Indian Statistical Institute

Least-Squares Estimation of a Step Function
Author(s): Yi-Ching Yao and S. T. Au
Source: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, Vol. 51, No. 3 (Oct., 1989), pp. 370-381
Published by: Springer on behalf of the Indian Statistical Institute
Stable URL: http://www.jstor.org/stable/25050759
Accessed: 29-07-2016 15:26 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://about.jstor.org/terms

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Springer, Indian Statistical Institute are collaborating with JSTOR to digitize, preserve and extend access to *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

Sankhyā : The Indian Journal of Statistics
1989, Volume **51**, Series A, Pt. 3, pp. 370-381.

LEAST-SQUARES ESTIMATION OF A STEP FUNCTION

By YI-CHING YAO
Colorado State University
and
S. T. AU
AT and T Bell Laboratories

SUMMARY. Consider the problem of estimating a step function in the presence of additive measurement noise. In the case that the number of jumps is known. the least-squares estimators for the locations of the jumps and the levels of the step function are studied and their limiting distributions are derived. When the number of jumps is unknown. an estimator is proposed which is consistent under the condition that the number of jumps is not greater than a given upper bound.

1. Introduction

We consider the problem of estimating a step function in the presence of additive measurement noise. Specifically, let $g : [0, 1] \to (-\infty, \infty)$ be a step function with R jumps ; that is, $g(0) = \mu_1$ and $g(t) = \mu_r$, $\tau_{r-1} < t \leqslant \tau_r (1 \leqslant r \leqslant R+1)$ with $\mu_r \neq \mu_{r+1}$ and $0 \equiv \tau_0 < \tau_1 < \ldots < \tau_R < \tau_{R+1} \equiv 1$. We observe

$$X_n(i) = g(i/n) + W_n(i), \; i = 1, \ldots, n$$

where $\{W_n(i) : 1 \leqslant i \leqslant n\}$ is an independent and identically distributed (i.i.d.) sequence of measurement errors with common distribution function F satisfying $\int_{-\infty}^{\infty} x dF(x) = 0$. It is assumed that the τ_r, μ_r and F are unknown but the number of jumps R is either known or unknown. It is desired to estimate the τ_r and μ_r as well as R (when it is unknown) based on $\{X_n(i) : 1 \leqslant i \leqslant n\}$.

This problem arises in seismology where $\tau_r - \tau_{r-1}$ denotes the thickness of the r-th sedimentary layer and $g(t)$ represents a physical quantity (e.g., density of earth's crust) at point t (t being the depth of the point). It is of importance to know the number of layers and the thickness of each layer.

This problem is related to the so-called change-point problem. See Shaban (1980) for a comprehensive list of references for the change-point problem. In particular, Hinkley (1970) investigated the behavior of the

AMS (1980) *subject classifications :* Primary 62G05 ; Secondary 62E20.

Key words and phrases : Change-point, limiting distribution, random walk.

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

maximum likelihood estimator for the single change-point case (i.e. $R = 1$), and Chernoff and Zacks (1964) and Yao (1984) considered the multiple change-point case from a Bayesian point of view. Yao (1988) considered the case of normal F and proposed a consistent estimator of R via Schwarz' criterion. He noted that this estimator tends to overestimate R when F is non-normal. In this paper, we assume that R is fixed (either known or unknown) and use the least-squares method to estimate the τ_r and μ_r. In the next section, the case of R known is considered and the limiting distributions (as $n \to \infty$) of the least-squares estimators of the τ_r and μ_r are derived which are related to the locations of the minima for certain random walks. In Section 3, the case of R unknown is treated and an estimator of R is proposed which is consistent under the condition that R is not greater than a given upper bound.

2. Least-squares estimators with R known

Let $\hat{\tau}_1, \ldots, \hat{\tau}_R$ be the values of $l_1/n, \ldots, l_R/n$ (with $1 \leqslant l_1 < \ldots < l_R \leqslant n-1$) which minimize the sum of squares

$$S_n(l_1, \ldots, l_R) = \sum_{r=1}^{R+1} \sum_{i=l_{r-1}+1}^{l_r} \{X_n(i) - \bar{X}_n(l_{r-1}, l_r)\}^2$$

where $l_0 = 0$, $l_{R+1} = n$ and $\bar{X}_n(i, j)$ denotes the average of $X_n(i+1), \ldots, X_n(j)$. It is convenient to define $\hat{\tau}_0 = 0$ and $\hat{\tau}_{R+1} = 1$. The $\hat{\tau}_r$ and $\hat{\mu}_r = \bar{X}_n(n\hat{\tau}_{r-1}, n\hat{\tau}_r)$ are called the least-squares estimators of τ_r and μ_r. To derive the limiting distributions of $\hat{\tau}_r$ and $\hat{\mu}_r$, we need the following assumptions :

(A1) Let τ_r^0 and μ_r^0 be the true values of τ_r and μ_r such that $\mu_r^0 \neq \mu_{r+1}^0$, $0 < \tau_1^0 < \ldots < \tau_R^0 < 1$, and we take the convention that $\tau_0^0 = 0$ and $\tau_{R+1}^0 = 1$;

(A2) The distribution F is continuous so that the $\hat{\tau}_r$ are uniquely defined with probability 1 ;

(A3) $\int_{-\infty}^{\infty} x^6 \, dF(x) < \infty$.

