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

\[ y_k = \mu_k + X_k, \quad k = 1, 2, \ldots \]  

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 \( \mu_k \) nonparametrically. The isotonic estimators may be described as

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

1
Alternatively, letting $\lfloor x \rfloor$ denote the greatest integer that is less than or equal to $x \in \mathbb{R}$, $Y_n$ the cumulative sum diagram,

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

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.

![Figure 1: Global Temperature Anomalies, 1850-1999](image)

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 $\mu_n$, 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 $\mu_n$ by

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

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 $\mu_n$, and our results hold conditionally given the starting values. The asymptotic distribution of the boundary corrected estimator in (20) 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) by $y_{\tau+1}, \ldots, y_{\tau+n}$, where $\tau$ 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 (3) 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 (3) 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, as above; let $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, $\mathcal{F}_n = \sigma\{\ldots, X_{n-1}, X_n\}$, and

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

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

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

then

$$\Gamma := \sum_{k=0}^{\infty} 2^{-\frac{1}{2}k} \| E(S_{2k} | \mathcal{F}_0) \| < \infty,$$

$$E \left[ \max_{k \leq n} S_k^2 \right] \leq 6 \left( E(X_1^2) + \Gamma \right) n; \quad (4)$$

also,

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

exists, and $B_n$ converges in distribution to $\sigma \mathcal{B}$. In fact, a stronger conclusion is possible. It will be shown that the conditional distributions of $B_n$ given $\mathcal{F}_0$ converge in probability to the distribution of $\sigma \mathcal{B}$. It is shown in Maxwell and Woodroofe (2000) that (3) is almost necessary (within a logarithmic term of necessity) for convergence of the conditional distributions in the following sense: If (5) holds and the conditional distribution of
\[ S_n / \sqrt{n} \] given \( F_0 \) converges in probability to the normal distribution with mean 0 and variance \( \sigma^2 \), then \( \| E(S_n | F_0) \| = o(\sqrt{n}) \).

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 \((X, d)\) denote a complete separable metric space, and let \( \rho \) be a metric that metrizes weak convergence of probability distributions on the Borel sets of \( X \), for example the metric (6) below. Next, let \( X_n, Y_n, n = 1, 2, \ldots \), be random elements assuming values in \( X \); suppose that \( X_n \) and \( Y_n \) are defined on the same probability space \((\Omega_n, A_n, P_n)\) say; let \( A_{n1} \subseteq A_n \) be sub sigma algebras; and let \( \mu_n \) and \( \nu_n \) be regular conditional distributions for \( X_n \) given \( A_{n1} \). If \( \rho(\mu, \nu_n) \to 0 \) in probability and \( d(X_n, Y_n) \to 0 \) 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), page 25, by considering subsequences which converge to \( \infty \) so rapidly that \( \rho(\mu, \mu_n) \to 0 \) and \( d(X_n, Y_n) \to 0 \) w.p.1 along the subsequence.

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

\[ \rho(\mu, \nu) = \sup_{\| g \|_{\text{Lip}} \leq 1} \left| \int_X g d\mu - \int_X g d\nu \right| \] (6)

for probability distributions \( \mu \) and \( \nu \) on the Borel sets of \( X \). Then \( \rho \) metrizes convergence in distribution (Dudley (2002), Theorem 11.3.3). Here is a useful feature of \( \rho \). Let \( A_{n1} \subseteq A_{n2} \) be sub sigma algebras of \( A_n \) and let \( \mu_{n1} \) and \( \mu_{n2} \) be regular conditional distributions for \( X_n \) given \( A_{n1} \) and \( A_{n2} \). Then \( \rho(\mu, \mu_{n1}) \leq E[\rho(\mu, \mu_{n2}) | A_{n1}] \) and, therefore,

\[ E[\rho(\mu, \mu_{n1})] \leq E[\rho(\mu, \mu_{n2})]. \] (7)

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

\[ X_{m, \gamma}(t) = \frac{S_{m+|\gamma|} - S_m}{\sqrt{\gamma}} \] (8)

for \( t \geq -m / \gamma \); and let \( X_{m, \gamma}^a = X_{m, \gamma}|[a, b] \) denote the restriction of \( X_{m, \gamma} \) to an interval \([a, b] \). Thus \( B_n = X_{0, \gamma}^{0,1} \). Let \( W \) denote a standard two-sided Brownian motion. Both \( X_{m, \gamma}^a \) and \( W^a = W|[a, b] \) are regarded as elements of \( D[a, b] \).
Proposition 1. Suppose that (3) 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 \( \mathcal{X}^{a,b}_{m_n,\gamma_n} \) given \( \mathcal{F}_0 \) converges in probability to the distribution of \( \sigma^{\psi a,b} \).

