Statistica Sinica **22** (2012), 359-378
doi:http://dx.doi.org/10.5705/ss.2009.153

ESTIMATING A MONOTONE TREND

Ou Zhao and Michael Woodroofe

University of South Carolina and University of Michigan

Abstract: The estimation of a monotone trend that has been obscured by stationary fluctuations is considered and illustrated by global temperature anomalies. At an interior point, the rescaled isotonic estimators are shown to converge in distribution to Chernoff's distribution under minimal conditions on the stationary errors; and two modifications for estimating the value at an end point are compared. The asymptotic results are shown to hold conditionally given the starting values and, so, allow some relaxation of the stationarity assumption. The quality of the implicit approximations is assessed by simulation and found to be quite good for several models.

Key words and phrases: Asymptotic distribution, Brownian motion, cumulative sum diagram, greatest convex minorant, least squares, maximal inequality, spiking problem, stationary processes.

1. Introduction

Consider a time series that consists of a nondecreasing trend observed with stationary fluctuations, say

$$y_k = \mu_k + X_k, \quad k = 1, 2, \ldots, \tag{1.1}$$

where $-\infty < \mu_1 \leq \mu_2 \leq \cdots$ and $\ldots X_{-1}, X_0, X_1, \ldots$ is a strictly stationary sequence with mean 0 and finite variance. The global temperature anomalies in Example 1 provide a particular example. Others are provided by the sizes of animal or plant populations following an environmental insult. If a segment of the series is observed, say $y_1, \ldots, y_n$, then isotonic methods suggest themselves for estimating the μ_k nonparametrically. The isotonic estimators may be described as

$$\tilde{\mu}_k = \max_{i \leq k} \min_{k \leq j \leq n} \frac{y_i + \cdots + y_j}{j - i + 1}. \tag{1.2}$$

Alternatively, letting $\lfloor x \rfloor$ denote the greatest integer less than or equal to $x \in \mathbb{R}$, Y_n the cumulative sum diagram,

$$Y_n(t) = \frac{y_1 + \cdots + y_{\lfloor nt \rfloor}}{n},$$

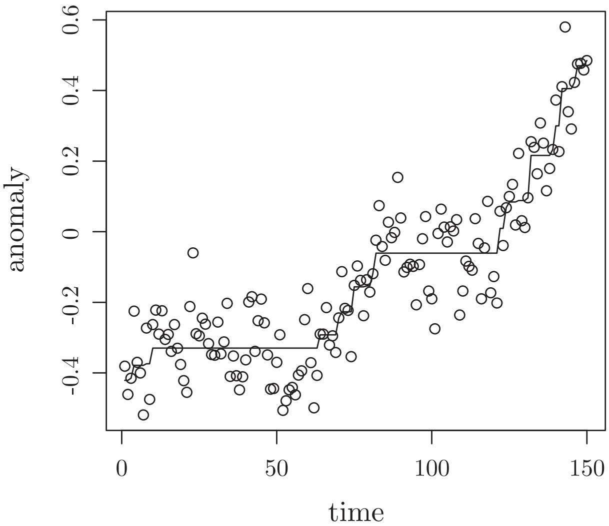

Figure 1. Global Temperature Anomalies, 1850-1999.

and $\tilde{Y}_n$ its greatest convex minorant, $\tilde{\mu}_k = \tilde{Y}_n'(k/n)$, the left hand derivative of $\tilde{Y}_n$ evaluated at $t = k/n$. See Chapter 1 of Robertson, Wright, and Dykstra (1988) for background on isotonic estimation.

Example 1. Annual global temperature anomalies from 1850-1999 are shown in Figure 1 with the isotonic estimator of trend superimposed.

In view of the *spiking problem*, described below, the modest increase at the very beginning is not impressive, and global warming does not appear to have begun until about 1915. This impression is refined by a formal confidence bound in Section 2.

With the global warming data, there is special interest in estimating μ_n, the current temperature anomaly, and there isotonic methods encounter the *spiking problem*, described in Section 7.2 of Robertson, Wright, and Dykstra (1988) for the closely related problem of estimating a monotone density. Briefly, the estimators are simply too big (small) at the the right (left) end point. We consider two methods for correcting this problem, the penalized estimators of Woodroofe and Sun (1993) and the boundary corrected estimator of Kulikov and Lopuhaä (2006), both introduced for monotone densities. The former estimates μ_n by

$$\hat{\mu}_{p,n} = \max_{i \le n} \frac{y_i + \cdots + y_n}{n - i + 1 + \lambda_n},$$

where $\lambda_n > 0$ is smoothing parameter, and the latter by $\hat{\mu}_{b,n} = \tilde{\mu}_{m_n}$, where $m_n < n$ is another smoothing parameter.

The main results of this paper obtain the asymptotic distributions of estimation errors, properly normalized, for the estimators described above. One of these results is well known for monotone regression with i.i.d. errors, and analogues of the others are known for monotone density estimation. Interest here is in extending these results to allow for dependence. Others have been interested in this question recently—notably Anevski and Hössjer (2006) and Dedecker, Merlevède, and Peligrad (2009). Our results go beyond theirs in two ways. We consider the boundary case, estimating μ_n, and our results hold conditionally given the starting values. The asymptotic distribution of the boundary corrected estimator in (3.12) is new. The conditional convergence is important because it allows observation to begin at a random time–that is, to replace $y_1, \ldots, y_n$ in (1.1) by $y_{\tau+1}, \ldots, y_{\tau+n}$, where τ is a random variable. The size of an animal population following an environmental insult, for example, starts at a random time. In addition, our conditions are weaker than those of Anevski and Hössjer (2006). Instead of the strong mixing condition, called (A9) in their paper, we use the condition (2.1) below, introduced in Maxwell and Woodroofe (2000) and further developed in Peligrad and Utev (2005). One objective of this paper is to show by example how recent results on the central limit question for sums of stationary processes can be used to weaken mixing conditions in statistical applications. Condition (2.1) is, in fact, nearly necessary for the conditional convergence, as explained in Section 2 below. Our conditions are not strictly comparable to those of Dedecker, Merlevède, and Peligrad (2009) who require a linear structure, but allow a more general class of limits. A preliminary version of this material appeared in Zhao (2008).

The main results are stated and proved in Section 3 and then illustrated by simulations and further consideration of Example 1 in Section 4. Section 2 contains some background material. Global temperature anomalies are reconsidered in Sections 2 and 4.

2. Preliminaries

A maximal inequality and conditional convergence. The main results of Peligrad and Utev (2005) are an important technical tool. To state them, let $\ldots X_{-1}, X_0, X_1, \ldots$ be a strictly stationary and ergodic sequence with mean 0 and finite variance; let $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, $\mathcal{F}_n = \sigma\{\ldots, X_{n-1}, X_n\}$, and

$$\mathbb{B}_n(t) = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}$$

for $0 \leq t \leq 1$; let $\mathbb{B}$ denote a standard Brownian motion. Both $\mathbb{B}$ and $\mathbb{B}_n$ are regarded as random elements with values in $D[0,1]$ endowed with the Skorohod topology, Billingsley (1968, Chap. 3),. Let $\|\cdot\|$ denote the norm in $L^2(P)$, $\|Y\| = \sqrt{E(Y^2)}$. It is shown in Peligrad and Utev (2005) that if

$$\sum_{n=1}^{\infty} n^{-3/2} \|E(S_n|\mathcal{F}_0)\| < \infty, \tag{2.1}$$

then

$$\Gamma := \sum_{k=0}^{\infty} 2^{-k/2} \|E(S_{2^k}|\mathcal{F}_0)\| < \infty,$$

$$E\left[\max_{k\le n} S_k^2\right] \le 6\left[E(X_1^2) + \Gamma\right] n; \tag{2.2}$$

also,

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} E(S_n^2) \tag{2.3}$$

exists, and $\mathbb{B}_n$ converges in distribution to $\sigma\mathbb{B}$. In fact, a stronger conclusion is possible. It will be shown that the conditional distributions of $\mathbb{B}_n$ given $\mathcal{F}_0$ converge in probability to the distribution of $\sigma\mathbb{B}$. It is shown in Maxwell and Woodroofe (2000) that (2.1) is almost necessary (within a logarithmic term of necessity) for convergence of the conditional distributions in the following sense: if (2.3) holds and the conditional distribution of $S_n/\sqrt{n}$ given $\mathcal{F}_0$ converges in probability to the normal distribution with mean 0 and variance σ^2, then $\|E(S_n|\mathcal{F}_0)\| = o(\sqrt{n})$.

