the proof.

Note 1. Condition (ii) is satisfied if Θ is locally convex at $\boldsymbol{\theta}^{(0)}$.

Note 2. Condition (ii), that C is convex, is a bit stronger than what is actually needed to prove the lemma. If C is a positively homogeneous set which can be divided into disjoint convex sets $\{C_j\}$ such that with large probability as $n \to \infty$ the estimates $\hat{\boldsymbol{\theta}}^*$, $\hat{\boldsymbol{\gamma}}^*$ and $\tilde{\boldsymbol{\theta}}^*$ (the closest point in Θ to $\bar{\boldsymbol{\theta}}^*$ in the metric determined by $\mathbf{G}$) each lie in the same C_j, then the proof goes through. This relaxed condition is necessary for the validity of example 3, case (iv) at the end of this section.

Lemmas 4.4 and 4.8 are the basic building blocks for the calculation of the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, the l.s.e. in the original problem. It now remains to show that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ do not differ by too much. The next sequence of lemmas leads up to the result that $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^* = o_p(n^{-\frac{1}{2}})$. This shows that the rate of convergence of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}^{(0)}$ is $n^{-\frac{1}{2}}$ rather than $n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}$ and that $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\theta}}$ (suitably normalized) have the same asymptotic distribution.

The following three lemmas lead to uniform bounds on the order of magnitude of certain expressions which occur in regression problems. The bounds will be precisely stated in Lemmas 4.10 and 4.11.

Lemma 4.9 concerns the behavior of partial sums of numbers. This is Theorem 3.1 of Sylwester [24] and is included here only for the sake of completeness. The proof given below is due to a referee.

LEMMA 4.9. *Let y_i, $i = 1, 2, \cdots, n$ be any numbers and w_i, $i = 1, 2, \cdots, n$ be any n numbers such that $1 \geqq w_1 \geqq w_2 \geqq \cdots \geqq w_n \geqq 0$. Then*

$$\max_{1\leqq k\leqq n} |\textstyle\sum_{i=1}^k y_i| \geqq \max_{1\leqq k\leqq n} |\textstyle\sum_{i=1}^k y_i w_i| \,.$$

PROOF. Define $w_{n+1} = 0$. Let $d_i = w_i - w_{i+1} \geqq 0$, $i = 1, \cdots, n$. Then $w_i = \sum_{j=i}^n d_j$. Thus

$$\begin{aligned}
&\max_{1\leqq k\leqq n} |\textstyle\sum_{i=1}^k y_i w_i| \\
&\quad \equiv \max_{1\leqq k\leqq n} |\textstyle\sum_{i=1}^k y_i(\sum_{j=i}^n d_j)| = \max_{1\leqq k\leqq n} |\sum_{j=1}^n d_j \sum_{i=1}^{\min(j,k)} y_i| \\
&\quad \leqq \textstyle\sum_{j=1}^n d_j[\max_{1\leqq k\leqq n} |\sum_{i=1}^{\min(j,k)} y_i|] = \sum_{j=1}^n d_j[\max_{1\leqq k\leqq j} |\sum_{i=1}^k y_i|] \\
&\quad \leqq (\textstyle\sum_{j=1}^n d_j)(\max_{1\leqq k\leqq n} |\sum_{i=1}^k y_i|) \leqq \max_{1\leqq k\leqq n} |\sum_{i=1}^k y_i| \,. \qquad \square
\end{aligned}$$

Let $\mathscr{F}_s \equiv \{f(t)\}$ be the class of functions defined on $0 \leqq t \leqq 1$ such that $f(t)$ is composed of at most s segment each of which is a differentiable function possessing at most z sign changes in derivative. Let e_{ni}, $i = 1, 2, \cdots, n$ be a double sequence of i.i.d. random variables having mean 0 and variance 1. Denote $f(t_{ni})$ by f_{ni}. For convenience arrange the t_{ni} so that $0 \leqq t_{n1} \leqq \cdots \leqq t_{nn} \leqq 1$.

LEMMA 4.10.

$$|\textstyle\sum_{i=1}^n f_{ni} e_{ni}| \leqq \{\max_{1\leqq i\leqq n} |f_{ni}|\} O_p(n^{\frac{1}{2}})$$

where $O_p(n^{\frac{1}{2}})$ applies uniformly over all $f \in \mathscr{F}_s$.

PROOF. Assume without loss of generality $\max |f_{ni}| = 1$. By definition of $\mathscr{F}_s$, the set $I \equiv \{t: |f(t)| \leqq 1\}$, which contains all the t_{ni}, consists of at most $s(z + 1)$ (possibly degenerate) intervals in each of which $df(t)/dt$ has at most z zeros. Therefore each of the $s(z + 1)$ intervals can be subdivided into $2(z + 1)$ or less subintervals, throughout which $f(t)$ is positive and increasing, negative and increasing, positive and decreasing, or negative and decreasing. Let $I(k)$, $k = 1, 2, \cdots, 2s(z + 1)^2$ denote the kth such subinterval. Obviously

$$|\textstyle\sum_{i=1}^n f_{ni} e_{ni}| \leqq \sum_k |\sum_{I(k)} f_{ni} e_{ni}| \,. \tag{4.8}$$

The f_{ni} can be assumed positive and decreasing within each subinterval, since otherwise the f_{ni} can be multiplied by minus one or the summation can be taken in reverse order. By Lemma 4.9

$$(4.9) \qquad |\sum_{I(k)} f_{ni} e_{ni}| \leqq \max_{1\leqq h\leqq k\leqq n} |\sum_{i=h}^{k} e_{ni}| \equiv A_n .$$

Equation (4.9) is valid for every nondegenerate interval.