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

If \( a \geq 0 \), then it suffices to consider the case \( a = 0 \), since then the convergence of \( \mathcal{X}^{0,b}_{m_n,\gamma_n} \) implies that of \( \mathcal{X}^{a,b}_{m_n,\gamma_n} \). 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 \( H_n^0 \) be a regular conditional distribution for \( \mathcal{X}^{0,b}_{0,\gamma_n} \) given \( \mathcal{F}_0 \), so that \( \lim_{n \to \infty} E[\rho(H,H_n^0)] = 0 \). Next, let \( H_n^* \) be a regular conditional distribution for \( \mathcal{X}_n \) given \( \mathcal{F}_{m_n} \). Then \( E[\rho(H,H_n)] \leq E[\rho(H,H_n^0)] \) by (7), and \( E[\rho(H,H_n^0)] = E[\rho(H,H_n^*)] \) 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

\[
M_n(t) = \frac{1}{\sqrt{\gamma_n}} M_{[\gamma_n t]} \quad \text{and} \quad \mathbb{R}_n(t) = \frac{1}{\sqrt{\gamma_n}} R_{[\gamma_n t]}
\]

for \( 0 \leq t \leq b \). Then, clearly \( \mathcal{X}_n = M_n + \mathbb{R}_n \) and \( \mathbb{R}_n(t) \to 0 \) in probability for each fixed \( 0 \leq t \leq b \). Let \( K_n \) denote a regular conditional distribution for \( M_n \) given \( \mathcal{F}_0 \). Then \( \rho(H,K_n) \to 0 \) w.p.1, by the functional version of the martingale central limit theorem, applied conditionally; see, for example, Hall and Heyde (1980), Section 4. From Peligrad and Utev (2005) the (unconditional) distributions of \( \mathcal{X}_n \) are tight. So, the ( unconditional) distributions of \( \mathbb{R}_n \) are tight and, therefore, \( \max_{0 \leq t \leq b} |\mathbb{R}_n(t)| \to 0 \) 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 \( m_n^* = m_n + [\gamma_n a] \) and let \( n \) be so large that \( m_n^* > 0 \). Then

\[
\mathcal{X}_n(t) = \mathcal{X}^{0,-a}_{m_n^*,\gamma_n} (t-a) - \mathcal{X}^{0,-a}_{m_n^*,\gamma_n} (-a) + \epsilon_n(t)
\]

for \( a \leq t \leq b \), where

\[
\max_{a \leq t \leq b} |\epsilon_n(t)| \leq 2 \max_{m_n^*+\gamma_n a-1 \leq k \leq m_n^*+\gamma_n b+1} \frac{|X_k|}{\sqrt{\gamma_n}} \to 0
\]
in probability. So, it suffices to show that the conditional distribution of $X^*_n := X^*_{m^*_n, \gamma_n}$ given $\mathcal{F}_0$ converges to the distribution of $\sigma \mathcal{W}$ in $D[0, b - a]$. Let $H^o$ be the distribution of $\sigma \mathcal{W}$ in $D[0, b - a]$, $H^*_n$ the RCD for $X^*_{0, m^*_n, \gamma_n}$ given $\mathcal{F}_0$, $H^*_n$ a RCD for $X^*_n$ given $\mathcal{F}_0$, and $H^{**}_n$ a RCD for $X^*_n$ given $\mathcal{F}_{m^*_n}$. Then, as above

$$E[\rho(H^o, H^*_n)] \leq E[\rho(H^o, H^{**}_n)] = E[\rho(H^o, H^o)] \to 0$$

by (7), 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’s (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 $D_n = \sup_{0 \leq t \leq 1} \sqrt{n}|\tilde{Y}_n(t) - \tilde{Y}_n(1)| = -\inf_{0 \leq t \leq 1}[Y_n(t) - \bar{y}_n t]$, 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 $j \leq k$, or equivalently that $\phi$ is constant on the interval $[0, k/n]$ for any $1 < k < n$.