Properties of weak convergence—for example, the Continuous Mapping Theorem and Slutzky's Theorem, extend easily to the convergence of conditional distributions. We illustrate with Slutzky's Theorem, Billingsley (1968). Let $(\mathcal{X}, d)$ denote a complete separable metric space, and let ρ be a metric that metrizes weak convergence of probability distributions on the Borel sets of $\mathcal{X}$, for example the metric (2.4) below. Next, let $\mathbb{X}_n, \mathbb{Y}_n, n = 1, 2, \ldots$, be random elements assuming values in $\mathcal{X}$; suppose that $\mathbb{X}_n$ and $\mathbb{Y}_n$ are defined on the same probability space $(\Omega_n, \mathcal{A}_n, P_n)$ say; let $\mathcal{A}_n^o \subseteq \mathcal{A}_n$ be sub sigma algebras; let μ_n and ν_n be regular conditional distributions for $\mathbb{X}_n$ and $\mathbb{Y}_n$ given $\mathcal{A}_n^o$. *If $\rho(\mu, \mu_n) \to 0$ in probability and $d(\mathbb{X}_n, \mathbb{Y}_n) \to 0$ in probability, then $\rho(\mu, \nu_n) \to 0$ in probability.* The assertion can be easily proved from the usual statement of Slutzky's Theorem, for example, Billingsley (1968, p.25), by considering subsequences which converge to ∞ so rapidly that $\rho(\mu, \mu_n) \to 0$ and $d(\mathbb{X}_n, \mathbb{Y}_n) \to 0$ *w.p.*1 along the subsequence.

There is a convenient choice of ρ. Write $\|g\|_{\mathrm{Lip}} = \sup_x |g(x)| + \sup_{x\neq y} |g(x) - g(y)|/d(x,y)$ for bounded Lipschitz continuous functions g, and let

$$\rho(\mu, \nu) = \sup_{\|g\|_{\mathrm{Lip}}\le 1} \left| \int_{\mathcal{X}} g d\mu - \int_{\mathcal{X}} g d\nu \right| \tag{2.4}$$

for probability distributions μ and ν on the Borel sets of $\mathcal{X}$. Then ρ metrizes convergence in distribution (Dudley (2002, Thm. 11.3.3)). Here is a useful feature of ρ. Let $\mathcal{A}_{n,1} \subseteq \mathcal{A}_{n,2}$ be sub sigma algebras of $\mathcal{A}_n$ and let $\mu_{n,1}$ and $\mu_{n,2}$ be regular conditional distributions for $\mathbb{X}_n$ given $\mathcal{A}_{n,1}$ and $\mathcal{A}_{n,2}$. Then $\rho(\mu, \mu_{n,1}) \leq E[\rho(\mu, \mu_{n,2})|\mathcal{A}_{n,1}]$ and, therefore,

$$E[\rho(\mu, \mu_{n,1})] \leq E[\rho(\mu, \mu_{n,2})]. \tag{2.5}$$

One more bit of preparation: if $\ldots X_{-1}, X_0, X_1, \ldots$ is any stationary sequence for which $E(X_n^2) < \infty$, then $X_n^2/n \to 0$ *a.s.* by an easy application of the Borel-Cantelli Lemmas, or the Ergodic Theorem, Petersen (1989), page 30; thus, $X_n/\sqrt{n} \to 0$ *a.s.*

If $\gamma > 0$, and $m \geq 0$ is an integer, let

$$\mathbb{X}_{m,\gamma}(t) = \frac{S_{m+\lfloor \gamma t \rfloor} - S_m}{\sqrt{\gamma}} \tag{2.6}$$

for $t \geq -m/\gamma$; and let $\mathbb{X}_{m,\gamma}^{a,b} = \mathbb{X}_{m,\gamma}|[a,b]$ denote the restriction of $X_{m,\gamma}$ to an interval $[a,b]$. Thus $\mathbb{B}_n = \mathbb{X}_{0,n}^{0,1}$. Let $\mathbb{W}$ denote a standard two-sided Brownian motion. Both $\mathbb{X}_{m,\gamma}^{a,b}$ and $\mathbb{W}^{a,b} = \mathbb{W}|[a,b]$ are regarded as elements of $D[a,b]$.

Proposition 1. *Suppose that* (2.1) *holds. Let $m_n \geq 0$ be integers, let $0 < \gamma_n \to \infty$, and let $-\infty < a < b < \infty$. If either $a \geq 0$ or $\gamma_n/m_n \to 0$, then the conditional distribution of $\mathbb{X}_{m_n,\gamma_n}^{a,b}$ given $\mathcal{F}_0$ converges in probability to the distribution of $\sigma\mathbb{W}^{a,b}$.*

Proof. For fixed a and b, write $\mathbb{X}_n = \mathbb{X}_{m_n,\gamma_n}^{a,b}$. Let H_n denote a regular conditional distribution for $\mathbb{X}_n$ given $\mathcal{F}_0$ and H the distribution of $\sigma\mathbb{W}^{a,b}$. Then it is necessary to show that $\rho[H, H_n] \to 0$ in probability.

If $a \geq 0$, it suffices to consider the case $a = 0$, since then the convergence of $\mathbb{X}_{m_n,\gamma_n}^{0,b}$ implies that of $\mathbb{X}_{m_n,\gamma_n}^{a,b}$. It also suffices to consider the case $m_n = 0$. To see why, suppose that the result is known for $m_n = 0$ and let H_n^o be a regular conditional distribution for $\mathbb{X}_{0,\gamma_n}^{0,b}$ given $\mathcal{F}_0$, so that $\lim_{n\to\infty} E[\rho(H, H_n^o)] = 0$. Next, let H_n^* be a regular conditional distribution for $\mathbb{X}_n$ given $\mathcal{F}_{m_n}$. Then $E[\rho(H, H_n)] \leq E[\rho(H, H_n^*)]$ by (2.5), and $E[\rho(H, H_n^*)] = E[\rho(H, H_n^o)]$ since the process is stationary. So, $\lim_{n\to\infty} E[\rho(H, H_n)] = 0$, as required.

So consider the case that $m_n = 0$ and $a = 0$. From Maxwell and Woodroofe (2000) there is a martingale M_n with stationary increments and a sequence $\{R_n\}$ for which $\|R_n\|/\sqrt{n} \to 0$, and $S_n = M_n + R_n$ *w.p.*1 for all n. Let

$$\mathbb{M}_n(t) = \frac{1}{\sqrt{\gamma_n}} M_{\lfloor \gamma_n t \rfloor} \quad \text{and} \quad \mathbb{R}_n(t) = \frac{1}{\sqrt{\gamma_n}} R_{\lfloor \gamma_n t \rfloor}$$

for $0 \le t \le b$. Then clearly $\mathbb{X}_n = \mathbb{M}_n + \mathbb{R}_n$ and $\mathbb{R}_n(t) \to 0$ in probability for each fixed $0 \le t \le b$. Let K_n denote a regular conditional distribution for $\mathbb{M}_n$ given $\mathcal{F}_0$. Then $\rho(H, K_n) \to 0$ with probability one by the functional version of the Martingale Central Limit Theorem, applied conditionally; see, for example, Hall and Heyde (1980, Sec. 4), From Peligrad and Utev (2005) the (unconditional) distributions of $\mathbb{X}_n$ are tight. So, the (unconditional) distributions of $\mathbb{R}_n$ are tight and, therefore, $\max_{0\le t\le b} |\mathbb{R}_n(t)| \to 0$ in probability. The special case follows from the conditional version of Slutzky's Theorem.