Lemma 1 : *Suppose that $V_1, \ldots, V_n$ are i.i.d. with $EV_1 = 0$ and $EV_1^{2m} < \infty$ for some positive integer m. Then as $n \to \infty$,*

$$\max_{0 \leqslant i < j \leqslant n} (V_{i+1} + \ldots + V_j)^2/(j-i) = O_p(n^{2/m}).$$

Proof : Clearly, for k large,

$$E(V_1 + \ldots + V_k)^{2m} = 2^{-m}(2m)! \binom{k}{m} (EV_1^2)^m + O(k^{m-1}).$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

That is, there exists $C > 0$ such that $E(V_1+\ldots+V_k)^{2m} \leqslant C\,k^m$ for all $k \geqslant 1$. Now for any $\xi > 0$,

$$\Pr\{\max_{0\leqslant i<j\leqslant n}(V_{i+1}+\cdots+V_j)^2/(j-i) > \xi n^{2/m}\}$$

$$\leqslant \sum_{0\leqslant i<j\leqslant n}\Pr\{(V_{i+1}+\cdots+V_j)^2/(j-i) > \xi n^{2/m}\}$$

$$\leqslant \sum_{0\leqslant i<j\leqslant n}(\xi n^{2/m})^{-m}E(V_{i+1}+\cdots+V_j)^{2m}/(j-i)^m$$

$$\leqslant \sum_{0\leqslant i<j\leqslant n}\xi^{-m}n^{-2}C \to 2^{-1}\xi^{-m}C \text{ as } n\to\infty.$$

Since $2^{-1}\xi^{-m}C$ can be made arbitrarily small by choosing a large ξ, the lemma follows. □

Lemma 2 : *Assume* $(A1)-(A3)$ *hold. Then* $\hat{\tau}_r-\tau_r^0 = O_p\,(\mathbf{1})$, $1 \leqslant r \leqslant R$; *that is,* $\hat{\tau}_r$ *is a consistent estimator of* τ_r^0.

Proof: Let ϵ be any positive number such that $\tau_s^0+2\epsilon < \tau_{s+1}^0-2\epsilon$, $s = 0, \ldots, R$, and let $A_r(n, \epsilon) = \{(l_1/n, \ldots, l_R/n) : 0 < l_1 < \ldots < l_R < n$ and $|l_s/n-\tau_r^0| > \epsilon$, $1 \leqslant s \leqslant R\}$, $1 \leqslant r \leqslant R$. We claim that, for $r = 1, \ldots, R$, $(\hat{\tau}_1, \ldots, \hat{\tau}_R)\,\epsilon\,A_r\,(n, \epsilon)$ with probability approaching 0 as $n\to\infty$. Together with the fact that $\tau_s^0+2\epsilon < \tau_{s+1}^0-2\epsilon(s = 0, \ldots, R)$, this claim implies that, with probability approaching 1, exactly one of $\hat{\tau}_1, \ldots, \hat{\tau}_R$ is between $\tau_r^0-\epsilon$ and $\tau_r^0+\epsilon$, $1 \leqslant r \leqslant R$. (Of course, this one must be $\hat{\tau}_r$). Since ϵ can be made arbitrarily small, $\hat{\tau}_r-\tau_r^0 = o_p(1)$.

We still have to prove that $\Pr(\{\hat{\tau}_1, \ldots, \hat{\tau}_R)\,\epsilon\,A_r(n, \epsilon)\} \to 0$ as $n\to\infty$. It suffices to show that, with probability approaching 1,

$$\min_{(l_1/n,\ \ldots,\ l_R/n)\epsilon A_r(n,\ \epsilon)} S_n(l_1, \ldots, l_R) > S_n([n\tau_1^0], \ldots, [n\tau_R^0]) \qquad \ldots \quad (1)$$

where $[x]$ denotes the largest integer not beyond x. To show this, it is convenient to generalize the definition of S_n to every subset $\{k_1, \ldots, k_m\}$ of $\{1, \ldots, n-1\}$:

$$S_n(k_1, \ldots, k_m) = \sum_{\alpha=1}^{m+1}\ \sum_{i=k_{(\alpha-1)}+1}^{k_{(\alpha)}}\{X_n(i)-\overline{X}_n(k_{(\alpha-1)}, k_{(\alpha)})\}^2$$

where $k_{(0)} = 0$, $k_{(m+1)} = n$ and $k_{(1)} < k_{(2)} < \ldots < k_{(m)}$ are the ordered version of $k_1, \ldots, k_m$. Now for any *fixed* $(l_1/n, \ldots, l_R/n)\,\epsilon\,A_r(n, \epsilon)$, it is clear that

$$S_n(l_1, \ldots, l_R) \geqslant S_n(l_1, \ldots, l_R, [n\tau_1^0], \ldots, [n\tau_{r-1}^0],$$
$$[n(\tau_r^0-\epsilon)], [n(\tau_r^0+\epsilon)], [n\tau_{r+1}^0], \ldots, [n\tau_R^0]) \qquad \ldots \quad (2)$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

The right hand side of (2) can be written as $T_1+\ldots+T_{R+2}$ where $T_s(s=1, \ldots, r-1, r+2, \ldots, R+1)$ is the sum of squares involving $X_i([n\tau^0_{s-1}]<i\leqslant[n\tau^0_s])$; T_r is that involving $X_i\,([n\tau^0_{r-1}]<i\leqslant[n(\tau^0_r-\epsilon)])$; T_{r+1} is that involving $X_i([n(\tau^0_r+\epsilon)]<i\leqslant[n\tau^0_{r+1}])$; and T_{R+2} is that involving $X_i\,([n(\tau^0_r-\epsilon)]<i\leqslant[n(\tau^0_r+\epsilon)])$. More precisely, let $\xi(1,s)<\ldots<\xi(J(s),s)$ denote the elements of the set $\{l_1,\ldots,l_R\}\bigcap\{x\,;[n\tau^0_{s-1}]<x<[n\tau^0_s]\}$. Thus, for $s=1,\ldots,r-1, r+2,\ldots,R+1$,

$$T_s=\sum_{j=1}^{J(s)+1}\sum_{i=\xi(j-1,s)+1}^{\xi(j,s)}\{X_n(i)-\overline{X}_n(\xi(j-1,s),\xi(j,s))\}^2$$

$$=\sum_{i=[n\tau^0_{s-1}]+1}^{[n\tau^0_s]}W^2_n(i)-\sum_{j=1}^{J(s)+1}\{\xi(j,s)-\xi(j-1,s)\}\,\overline{W}^2_n(\xi(j-1,s),\,\xi(j,s))$$

$$\geqslant\sum_{i=[n\tau^0_{s-1}]+1}^{[n\tau^0_s]}W^2_n(i)-(R+1)\max_{[n\tau^0_{s-1}]\leqslant i<j\leqslant[n\tau^0_s]}(j-i)\,\overline{W}^2_n(i,j)\quad\ldots\quad(3)$$

where $\xi(0,s)=[n\tau^0_{s-1}]$, $\xi(J(s)+1,s)=[n\tau^0_s]$ and $\overline{W}_n(i,j)$ denotes the average of $W_n(i+1),\ldots,W_n(j)$. Similarly,

$$T_r\geqslant\sum_{i=[n\tau^0_{r-1}]+1}^{[n(\tau^0_r-\epsilon)]}W^2_n(i)-(R+1)\max_{[n\tau^0_{r-1}]\leqslant i<j\leqslant[n(\tau^0_r-\epsilon)]}(j-i)\overline{W}^2_n(i,j)\quad\ldots\quad(4)$$

$$T_{r+1}\geqslant\sum_{i=[n(\tau^0_r+\epsilon)]+1}^{[n\tau^0_{r+1}]}W^2_n(i)-(R+1)\max_{[n(\tau^0_r+\epsilon)]\leqslant i<j\leqslant[n\tau^0_{r+1}]}(j-i)\overline{W}^2_n(i,j)\quad\ldots\quad(5)$$