Thus, let

$$D^k_n = -\inf_{0 \leq j \leq k} \sqrt{k} \sum_{i=1}^j (y_i - \bar{y}_k)$$

and consider the test that rejects $H^0_k$ for large values of $D^k_n$. If $k = k_n \sim qn$, where $0 < q < 1$, then $\lim_{n \to \infty} P[D^k_n > \lambda] = e^{-2\lambda^2}$, by a simple application of Proposition 1, and an asymptotic level-$\alpha$ test is to reject if

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

For $\alpha = .05$, the hypothesis is accepted for $k \leq 75$ 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 (3) 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 $\epsilon > 0$,

$$E(X_1^4) < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha_n^{\frac{1}{2} - \epsilon} < \infty. \quad (A9)$$
Proposition 2. If (A9) holds, then (3) holds.

Proof. First write
\[ \|E(S_n|F_0)\| = \sup_Y E\{E(S_n|F_0)Y\} = \sup_Y E[S_nY], \]
where the supremum is taken over all \( F_0 \)-measurable functions \( Y \) for which \( \|Y\| \leq 1 \). By standard mixing inequalities, e.g., Corollary A.2 of Hall and Heyde (1980), Appendix III,
\[ |E(S_nY)| \leq \sum_{k=1}^{n} |E(X_kY)| \leq \sum_{k=1}^{n} 8\|X_k\|_4\|Y\|_2^\frac{1}{2} \leq 8\|X_0\|_4 \sum_{k=1}^{n} \alpha_k^\frac{1}{4} \]
for \( F_0 \)-measurable function \( Y \) with \( \|Y\|_2 \leq 1 \), where \( \cdot \| \) denotes the norm in \( L^p \). So, \( \|E(S_n|F_0)\| \leq 8\|X_0\|_4 \sum_{k=1}^{n} \alpha_k^\frac{1}{4} \) and
\[ \sum_{n=1}^{\infty} n^{-\frac{3}{2}} \|E(S_n|F_0)\| \leq 24\|X_0\|_4 \sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \alpha_k^\frac{1}{4}. \]
Now let \( \epsilon \) 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
\[ 24\|X_0\|_4 \left[ \sum_{k=1}^{\infty} \left( \frac{1}{k} \right)^\frac{1}{p} \right]^{\frac{1}{2}} \left[ \sum_{k=1}^{\infty} \alpha_k^\frac{1}{4q} \right]^{\frac{1}{2}}, \]
which is finite. \( \square \)

Figure 2 shows the autocorrelation plot of the residual global temperature anomalies. This is consistent with a low order autoregressive model for which (3) 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 \( \mu_k \) changes gradually in the sense that
\[ \mu_k = \phi \left( \frac{k}{n} \right), \quad (9) \]
where $\phi$ is a continuous, nondecreasing function on $[0, 1]$. Thus, $\mu_k$ depends on $n$ as well as $k$, but the dependence on $n$ will be suppressed in the notation. Let

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

$$
\Phi(t) = \int_0^t \phi(s) ds
$$

and

$$
\tilde{\phi}_n(t) = \tilde{Y}_n'(t),
$$

the left hand derivative of the greatest convex minorant of $Y_n$, for $0 \leq t \leq 1$. Then $\tilde{\mu}_k = \tilde{\phi}_n(k/n)$ and

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

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 $n^{\frac{1}{3}}(1-t_n) \to \ell \in (0, \infty]$. The first theorem obtains the asymptotic distribution of

$$
\Xi_n = n^{\frac{1}{2}} [\tilde{\phi}_n(t_n) - \phi(t_n)] \quad (10)
$$
for regular sequences \( \{t_n\} \). Observe that if \( \phi \) 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,

\[
\mathbb{Z}(s) = \sigma \mathbb{W}(s) + \frac{1}{2} \phi'(t_0)s^2
\]

for \( s \in \mathbb{R} \), and

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

for \( s \in I_n := [-n^{\frac{1}{3}}t_n, n^{\frac{1}{3}}(1 - t_n)] \). 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 (3) and (9) hold, that \( \phi \) is continuously differentiable near \( t_0 \in (0, 1] \), and that \( \phi'(t_0) > 0 \). Let \( t_n \to t_0 \) be regular and \( 0 < \ell = \lim_{n \to \infty} n^{\frac{1}{3}}(1 - t_n) \leq \infty \). 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

\[
\mathbb{Z}_n(s) = \Psi_n(s) + \mathbb{W}_n(s) + R_n(s),
\]

where

\[
\Psi_n(s) = n^{\frac{2}{3}} \left[ \Phi(t_n + n^{-\frac{1}{3}}s) - \Phi(t_n) - \phi(t_n)n^{-\frac{1}{3}}s \right],
\]