Suppose now that $\gamma_n/m_n \to 0$ and $a < 0$. Then, as above we may suppose $b > 0$. Let $m_n^* = m_n + \lfloor \gamma_n a \rfloor$ and let n be so large that $m_n^* > 0$. Then

$$\mathbb{X}_n(t) = \mathbb{X}^{0,b-a}_{m_n^*,\gamma_n}(t-a) - \mathbb{X}^{0,b-a}_{m_n^*,\gamma_n}(-a) + \epsilon_n(t)$$

for $a \le t \le b$, where

$$\max_{a\le t\le b} |\epsilon_n(t)| \le 2 \max_{m_n+\gamma_n a-1\le k\le m_n+\gamma_n b+1} \frac{|X_k|}{\sqrt{\gamma_n}} \to 0$$

in probability. So, it suffices to show that the conditional distribution of $\mathbb{X}_n^* := \mathbb{X}^{0,b-a}_{m_n^*,\gamma_n}$ given $\mathcal{F}_0$ converges to the distribution of $\sigma\mathbb{W}$ in $D[0, b-a]$. Let H^o be the distribution of $\sigma\mathbb{W}$ in $D[0, b-a]$, H_n^o the RCD for $\mathbb{X}^{0,b-a}_{0,\gamma_n}$ given $\mathcal{F}_0$, H_n^* a RCD for $\mathbb{X}_n^*$ given $\mathcal{F}_0$, and H_n^{**} a RCD for $\mathbb{X}_n^*$ given $\mathcal{F}_{m_n^*}$. Then, as above,

$$E\left[\rho(H^o, H_n^*)\right] \le E\left[\rho(H^o, H_n^{**})\right] = E\left[\rho(H^o, H_n^o)\right] \to 0$$

by (2.5), stationarity, and the special case.

Example 2. (Global Temperatures Revisited). Wu, Woodroofe, and Mentz (2001) tested the hypothesis $\mu_k = c$ for all k using a likelihood ratio test and found a highly significant result, well beyond the range of their tables. The same conclusion is reached using an analog of Hartigan and Hartigan (1985) Dip test. For variety and simplicity, we use a dip-like test in what follows. For testing $\mu_k = c$ for all k, the latter uses the test statistic $D_n = \sup_{0\le t\le 1} \sqrt{n}|\tilde{Y}_n(t) - \tilde{Y}_n(1)t| = -\inf_{0\le t\le 1}[Y_n(t) - \bar{y}_n t]$, where $\bar{y}_n = (y_1 + \cdots + y_n)/n$. A similar test may be used to test the hypothesis H_0^k that $\mu_j = c$ for $j \le k$, or equivalently that ϕ is constant on the interval $[0, k/n]$ for any $1 < k < n$. Thus, let

$$D_n^k = -\inf_{0\le j\le k} \left[\frac{1}{\sqrt{k}} \sum_{i=1}^{j} (y_i - \bar{y}_k)\right]$$

and consider the test that rejects H_0^k for large values of D_n^k. If $k = k_n \sim qn$, where $0 < q < 1$, then $\lim_{n\to\infty} P[D_n^k > \lambda] = e^{-2\lambda^2}$ by a simple application of

Proposition 1, and an asymptotic level-α test is to reject if

$$D_n^k > \sqrt{\frac{1}{2}\log\left(\frac{1}{\alpha}\right)}.$$

For $\alpha = 0.05$, the hypothesis is accepted for $k \le 75$ and rejected for larger values of k. So, 1925 is an asymptotic lower confidence bound for the start of global warming.

Relation to strong mixing. The condition (2.1) may be compared with mixing conditions. Let $\mathcal{G}_n = \sigma\{X_n, X_{n+1}, \ldots\}$ and recall that the strong mixing coefficients are defined by

$$\alpha_n = \sup_{A\in\mathcal{F}_0, B\in\mathcal{G}_n} |P(A\cap B) - P(A)P(B)|.$$

Then the condition of Anevski and Hössjer (2006) may be stated: for some $\epsilon > 0$,

$$E(X_1^4) < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha_n^{1/2-\epsilon} < \infty. \tag{A9}$$

Proposition 2. *If* (A9) *holds, then* (2.1) *holds.*

Proof. First write

$$\|E(S_n|\mathcal{F}_0)\| = \sup_Y E\left[E(S_n|\mathcal{F}_0)Y\right] = \sup_Y E[S_nY],$$

where the supremum is taken over all $\mathcal{F}_0$-measurable functions Y for which $\|Y\| \le 1$. By standard mixing inequalities, e.g., Corollary A.2 of Hall and Heyde (1980), Appendix III,

$$|E(S_nY)| \le \sum_{k=1}^{n} |E(X_kY)| \le \sum_{k=1}^{n} 8\|X_k\|_4\|Y\|_2\alpha_k^{1/4} \le 8\|X_0\|_4 \sum_{k=1}^{n} \alpha_k^{1/4}$$

for $\mathcal{F}_0$-measurable function Y with $\|Y\|_2 \le 1$, where $\|\cdot\|_p$ denotes the norm in L^p. So, $\|E(S_n|\mathcal{F}_0)\| \le 8\|X_0\|_4 \sum_{k=1}^{n} \alpha_k^{1/4}$ and

$$\sum_{n=1}^{\infty} n^{-3/2}\|E(S_n|\mathcal{F}_0)\| \le 24\|X_0\|_4 \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}\alpha_k^{1/4}.$$

Now let ϵ be as in (A9). Take $\max\{1, 2-4\epsilon\} < q < 2$, and let $p = q/(q-1)$. Then $p > 2$, and the right hand side of last line is at most

$$24\|X_0\|_4 \left[\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)^{p/2}\right]^{1/p} \left[\sum_{k=1}^{\infty} \alpha_k^{q/4}\right]^{1/q},$$

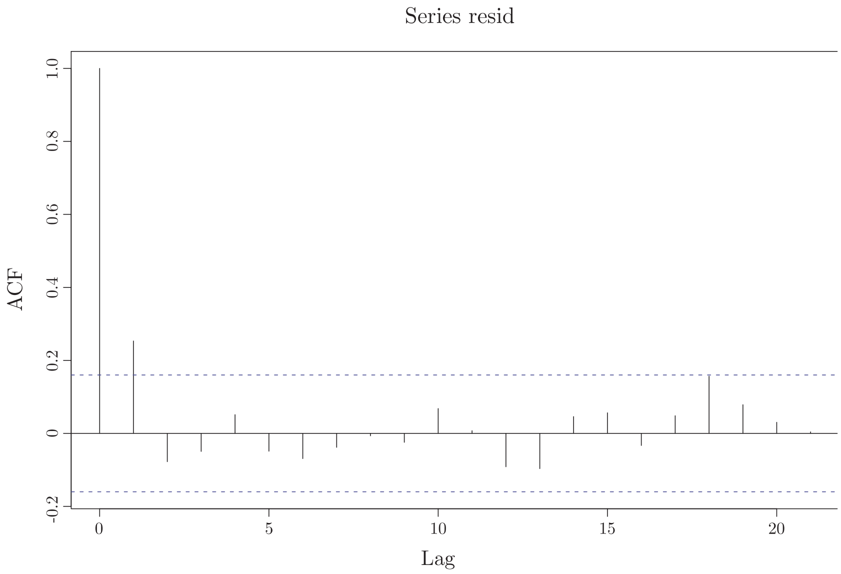

Figure 2. Global Temperature Anomalies.

which is finite.

Figure 2 shows the autocorrelation plot of the residual global temperature anomalies. This is consistent with a low order autoregressive model for which (2.1) is easily verified. By way of contrast, a low order autoregressive process need not be strongly mixing. (The Bernoulli shift process in Maxwell and Woodroofe (2000) provides an example.)

3. Asymptotic Distributions

The LSE's. Throughout this section, we suppose that the trend μ_k changes gradually in the sense that

$$\mu_k = \phi\left(\frac{k}{n}\right), \tag{3.1}$$

where ϕ is a continuous, nondecreasing function on $[0, 1]$. Thus, μ_k depends on n as well as k, but the dependence on n will be suppressed in the notation. Let

$$\Phi_n(t) = \frac{1}{n}\sum_{j=1}^{\lfloor nt \rfloor} \phi\left(\frac{j}{n}\right),$$

$$\Phi(t) = \int_0^t \phi(s)ds, \qquad \text{and} \qquad \tilde{\phi}_n(t) = \tilde{Y}_n'(t),$$

the left hand derivative of the greatest convex minorant of Y_n, for $0 \le t \le 1$.