Also,

$$T_{R+2}=\sum_{i=[n(\tau^0_r-\epsilon)]+1}^{[n(\tau^0_r+\epsilon)]}\{X_n(i)-\overline{X}_n([n(\tau^0_r-\epsilon)],[n(\tau^0_r+\epsilon)])\}^2$$

$$=\sum_{i=[n(\tau^0_r-\epsilon)]+1}^{[n(\tau^0_r+\epsilon)]}W^2_n(i)+T'\quad\ldots\quad(6)$$

where

$$T'=\{[n\tau^0_r]-[n(\tau^0_r-\epsilon)]\}\{\overline{X}_n([n(\tau^0_r-\epsilon)],[n\tau^0_r])-\overline{X}_n([n(\tau^0_r-\epsilon)],[n(\tau^0_r+\epsilon)])\}^2$$
$$+\{[n(\tau^0_r+\epsilon)]-[n\tau^0_r]\}\{\overline{X}_n([n\tau^0_r],[n(\tau^0_r+\epsilon)])-\overline{X}_n([n(\tau^0_r-\epsilon)],[n(\tau^0_r+\epsilon)])\}^2$$
$$-\{[n\tau^0_r]-[n(\tau^0_r-\epsilon)]\}\overline{W}^2_n([n(\tau^0_r-\epsilon)],[n\tau^0_r])-\{[n(\tau^0_r+\epsilon)]$$
$$-[n\tau^0_r]\}\overline{W}^2_n([n\tau^0_r],[n(\tau^0_r+\epsilon)])=2^{-1}n\epsilon(\mu^0_{r+1}-\mu^0_r)^2+o_p(n)\quad\ldots\quad(7)$$

Note that the right hand sides of (3)—(6) do not depend on $(l_1/n,\ldots,l_R/n)$ which is assumed to belong to $A_r(n,\epsilon)$. By Lemma 1, the three maxima on

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

the right hand sides of (3)—(5) are $o_p(n)$ since $\int_{-\infty}^{\infty} x^6 dF(x) < \infty$. So from (2)—(7), we have

$$\min_{(l_1/n,\ldots\ldots,l_R/n)\epsilon A_r(n,\epsilon)} S_n(l_1, \ldots, l_R) \geqslant \sum_{i=1}^{n} W_n^2(i) + 2^{-1}n\epsilon(\mu_{r+1}^0 - \mu_r^0)^2 + o_p(n).$$

Since

$$S_n([n\tau_1^0], \ldots, [n\tau_R^0]) = \sum_{i=1}^{n} W_n^2(i) - \sum_{s=1}^{R+1} ([n\tau_s^0] - [n\tau_{s-1}^0]) \overline{W}_n^2([n\tau_{s-1}^0], [n\tau_s^0])$$

$$= \sum_{i=1}^{n} W_n^2(i) + O_p(1),$$

the inequality (1) holds with probability approaching 1, completing the proof. □

Lemma 3. *Assume that* $(A1)$—$(A3)$ *hold. Then* $n(\hat{\tau}_r - \tau_r^0) = O_p(1)$, $1 \leqslant r \leqslant R$.

Proof: We will only consider $2 \leqslant r \leqslant R-1$. The cases $r = 1$ and $r = R$ can be treated with obvious modification. For any fixed $\epsilon > 0$, it suffices to show that there exists $M > 0$ such that $\Pr(n|\hat{\tau}_r - \tau_r^0| > M) < \epsilon$ for all sufficiently large n. Let $\delta_1 = |\mu_{r+1}^0 - \mu_r^0|/20 > 0$. By the strong law of large numbers, there exist $C > 0$ and integer $M > 0$ such that the event $E_r(n, C, M)$:

$$|\overline{X}_n([n\tau_s^0], [n\tau_s^0]+i)| < C, i = 1, \ldots, [n\tau_{s+1}^0] - [n\tau_s^0]\ ; s = 0, \ldots, R\ ; \quad \ldots \quad (8)$$

$$|\overline{X}_n([n\tau_{s+1}^0]-i, [n\tau_{s+1}^0])| < C, i = 1, \ldots, [n\tau_{s+1}^0] - [n\tau_s^0]\ ; s = 0, \ldots, R\ ; \ldots \quad (9)$$

$$|\overline{X}_n([n\tau_r^0]-i, [n\tau_r^0]) - \mu_r^0| < \delta_1, i = M, \ldots, [n\tau_r^0] - [n\tau_{r-1}^0]\ ; \quad \ldots \quad (10)$$

$$|\overline{X}_n([n\tau_r^0], [n\tau_r^0]+i) - \mu_{r+1}^0| < \delta_1, i = M, \ldots, [n\tau_{r+1}^0] - [n\tau_r^0]\ ; \quad \ldots \quad (11)$$

occurs with probability $> 1-\epsilon/4$ for all sufficiently large n. Choose $\delta_2 > 0$ so that $\tau_s^0 + 2\delta_2 < \tau_{s+1}^0 - 2\delta_2 (0 \leqslant s \leqslant R)$, and

$$\{2\delta_2(C + |\mu_{r+1}^0|) + \delta_1(\tau_{r+1}^0 - \tau_r^0)\}/(\tau_{r+1}^0 - \tau_r^0 - \delta_2) < |\mu_{r+1}^0 - \mu_r^0|/8, \quad \ldots \quad (12)$$

$$\{\delta_2(C + |\mu_r^0|) + \delta_1(\tau_r^0 - \tau_{r-1}^0)\}/(\tau_r^0 - \tau_{r-1}^0 - \delta_2) < |\mu_{r+1}^0 - \mu_r^0|/16, \quad \ldots \quad (13)$$

$$\delta_2 C^2\{1 + 2(\tau_r^0 - \tau_{r-1}^0 + 2\delta_2)/(\tau_r^0 - \tau_{r-1}^0 - 2\delta_2)\}^2/(\tau_r^0 - \tau_{r-1}^0 - \delta_2) < (\mu_{r+1}^0 - \mu_r^0)^2/16. \ldots (14)$$