The set I may also contain degenerate one-point intervals in which all $f(t_{ni})$ are $+1$ or all are -1. For these intervals (4.9) holds trivially. Hence from (4.8) and (4.9)

$$(4.10) \qquad |\sum_{i=1}^{n} f_{ni} e_{ni}| \leqq 2s(z+1)^2 A_n .$$

Thus far in the proof probability has not been mentioned. The only random variable on the right-hand side of (4.10) is A_n, which does not involve $\mathscr{F}_s$. By Kolmogorov's inequality $A_n = O_p(n^{\frac{1}{2}})$. $\square$

The proof of Lemma 4.10 can easily be extended to yield

LEMMA 4.11. *Let $\mathscr{S}_m$ denote the collection of subsets, S, of* [0, 1] *each of which consists of at most m intervals. Then*

$$|\sum_S f_{nk} e_{ni}| \leqq \{\max_S |f_{ni}|\} O_p(n^{\frac{1}{2}})$$

uniformly for $f \in \mathscr{F}_s$ and $S \in \mathscr{S}_m$.

LEMMA 4.12. *If* (i) *$\boldsymbol{\theta}$ is well-identified at $\boldsymbol{\mu}^{(0)}$ by* $\mathbf{t}$ *and the components of* $\mathbf{t}$ *are centers of observations,* (ii) *the conditions of Lemma* 3.13 *are satisfied* (*in both the original and pseudo problems*), (iii) Θ *is locally convex at $\boldsymbol{\theta}^{(0)}$, then*

$$(4.11) \qquad \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^* = o_p(n^{-\frac{1}{2}}) .$$

PROOF. It follows from the definition of identification that $\boldsymbol{\theta}$ is also identified at $\boldsymbol{\mu}^{(0)}$ by $\mathbf{t}$ in the pseudo problem and the components of $\mathbf{t}$ are centers of observations with respect to the pseudo problem. It follows from Theorems 3.16 and 4.1 that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ and $\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$.

Select $a_n > 0$ such that $a_n/(\log\log n)^{\frac{1}{2}} \to \infty$ and $a_n = o((n/n^{**})^{\frac{1}{2}})$. Let $\mathscr{U}_n = \{\boldsymbol{\xi} \in \Xi : |\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}| < a_n n^{-\frac{1}{2}},\ \tau_j \in L_j(n), j = 1, \cdots, r-1\}$. Then $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^*$ both lie in $\mathscr{U}_n$ with large probability as $n \to \infty$. Note that the function $s^*(\boldsymbol{\xi})$ depends only on $\boldsymbol{\theta}$ for $\boldsymbol{\xi} \in \mathscr{U}_n$, so that $s^*(\boldsymbol{\xi}) = s^*(\boldsymbol{\theta})$.

Recall $s(\boldsymbol{\xi}) = n^{-1}\sum_{i=1}^{n}(e_{ni} + \nu_{ni})^2$, $s^*(\boldsymbol{\xi}) = n^{-1}\sum^{*}(e_{ni} + \nu_{ni})^2$. Thus

$$(4.12) \qquad \begin{aligned} s(\boldsymbol{\xi}) &= s^*(\boldsymbol{\xi}) + n^{-1}\textstyle\sum^{**}(e_{ni} + \nu_{ni})^2 \\ &= s^*(\xi) + n^{-1}\textstyle\sum^{**} e_{ni}^2 + 2n^{-1}\sum^{**} e_{ni}\nu_{ni} + n^{-1}\sum^{**}\nu_{ni}^2 . \end{aligned}$$

It follows from the definition of $\mathscr{U}_n$ that

$$\sup_{\boldsymbol{\xi} \in \mathscr{U}_n} \max_{t \in \cup L_j(n)} |\nu(\boldsymbol{\xi}; t)| = O(a_n n^{-\frac{1}{2}}) .$$

Thus $\sup_{\boldsymbol{\xi} \in \mathscr{U}_n} |n^{-1}\sum^{**}\nu_{ni}^2| = o(n^{-1})$ and Lemma 4.11 with n^{**} substituted for n implies

$$\sup_{\boldsymbol{\xi} \in \mathscr{U}_n} |n^{-1}\sum^{**} e_{ni}\nu_{ni}| = o_p(n^{-1}) .$$

Equation (4.12) thus implies

$$s(\boldsymbol{\xi}) = s^*(\boldsymbol{\xi}) + n^{-1} \textstyle\sum^{**} e_{ni}^2 + o_p(n^{-1}) \tag{4.13}$$

where $o_p(n^{-1})$ is uniformly small for $\boldsymbol{\xi} \in \mathscr{U}_n$.

Since $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^*$ are the l.s.e. and p.l.s.e. respectively,

$$s(\hat{\boldsymbol{\xi}}) \leqq s(\hat{\boldsymbol{\xi}}^*)\,, \qquad s^*(\hat{\boldsymbol{\xi}}^*) \leqq s^*(\hat{\boldsymbol{\xi}})\,. \tag{4.14}$$

Equations (4.13) and (4.14) imply

$$0 \leqq s(\hat{\boldsymbol{\xi}}^*) - s(\hat{\boldsymbol{\xi}}) = s^*(\hat{\boldsymbol{\xi}}^*) - s^*(\hat{\boldsymbol{\xi}}) + o_p(n^{-1}) \leqq o_p(n^{-1})\,. \tag{4.15}$$

Therefore $s^*(\hat{\boldsymbol{\xi}}) - s^*(\hat{\boldsymbol{\xi}}^*) = o_p(n^{-1})$. Since Θ is locally convex at $\boldsymbol{\theta}^{(0)}$ and $\hat{\boldsymbol{\theta}}^*$ is the minimizing value in Θ, it follows that the directional derivative of $s^*(\boldsymbol{\xi})$ at $\hat{\boldsymbol{\theta}}^*$ in any direction into Θ is nonnegative. Now

$$\begin{aligned} s^*(\hat{\boldsymbol{\xi}}) = s^*(\hat{\boldsymbol{\xi}}^*) &+ n^{-\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*)'\left[n^{\frac{1}{2}} \frac{\partial s^*(\hat{\boldsymbol{\xi}}^*)}{\partial \boldsymbol{\theta}}\right] \\ &+ \tfrac{1}{2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*)' \frac{\partial^2 s(\hat{\boldsymbol{\xi}}^*)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*)\,. \end{aligned} \tag{4.16}$$