Then $\tilde{\mu}_k = \tilde{\phi}_n(k/n)$ and

$$\sup_{0 \le t \le 1} |\Phi_n(t) - \Phi(t)| = O\left(\frac{1}{n}\right).$$

A sequence $\{t_n\} \subset (0,1)$ is called *regular* if either $t_n \to t_0 \in (0,1)$ as $n \to \infty$, or $t_n \to 1$ and $n^{1/3}(1-t_n) \to \ell \in (0,\infty]$. The first theorem obtains the asymptotic distribution of

$$\Xi_n = n^{1/3}[\tilde{\phi}_n(t_n) - \phi(t_n)] \tag{3.2}$$

for regular sequences $\{t_n\}$. Observe that if ϕ is continuously differentiable near t_0, then the asymptotic distribution, if any, is unchanged if t_n is replaced by $[nt_n]/n$.

Let $\mathbb{W}$ be a standard two-sided Brownian motion as in Section 2,

$$\mathbb{Z}(s) = \sigma\mathbb{W}(s) + \frac{1}{2}\phi'(t_0)s^2 \tag{3.3}$$

for $s \in \mathbb{R}$, and

$$\mathbb{Z}_n(s) = n^{2/3}\left[Y_n(t_n + n^{-1/3}s) - Y_n(t_n) - \phi(t_n)n^{-1/3}s\right] \tag{3.4}$$

for $s \in I_n := [-n^{1/3}t_n, n^{1/3}(1-t_n)]$. Then $\Xi_n = \tilde{\mathbb{Z}}_n'(0)$, the left hand derivative of the greatest convex minorant of $\mathbb{Z}_n$ at $s = 0$.

Let $f|J$ denote the restriction of a function f to a subset J of its domain.

Proposition 3. *Suppose that* (2.1) *and* (3.1) *hold, that ϕ is continuously differentiable near $t_0 \in (0,1]$, and that $\phi'(t_0) > 0$. Let $t_n \to t_0$ be regular and $0 < \ell = \lim_{n\to\infty} n^{1/3}(1-t_n) \le \infty$. Then for $-\infty < a < b < \ell$, the conditional distributions of $\mathbb{Z}_n|[a,b]$ given $\mathcal{F}_0$ converge in $D[a,b]$ to the* (*unconditional*) *distribution of $\mathbb{Z}|[a,b]$.*

Proof. To begin, write

$$\mathbb{Z}_n(s) = \Psi_n(s) + \mathbb{W}_n(s) + R_n(s), \tag{3.5}$$

where

$$\Psi_n(s) = n^{2/3}\left[\Phi(t_n + n^{-1/3}s) - \Phi(t_n) - \phi(t_n)n^{-1/3}s\right],$$
$$\mathbb{W}_n(s) = n^{\frac{1}{6}}\left[\mathbb{B}_n(t_n + n^{-1/3}s) - \mathbb{B}_n(t_n)\right],$$

and

$$R_n(s) = n^{2/3}\left[(\Phi_n - \Phi)(t_n + n^{-1/3}s) - (\Phi_n - \Phi)(t_n)\right].$$

It is clear that $\sup_{s\in I_n}|R_n(s)| \le 2n^{2/3}\sup_{0\le t\le 1}|\Phi_n(t)-\Phi(t)| = O(n^{-1/3}) \to 0$ as $n\to\infty$, and that

$$\lim_{n\to\infty}\Psi_n(s) = \frac{1}{2}\phi'(t_0)s^2$$

uniformly on compactas. So, it suffices to show that the conditional distribution of $\mathbb{W}_n|[a,b]$ given $\mathcal{F}_0$ converges in probability to the distribution of $\sigma\mathbb{W}|[a,b]$ in $D[a,b]$ for all compact subintervals $[a,b]\subseteq(-\infty,\ell]$; this follows easily from Proposition 1. To see how, let $m_n=\lfloor nt_n\rfloor$, $\gamma_n=n^{2/3}$, and observe that

$$\mathbb{W}_n(s) = \left[\frac{S_{m_n+\lfloor\gamma_n s\rfloor}-S_{m_n}}{\sqrt{\gamma_n}}\right] + \epsilon_n'(s) = \mathbb{X}^{a,b}_{m_n,\gamma_n}(s)+\epsilon_n'(s),$$

where $\max_{a\le s\le b}|\epsilon_n'(s)|\to 0$ in probability.

Unfortunately, Ξ_n is not quite a continuous functional of $\mathbb{Z}_n$. Two lemmas are needed to obtain its limiting distribution. The first is simply a restatement of Lemmas 5.1 and 5.2 of Wang and Woodroofe (2007). If $f: I\to\mathbb{R}$ is a bounded function and $J\subseteq I$ is a subinterval, let G_Jf denote the greatest convex minorant of $f|J$.

Lemma 1. *Let f be a bounded piecewise continuous function on a closed interval I and $[a_1,a_2]\subseteq[b_1,b_2]\subseteq I$. If*

$$f\left(\frac{a_i+b_i}{2}\right) < \frac{G_If(a_i)+G_If(b_i)}{2},\ i=1,2,$$

then $G_If = G_{[b_1,b_2]}f$ on $[a_1,a_2]$.

Lemma 2. *With the notations and conditions of Proposition* 3, $\sup_{s\in I_n}[|\mathbb{W}_n(s)| - \epsilon\min(s^2,|s|)]$ *is stochastically bounded for any* $\epsilon>0$.

Proof. Let $I_n^+=[0,n^{1/3}(1-t_n)]$ and $I_n^-=[-n^{1/3}t_n,0]$. It will be shown that $\sup_{s\in I_n^+}|W_n(s)|-\epsilon\min(s^2,|s|)$ is stochastically bounded, the treatment of $\sup_{s\in I_n^-}[|W_n(s)|-\epsilon\min(s^2,|s|)]$ being similar. Let $m_n(s)=\lfloor nt_n+n^{2/3}s\rfloor-\lfloor nt_n\rfloor$ for $s\in I_n^+$ and observe that $n^{2/3}s-1\le m_n(s)\le n^{2/3}s$. Then

$$P\left[\sup_{s\in I_n^+}|\mathbb{W}_n(s)|-\epsilon\min(s^2,|s|)>c\right] = P\left[\sup_{s\in I_n^+}\frac{|S_{m_n(s)}|}{n^{1/3}}-\epsilon\min(s^2,|s|)>c\right]$$

for fixed $n\ge 1$ and $c>0$. If $c\ge 2$, the term on the right is at most

$$P\left[\max_{m\le n^{2/3}}|S_m|>n^{1/3}c\right]+P\left[\max_{n^{2/3}\le m\le n}|S_m|-n^{-1/3}\epsilon m>\frac{1}{2}n^{1/3}c\right]. \tag{3.6}$$

Then, using the maximal inequality (2.2), it follows that (3.6) is majorized by

$$
\begin{aligned}
&P\left[\max_{m\le n^{2/3}}|S_m|>\frac{1}{2}n^{1/3}c\right]+\sum_{k=1}^{\infty}P\left[\max_{m\le 2^k n^{2/3}}|S_m|>\frac{1}{2}\left(c+\epsilon 2^k\right)n^{1/3}\right]\\
&\quad\le 24\left[\|X_1\|^2+\Gamma\right]\sum_{k=0}^{\infty}\frac{2^k}{[c+\epsilon(2^k-1)]^2},
\end{aligned}
$$

which is independent of n, and approaches 0 as $c\to\infty$.

Proposition 4. *If the assumptions of Proposition* 3 *hold with $\ell=\infty$, then for any compact interval $[a_1,a_2]\subseteq\mathbb{R}$ and any $\epsilon>0$, there is a compact interval $[b_1,b_2]\supseteq[a_1,a_2]$ such that*

$$
P\left[\tilde{\mathbb{Z}}_n=G_{[b_1,b_2]}\mathbb{Z}_n \text{ on } [a_1,a_2]\right]\ge 1-\epsilon\le P\left[G_{\mathbb{R}}\mathbb{Z}=G_{[b_1,b_2]}\mathbb{Z}\quad\text{on}\quad[a_1,a_2]\right] \tag{3.7}
$$

for all large n.