Let $B(n, \delta_2) = \{(l_1/n, \ldots, l_R/n) : \ 0 < l_1 < \ldots < l_R < n,\ |l_s/n - \tau_s^0| < \delta_2, 1 \leqslant s \leqslant R\}$ and $B_r(n, \delta_2, M) = \{(l_1/n, \ldots, l_R/n) \,\epsilon\, B(n, \delta_2) : \ l_r - n\tau_r^0 < -M\}$. By Lemma 1, $(\hat{\tau}_1, \ldots, \hat{\tau}_R) \,\epsilon\, B(n, \delta_2)$ with probability $> 1-\epsilon/4$ for large n. We claim that $(\hat{\tau}_1, \ldots, \hat{\tau}_R) \,\epsilon\, B_r(n, \delta_2, M)$ with probability $< \epsilon/4$ for large n. Assuming that this claim holds, we have, for large n,

$$\Pr\{n(\hat{\tau}_r - \tau_r^0) < -M\} \leqslant \Pr\{(\hat{\tau}_1, \ldots, \hat{\tau}_R) \notin B(n, \delta_2)\}$$

$$+ \Pr\{(\hat{\tau}_1, \ldots, \hat{\tau}_R) \,\epsilon\, B_r(n, \delta_2, M)\} < \epsilon/4 + \epsilon/4 = \epsilon/2$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

Similarly, it can be shown that $\Pr\{n(\hat{\tau}_r-\tau_r^0) > M\} < \epsilon/2$ for large n, and so $\Pr\{n|\hat{\tau}_r-\tau_r^0| > M\} < \epsilon$ for large n.

It remains to prove the claim. For every $(l_1/n, \ldots, l_R/n) \epsilon B_r(n, \delta_2, M)$, let $(l_1'/n, \ldots, l_R'/n) \epsilon B(n, \delta_2)$ be such that $l_s' = l_s$ for $s \neq r$ and $l_r' = [n\tau_r^0]$. Clearly, for $(l_1/n, \ldots, l_R/n) \epsilon B_r(n, \delta_2, M)$,

$$\begin{aligned}
S_n(l_1, \ldots, l_R)-S_n(l_1', \ldots, l_R') &\geqslant \sum_{i=l_{r-1}+1}^{l_r} \{X_n(i)-\bar{X}_n(l_{r-1}, l_r)\}^2 \\
&+ \sum_{i=l_r+1}^{[n\tau_r^0]} \{X_n(i)-\bar{X}_n(l_r, l_{r+1})\}^2 \\
&- \sum_{i=l_{r-1}+1}^{[n\tau_r^0]} \{X_n(i)-\bar{X}_n(l_{r-1}, [n\tau_r^0])\}^2 \\
&= \sum_{i=l_r+1}^{[n\tau_r^0]} (\{X_n(i)-\bar{X}_n(l_r, l_{r+1})\}^2-\{X_n(i)-\bar{X}_n(l_{r-1}, [n\tau_r^0])\}^2) \\
&- \sum_{i=l_{r-1}+1}^{l_r} (\{X_n(i)-\bar{X}_n(l_{r-1}, [n\tau_r^0])\}^2-\{X_n(i)-\bar{X}_n(l_{r-1}, l_r)\}^2) \\
&= ([n\tau_r^0]-l_r)(\{\bar{X}_n(l_r, l_{r+1})-\bar{X}_n(l_r, [n\tau_r^0])\}^2-\{\bar{X}_n(l_{r-1}, [n\tau_r^0])-\bar{X}_n(l_r, [n\tau_r^0])\}^2) \\
&- (l_r-l_{r-1})\{\bar{X}_n(l_{r-1}, [n\tau_r^0])-\bar{X}_n(l_{r-1}, l_r)\}^2 \qquad \ldots \quad (15)
\end{aligned}$$

In the event $E_r(n, C, M)$, for $(l_1/n, \ldots, l_R/n) \epsilon B_r(n, \delta_2, M)$,

$$|\bar{X}_n(l_r, [n\tau_r^0])-\mu_r^0| < \delta_1 < |\mu_{r+1}^0-\mu_r^0|/8, \text{ by (10)},$$

$$\begin{aligned}
|\bar{X}_n(l_r, l_{r+1})-\mu_{r+1}^0| &\leqslant \left| \frac{[n\tau_r^0]-l_r}{l_{r+1}-l_r} \{\bar{X}_n(l_r, [n\tau_r^0])-\mu_{r+1}^0\} \right| \\
&+ \left| \frac{[n\tau_{r+1}^0]-[n\tau_r^0]}{l_{r+1}-l_r} \{\bar{X}_n([n\tau_r^0], [n\tau_{r+1}^0])-\mu_{r+1}^0\} \right| \\
&+ \left| \frac{l_{r+1}-[n\tau_{r+1}^0]}{l_{r+1}-l_r} \{\bar{X}_n([n\tau_{r+1}^0], l_{r+1})-\mu_{r+1}^0\} \right| \\
&\leqslant \frac{[n\tau_r^0]-l_r}{l_{r+1}-l_r} (C+|\mu_{r+1}^0|)+\frac{[n\tau_{r+1}^0]-[n\tau_r^0]}{l_{r+1}-l_r} \delta_1 \\
&+ \left| \frac{l_{r+1}-[n\tau_{r+1}^0]}{l_{r+1}-l_r} \right| (C+|\mu_{r+1}^0|) \\
&< 2\{\delta_2(C+|\mu_{r+1}^0|)+(\tau_{r+1}^0-\tau_r^0)\delta_1+\delta_2(C+|\mu_{r+1}^0|)\}/ \\
&\qquad (\tau_{r+1}^0-\tau_r^0-\delta_2), \text{ for large } n, \\
&< |\mu_{r+1}^0-\mu_r^0|/4, \text{ by (12)},
\end{aligned}$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