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

and

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

It is clear that \( \sup_{s \in I_n} |R_n(s)| \leq 2n^{\frac{2}{3}} \sup_{0 \leq t \leq 1} |\Phi_n(t) - \Phi(t)| = O(n^{-\frac{1}{3}}) \to 0 \) 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 \([a, b] \subseteq (-\infty, \ell] \); and this follows easily from Proposition 1. To see how, let \(m_n = |nt_n|, \gamma_n = n^{2/3} \); and observe that

\[
\mathbb{W}_n(s) = \left[ \frac{S_{m_n+\lfloor \gamma_ns \rfloor} - S_{m_n}}{\sqrt{\gamma_n}} \right] + \epsilon'_n(s) = \mathbb{W}_{m_n,\gamma_n}(s) + \epsilon'_n(s),
\]

where \(\max_a \leq s \leq b |\epsilon'_n(s)| \to 0 \) in probability. □

Unfortunately, \(\Xi_n\) is not quite a continuous functional of \(Z_n\). The following 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 \([a, b] \subseteq [b_1, b_2] \subseteq I\). If

\[
 f \left( \frac{a_i + b_i}{2} \right) < \frac{G_Jf(a_i) + G_Jf(b_i)}{2}, \quad i = 1, 2,
\]

then \(G_Jf = G_{[b_1, b_2]}f\) 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 \(\sup_{s \in I_n^+} |\mathbb{W}_n(s)| - \epsilon \min(s^2, |s|)\) is stochastically bounded, the treatment of \(\sup_{s \in I_n^-} |\mathbb{W}_n(s)| - \epsilon \min(s^2, |s|)\) being similar. Let \(m_n(s) = \lfloor nt_n + n^{2/3}\ell \rfloor - \lfloor nt_n \rfloor\) for \(s \in I_n^+\) and observe that \(n^{2/3} - 1 \leq m_n(s) \leq n^{2/3} s\). Then

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

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

\[
P \left[ \max_{m \leq n^{2/3}} |S_m| > n^{1/3}c \right] + P \left[ \max_{n^{2/3} \leq m \leq n} |S_m| - n^{-1/3} \epsilon m > c/2 \right]
\]

Then, using the maximal inequality (4), it follows that (14) is majorized by

\[
P \left[ \max_{m \leq n^{2/3}} |S_m| > \frac{1}{2} n^{1/3}c \right] + \sum_{k=1}^{\infty} P \left[ \max_{m \leq 2^k n^{2/3}} |S_m| > \frac{1}{2} \left( c + \epsilon 2^k \right) n^{1/3} \right] + \frac{1}{2} \left( c + \epsilon (2^k - 1) \right) ^2 \]

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

$$P \left[ \tilde{Z}_n = G_{[b_1, b_2]Z_n} \text{ on } [a_1, a_2] \right] \geq 1 - \epsilon \leq P \left[ G_{Z} = G_{[b_1, b_2]}Z \text{ on } [a_1, a_2] \right]$$

for all large $n$.

**Proof.** Observe that $\Psi_n$ is convex in (13) and let $\gamma = \frac{1}{2} \phi'(t_0)$. Then there are $n_0 \geq 1$ and $\delta > 0$ for which $n_0 > 1/\delta^3$ and

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

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

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

for $|s| \leq \delta n^3$ and

$$\Psi_n(s) \geq \frac{9}{10} \gamma \min(s^2, |s|)$$

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

$$P \left[ \min(\psi_n(s), |R_n(s)|) \leq \frac{1}{10} \gamma \min(s^2, |s|) + c \text{ for all } s \in I_n \right] \geq 1 - \epsilon$$

by Lemma 2. Let $B_n$ be the event defined on the left side of (17). Then $B_n$ implies $Z_n(s) \geq 8\Psi_n(s)/9 - c$ for all $s \in I_n$ and, therefore, $\tilde{Z}_n(s) \geq 8\Psi_n(s)/9 - c$ for all $s \in I_n$, since $\Psi_n$ 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

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

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

$$2Z_n \left( \frac{a_2 + b_2}{2} \right) \geq \tilde{Z}_n(a_2) + \tilde{Z}_n(b_2)$$

$$\leq \frac{12}{5} \gamma \left( \frac{a_2 + b_2}{2} \right)^2 + \frac{4}{5} \gamma \left[ a_2^2 + b_2^2 \right] + 4c$$

$$= -\frac{7}{5} \left[ a_2^2 + b_2^2 + 6a_2 b_2 \right] + 4c,$$

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 $2Z_n \left[ \frac{1}{2}(a_1 + b_1) \right] < \tilde{Z}_n(a_1) + \tilde{Z}_n(b_1)$. The left side of (15) 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 \((G_{I_n}Z_n)|J\) given \(F_0\) converge in probability to the distribution of \((G_{(-\infty, \ell})Z)|J\) for every compact interval \(J \subseteq (-\infty, \ell)\), and the conditional distributions of \(\Xi_n\) given \(F_0\) converge in probability to the distribution of \([G_{(-\infty, \ell)}Z]'(0)\).