Proof. Observe that Ψ_n is convex in (3.5), and let $\gamma=\frac{1}{2}\phi'(t_0)$. Then there are $n_0\ge 1$ and $\delta>0$ for which $n_0>1/\delta^3$ and

$$
\frac{9}{5}\gamma\le\phi'(t)\le\frac{11}{5}\gamma
$$

whenever $|t-t_n|\le\delta$ and $n\ge n_0$. It then follows from a Taylor series expansion and convexity that for $n\ge n_0$

$$
\frac{9}{10}\gamma s^2\le\Psi_n(s)\le\frac{11}{10}\gamma s^2
$$

for $|s|\le\delta n^{1/3}$ and

$$
\Psi_n(s)\ge\frac{9}{10}\gamma\min\left(s^2,|s|\right) \tag{3.8}
$$

for all $s\in I_n$. Given ϵ, there is a c such that for all large n,

$$
P\left[|\mathbb{W}_n(s)|+|R_n(s)|\le\frac{1}{10}\gamma\min(s^2,|s|)+c \text{ for all } s\in I_n\right]\ge 1-\epsilon \tag{3.9}
$$

by Lemma 2. Let B_n be the event defined on the left side of (3.9). Then B_n implies $\mathbb{Z}_n(s)\ge 8\Psi_n(s)/9-c$ for all $s\in I_n$ and, therefore, $\tilde{\mathbb{Z}}_n(s)\ge 8\Psi_n(s)/9-c$ for all $s\in I_n$, since Ψ_n is convex. Let $[a_1,a_2]$ be as in the statement of the proposition; let $b_2>\max\{0,a_2\}$ be so large that

$$
\gamma\left[a_2^2+b_2^2-6a_2b_2\right]>20c;
$$

let $n > n_0$ be so large that $a_2, b_2 \in I_n$ and $\max\{|a_2|, |b_2|\} \leq \delta n^{1/3}$. Then B_n implies

$$2\mathbb{Z}_n\left(\frac{a_2+b_2}{2}\right) - \left[\tilde{\mathbb{Z}}_n(a_2) + \tilde{\mathbb{Z}}_n(b_2)\right] \leq \frac{12}{5}\gamma\left(\frac{a_2+b_2}{2}\right)^2 - \frac{4}{5}\gamma\left[a_2^2+b_2^2\right] + 4c$$
$$= -\frac{\gamma}{5}\left[a_2^2+b_2^2-6a_2b_2\right] + 4c,$$

which is negative by the choice of b_2. Similarly, for large n, B_n implies the existence of $b_1 < a_1$ for which $2\mathbb{Z}_n[(1/2)(a_1+b_1)] < \tilde{\mathbb{Z}}_n(a_1) + \tilde{Z}_n(b_1)$. The left side of (3.7) then follows from Lemma 1; the right hand inequality is similar, but simpler.

Theorem 1. *If the assumptions of Proposition* 3 *hold, then the conditional distributions of* $(G_{I_n}\mathbb{Z}_n)|J$ *given* $\mathcal{F}_0$ *converge in probability to the distribution of* $(G_{(-\infty,\ell)}\mathbb{Z})|J$ *for every compact interval* $J \subseteq (-\infty, \ell)$*, and the conditional distributions of* Ξ_n *given* $\mathcal{F}_0$ *converge in probability to the distribution of* $[G_{(-\infty,\ell)}\mathbb{Z}]'(0)$.

Proof. We first consider the case $\ell = \infty$. If J is any compact interval and $\epsilon > 0$, then there is a compact K such that

$$P\left[G_{I_n}\mathbb{Z}_n = G_K\mathbb{Z}_n \text{ on } J\right] \geq 1 - \epsilon \leq P\left[G_{\mathbb{R}}\mathbb{Z} = G_K\mathbb{Z} \text{ on } J\right] \tag{3.10}$$

for all large n. Let H and H^o denote the distributions of $(G_{\mathbb{R}}\mathbb{Z})|J$ and $(G_K\mathbb{Z})|J$, and let H_n and H_n^o denote regular conditional distributions for $(G_{I_n}\mathbb{Z}_n)|J$ and $(G_K\mathbb{Z}_n)|J$ given $\mathcal{F}_0$. Recalling ρ as defined in (2.4) with $\mathcal{X} = D(J)$, then $E[\rho(H^o, H_n^o)] \to 0$ by the Continuous Mapping Theorem since the conditional distribution of $\mathbb{Z}_n|K$ converges to the distribution of $\mathbb{Z}|K$. It follows that

$$E[\rho(H, H_n)] \leq \rho(H, H^o) + E[\rho(H^o, H_n^o)] + E[\rho(H_n^o, H_n)] \leq 3\epsilon$$

for sufficiently large n, since $\rho(H, H^o) \leq \epsilon$ and $\rho(H_n^o, H_n) \leq \epsilon$ with probability one, by Proposition 4.

Now suppose $\ell < \infty$, and consider $J = [a_1, a_2] \subseteq (-\infty, \ell]$. Following the proof of Proposition 4, there is $b_1 < a_1$ for which (3.10) holds with $K = [b_1, n^{1/3}(1-t_n)]$ for all large n, then the rest of the argument is similar as above. The second assertion of the theorem is an immediate consequence of the Continuous Mapping Theorem.

The next corollary gives the asymptotic distributions of rescaled estimation at an interior point in (3.11) and for the boundary estimator $\hat{\mu}_{b,n} = \tilde{\mu}_{m_n}$ in (3.12) if $m_n = n - n^{2/3}\ell$, where $0 < \ell < \infty$.

Corollary 1. *Suppose that the assumptions of Proposition* 3 *hold and let* $\kappa = [(1/2)\sigma^2\phi'(t)]^{1/3}$. *If* $t \in (0,1)$, *then*

$$n^{1/3}\left(\frac{\tilde{\phi}_n(t)-\phi(t)}{\kappa}\right) \Rightarrow 2\operatorname*{arg\,min}_{-\infty<s<\infty}\left[\mathbb{W}(s)+s^2\right]; \tag{3.11}$$

and, if $0<\ell<\infty$,

$$n^{1/3}\left[\tilde{\phi}_n(1-\ell n^{-1/3})-\phi(1)\right] \Rightarrow \left[G_{(-\infty,\ell]}\left(\sigma\mathbb{W}(s)+\frac{1}{2}\phi'(1)s^2\right)\right]'(0)-\ell\phi'(1). \tag{3.12}$$

Proof. The convergence follows directly from Theorem 1 since the left side of (3.11), for example, is simply Ξ_n/κ by taking $t_n \equiv t$. That $[G_{\mathbb{R}}\mathbb{Z}]'(0) = 2\kappa\ \mathrm{argmin}_s[\mathbb{W}(s)+s^2]$ in distribution follows from rescaling properties of Brownian motion.

The penalized LSE. Now consider the penalized LSE. Clearly,

$$\hat{\mu}_{p,n}-\mu_n = \max_{1\le k\le n}\frac{y_{n-k+1}+\cdots+y_n-(\lambda_n+k)\mu_n}{k+\lambda_n}.$$

The numerator here may be written as

$$y_{n-k+1}+\cdots+y_n-(\lambda_n+k)\mu_n = n^{1/3}\left[\mathbb{W}_{p,n}\left(\frac{k}{n^{2/3}}\right)-\Delta_n\left(\frac{k}{n^{2/3}}\right)-n^{-1/3}\lambda_n\mu_n\right],$$

where

$$\mathbb{W}_{p,n}(t) = n^{-1/3}\sum_{j=1}^{\lfloor n^{2/3}t\rfloor}X_{n-j+1},$$

and

$$\Delta_n(t) = n^{-1/3}\sum_{j=1}^{\lfloor n^{2/3}t\rfloor}(\mu_n-\mu_{n-j+1}).$$

It is clear that the conditional distribution of $\mathbb{W}_{p,n}$ converges to the distribution of $\sigma\mathbb{W}$ in $D[0,a]$ for all $0<a<\infty$. If (3.1) holds and ϕ is continuously differentiable near 1, then

$$\lim_{n\to\infty}\Delta_n(t) = \frac{1}{2}\phi'(1)t^2$$

uniformly on compact subintervals of $[0,\infty)$; if $\phi'(1)>0$, then there is an $\eta>0$ for which $\Delta_n(t)\ge 2\eta t^2$ for all $6n^{-2/3}\le t\le n^{1/3}$ and $n\ge 6$. Suppose now that $\lambda_n = \alpha n^{1/3}$ for some $0<\alpha<\infty$ and let

$$\mathbb{Z}_{p,n}(t) = \frac{\mathbb{W}_{p,n}(t)-\Delta_n(t)-\alpha\phi(1)}{t+\alpha/n^{1/3}}$$

for $0\le t\le n^{1/3}$. Then $n^{1/3}(\hat{\mu}_{p,n}-\mu_n) = \max_{n^{-2/3}\le t\le n^{1/3}}\mathbb{Z}_{p,n}(t)$.