(We have used the convention that $\bar{X}_n(i, j) = \bar{X}_n(j, i)$ for $j < i$ and $\bar{X}_n(i, i) = 0$ in the above derivation.)

$$
\begin{aligned}
& |\bar{X}_n(l_{r-1}, [n\tau_r^0]) - \mu_r^0| \\
& \leqslant \left| \frac{[n\tau_{r-1}^0] - l_{r-1}}{[n\tau_r^0] - l_{r-1}} \{\bar{X}_n(l_{r-1}, [n\tau_{r-1}^0]) - \mu_r^0\} \right| \\
& \quad + \left| \frac{[n\tau_r^0] - [n\tau_{r-1}^0]}{[n\tau_r^0] - l_{r-1}} \{\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0]) - \mu_r^0\} \right| \\
& \leqslant \{|[n\tau_{r-1}^0] - l_{r-1}|(C + |\mu_r^0|) + ([n\tau_r^0] - [n\tau_{r-1}^0])\delta_1\}/([n\tau_r^0] - l_{r-1}) \\
& \leqslant 2\{\delta_2(C + |\mu_r^0|) + (\tau_r^0 - \tau_{r-1}^0)\delta_1\}/(\tau_r^0 - \tau_{r-1}^0 - \delta_2), \text{ for large } n \\
& < |\mu_{r+1}^0 - \mu_r^0|/8, \text{ by (13).}
\end{aligned}
$$

So, for large n, in $E_r(n, C, M)$, for every $(l_1/n, \ldots, l_R/n) \in B_r(n, \delta_2, M)$

$$
\{\bar{X}_n(l_r, l_{r+1}) - \bar{X}_n(l_r, [n\tau_r^0])\}^2 - \{\bar{X}_n(l_{r-1}, [n\tau_r^0]) - \bar{X}_n(l_r, [n\tau_r^0])\}^2 > (3/16)(\mu_{r+1}^0 - \mu_r^0)^2 \quad \ldots \quad (16)
$$

Again, for large n, in $E_r(n, C, M)$, for every $(l_1/n, \ldots, l_R/n) \in B_r(n, \delta_2, M)$,

$$
|\bar{X}_n(l_r, [n\tau_r^0])| < C,
$$

$$
\begin{aligned}
|\bar{X}_n(l_{r-1}, l_r)| & \leqslant \left| \frac{[n\tau_{r-1}^0] - l_{r-1}}{l_r - l_{r-1}} \bar{X}_n(l_{r-1}, [n\tau_{r-1}^0]) \right| \\
& \quad + \left| \frac{l_r - [n\tau_{r-1}^0]}{l_r - l_{r-1}} \bar{X}_n([n\tau_{r-1}^0], l_r) \right| \\
& \leqslant \{2\delta_2 C + (\tau_r^0 - \tau_{r-1}^0 + 2\delta_2)C\}/(\tau_r^0 - \tau_{r-1}^0 - 2\delta_2) \\
& \leqslant 2C(\tau_r^0 - \tau_{r-1}^0 + 2\delta_2)/(\tau_r^0 - \tau_{r-1}^0 - 2\delta_2),
\end{aligned}
$$

and so

$$
\begin{aligned}
& (l_r - l_{r-1})\{\bar{X}_n(l_{r-1}, [n\tau_r^0]) - \bar{X}_n(l_{r-1}, l_r)\}^2 \\
& = (l_r - l_{r-1}) \left\{ \frac{l_r - [n\tau_r^0]}{[n\tau_r^0] - l_{r-1}} \bar{X}_n(l_{r-1}, l_r) + \frac{[n\tau_r^0] - l_r}{[n\tau_r^0] - l_{r-1}} \bar{X}_n(l_r, [n\tau_r^0]) \right\}^2 \\
& = ([n\tau_r^0] - l_r) \cdot \frac{l_r - l_{r-1}}{[n\tau_r^0] - l_{r-1}} \cdot \frac{[n\tau_r^0] - l_r}{[n\tau_r^0] - l_{r-1}} \{|\bar{X}_n(l_{r-1}, l_r)| + |\bar{X}_n(l_r, [n\tau_r^0])|\}^2 \\
& \leqslant ([n\tau_r^0] - l_r) \cdot 1 \cdot \frac{2\delta_2}{\tau_r^0 - \tau_{r-1}^0 - \delta_2} \{2C(\tau_r^0 - \tau_{r-1}^0 + 2\delta_2)/(\tau_r^0 - \tau_{r-1}^0 - 2\delta_2) + C\}^2 \\
& < ([n\tau_r^0] - l_r)(\mu_{r+1}^0 - \mu_r^0)^2/8, \text{ by (14).} \quad \ldots \quad (17)
\end{aligned}
$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

From (15)—(17), for large n, in $E_r(n, C, M)$, we have

$$\min_{(l_1/n, \ldots, l_R/n) \in B_r(n, \delta_2, M)} \{S_n(l_1, \ldots, l_R) - S_n(l'_1, \ldots, l'_R)\}$$

$$\geqslant (16)^{-1}(\mu^0_{r+1} - \mu^0_r)^2 \min_{(l_1/n, \ldots, l_R/n) \in B_r(n, \delta_2, M)} ([n\tau^0_r] - l_r)$$

$$\geqslant (16)^{-1}(\mu^0_{r+1} - \mu^0_r)^2 \, M > 0,$$

implying that $(\hat{\tau}_1, \ldots, \hat{\tau}_R) \notin B_r(n, \delta_2, M)$. So, for large n,

$$\Pr\{(\hat{\tau}_1, \ldots, \hat{\tau}_R) \in B_r(n, \delta_2, M)\} \leqslant 1 - \Pr\{E_r(n, C, M)\} < \epsilon/4.$$