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
\[
P\{G_{I_n}Z_n = G_KZ_n \text{ on } J\} \geq 1 - \epsilon \leq P\{G_RZ = G_KZ \text{ on } J\} \quad (18)
\]
for all large \(n\). Let \(H\) and \(H^0\) denote the distributions of \((G_RZ)|J\) and \((G_KZ)|J\); and let \(H_n\) and \(H^0_n\) denote regular conditional distributions for \((G_{I_n}Z_n)|J\) and \((G_KZ_n)|J\) given \(F_0\). Recalling \(\rho\) as defined in (6) with \(X = D(J)\), then
\[
E[\rho(H_n, H^0_n)] \to 0,
\]
by the continuous mapping theorem since the conditional distribution of \(Z_n|K\) converges to the distribution of \(Z|K\). It follows that
\[
E[\rho(H, H_n)] \leq \rho(H, H^0) + E[\rho(H^0, H^0_n)] + E[\rho(H^0_n, H_n)] \leq 3\epsilon
\]
for sufficiently large \(n\), since \(\rho(H, H^0) \leq \epsilon\) and \(\rho(H^0_n, H_n) \leq \epsilon\) 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 (18) 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 (19) and for the boundary estimator \(\hat{\mu}_{b,n} = \hat{\mu}_{m,n}\) in (20) if \(m_n = n - n^{2/3}\ell\), where \(0 < \ell < \infty\).

Corollary 1. Suppose that the assumptions of Proposition 3 hold and let \(\kappa = [\frac{1}{2} \sigma^2 \phi'(t)]^{\frac{1}{3}}\). If \(t \in (0, 1)\), then
\[
n^{\frac{1}{3}} \left( \frac{\tilde{\phi}_n(t) - \phi(t)}{\kappa} \right) \Rightarrow 2 \arg \min_{-\infty < s < \infty} \left[ W(s) + \frac{s^2}{2} \right]; \quad (19)
\]
and if \(0 < \ell < \infty\),
\[
n^{\frac{1}{3}} \left[ \tilde{\phi}_n(1 - \ell n^{-\frac{1}{3}}) - \phi(1) \right] \Rightarrow \left[ G_{(-\infty, \ell)} \left( \sigma W(s) + \frac{1}{2} \phi'(1)s^2 \right) \right]'(0) - \ell \phi'(1). \quad (20)
\]
Proof. The convergence follows directly from Theorem 1 since the left side of (19), for example, is simply $\Xi_n/k$ by taking $t_n \equiv t$. That $[G_RZ]'(0) = 2\kappa \argmin_s[W(s) + s^2]$ in distribution follows from rescaling properties of Brownian motion. □

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

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

The numerator here may be written as

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

where

$$W_{p,n}(t) = n^{-1/3} \sum_{j=1}^{[n^{2/3}t]} X_{n-j+1}$$

and

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

It is clear that the conditional distribution of $W_{p,n}$ converges to the distribution of $\sigma W$ in $D[0,a]$ for all $0 < a < \infty$. If (9) holds and $\phi$ 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 $[0, \infty)$; and if $\phi'(1) > 0$, then there is an $\eta > 0$ for which $\Delta_n(t) \geq 2\eta t^2$ for all $6n^{-2/3} \leq t \leq n^{2/3}$ and $n \geq 6$. Suppose now that $\lambda_n = \alpha n^{2/3}$ for some $0 < \alpha < \infty$ and let

$$Z_{p,n}(t) = \frac{W_{p,n}(t) - \Delta_n(t) - \alpha \phi(1)}{t + \alpha/n^{1/3}}$$

for $0 \leq t \leq n^{1/3}$. Then $n^{1/3}(\hat{\mu}_{p,n} - \mu_n) = \max_{6n^{-2/3} \leq t \leq n^{1/3}} Z_{p,n}(t)$.

Theorem 2. Suppose that (3) and (9) hold, that $\phi$ is continuously differentiable near 1, and that $\phi(1) \phi'(1) > 0$. Then

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

where

$$Z_{p,\infty}(t) = \frac{\sigma W(t) - \alpha \phi(1) - \frac{1}{2} \phi'(1) t^2}{t}$$

for $0 < t < \infty$.  