Theorem 2. *Suppose that* (2.1) *and* (3.1) *hold, that* ϕ *is continuously differentiable near* 1, *and that* $\phi(1)\phi'(1) > 0$. *Then*

$$n^{1/3}(\hat{\mu}_{p,n} - \mu_n) \Rightarrow \sup_{0<t<\infty} \mathbb{Z}_{p,\infty}(t), \tag{3.13}$$

where, for $0 < t < \infty$,

$$\mathbb{Z}_{p,\infty}(t) = \frac{\sigma\mathbb{W}(t) - \alpha\phi(1) - \phi'(1)t^2/2}{t}.$$

Proof. Clearly, $\mathbb{Z}_{p,n} \Rightarrow \mathbb{Z}_{p,\infty}$ in $D(K)$ for all compact subintervals $K \subseteq (0,\infty)$. So, it suffices to show that for every $\epsilon \in (0,1)$ there is a $\delta > 0$ for which

$$P\left[\sup_{n^{-2/3}<t<n^{1/3}} \mathbb{Z}_{p,n}(t) = \sup_{\delta<t<\delta^{-1}} \mathbb{Z}_{p,n}(t)\right] \geq 1-\epsilon$$

for all large n; for this it suffices to show that for every $\epsilon \in (0,1)$, there is a $\delta > 0$ for which

$$P\left[\sup_{n^{-2/3}<t<\delta} \mathbb{Z}_{p,n}(t) > -\frac{1}{\epsilon}\right] + P\left[\sup_{\delta^{-1}\leq t\leq n^{1/3}} \mathbb{Z}_{p,n}(t) > -\frac{1}{\epsilon}\right] \leq \epsilon \tag{3.14}$$

for all large n. The first term on the left side of (3.14) is easy. If $\delta < \alpha\epsilon/2$, then

$$P\left[\sup_{n^{-2/3}<t<\delta} \mathbb{Z}_{p,n}(t) > -\frac{1}{\epsilon}\right] \leq P\left[\sup_{0\leq t\leq\delta} \mathbb{W}_{p,n}(t) > \frac{1}{2}\alpha\right],$$

which is less than $\epsilon/2$ for all large n if δ is sufficiently small, since $\mathbb{W}_{p,n} \Rightarrow \sigma\mathbb{W}$ in $D[0,1]$ and $P[\sup_{0\leq t\leq\delta}\mathbb{W}(t) > \alpha/2] = 2[1-\Phi(\alpha/(2\sqrt{\delta}))]$. For the second, recall that there is an $\eta > 0$ for which $\Delta_n(t) \geq 2\eta t^2$ for all $t \leq n^{1/3}$, and consider $\delta < \epsilon\eta$ and $n > [\epsilon\phi(1)]^{-3}$. Then

$$\begin{aligned} P\left[\sup_{\delta^{-1}<t\leq n^{1/3}} \mathbb{Z}_{p,n}(t) > -\frac{1}{\epsilon}\right] &= P\left[\mathbb{W}_{p,n}(t) > \Delta_n(t) - \frac{t}{\epsilon},\ \text{for some } t \in [\frac{1}{\delta}, n^{1/3}]\right] \\ &\leq P\left[S_n - S_{n-j} > \eta\left(\frac{j^2}{n}\right) \text{ for some } j \in [\frac{1}{\delta}n^{2/3}, n]\right]. \end{aligned}$$

Let $m = \lfloor \delta^{-1}n^{2/3} \rfloor$. Then by stationarity, the last term is at most

$$\begin{aligned} \sum_{k=1}^{\infty} P\left[\max_{j\leq m2^k} S_j > \frac{\eta m^2 2^{2k-2}}{n}\right] &\leq 6\left[\|X_0\|^2 + \Gamma\right] \sum_{k=1}^{\infty} \frac{n^2}{\eta^2 m^3 2^{3k-4}} \\ &\leq 96\left[\|X_0\|^2 + \Gamma\right] \frac{\delta^3}{\eta^2} \end{aligned}$$

for large n, and this may be made less than $\epsilon/2$ by taking δ sufficiently small.

Remark. Let $c = [(1/2)\sigma^{-4}\phi'(1)]^{1/3}\phi(1)\alpha$. Then, from simple rescaling properties of Brownian motion, the right side of (3.13) has the same distribution as

$$\kappa\left[\sup_{0<t<\infty}\frac{\mathbb{W}(t)-c-t^2}{t}\right]=\kappa U_c,\ \text{say.} \tag{3.15}$$

Similarly, letting $b = [\phi'(1)/(2\sigma)]^{2/3}\ell$, the right side of (3.12) has the same distribution as

$$\kappa\left\{\sup_{-\infty<s<0}\inf_{0<t<b}\frac{[\mathbb{W}(t)-\mathbb{W}(s)]+(t^2-s^2)}{t-s}-2b\right\}=\kappa T_b\ \text{say.} \tag{3.16}$$

In applications, it is important to choose the right smoothing parameters, α and ℓ. We content ourselves with first choosing b and c to minimize $E[T_b^2]$ and $E[U_c^2]$, and then solving for α and ℓ by plugging in consistent estimates of $\phi(1), \phi'(1)$, and σ into the equations for b and c given above. The minimizing values of b^* and c^* will be given in Section 4 by simulation. Below we consider estimating σ^2 and $\phi'(t)$ for $t \in (0, 1]$.

Nuissance Parameters. The asymptotic distributions in Corollary 1 and Theorem 2 depend on the unknowns $\phi'(t)$, and σ^2, and consistent estimations of these quantities are of interest. That $\tilde{\phi}_n(t)$ is a consistent estimator of $\phi(t)$ follows from Theorem 1. Consistent estimators of σ^2 are supplied by

$$\hat{\sigma}_n^2 = \hat{\gamma}_n(0) + 2\sum_{k\le\sqrt{n}}\left(1-\frac{k}{\sqrt{n}}\right)\hat{\gamma}_n(k),$$

where

$$\hat{\gamma}_n(k) = \frac{1}{n}\sum_{j=1}^{n-k}(y_j-\tilde{\mu}_j)(y_{j+k}-\tilde{\mu}_{j+k}).$$

See Wu, Woodroofe, and Mentz (2001) for the details. It is also possible to consistently estimate $\phi'(t)$. Let $\epsilon_1, \epsilon_2, \ldots$ be a positive sequence for which $\lim_{n\to\infty}\epsilon_n = 0$ and $\lim_{n\to\infty} n^{1/3}\epsilon_n = \infty$.

Corollary 2. *Suppose that ϕ is continuously differentiable near $t_0 \in (0, 1]$ and let t_n, $n = 1, 2, \ldots$, be a regular sequence for which $t_n \to t_0$. Then*

$$\frac{\tilde{\phi}_n(t_n)-\tilde{\phi}_n(t_n-\epsilon_n)}{\epsilon_n}\to\phi'(t_0) \tag{3.17}$$

in probability as $n \to \infty$.

Table 1. $\phi(t) = t^2$.