This proves the claim and the lemma follows. □

Theorem 1: *Assume that* (A1)—(A3) *hold. As* $n \to \infty$, $\hat{\tau}_1, \ldots, \hat{\tau}_R$ *are asymptotically independent and* $n\hat{\tau}_r - [n\tau^0_r]$ *converges in distribution to* L_r, *the location of the minimum for the random walk* $\{\ldots, Z^{(r)}_{-1}, Z^{(r)}_0, Z^{(r)}_1, \ldots\}$ *where*

$Z^{(r)}_0 = 0,$

$$Z^{(r)}_j = \begin{cases} \sum_{i=1}^{j} \{\operatorname{sign}(\mu^0_{r+1} - \mu^0_r) \, U^{(r)}_i + |\mu^0_{r+1} - \mu^0_r|/2\}, & j = 1, 2, \ldots, \\ \sum_{i=j+1}^{0} \{\operatorname{sign}(\mu^0_r - \mu^0_{r+1}) \, U^{(r)}_i + |\mu^0_{r+1} - \mu^0_r|/2\}, & j = -1, -2, \ldots, \end{cases}$$

and the $U^{(r)}_i$ $(-\infty < i < \infty, 1 \leqslant r \leqslant R)$ *are i.i.d with common distribution* F.

Proof: Note that $EZ^{(r)}_j = |2^{-1} j(\mu^0_{r+1} - \mu^0_r)|$ so that the random walk $\{Z^{(r)}_j\}$ converges to $+\infty$ as $|j| \to +\infty$. This fact together with Lemma 3 implies that for any fixed $\epsilon > 0$, there exists an integer $M > 0$ such that

$$\Pr(|L_r| \leqslant M, 1 \leqslant r \leqslant R) > 1 - \epsilon, \quad \ldots \quad (18)$$

$$\Pr(|n\hat{\tau}_r - [n\tau^0_r]| \leqslant M, 1 \leqslant r \leqslant R) > 1 - \epsilon, \text{ for large } n. \quad \ldots \quad (19)$$

We claim that $S_n([n\tau^0_1] + i_1, \ldots, [n\tau^0_R] + i_R) - S_n([n\tau^0_1], \ldots, [n\tau^0_R])$ converge, jointly in i_r with $|i_r| \leqslant M$ $(1 \leqslant r \leqslant R)$, in distribution to $2 \sum_{r=1}^{R} |\mu^0_{r+1} - \mu^0_r| \, Z^{(r)}_{i_r}$. Let $\hat{l}_{r,M}$ be the values of l_r $(1 \leqslant r \leqslant R)$ that minimize $S_n(l_1, \ldots, l_R)$ subject to $|l_r - [n\tau^0_r]| \leqslant M$ $(1 \leqslant r \leqslant R)$. Let $L_{r,M}$ be the value of j that minimizes $Z^{(r)}_j$ subject to $|j| \leqslant M$. The above claim implies that, for sufficiently large n and for all i_r with $|i_r| \leqslant M$,

$$|\Pr(\hat{l}_{r,M} - [n\tau^0_r] = i_r, 1 \leqslant r \leqslant R) - \Pr(L_{r,M} = i_r, 1 \leqslant r \leqslant R)| < \epsilon \quad \ldots \quad (20)$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

Note that $|n\hat{\tau}_r-[n\tau_r^0]| \leqslant M$ $(1 \leqslant r \leqslant R)$ implies $\hat{l}_{r,M} = n\hat{\tau}_r (1 \leqslant r \leqslant R)$ and $|L_r| \leqslant M$ implies $L_{r,M} = L_r$. It follows from (18)—(20) that for large n and for all i_r with $|i_r| \leqslant M$,

$$|\Pr(n\hat{\tau}_r-[n\tau_r^0] = i_r,\ 1 \leqslant r \leqslant R)-\Pr(L_r = i_r,\ 1 \leqslant r \leqslant R)| < 3\epsilon.$$

The theorem follows by letting $\epsilon \to 0$ and $M \to \infty$.

It remains to prove the claim. We will consider only the case that $0 \leqslant i_r \leqslant M$ $(1 \leqslant r \leqslant R)$. The case that some of the i_r are negative can be treated similarly. Note that

$$\begin{aligned}
&S_n([n\tau_1^0]+i_1, \ldots, [n\tau_R^0]+i_R)-S_n([n\tau_1^0], \ldots, [n\tau_R^0]) \\
&= \sum_{r=1}^{R+1} \sum_{i=[n\tau_{r-1}^0]+i_{r-1}+1}^{[n\tau_r^0]} (\{X_n(i)-\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0]+i_r)\}^2 \\
&-\{X_n(i)-\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0])\}^2) \\
&+\sum_{r=1}^{R} \sum_{i=[n\tau_r^0]+1}^{[n\tau_r^0]+i_r} (\{X_n(i)-\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0]+i_r)\}^2 \\
&-\{X_n(i)-\bar{X}_n([n\tau_r^0], [n\tau_{r+1}^0])\}^2) \qquad \ldots \quad (21)
\end{aligned}$$

where $i_0 = 0$ and $i_{R+1} = 0$. Since M is fixed and $|i_r| \leqslant M$, $\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0]+i_r)$ and $\bar{X}_n([n\tau_r^0], [n\tau_{r+1}^0])$ converge in probability to μ_r^0 and μ_{r+1}^0, respectively, as $n \to \infty$, so that the second term on the right hand side of (21) converges in distribution to $2\sum_{r=1}^{R} |\mu_{r+1}^0-\mu_r^0| Z_{i_r}^{(r)}$. The first term equals $\sum_{r=1}^{R+1} \{[n\tau_r^0]-[n\tau_{r-1}^0]-i_{r-1}\}\{\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0])-\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0]+i_r)\}\{2\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0])-\bar{X}_n([n\tau_{r-1}^0]+i_{r-1}, [n\tau_r^0]+i_r)-\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0])\}$ which is $O_p(1/n)$. This proves the claim. □