Note that the last two terms on the right side of the above equation are positive. Equations (4.15) and (4.16) imply $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^* = o_p(n^{-\frac{1}{2}})$. □

Note. Condition (iii) is more than what is actually needed to prove the lemma. If Θ can be partitioned into disjoint locally convex sets $\{\Theta_j\}$ such that with large probability as $n \to \infty$ $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^*$ lie in the same Θ_j, then the directional derivative in (4.16) is positive and the proof of the lemma goes through. This relaxed condition is necessary for the validity of example 3, case (iv) at the end of this section.

Lemma 4.12 implies that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})$ and $n^{\frac{1}{2}}(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^{(0)})$ have the same asymptotic distribution. Thus

THEOREM 4.13. *Suppose* (i) $\boldsymbol{\theta}$ *is well-identified at* $\boldsymbol{\mu}^{(0)}$ *by* $\mathbf{t}$ *and the components of* $\mathbf{t}$ *are centers of observations,* (ii) *the conditions of Lemma* 3.13 *are satisfied.*

(a) *If* $\boldsymbol{\theta}^{(0)}$ *is an interior point of* Θ *then* $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})] \to \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$ *where* $\mathbf{G}$ *is the positive definite information matrix and*

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

(b) *If* $\boldsymbol{\theta}^{(0)}$ *is a boundary point of* Θ *and* $\Theta - \boldsymbol{\theta}^{(0)}$ *is approximated by the convex, positively homogeneous set* C, *then* $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})]$ *converges to the distribution of* $\boldsymbol{\pi}(\mathbf{Z})$ *where* $\boldsymbol{\pi}(\mathbf{Z})$ *minimizes* $(\mathbf{Z} - \boldsymbol{\theta})'\mathbf{G}(\mathbf{Z} - \boldsymbol{\theta})$ *among all* $\boldsymbol{\theta} \in C$, *and* $\mathscr{L}(\mathbf{Z}) = \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$.

Note. The convexity assumption in (b) can be relaxed in the manner discussed in Note 2 following Lemma 4.8.

COROLLARY 4.14. *If* $D^+(j, j, m_j) = D^-(j, j, m_j)$, $D^+(j+1, j, m_j) = D^-(j+1, j, m_j)$, $j = 1, 2, \cdots, r-1$ *and* $m_1, \cdots, m_{r-1}$ *are odd, then* $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})] \to \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$.

The conditions of the corollary are satisfied in the important special case of broken line regression.

REMARK. If $Ee^4 < \infty$ then $\hat{\sigma}^2 - \hat{\sigma}^{*2} = o_p(n^{-\frac{1}{2}})$. To see this, note that $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (e_{ni} + \hat{\nu}_{ni})^2 = n^{-1}\sum_{i=1}^n e_{ni}^2 + O_p(n^{-1})$ and $\hat{\sigma}^{*2} = n^{*-1}\sum^*(e_{ni} + \hat{\nu}_{ni}^*)^2 = n^{*-1}\sum^* e_{ni}^2 + O_p(n^{-1})$. The central limit theorem implies that $\sum^* e_{ni}^2 = \sigma_0^2 n^* + X$, $\sum_{i=1}^n e_{ni}^2 = \sigma_0^2 n + X + Y$ where $X = O_p(n^{\frac{1}{2}})$, $Y = o_p(n^{\frac{1}{2}})$. Thus $\hat{\sigma}^2 - \hat{\sigma}^{*2} = o_p(n^{-\frac{1}{2}})$.

The behavior of the $\hat{\tau}_j$'s will now be discussed. First the asymptotic behavior of the $\hat{\tau}_j^*$'s will be considered. It will then be shown that $\hat{\tau}_j$ and $\hat{\tau}_j^*$ are close and thus have the same asymptotic distribution.

Refer back to equation (3.13) for the intersection point of $f_j(\boldsymbol{\theta}_j; t)$ and $f_{j+1}(\boldsymbol{\theta}_{j+1}; t)$, for all $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$ sufficiently close to $\boldsymbol{\theta}_j^{(0)}$, $\boldsymbol{\theta}_{j+1}^{(0)}$, $j = 1, 2, \cdots, r-1$. Denote $(1/m_j!)[D^\pm(j+1, j, m_j) - D^\pm(j, j, m_j)]$ by $D_j^\pm$, $\partial f_j(\boldsymbol{\theta}_j^{(0)}, \tau_k^{(0)})/\partial\boldsymbol{\theta}_j$ by ${}_k\mathbf{f}_j$, and delete the $\pm$ from $D_j^\pm$ if the mth t-derivatives are continuous at $\tau_j^{(0)}$. Since $\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)} = O_p(n^{-\frac{1}{2}})$, equation (3.13) implies that $(\hat{\tau}_j^* - \tau_j^{(0)})^{m_j} = O_p(n^{-\frac{1}{2}})$ and so (3.13) may be rewritten as

$$n^{\frac{1}{2}}\mathbf{B}^* = \mathbf{A}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})] \tag{4.17}$$

where $\mathbf{B}^*$ is the $r-1$ dimensional column vector with ith coordinate $(\hat{\tau}_j^* - \tau_i^{(0)})^{m_i}$ and $\mathbf{A} \equiv \mathbf{A}(\boldsymbol{\tau}^{(0)})$ is the $(r-1) \times q$ matrix whose (i, j)th components are ${}_i\mathbf{f}_i'/D_i^\pm$ for $K(1) + \cdots + K(i-1) + 1 \leqq j \leqq K(1) + \cdots + K(i)$, $-{}_i\mathbf{f}_{i+1}'/D_i^\pm$ for $K(1) + \cdots + K(i) + 1 \leqq j \leqq K(1) + \cdots + K(i+1)$, and 0 otherwise, where i runs from 1 to $r-1$.