		$\rho = 0.5$					$\rho = 0.9$				
		t_0					t_0				
p	$\pm$	1/3	1/2	2/3	min	max	1/3	1/2	2/3	min	max
0.025	0.0016	0.0142	0.0181	0.0225	0.0142	0.0225	0.0219	0.0104	0.0119	0.0102	0.0219
0.050	0.0022	0.0341	0.0406	0.0448	0.0323	0.0448	0.0443	0.0275	0.0303	0.0262	0.0443
0.100	0.0030	0.0773	0.0898	0.0939	0.0762	0.0950	0.0878	0.0676	0.0734	0.0673	0.0878
0.200	0.0040	0.1725	0.1907	0.1892	0.1717	0.1920	0.1808	0.1631	0.1660	0.1584	0.1808
0.250	0.0043	0.2231	0.2373	0.2398	0.2219	0.2440	0.2271	0.2140	0.2193	0.2067	0.2271
0.300	0.0046	0.2745	0.2906	0.2892	0.2706	0.2951	0.2761	0.2643	0.2723	0.2563	0.2761
0.400	0.0049	0.3781	0.3864	0.3902	0.3762	0.3962	0.3719	0.3683	0.3827	0.3581	0.3827
0.500	0.0050	0.4834	0.4915	0.4930	0.4786	0.4963	0.4738	0.4718	0.4925	0.4673	0.4929
0.600	0.0049	0.5908	0.5921	0.5962	0.5826	0.5990	0.5769	0.5805	0.5955	0.5729	0.6002
0.700	0.0046	0.6957	0.6893	0.6986	0.6852	0.6997	0.6765	0.6847	0.7068	0.6765	0.7068
0.750	0.0043	0.7448	0.7381	0.7470	0.7381	0.7521	0.7271	0.7400	0.7579	0.7271	0.7591
0.800	0.0040	0.7943	0.7942	0.7977	0.7889	0.8021	0.7818	0.7924	0.8087	0.7811	0.8121
0.900	0.0030	0.8947	0.8985	0.9014	0.8888	0.9021	0.8863	0.8980	0.9050	0.8863	0.9093
0.950	0.0022	0.9455	0.9491	0.9508	0.9442	0.9532	0.9383	0.9502	0.9556	0.9382	0.9556
0.975	0.0016	0.9709	0.9746	0.9763	0.9709	0.9781	0.9663	0.9759	0.9801	0.9660	0.9801

Note. Columns three, four, and five show the empirical distribution function of Ξ_n/κ at the p^{th} percentile of Chernoff's distribution for $t_0 = 1/3,\ 1/2$, and $2/3$ in Tables 1 and 3 and $t_0 = 0.6,\ 0.75$, and 0.9 in Table 2. The value of p is in column one, and column two lists the standard errors of the simulations. Columns six and seven list the minimum and maximum of the empirical distribution function over $1/3 \le t_0 \le 2/3$. Columns eight through twelve provide the same information for $\rho = 0.9$

Proof. The difference between the left side of (3.17) and $\phi'(t_0)$ is at most

$$\left|\frac{\tilde{\phi}_n(t_n) - \phi(t_n)}{\epsilon_n}\right| + \left|\frac{\tilde{\phi}_n(t_n - \epsilon_n) - \phi(t_n - \epsilon_n)}{\epsilon_n}\right| + \left|\frac{\phi(t_n - \epsilon_n) - \phi(t_n)}{\epsilon_n} - \phi'(t_0)\right|.$$

The first two terms in the display approach zero in probability by applying Theorem 1 to the regular sequences $\{t_n\}$ and $\{t_n - \epsilon_n\}$, and so does the third since ϕ was assumed to be continuously differentiable near t_0.

4. Simulations

Simulations were conducted to assess the accuracy of the approximation implicit in (3.11). Several things affect this, including the nature of the process $\ldots X_{-1}, X_0, X_1, \ldots$, the function ϕ, the choice of t_0, and the sample size. For the fluctuations, we considered an autoregressive process $X_k = \rho X_{k-1} + \epsilon_k$, where $\ldots \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ are i.i.d. normally distributed random variables with mean 0. In Tables 1, 2, and 3, we considered two values of ρ, $\rho = 0.5$ and 0.9, representing moderate and strong dependence, three ϕ's, $\phi(t) = t^2$, $(t-0.5)_+ = \max(0, t-0.5)$,

Table 2. $\phi(t) = (t - 0.5)_+$.

		$\rho = 0.5$					$\rho = 0.9$				
			t_0					t_0			
p	$\pm$	0.6	0.75	0.9	min	max	0.6	0.75	0.9	min	max
0.025	0.0016	0.0091	0.0232	0.0232	0.0091	0.0264	0.0033	0.0102	0.0151	0.0033	0.0155
0.050	0.0022	0.0278	0.0483	0.0475	0.0278	0.0505	0.0087	0.0269	0.0357	0.0087	0.0370
0.100	0.0030	0.0783	0.0990	0.0958	0.0783	0.1007	0.0310	0.0703	0.0820	0.0310	0.0820
0.200	0.0040	0.1820	0.1973	0.1959	0.1820	0.2014	0.1065	0.1644	0.1772	0.1065	0.1809
0.250	0.0043	0.2391	0.2436	0.2472	0.2391	0.2503	0.1543	0.2194	0.2255	0.1543	0.2316
0.300	0.0046	0.2923	0.2921	0.2977	0.2909	0.3025	0.2073	0.2735	0.2677	0.2073	0.2829
0.400	0.0049	0.3979	0.3942	0.4021	0.3925	0.4058	0.3164	0.3843	0.3651	0.3164	0.3894
0.500	0.0050	0.5029	0.4960	0.5007	0.4945	0.5082	0.4390	0.4932	0.4644	0.4390	0.4938
0.600	0.0049	0.6038	0.5996	0.6004	0.5944	0.6093	0.5652	0.5970	0.5622	0.5622	0.6004
0.700	0.0046	0.7049	0.7015	0.7048	0.6962	0.7140	0.6839	0.7013	0.6606	0.6606	0.7050
0.750	0.0043	0.7550	0.7526	0.7548	0.7461	0.7657	0.7424	0.7503	0.7099	0.7099	0.7565
0.800	0.0040	0.8067	0.8072	0.8018	0.7991	0.8133	0.7985	0.8015	0.7608	0.7608	0.8116
0.900	0.0030	0.9050	0.9121	0.9000	0.8989	0.9124	0.9092	0.9049	0.8655	0.8646	0.9114
0.950	0.0022	0.9530	0.9568	0.9503	0.9495	0.9576	0.9587	0.9521	0.9246	0.9241	0.9589
0.975	0.0016	0.9753	0.9775	0.9743	0.9728	0.9803	0.9805	0.9748	0.9553	0.9553	0.9816

See note to Table 1.

Table 3. $\phi(t) = \sqrt{t}$.

		$\rho = 0.5$					$\rho = 0.9$				
			t_0					t_0			
p	$\pm$	1/3	1/2	2/3	min	max	1/3	1/2	2/3	min	max
0.025	0.0016	0.0392	0.0130	0.0046	0.0046	0.0392	0.0444	0.0155	0.0040	0.0040	0.0444
0.050	0.0022	0.0695	0.0309	0.0151	0.0146	0.0695	0.0758	0.0328	0.0138	0.0138	0.0758
0.100	0.0030	0.1259	0.0770	0.0456	0.0456	0.1259	0.1292	0.0731	0.0422	0.0422	0.1292
0.200	0.0040	0.2270	0.1759	0.1354	0.1346	0.2270	0.2291	0.1742	0.1371	0.1371	0.2291
0.250	0.0043	0.2709	0.2284	0.1909	0.1892	0.2716	0.2771	0.2209	0.1907	0.1907	0.2780
0.300	0.0046	0.3193	0.2840	0.2496	0.2478	0.3219	0.3288	0.2819	0.2484	0.2484	0.3288
0.400	0.0049	0.4176	0.3925	0.3772	0.3759	0.4211	0.4263	0.3940	0.3723	0.3723	0.4271
0.500	0.0050	0.5088	0.5037	0.5060	0.5021	0.5194	0.5216	0.5146	0.5039	0.5028	0.5217
0.600	0.0049	0.6071	0.6154	0.6375	0.6050	0.6375	0.6111	0.6261	0.6296	0.6111	0.6343
0.700	0.0046	0.6979	0.7291	0.7581	0.6979	0.7620	0.7048	0.7386	0.7633	0.7048	0.7633
0.750	0.0043	0.7478	0.7831	0.8180	0.7478	0.8180	0.7532	0.7890	0.8198	0.7532	0.8198
0.800	0.0040	0.7966	0.8334	0.8730	0.7966	0.8734	0.8052	0.8421	0.8717	0.8052	0.8717
0.900	0.0030	0.8881	0.9312	0.9595	0.8881	0.9599	0.9029	0.9360	0.9584	0.9029	0.9584
0.950	0.0022	0.9424	0.9741	0.9869	0.9424	0.9875	0.9480	0.9759	0.9839	0.9480	0.9850
0.975	0.0016	0.9686	0.9897	0.9966	0.9686	0.9970	0.9728	0.9909	0.9948	0.9728	0.9948

See note to Table 1.

and $\sqrt{t}$, and three values of t_0, $t_0 = 1/3,\ 1/2$, and $2/3$ in Tables 1 and 3 and $t_0 = 0.6,\ 0.75$, and 0.9 in Table 2. In each case the variance of $\ldots \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$

Table 4. The Boundary Corrected Estimator.