Theorem 2 : *Assume that* (A1)—(A3) *hold.* (*i*) *As* $n \to \infty$, $n^{1/2}(\hat{\mu}_r-\mu_r^0)$, $1 \leqslant r \leqslant R+1$ *are asympototically independent and normally distributed with means* 0 *and variances* $\sigma^2/(\tau_r^0-\tau_{r-1}^0)$ *where* $\sigma^2 = \int_{-\infty}^{\infty} x^2 dF(x)$. (*ii*) $n^{1/2}(\hat{\mu}_r-\mu_r^0)$, $(r = 1, \ldots, R+1)$, $n\hat{\tau}_r-[n\tau_r^0]$, $(r = 1, \ldots, R)$ *are asymptotically independent.*

Proof : (i) Since $\hat{\tau}_r-\tau_r^0 = O_p(n^{-1})$,

$$\bar{X}_n(n\hat{\tau}_{r-1}, n\hat{\tau}_r)-\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0]) = O_p(n^{-1}).$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

Obviously, $n^{1/2}\{\bar{X}_n([n\tau_{r-1}^0], [n\tau_r^0])-\mu_r^0\}$, $1 \leqslant r \leqslant R+1$, are asymptotically independent and normally distributed with means 0 and variances $\sigma^2/(\tau_r^0-\tau_{r-1}^0)$. Part (i) is complete by noting $\hat{\mu}_r = \bar{X}_n(n\hat{\tau}_{r-1}, n\hat{\tau}_r)$.

(ii) By the proof of Theorem 1, for any $\epsilon > 0$, there exists $M > 0$ such that (18) and (19) hold. It can be shown (cf (21)) that for $|i_r| \leqslant M$, $r = 1, \ldots, R$,

$$S_n([n\tau_1^0]+i_1, \ldots, [n\tau_R^0]+i_R)-S_n([n\tau_1^0], \ldots, [n\tau_R^0])$$
$$= 2\sum_{r=1}^{R} |\mu_{r+1}^0-\mu_r^0|\, Q_n^{(r)}(i_r)+O_p(n^{-1/2})$$

where $Q_n^{(r)}(0) = 0$ and

$$Q_n^{(r)}(j) = \begin{cases} \sum\limits_{i=1}^{j} \{\mathrm{sign}(\mu_{r+1}^0-\mu_r^0)W_n([n\tau_r^0]+i)+|\mu_{r+1}^0-\mu_r^0|/2\}, & j = 1, 2, \ldots \\ \sum\limits_{i=j+1}^{0} \{\mathrm{sign}(\mu_r^0-\mu_{r+1}^0)W_n([n\tau_r^0]+i)+|\mu_{r+1}^0-\mu_r^0|/2\}, & j = -1, -2, \ldots \end{cases}$$

Let $\hat{l}'_{r,M}$ be the value of j that minimizes $Q_n^{(r)}(j)$ subject to $|j| \leqslant M$. Clearly, for fixed M, as $n \to \infty$,

$$\Pr(\hat{l}'_{r,M} = \hat{l}_{r,M}-[n\tau_r^0],\ r = 1, \ldots, R) \to 1$$

where $\hat{l}_{r,M}$ is defined in the proof of Theorem 1. Let

$$\hat{\mu}'_r = \begin{cases} \bar{X}_n(0,[n\tau_r^0]-M-1), & r = 1 \\ \bar{X}_n([n\tau_{r-1}^0]+M+1, [n\tau_r^0]-M-1), & r = 2, \ldots, R \\ \bar{X}_n([n\tau_R^0]+M+1, n), & r = R+1. \end{cases}$$

Clearly, $\hat{\mu}'_r$, $r = 1, \ldots, R+1$, $\hat{l}'_{r,m}$, $r = 1, \ldots, R$ are independent. But $n^{1/2}(\hat{\mu}'_r-\hat{\mu}_r) = o_p(1)$ and $\Pr(\hat{l}_{r,M} = n\hat{\tau}_r,\ r = 1, \ldots, R) > 1-\epsilon$ for large n so that $\Pr(\hat{l}'_{rM}=n\hat{\tau}_r-[n\tau_r^0],\ r = 1, \ldots, R) > 1-2\epsilon$ for large n. Letting $\epsilon \to 0$ and $M \to \infty$ completes the proof. □

3. The case that R is unknown

Throughout this section, R_0 is assumed to be the true unknown value of R. To construct a consistent estimator of R_0, we need the assumption that R_0 is not greater than a known upper bound R_U. For every R, let $\hat{\tau}_{1,R}, \ldots, \hat{\tau}_{R,R}$ be the values of $l_1/n, \ldots, l_R/n$ that minimize $S_n(l_1, \ldots, l_R)$, and let $\hat{\sigma}_R^2 = S_n(n\hat{\tau}_{1,R} \ldots, n\hat{\tau}_{R,R})/n$.

Theorem 3 : *Assume that* $(A1)-(A2)$ *hold and that* $\int_{-\infty}^{\infty} x^{2m}\, dF(x) < \infty$, $m \geqslant 3$. *Let* $\hat{R}$ *be the value of* R *that minimizes* $n \log \hat{\sigma}_R^2+RC_n$ *subject to* $R \leqslant R_U$

A 3-17

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

where $\{C_n\}$ *is any sequence satisfying* $C_n n^{-2/m} \to \infty$ *and* $C_n n^{-1} \to 0$ *as* $n \to \infty$. *Then* $Pr(\hat{R} = R_0)$ *converges to* 1 *as* $n \to \infty$.

Proof: Clearly, $\hat{\sigma}^2_{R_0}$ converges in probability to $\sigma^2 = \int_{-\infty}^{\infty} x^2\, dF(x)$, and for $R < R_0$ there exists $\epsilon > 0$ such that $\hat{\sigma}^2_R > \sigma^2 + \epsilon$ with probability approaching 1. Therefore, we have $\Pr(\hat{R} \geqslant R_0) \to 1$ as $n \to \infty$ since $C_n = o(n)$. For every R with $R_0 < R \leqslant R_U$,

$$\begin{aligned} S_n([n\tau_1^0], \ldots, [n\tau_{R_0}^0]) &\geqslant S_n(n\hat{\tau}_{1,R_0}, \ldots, n\hat{\tau}_{R_0,R_0}) \geqslant S_n(n\hat{\tau}_{1,R}, \ldots, n\hat{\tau}_{R,R}) \\ &\geqslant S_n(n\hat{\tau}_{1,R}, \ldots, n\hat{\tau}_{R,R}, [n\tau_1^0], \ldots, [n\tau_{R_0}^0]) \\ &\geqslant S_n([n\tau_1^0], \ldots, [n\tau_{R_0}^0]) - 2R \max_{1 \leqslant r \leqslant R_0+1} \\ &\qquad \max_{[n\tau_{r-1}^0] < i < j \leqslant [n\tau_r^0]} (j-i)\ \overline{W}_n^2(i,j) \\ &= S_n([n\tau_1^0], \ldots, [n\tau_{R_0}^0]) - O_p(n^{2/m}), \text{ by Lemma 1.} \end{aligned}$$