The asymptotic distribution of $\hat{\tau}_j^*$ depends on the value of m_j and on whether or not $D_j^+ = D_j^-$. Suppose first that all of the m_j are odd and $D_j^+ = D_j^- = D_j$, $j = 1, 2, \cdots, r-1$. Lemma 4.2 and Theorem 4.13 (a) imply that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})$ is asymptotically normal. Equation (4.17) then implies

$$\mathscr{L}\{n^{\frac{1}{2}}((\hat{\tau}_1^* - \tau_1^{(0)})^{m_1}, \cdots, (\hat{\tau}_{r-1}^* - \tau_{r-1}^{(0)})^{m_{r-1}})'\} \to \mathscr{N}(\mathbf{0}, \mathbf{A}\mathbf{G}^{-1}\mathbf{A}') .$$

If $D_j^+ \neq D_j^-$ then $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})$ may or may not be asymptotically normal, depending on just how the segments intersect. This will be illustrated by example. It is apparent though from (4.17) that $n^{\frac{1}{2}}(\hat{\tau}_j^* - \tau_j^{(0)})^{m_j}$ is not asymptotically normal. Its asymptotic distribution is a mixture of half normal distributions.

If m_j is even, then it can be seen from (3.13) that $n^{\frac{1}{2}}((\hat{\boldsymbol{\theta}}_j^* - \boldsymbol{\theta}_j^{(0)})', (\hat{\boldsymbol{\theta}}_{j+1}^* - \boldsymbol{\theta}_{j+1}^{(0)})')$ does not generally have an asymptotic normal distribution. However, if $D_k^+ = D_k^-$ and m_k is odd for $k \neq j$, then $\hat{\boldsymbol{\theta}}_k^*$ is the unrestrained l.s.e. for $k \neq j-1, j, j+1$. Thus, $n^{\frac{1}{2}}(\hat{\tau}_k^* - \tau_k^{(0)})^{m_k}$ is asymptotically normally distributed for all $k \neq j-1, j, j+1$.

This discussion is summarized in the following theorem.

THEOREM 4.15. (i) *If m_j is odd and $D_j^+ = D_j^-$, $j = 1, 2, \cdots, r-1$, then*

$$\mathscr{L}\{n^{\frac{1}{2}}(\hat{\tau}_1{}^* - \tau_1{}^{(0)})^{m_1}, \cdots, n^{\frac{1}{2}}(\hat{\tau}_{r-1}^* - \tau_{r-1}^{(0)})^{m_{r-1}}\} \to \mathscr{N}(\mathbf{0}, \mathbf{AG}^{-1}\mathbf{A}') \tag{4.18}$$

where $\mathbf{A} \equiv \mathbf{A}(\boldsymbol{\tau}^{(0)})$ is the $(r-1) \times q$ matrix on the right side of (4.17). *If m_j is odd but $D_j^+ \neq D_j^-$, then the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\tau}_j{}^* - \tau_j{}^{(0)})^{m_j}$ may be nonnormal.*

(ii) *If m_k is even, then with the exception of $j = k-1, k, k+1$ the statements in* (i) *are applicable.*

It will now be shown that $\hat{\tau}_j$ and $\hat{\tau}_j{}^*$ have the same asymptotic distribution.

LEMMA 4.16. *Let $\mathbf{B}^*$ be the $r-1$ dimensional column vector defined in* (4.17) *and let $\mathbf{B}$ be defined similarly, but with $\hat{\tau}_i{}^*$ replaced by $\hat{\tau}_i$. Then $n^{\frac{1}{2}}(\mathbf{B} - \mathbf{B}^*) = o_p(1)$.*

PROOF. Equation (4.17) is valid at $\hat{\boldsymbol{\xi}}$ as well as $\hat{\boldsymbol{\xi}}^*$. Thus

$$n^{\frac{1}{2}}(\mathbf{B} - \mathbf{B}^*) = \mathbf{A}(\boldsymbol{\tau}^{(0)})n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^*) + o_p(1) = o_p(1)\,. \qquad \square$$

Theorem 4.15 and Lemma 4.16 imply

THEOREM 4.17. *The asymptotic distribution of $\hat{\boldsymbol{\tau}}$ as $n \to \infty$ is the same as that stated for $\hat{\boldsymbol{\tau}}^*$ in Theorem* 4.15. *In particular, if $m_1, \cdots, m_{r-1}$ are odd and $D_j^+ = D_j^- = D_j$, $j = 1, \cdots, r-1$, then $\mathscr{L}(n^{\frac{1}{2}}\mathbf{B}) \to \mathscr{N}(\mathbf{0}, \mathbf{AG}^{-1}\mathbf{A}')$ where $\mathbf{B}$ is defined in Lemma* 4.16.

The section will be concluded with several examples that illustrate the results stated in the theorems.