14
Clearly, $Z_{p,n} \Rightarrow Z_{p,\infty}$ 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

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

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

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

(22)

for all large $n$. The first term on the left side of (22) is easy. If $\delta < \alpha \epsilon / 2$, then

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

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

$$P \left[ \sup_{\delta^{-1} \leq t \leq n^{3}} Z_{p,n}(t) > -\frac{1}{\epsilon} \right] = P \left[ W_{p,n}(t) > \Delta_{n}(t) - \frac{t}{\epsilon}, \text{ for some } t \in \left[ \frac{1}{\delta}, n^{3}\right] \right]$$

$$\leq P \left[ S_{n} - S_{n-j} > \eta \left( \frac{j^2}{n} \right), \text{ for some } j \in \left[ \frac{1}{\delta} n^{3}, n\right] \right]$$

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

$$\sum_{k=1}^{\infty} P \left[ \max_{j \leq m 2^k} S_{j} > \frac{\eta m^{2} 2^{2k-2}}{n} \right] \leq 6 \left[ \|X_{0}\|^{2} + \Gamma \right] \sum_{k=1}^{\infty} \frac{n^{2}}{\eta^{2} m^{3} 2^{3k-4}}$$

$$\leq 96 \left[ \|X_{0}\|^{2} + \Gamma \right] \frac{\delta^{3}}{\eta^{2}}$$

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

**Remark.** Let $c = \lfloor \frac{1}{2} \sigma^{-4} \phi'(1) \rfloor^{\gamma} \phi(1) \alpha$. Then, from simple rescaling properties of Brownian motion, the right side of (21) has the same distribution as

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

(23)
Similarly, letting $b = [\phi'(1)/(2\sigma)]^{2}t$, the right side of (20) has the same distribution as
\[
\kappa \left\{ \sup_{-\infty < s < 0} \inf_{0 < t < b} \frac{[W(t) - W(s)] + (t^{2} - s^{2})}{t - s} - 2b \right\} = \kappa T_{b} \text{ say.} \quad (24)
\]

**Nuissance Parameters.** The asymptotic distributions in Corollary 1 and Theorem 2 depend on the unknowns $\phi'(t)$, and $\sigma^{2}$, and consistent estimation 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 $\sigma^{2}$ are supplied by
\[
\hat{\sigma}^{2}_{n} = \hat{\gamma}_{n}(0) + 2 \sum_{k \leq \sqrt{n}} \left( 1 - \frac{k}{\sqrt{n}} \right) \hat{\gamma}_{n}(k),
\]
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 $\phi'$. Let $\epsilon_{1}, \epsilon_{2}, \ldots$ be a positive sequence for which $\lim_{n \to \infty} \epsilon_{n} = 0$ and $\lim_{n \to \infty} n^{\frac{3}{2}} \epsilon_{n} = \infty$.

**Corollary 2.** Suppose that $\phi$ 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
\[
\frac{\tilde{\phi}_{n}(t_{n}) - \tilde{\phi}_{n}(t_{n} - \epsilon_{n})}{\epsilon_{n}} \to \phi'(t_{0}) \quad (25)
\]
in probability as $n \to \infty$.