So,

$$0 \leqslant \hat{\sigma}^2_{R_0} - \hat{\sigma}^2_R = n^{-1}\{S_n(n\hat{\tau}_{1,R_0}, \ldots, n\hat{\tau}_{R_0,R_0}) - S_n(n\hat{\tau}_{1,R}, \ldots, n\hat{\tau}_{R,R})\} = O_p(n^{2/m-1}).$$

$$n \log \hat{\sigma}^2_{R_0} - n \log \hat{\sigma}^2_R = -n \log\{1 + (\hat{\sigma}^2_R - \hat{\sigma}^2_{R_0})/\hat{\sigma}^2_{R_0}\} = -n \log\{1 + O_p(n^{2/m-1})\}$$
$$= O_p(n^{2/m}).$$

Since $C_n n^{-2/m} \to \infty$ as $n \to \infty$, with probability approaching 1,

$$n \log \hat{\sigma}^2_R + RC_n > n \log \hat{\sigma}^2_{R_0} + R_0 C_n,$$

completing the proof. □

Let $\hat{\mu}_{r,R} = \overline{X}_n(n\hat{\tau}_{r-1,R}, n\hat{\tau}_{r,R})$, $r = 1, \ldots, R+1$ where $\hat{\tau}_{0,R} \equiv 0$ and $\hat{\tau}_{R+1,R} \equiv 1$. The dimension of the random vector

$$(\hat{\mu}_{1,\hat{R}}, \ldots, \hat{\mu}_{\hat{R}+1,\hat{R}}, \hat{\tau}_{1,\hat{R}}, \ldots, \hat{\tau}_{\hat{R},\hat{R}})$$

equals $2\hat{R}+1$, a random variable. To consider the limiting distribution of this random vector, we take the convention that

$$\mu_r^0 \equiv 0,\ r = R_0+2, \ldots, R_U+1\ ;\ \tau_r^0 \equiv 1,\ r = R_0+1, \ldots, R_U\ ;$$

$$\hat{\mu}_{r,\hat{R}} \equiv 0,\ r = \hat{R}+2, \ldots, R_U+1\ ;\ \hat{\tau}_{r,\hat{R}} \equiv 1,\ r = \hat{R}+1, \ldots, R_U.$$

Theorem 4 : *Assume that* $(A1)$—$(A3)$ *hold. Let* $\hat{R}$ *be defined as in Theorem* 3. *Then*

$$n^{1/2}(\hat{\mu}_{r,\hat{R}} - \mu_r^0),\ r = 1, \ldots, R_U+1\ ;\ n\hat{\tau}_{r,\hat{R}} - [n\tau_r^0],\ r = 1, \ldots, R_U$$

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms

are asymptotically independent,

$$n^{1/2}(\hat{\mu}_{r,\hat{R}}-\mu_r^0)\xrightarrow{D} N(0, \sigma^2/(\tau_r^0-\tau_{r-1}^0)),\ r = 1, \ldots, R_0+1$$

$$n\hat{\tau}_{r,\hat{R}}-[n\tau_r^0]\xrightarrow{D} L_r,\ r = 1, \ldots, R_0$$

and

$$n^{1/2}(\hat{\mu}_{r,\hat{R}}-\mu_r^0),\ r = R_0+2, \ldots, R_U+1\ ;\ n\hat{\tau}^0_{r,\hat{R}}-[n\tau_r^0],\ r = R_0+1, \ldots, R_U$$

converge to the degenerate distribution with unit mass at 0, *where* L_r, $r = 1, \ldots, R_0$ *are defined in Theorem* 1.

Proof: By Theorem 3, as $n\to\infty$,

$$\Pr(\hat{\mu}_{r,\hat{R}} = \hat{\mu}_{r,R_0},\ r = 1, \ldots, R_U+1\ ;\ \hat{\tau}_{r,\hat{R}} = \hat{\tau}_{r,R_0},\ r = 1, \ldots, R_U)\to 1$$

where $\hat{\mu}_{r,R_0} \equiv 0, r = R_0+2, \ldots, R_U+1\ ;\ \hat{\tau}_{r,R_0} \equiv 1,\ r = R_0+1, \ldots, R_U$. The theorem follows immediately from Theorems 1 and 2. □

In closing, we mention a generalization of the function g :

$$g(0) = \mu_1$$

$$g(t) = \mu_r+\beta t,\ \tau_{r-1} < t \leqslant \tau_r,\ r = 1, \ldots, R+1$$

where $0 = \tau_0 < \tau_1 < \ldots < \tau_R < \tau_{R+1} = 1$. This is a linear regression model with (unknown) common slope β and varying intercepts. (The model discussed in the paper is a special case with known $\beta = 0$). It is of interest to study the behavior of the least-squares estimators, both when F is known (say normal), and when F is unknown. But we have not studied this generalization.

Acknowledgments. The authors would like to thank a referee for many useful comments.

REFERENCES

CHERNOFF, H. and ZACKS, S. (1964): Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.*, **35**, 999-1018.

HINKLEY, D. V. (1970): Inference about the change-point in a sequence of random variables. *Biometirka*, **57**, 1-17.

SHABAN, S. A. (1980): Change point problem and two-phase regression: an annotated bibliography. *International Statistical Review*, **48**, 83-93.

YAO, Y. C. (1984): Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *Ann. Statist.*, **12**, 1434-1447.

——— (1988): Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, **6**, 181-189.

Paper received: *July*, 1987.
Revised: *July*, 1988.

This content downloaded from 137.132.123.69 on Fri, 29 Jul 2016 15:26:49 UTC
All use subject to http://about.jstor.org/terms