1. Consider the two-segment broken line regression model.

$$\begin{aligned} \mu(\boldsymbol{\xi}; t) &= \theta_{11} + \theta_{12}t \qquad & 0 \leqq t \leqq \tau^{(0)} \\ &= \theta_{21} + \theta_{22}t \qquad & \tau^{(0)} \leqq t \leqq 1\,. \end{aligned}$$

Here $r = 2$, $K(1) = K(2) = 2$, $m_1 = 1$, $D_1^+ = D_1^- = \theta_{22}^{(0)} - \theta_{12}^{(0)}$. If $\mu(\boldsymbol{\xi}^{(0)}; t)$ is identified by centers of observations, Corollary 4.14 implies that $\{n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})\} \to \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$, where $\sigma_0{}^2\mathbf{G}$ is the 4×4 information matrix with components

$$\begin{aligned} \sigma_0{}^2G_{ij} &= \textstyle\int_0^{\tau^{(0)}} t^{i+j-2}\,dH(t) \qquad & &\text{for } i = 1, 2\,;\ j = 1, 2 \\ &= 0 & &\text{for } i = 1, 2\,;\ j = 3, 4 \ \text{ or } \ i = 3, 4\,;\ j = 1, 2 \\ &= \textstyle\int_{\tau(0)}^1 t^{i+j-6} & &\text{for } i = 3, 4\,;\ j = 3, 4\,. \end{aligned}$$

Equation (4.17) (substituting $\hat{\tau}$ for $\hat{\tau}^*$) asserts that

$$n^{\frac{1}{2}}(\hat{\tau} - \tau^{(0)}) = n^{\frac{1}{2}}\mathbf{A}(\tau^{(0)})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)}) + o_p(1)$$

where $\mathbf{A}(\tau^{(0)}) = (1, \tau^{(0)}, -1, -\tau^{(0)})/(\theta_{22}^{(0)} - \theta_{12}^{(0)})$. Thus $\mathscr{L}\{n^{\frac{1}{2}}(\hat{\tau} - \tau^{(0)})\} \to \mathscr{N}(\mathbf{0}, \mathbf{AG}^{-1}\mathbf{A}')$. This concludes the example.

Hinkley [11] reports for this special case, based on an empirical study, that the asymptotic normality of $\hat{\tau}$ is not a good approximation for small sample sizes. The estimated change-over point is given by

$$\hat{\tau} = \frac{\hat{\theta}_{11} - \hat{\theta}_{21}}{\hat{\theta}_{22} - \hat{\theta}_{12}} = \frac{(\theta_{11}^{(0)} - \theta_{21}^{(0)}) + [(\hat{\theta}_{11} - \hat{\theta}_{21}) - (\theta_{11}^{(0)} - \theta_{21}^{(0)})]}{(\theta_{22}^{(0)} - \theta_{21}^{(0)}) + [(\hat{\theta}_{22} - \hat{\theta}_{12}) - (\theta_{22}^{(0)} - \theta_{12}^{(0)})]}\,.$$

The denominator can be expanded in a geometric series to as high an order term as desired. Thus

$$\hat{\tau} - \tau^{(0)} = \tau^{(0)}\left[\frac{(\hat{\theta}_{11} - \hat{\theta}_{21})}{(\theta_{11}^{(0)} - \theta_{21}^{(0)})} - \frac{(\hat{\theta}_{22} - \hat{\theta}_{12})}{(\theta_{22}^{(0)} - \theta_{12}^{(0)})}\right]$$
$$- \tau^{(0)}\left(\frac{\hat{\theta}_{22} - \hat{\theta}_{12}}{\theta_{22}^{(0)} - \theta_{12}^{(0)}} - 1\right)\left[\frac{(\hat{\theta}_{11} - \hat{\theta}_{21})}{(\theta_{11}^{(0)} - \theta_{21}^{(0)})} - \frac{(\hat{\theta}_{22} - \hat{\theta}_{12})}{(\theta_{22}^{(0)} - \theta_{12}^{(0)})}\right] + O_p(n^{-\frac{3}{2}}) .$$

The first term on the right-hand side of the above expression, when suitably normalized, converges to the normal distribution discussed above. The second (correction) term is $O_p(n^{-1})$. Additional terms could be included if desired. Hinkley [12] discusses the distribution of the ratio of two correlated normal random variables, of which the distribution of $\hat{\tau}$ is an example.

2. In this example $\boldsymbol{\theta}^{(0)}$ lies at a boundary of Θ. Consider the two segment regression model

$$\begin{aligned} \mu(\boldsymbol{\xi}; t) &= \theta_{11} & 0 \leqq t \leqq \tau \\ &= \theta_{21}(t - \tfrac{1}{2})^2 & \tau \leqq t \leqq 1 . \end{aligned}$$

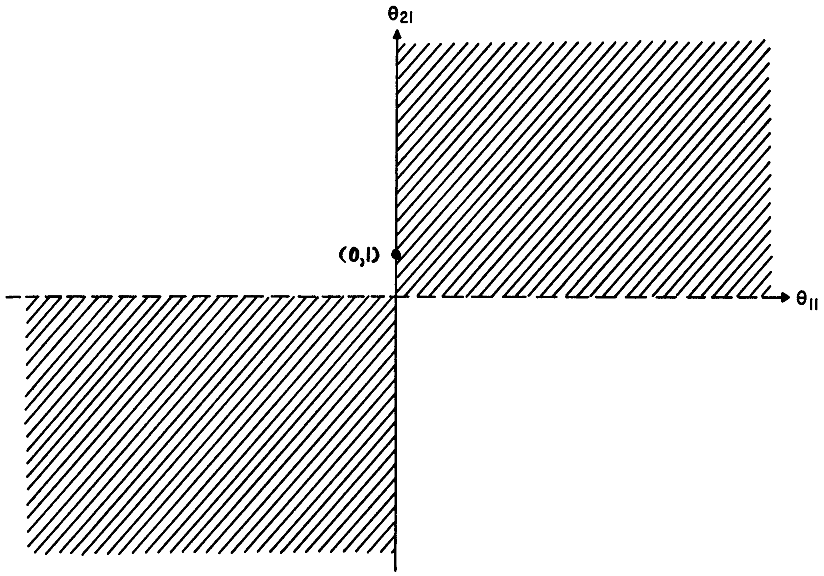

FIG. 4.1.