$\phi =$		t^2			$(t-\frac{1}{2})_+$			$\sqrt{t}$		
z	$n=\infty$	$\rho=0$	$\rho=0.5$	$\rho=0.9$	$\rho=0$	$\rho=0.5$	$\rho=0.9$	$\rho=0$	$\rho=0.5$	$\rho=0.9$
0.5	0.284	0.248	0.304	0.238	0.268	0.293	0.257	0.260	0.279	0.242
1.0	0.532	0.478	0.558	0.477	0.506	0.548	0.511	0.503	0.533	0.481
1.5	0.728	0.682	0.765	0.706	0.712	0.747	0.732	0.704	0.732	0.706
2.0	0.866	0.834	0.898	0.870	0.857	0.884	0.886	0.854	0.868	0.861
2.5	0.945	0.927	0.966	0.960	0.942	0.954	0.962	0.941	0.951	0.951
3.0	0.978	0.976	0.992	0.992	0.983	0.986	0.992	0.981	0.985	0.988
MSE	1.790	2.022	1.501	1.772	1.842	1.630	1.658	1.866	1.718	1.829

Table 5. The Penalized Estimator.

$\phi =$		t^2			$(t-\frac{1}{2})_+$			$\sqrt{t}$		
z	$n=\infty$	$\rho=0$	$\rho=0.5$	$\rho=0.9$	$\rho=0$	$\rho=0.5$	$\rho=0.9$	$\rho=0$	$\rho=0.5$	$\rho=0.9$
0.5	0.265	0.267	0.238	0.141	0.274	0.257	0.175	0.266	0.242	0.194
1.0	0.511	0.529	0.477	0.364	0.521	0.511	0.413	0.507	0.481	0.405
1.5	0.706	0.739	0.706	0.648	0.737	0.732	0.685	0.713	0.706	0.637
2.0	0.853	0.885	0.870	0.869	0.886	0.886	0.876	0.863	0.861	0.824
2.5	0.940	0.959	0.950	0.968	0.960	0.963	0.970	0.945	0.951	0.939
3.0	0.989	0.989	0.992	0.996	0.990	0.992	0.995	0.984	0.987	0.984
MSE	1.923	1.648	1.772	1.983	1.647	1.658	1.853	1.820	1.829	2.149

Notes. Monte Carlo estimates of the distribution function of scaled estimation errors for the boundary and penalized estimators are reported in the first six rows of Tables 4 and 5; the last row lists a Monte Carlo estimate of the mean squared error.

was chosen to make $\sqrt{E[X_k^2]} = 0.10$. The sample size was $n = 150$. In many ways, these choices are consistent with the global warming example, especially in Table 2. For each choice of these values 10,000 time series were generated and the empirical distribution function of Ξ_n/κ was computed at selected percentiles of Chernoff's distribution; the latter are available in Groeneboom and Wellner (2001).

In Table 1, the agreement between the empirical distribution function and the limiting distribution seems generally better in the right tail than the left where the empirical is consistently less than the limiting distribution. In Table 2, the agreement is very good for $t_0 = 0.75$ and 0.9 and also for $t_0 = 0.6$ except for the left tail. In Table 3, the empirical distribution of the absolute value appears to be stochastically smaller than the corresponding limit in all but two columns ($t_0 = 1/3$). In all tables the difference between moderate and strong dependence is modest, suggesting that the effect of dependence is adequately captured in the calculation of σ.

Next consider the boundary estimators $\hat{\mu}_{b,n}$ and $\hat{\mu}_{p,n}$. To use (or even simulate) them, the smoothing parameters m_n and λ_n must be specified. By (3.12) and (3.13), the expected squares of the asymptotic distributions of $\hat{\mu}_{b,n}$ and $\hat{\mu}_{p,n}$ are $\kappa^2 E(T_b^2)$ and $\kappa^2 E(U_c^2)$, where T_b and U_c are as in (3.16) and (3.15). From simulations these are minimized by $b^* \approx 0.31$ and $c^* \approx 0.61$. So,

$$\alpha^* = \frac{1}{\phi(1)} \left[\frac{2\sigma^4}{\phi'(1)} \right]^{1/3} c^* \quad \text{and} \quad \ell^* = \left[\frac{2\sigma}{\phi'(1)} \right]^{2/3} b^*$$

are suggested as asymptotically optimal choices in (3.15) and (3.16). These depend on parameters, but can be estimated as explained in the last part of Section 3. From the same simulations, the minimized values are $E(T_{b^*}^2) \approx 1.79$ and $E(U_{c^*}^2) \approx 1.92$ and, so, the boundary corrected estimator appears to be more efficient by about 5% asymptotically.

The asymptotic distributions of the normalized boundary corrected and penalized estimators do not agree well with the values for $n = 150$, especially for $\rho = 0.9$. However, the distributions of the absolute values when $n = 150$ are much closer to the asymptotic values; in most cases, the asymptotic values are at most the finite sample values or insignificantly larger. The MSE comparisons do not produce a clear winner: when $\rho = 0$, the boundary corrected estimator has a larger MSE; for $\rho = 0.9$, the penalized estimator has the larger MSE. The smoothing parameter for the boundary corrected estimator is easier to interpret, however: $\hat{\mu}_{b,n} = \tilde{\mu}_{m_n}$ is the least squares estimator of μ_{m_n}. The use of $\hat{\mu}_{b,n}$ is illustrated in the following example.

Example 3. (Global Temperature Anomalies). Letting $t_n = 140/150$ and $\epsilon_n = 15/150$ in (3.17) yields $\hat{\phi}'_1 = 2.1567$. Using this value together with $\hat{\sigma} = 0.1248$ from Wu, Woodroofe, and Mentz (2001) then yields $\hat{m}_n = 148$ (to the nearest integer) as an estimate of the asymptotically optimal m_n and $\hat{\mu}_{b,n} = 0.4700$. The choices of t_n and ϵ_n in (3.17) were arbitrary, but the final estimate is not highly sensitive to these. The estimate was unchanged when t_n was changed to any of $130/150, \ldots, 145/150$, or when ϵ_n was changed to 0.2. With this data, the boundary corrected estimator is not doing much correcting. This is appropriate, since there is no evidence of spiking.

References

Anevski, D. and Hössjer, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34**, 1874-1930.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Dedecker, J., Merlevède, F. and Peligrad, M. (2009). Invariance principles for linear processes: Applications to isotonic regression. *Bernoulli*, to appear.

Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge Univ. Press.

Groeneboom, P. and Wellner, J. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10**, 388-400.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

Hartigan, J. and Hartigan, P. (1985). The Dip test of unimodality. *Ann. Statist.* **13**, 70-84.

Kulikov, V. N. and Lopuhaä, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34**, 742-768.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28**, 713-724.

Peligrad, M. and Utev, S. (2005). A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* **33**, 798-815.

Petersen, K. (1989). *Ergodic Theory.* Paperback edition (with corrections), Cambridge Univ. Press.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Wang, X. and Woodroofe, M. (2007). A Kiefer-Wolfowitz comparison theorem for Wicksell's problem. *Ann. Statist.* **35**, 1559-1575.

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statist. Sinica* **3**, 505-515.

Wu, W. B., Woodroofe, M. and Mentz, G. (2001). Isotonic regression: another look at the change point problem. *Biometrika* **88**, 793-804.

Zhao, O. (2008). Stationary random walks and isotonic regression. Ph.D. thesis, Department of Statistics, University of Michigan.

Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

E-mail: ouzhao@stat.sc.edu

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.

E-mail: michaelw@umich.edu

(Received June 2009; accepted September 2010)