**Proof.** The difference between the left side of (25) 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 $\phi$ was assumed to be continuously differentiable near $t_{0}$. □
Table 1: $\phi(t) = t^2$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\pm$</th>
<th>$t_0 = \frac{1}{3}$</th>
<th>$t_0 = \frac{1}{2}$</th>
<th>$t_0 = \frac{2}{3}$</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.0016</td>
<td>.0142</td>
<td>.0181</td>
<td>.0225</td>
<td>.0142</td>
<td>.0225</td>
</tr>
<tr>
<td>.050</td>
<td>.0022</td>
<td>.0341</td>
<td>.0406</td>
<td>.0448</td>
<td>.0323</td>
<td>.0448</td>
</tr>
<tr>
<td>.100</td>
<td>.0030</td>
<td>.0773</td>
<td>.0898</td>
<td>.0939</td>
<td>.0762</td>
<td>.0950</td>
</tr>
<tr>
<td>.200</td>
<td>.0040</td>
<td>.1725</td>
<td>.1907</td>
<td>.1892</td>
<td>.1717</td>
<td>.1920</td>
</tr>
<tr>
<td>.250</td>
<td>.0043</td>
<td>.2231</td>
<td>.2373</td>
<td>.2398</td>
<td>.2219</td>
<td>.2440</td>
</tr>
<tr>
<td>.300</td>
<td>.0046</td>
<td>.2745</td>
<td>.2906</td>
<td>.2892</td>
<td>.2706</td>
<td>.2951</td>
</tr>
<tr>
<td>.400</td>
<td>.0049</td>
<td>.3781</td>
<td>.3864</td>
<td>.3902</td>
<td>.3762</td>
<td>.3962</td>
</tr>
<tr>
<td>.500</td>
<td>.0050</td>
<td>.4834</td>
<td>.4915</td>
<td>.4930</td>
<td>.4786</td>
<td>.4963</td>
</tr>
<tr>
<td>.600</td>
<td>.0049</td>
<td>.5908</td>
<td>.5921</td>
<td>.5962</td>
<td>.5826</td>
<td>.5990</td>
</tr>
<tr>
<td>.700</td>
<td>.0046</td>
<td>.6957</td>
<td>.6893</td>
<td>.6986</td>
<td>.6852</td>
<td>.6997</td>
</tr>
<tr>
<td>.750</td>
<td>.0043</td>
<td>.7448</td>
<td>.7381</td>
<td>.7470</td>
<td>.7381</td>
<td>.7521</td>
</tr>
<tr>
<td>.800</td>
<td>.0040</td>
<td>.7943</td>
<td>.7942</td>
<td>.7977</td>
<td>.7889</td>
<td>.8021</td>
</tr>
<tr>
<td>.900</td>
<td>.0030</td>
<td>.8947</td>
<td>.8985</td>
<td>.9014</td>
<td>.8888</td>
<td>.9021</td>
</tr>
<tr>
<td>.950</td>
<td>.0022</td>
<td>.9455</td>
<td>.9491</td>
<td>.9508</td>
<td>.9442</td>
<td>.9532</td>
</tr>
<tr>
<td>.975</td>
<td>.0016</td>
<td>.9709</td>
<td>.9746</td>
<td>.9763</td>
<td>.9709</td>
<td>.9781</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\pm$</th>
<th>$t_0 = \frac{1}{3}$</th>
<th>$t_0 = \frac{1}{2}$</th>
<th>$t_0 = \frac{2}{3}$</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.0016</td>
<td>.0142</td>
<td>.0181</td>
<td>.0225</td>
<td>.0142</td>
<td>.0225</td>
</tr>
<tr>
<td>.050</td>
<td>.0022</td>
<td>.0341</td>
<td>.0406</td>
<td>.0448</td>
<td>.0323</td>
<td>.0448</td>
</tr>
<tr>
<td>.100</td>
<td>.0030</td>
<td>.0773</td>
<td>.0898</td>
<td>.0939</td>
<td>.0762</td>
<td>.0950</td>
</tr>
<tr>
<td>.200</td>
<td>.0040</td>
<td>.1725</td>
<td>.1907</td>
<td>.1892</td>
<td>.1717</td>
<td>.1920</td>
</tr>
<tr>
<td>.250</td>
<td>.0043</td>
<td>.2231</td>
<td>.2373</td>
<td>.2398</td>
<td>.2219</td>
<td>.2440</td>
</tr>
<tr>
<td>.300</td>
<td>.0046</td>
<td>.2745</td>
<td>.2906</td>
<td>.2892</td>
<td>.2706</td>
<td>.2951</td>
</tr>
<tr>
<td>.400</td>
<td>.0049</td>
<td>.3781</td>
<td>.3864</td>
<td>.3902</td>
<td>.3762</td>
<td>.3962</td>
</tr>
<tr>
<td>.500</td>
<td>.0050</td>
<td>.4834</td>
<td>.4915</td>
<td>.4930</td>
<td>.4786</td>
<td>.4963</td>
</tr>
<tr>
<td>.600</td>
<td>.0049</td>
<td>.5908</td>
<td>.5921</td>
<td>.5962</td>
<td>.5826</td>
<td>.5990</td>
</tr>
<tr>
<td>.700</td>
<td>.0046</td>
<td>.6957</td>
<td>.6893</td>
<td>.6986</td>
<td>.6852</td>
<td>.6997</td>
</tr>
<tr>
<td>.750</td>
<td>.0043</td>
<td>.7448</td>
<td>.7381</td>
<td>.7470</td>
<td>.7381</td>
<td>.7521</td>
</tr>
<tr>
<td>.800</td>
<td>.0040</td>
<td>.7943</td>
<td>.7942</td>
<td>.7977</td>
<td>.7889</td>
<td>.8021</td>
</tr>
<tr>
<td>.900</td>
<td>.0030</td>
<td>.8947</td>
<td>.8985</td>
<td>.9014</td>
<td>.8888</td>
<td>.9021</td>
</tr>
<tr>
<td>.950</td>
<td>.0022</td>
<td>.9455</td>
<td>.9491</td>
<td>.9508</td>
<td>.9442</td>
<td>.9532</td>
</tr>
<tr>
<td>.975</td>
<td>.0016</td>
<td>.9709</td>
<td>.9746</td>
<td>.9763</td>
<td>.9709</td>
<td>.9781</td>
</tr>
</tbody>
</table>