Suppose $\theta_{11}^{(0)} = 0$, $\theta_{21}^{(0)} = 1$. Then $r = 2$, $f_1(\theta_1; t) = \theta_1$, $f_2(\theta_1; t) = \theta_2(t - \frac{1}{2})^2$, $\tau_1^{(0)} = \frac{1}{2}$, $K(1) = K(2) = 1$, $m_1 = 2$. Assume that $\sigma_0^2 = 1$. The parameter space Θ consists of the first and third quadrants of two-dimensional Euclidean space, including the θ_{21} axis but excluding the θ_{11} axis. The state of nature, (0, 1), is at a boundary of Θ.

Suppose, for definiteness that observations are equally spaced along the interval [0, 1]. It is not hard to show that the conditions of Lemma 4.12 are satisfied and so $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^* = o_p(n^{-\frac{1}{2}})$, etc. It thus suffices to discuss the asymptotic distribution of $\hat{\boldsymbol{\theta}}^*$.

Calculate the unrestrained estimators $\bar{\theta}_{11}^*$, $\bar{\theta}_{21}^*$ based on the observations within each segment separately. Θ is locally convex at (0, 1) and so the conditions of Lemma 4.8 are satisfied, taking for C the half space $\{\boldsymbol{\theta} : \theta_{11} \geqq 0\}$. In this example $\mathbf{G}$ is the 2×2 matrix with diagonal elements

$\int_0^{\frac{1}{2}} dt = \frac{1}{2}$ and $\int_{\frac{1}{2}}^1 (t - \frac{1}{2})^4 \, dt = \frac{1}{160}$ and off diagonal elements equal to 0.

Lemma 4.8 asserts that $n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^{(0)})$ has the same asymptotic distribution as the point in C closest to $\mathbf{Z}$, where $\mathscr{L}(\mathbf{Z}) = \mathscr{N}(\mathbf{0}, \mathbf{G}^{-1})$ and closeness is measured in the metric determined by $\mathbf{G}$. Thus

$$\mathscr{L}\{n^{\frac{1}{2}}(\hat{\theta}_{11}^*, \hat{\theta}_{21}^* - 1)\} \to \mathscr{L}(Z_1^+, Z_2)$$

where $Z_1^+ \equiv \max(0, Z_1)$ and $\mathscr{L}\{Z_1, Z_2)\} = \mathscr{N}(\mathbf{0}, \text{diag}(2, 160))$.

The change-over point, $\hat{\tau}^*$, is estimated in the obvious manner. Thus $(\hat{\tau}^* - \frac{1}{2})^2 = \hat{\theta}_{11}^*$ and so

$$\mathscr{L}\{n^{\frac{1}{2}}(\hat{\tau} - \tfrac{1}{2})^2\} \to \mathscr{L}\{Z_1^+\} \qquad \text{where} \quad \mathscr{L}\{Z_1\} = \mathscr{N}(0, 2)\,.$$

3. In this example $\boldsymbol{\theta}^{(0)}$ is sometimes a boundary point and sometimes an interior point of Θ. However, even when $\boldsymbol{\theta}^{(0)}$ is interior, $\hat{\tau}$ is not normally distributed. Suppose $\mu(\boldsymbol{\xi}; t)$ is represented by the two segment model $\mu(\boldsymbol{\xi}; t) = |t - \frac{1}{2}|$ if $0 \leqq t \leqq \tau$ and $\mu(\boldsymbol{\xi}; t) = \theta_1 + \theta_2 t$ if $\tau \leqq t \leqq 1$.

The parameter space Θ is pictured below.

Suppose $\boldsymbol{\theta}^{(0)}$ is such that $\theta_1^{(0)} + \theta_2^{(0)}/2 = 0$. Then $\boldsymbol{\theta}^{(0)}$ lies along the dashed line and is thus an interior point of Θ if $|\theta_2^{(0)}| > 1$ and is a boundary point if

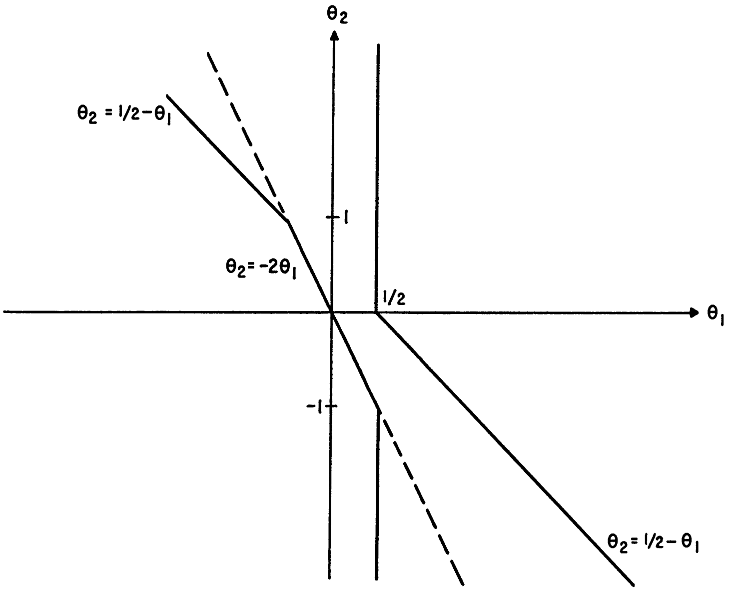

FIG. 4.2.

$|\theta_2^{(0)}| < 1$. In either case, $\tau^{(0)} = \frac{1}{2}$. The case $|\theta_2^{(0)}| = 1$ will be discussed separately.

Suppose that the observations are equally spaced along the interval [0, 1]. If $|\theta_2^{(0)}| \neq 1$, $\boldsymbol{\theta}$ is well-identified, if $\theta_2^{(0)} = -1$, $\boldsymbol{\theta}$ is identified but not well-identified, and if $\theta_2^{(0)} = 1$, $\boldsymbol{\theta}$ is unidentified. When $|\theta_2^{(0)}| \neq 1$ the conditions of Theorem 4.13 are satisfied. We consider this situation first.

Case (i). $|\theta_2^{(0)}| < 1$. The parameter space $\Theta - \boldsymbol{\theta}^{(0)}$ is approximated by the half space $2\theta_1 + \theta_2 \geqq 0$ as defined in Definition 4.7. Theorem 4.13 (b) implies $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})] \rightarrow \mathscr{L}[\boldsymbol{\pi}(Z)]$ where $\boldsymbol{\pi}(\mathbf{Z})$ minimizes $(\mathbf{Z} - \boldsymbol{\theta})'\mathrm{G}(\mathbf{Z} - \boldsymbol{\theta})$ among all $\boldsymbol{\theta}$ such that $2\theta_1 + \theta_2 \geqq 0$ and $\mathscr{L}(\mathbf{Z}) = \mathscr{N}(\mathbf{0}, \mathrm{G}^{-1})$. It is easily seen that

$$\begin{aligned} \boldsymbol{\pi}(\mathbf{Z}) &= \mathbf{Z} && \text{if } \; 2Z_1 + Z_2 \geqq 0 \\ &= \frac{\mathbf{Z}'\mathbf{G}(1, -2)'}{(1, -2)\mathbf{G}(1, -2)'}\begin{pmatrix} 1 \\ -2 \end{pmatrix} && \text{if } \; 2Z_1 + Z_2 < 0 \,. \end{aligned}$$

Thus $2\pi_1(\mathbf{Z}) + \pi_2(\mathbf{Z}) = \max(0, 2Z_1 + Z_2)$.

The estimated change-over point, $\hat{\tau}$, must satisfy the equation

$$|\hat{\tau} - \tfrac{1}{2}| = \hat{\theta}_1 + \hat{\tau}\hat{\theta}_2 = (\hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2) + (\hat{\tau} - \tfrac{1}{2})\hat{\theta}_2 \,.$$

Therefore

$$|\hat{\tau} - \tfrac{1}{2}| - (\hat{\tau} - \tfrac{1}{2})\theta_2^{(0)} = [(\hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2) - (\theta_1^{(0)} + \tfrac{1}{2}\theta_2^{(0)})] + O_p(n^{-1}) \,.$$

This equation has two solutions $\hat{\tau}_+$, $\hat{\tau}_-$, where $\hat{\tau}_+ \geqq \frac{1}{2}$, $\hat{\tau}_- \leqq \frac{1}{2}$.

$$n^{\frac{1}{2}}(\hat{\tau}_+ - \tfrac{1}{2}) = \frac{n^{\frac{1}{2}}[(\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2) - (\theta_1^{(0)} + \frac{1}{2}\theta_2^{(0)})]}{1 - \theta_2^{(0)}} + O_p(n^{-\frac{1}{2}})$$

$$n^{\frac{1}{2}}(\hat{\tau}_- - \tfrac{1}{2}) = -\frac{n^{\frac{1}{2}}[(\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2) - (\theta_1^{(0)} + \frac{1}{2}\theta_2^{(0)})]}{1 + \theta_2^{(0)}} + O_p(n^{-\frac{1}{2}}) \,.$$

Thus

$$\mathscr{L}[n^{\frac{1}{2}}(\hat{\tau}_+ - \tfrac{1}{2})] \rightarrow \mathscr{L}[(\pi_1(\mathbf{Z}) + \tfrac{1}{2}\pi_2(\mathbf{Z}))/(1 - \theta_2^{(0)})] \,,$$
$$\mathscr{L}[n^{\frac{1}{2}}(\hat{\tau}_- - \tfrac{1}{2})] \rightarrow \mathscr{L}[-(\pi_1(\mathbf{Z}) + \tfrac{1}{2}\pi_2(\mathbf{Z}))/(1 + \theta_2^{(0)})] \,.$$

Case (ii). $|\theta_2^{(0)}| > 1$. The underlying parameter, $\boldsymbol{\theta}^{(0)}$ is an interior point of Θ. Thus Theorem 4.13 (a) implies $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})] \rightarrow \mathscr{L}(\mathbf{Z}) = \mathscr{N}(\mathbf{0}, \mathrm{G}^{-1})$. If $\theta_2^{(0)} > 1$

$$\begin{aligned} \hat{\tau} - \tfrac{1}{2} &= -[(\hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2) - (\theta_1^{(0)} + \tfrac{1}{2}\theta_2^{(0)})]/(1 + \theta_2^{(0)}) + O_p(n^{-1}) \\ &\qquad\qquad \text{if } \; \hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2 \geqq 0 \\ &= [(\hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2) - (\theta_1^{(0)} + \tfrac{1}{2}\theta_2^{(0)})]/(1 - \theta_2^{(0)}) + O_p(n^{-1}) \\ &\qquad\qquad \text{if } \; \hat{\theta}_1 + \tfrac{1}{2}\hat{\theta}_2 \leqq 0 \,. \end{aligned}$$

This implies $\mathscr{L}[n^{\frac{1}{2}}(\hat{\tau} - \frac{1}{2})] \rightarrow \mathscr{L}(\tilde{\tau})$ where

$$\begin{aligned} \tilde{\tau} &= -(Z_1 + \tfrac{1}{2}Z_2)/(1 + a) && \text{if } \; Z_1 + \tfrac{1}{2}Z_2 \geqq 0 \\ &= (Z_1 + \tfrac{1}{2}Z_2)/(1 - a) && \text{if } \; Z_1 + \tfrac{1}{2}Z_2 \leqq 0 \end{aligned}$$

where $a = |\theta_2^{(0)}|$. Similarly, if $\theta_2^{(0)} < -1$, $\mathscr{L}[n^{\frac{1}{2}}(\hat{\tau} - \frac{1}{2})] \rightarrow \mathscr{L}(-\tilde{\tau})$. In both instances $\hat{\boldsymbol{\theta}}$, suitably normalized, is asymptotically normal; however $\hat{\tau}$ is not.

Case (iii). $\theta_2^{(0)} = 1$. The parameter $\boldsymbol{\theta}$ is not identified and so there is no reason that $\hat{\boldsymbol{\theta}}$ should even be consistent. However Corollary 3.22 is applicable.

Case (iv). $\theta_2^{(0)} = -1$. The regression is identified but not well-identified. The parameter space $\Theta - \boldsymbol{\theta}^{(0)}$ is approximated by the wedge-shaped positively homogeneous region, C, of apex angle 206.6 degrees between the half lines $\{\theta_1 - \frac{1}{2} = 0, \theta_2 + 1 \leqq 0\}$ and $\{\theta_2 + 2\theta_1 = 0, \theta_1 - \frac{1}{2} \leqq 0\}$. Note that C is not convex here. However Θ and C satisfy the relaxed condition mentioned in the notes following Lemmas 4.8 and 4.12. The region C can be divided into two convex regions by a half line beginning at the point $(\frac{1}{2}, -1)$. The appropriate angle of the line depends on $\mathbf{G}$. The extension of the line into the complement of C should contain all points equidistant (in the metric determined by $\mathbf{G}$) to both arms of the wedge. Thus, from Lemmas 4.8 and 4.12 $\mathscr{L}[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(0)})] \to \mathscr{L}[\boldsymbol{\pi}(\mathbf{Z})]$ where $(\mathbf{Z} - \boldsymbol{\pi}(\mathbf{Z}))'\mathbf{G}(\mathbf{Z} - \boldsymbol{\pi}(\mathbf{Z})) = \min_{\boldsymbol{\theta} \in C} (\mathbf{Z} - \boldsymbol{\theta})'\mathbf{G}(\mathbf{Z} - \theta)$. The estimate of the change-over point, τ, is not even consistent.

5. Some unresolved problems. This paper provides a derivation of the asymptotic distribution theory in the identified case, but leaves several important questions unanswered.

(i) Hinkley [11] has done numerical work which shows lack of agreement between the empirical and asymptotic distributions for moderate sample sizes, in a special case. Work should be done to assess the sample sizes necessary for the asymptotics to be valid and to obtain moderate sample size approximations to the distributions of the least squares estimators.

(ii) The distribution of the likelihood ratio statistic, λ, is of interest. It may be shown by arguments corollary to those of this paper that in the identified case $-2 \log \lambda$ has the usual chi-square asymptotic distribution. However these arguments break down in the unidentified case. This is precisely the situation where one wants to test whether or not there is a change in the regression. Quandt [20] reports on empirical grounds that the distribution of $-2 \log \lambda$ does not appear to be chi-square. This problem is considered in Feder [9], where it is shown by example that the asymptotic distribution varies with the spacing of the t-values.

(iii) The design question is of interest. That is, how should one select the independent variables, either sequentially or nonsequentially, to obtain the most precise parameter estimates and the most powerful tests?

6. Acknowledgments. I wish to thank many people who generously gave me some of their time to discuss ideas pertaining to this manuscript. In particular, Professors Herman Chernoff and the late L. J. Savage provided some valuable insights that kept me from sinking into the quicksand of definitions, theorems, and corollaries.

I thank Professor Herman Chernoff at Stanford University, Professor Francis J. Anscombe at Yale University, Professor John Tukey at Princeton University, and Dr. Richard L. Shuey at General Electric Corporate Research and Development for their support of this work, and the referee for his very helpful comments and suggestions.

I would also like to note that Sylwester's thesis, [24], was the first paper that came to grips with some of the difficulties inherent in regression problems having "kinky" sums of squares functions. His approach provided motivation for me and influenced the format of the present manuscript.

REFERENCES

[1] Bellman, R. and Roth, R. (1969). Curve fitting by segmented straight lines. *J. Amer. Statist. Assoc.* **64** 1079–1084.

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

[3] Cramér, H. (1966). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[4] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[5] Dunicz, B. L. (1969). Discontinuities in the surface structure of alcohol-water mixtures. *Sond. aus der Kolloid-Zeitschrift und Zeitschrift für Polymere.* **230** 346–357.

[6] 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.

[7] Farley, J. U., Hinich, M. and McGuire, T. W. (1971). Testing for a shift in the slopes of a multivariate linear time series model. Technical Report WP 77-70-1, Graduate School of Industrial Administration, Carnegie-Mellon Univ.

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

[9] Feder, P. I. (1975). The log likelihood ratio in segmented regression. *Ann. Statist.* **3** 84–97.

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

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

[12] Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika* **56** 635–639.

[13] Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17.

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

[15] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

[16] Loève, M. (1960). *Probability Theory*, (2nd ed.). Van Nostrand, New York.

[17] Poirer, D. J. (1973). Piecewise regression using cubic splines. *J. Amer. Statist. Assoc.* **68** 515–524.

[18] Pratt, J. W. (1959). On a general concept of "In probability." *Ann. Math. Statist.* **30** 549–558.

[19] Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc.* **53** 873–880.

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

[21] Robison, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. *J. Amer. Statist. Assoc.* **59** 214–224.

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

[23] SPRENT, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics* **17** 634–645.

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

[25] SYLWESTER, D. L. (1972). Asymptotic distribution theory for two-phase linear regression models. Personal communication.

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