Note: Columns three, four, and five show the empirical distribution function of $\Phi_n / \kappa$ at the $p^{th}$ percentile of Chernoff’s distribution for $t_0 = \frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ in Tables 1 and 3 and $t_0 = .6$, .75, and .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 $\frac{1}{3} \leq t_0 \leq \frac{2}{3}$. Columns eight through twelve provide the same information for $\rho = .9$

4 Simulations

Simulations were conducted to assess the accuracy of the approximation implicit in (19). Several things affect this, including the nature of the process $\ldots X_{-1}, X_0, X_1, \ldots$, the function $\phi$, 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 $\rho$, $\rho = .5$ and .9 representing moderate and strong dependence, three $\phi$’s, $\phi(t) = t^2$, $(t - .5)_{+} = \max(0, t - .5)$, and $\sqrt{t}$, and three values of $t_0$, $t_0 = \frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ in Tables 1 and 3 and $t_0 = .6$, .75, and .9 in Table 2. In each case the variance of $\ldots \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ was chosen to make $\sqrt{E[X_k^2]} = .10$. 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
Table 2: \( \phi(t) = (t - .5)_{+} \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \pm )</th>
<th>( \rho = .5 )</th>
<th>( \rho = .9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.025</td>
<td>.0016</td>
<td>( .0091, .0232, .0232, .0091, .0264 )</td>
<td>( .0033, .0102, .0151, .0033, .0155 )</td>
</tr>
<tr>
<td>.050</td>
<td>.0022</td>
<td>( .0278, .0483, .0475, .0278, .0505 )</td>
<td>( .0087, .0269, .0357, .0087, .0370 )</td>
</tr>
<tr>
<td>.100</td>
<td>.0030</td>
<td>( .0783, .0990, .0958, .0783, .1007 )</td>
<td>( .0310, .0703, .0820, .0310, .0820 )</td>
</tr>
<tr>
<td>.250</td>
<td>.0043</td>
<td>( .2391, .2436, .2472, .2391, .2503 )</td>
<td>( .1543, .2194, .2255, .1543, .2316 )</td>
</tr>
<tr>
<td>.300</td>
<td>.0046</td>
<td>( .2923, .2921, .2977, .2909, .3025 )</td>
<td>( .2073, .2735, .2677, .2073, .2829 )</td>
</tr>
<tr>
<td>.400</td>
<td>.0049</td>
<td>( .3979, .3942, .4021, .3925, .4058 )</td>
<td>( .3164, .3843, .3651, .3164, .3894 )</td>
</tr>
<tr>
<td>.500</td>
<td>.0050</td>
<td>( .5029, .4960, .5007, .4945, .5082 )</td>
<td>( .4390, .4932, .4644, .4390, .4938 )</td>
</tr>
<tr>
<td>.600</td>
<td>.0049</td>
<td>( .6038, .5996, .6004, .5944, .6093 )</td>
<td>( .5652, .5970, .5622, .5622, .6004 )</td>
</tr>
<tr>
<td>.700</td>
<td>.0046</td>
<td>( .7049, .7015, .7048, .6962, .7140 )</td>
<td>( .6839, .7013, .6606, .6606, .7050 )</td>
</tr>
<tr>
<td>.750</td>
<td>.0043</td>
<td>( .7550, .7526, .7548, .7461, .7657 )</td>
<td>( .7424, .7503, .7099, .7099, .7565 )</td>
</tr>
<tr>
<td>.800</td>
<td>.0040</td>
<td>( .8067, .8072, .8018, .7991, .8133 )</td>
<td>( .7985, .8015, .7608, .7608, .8116 )</td>
</tr>
<tr>
<td>.900</td>
<td>.0030</td>
<td>( .9050, .9121, .9000, .8989, .9124 )</td>
<td>( .9092, .9049, .8655, .8646, .9114 )</td>
</tr>
<tr>
<td>.950</td>
<td>.0022</td>
<td>( .9530, .9568, .9503, .9495, .9576 )</td>
<td>( .9587, .9521, .9246, .9241, .9589 )</td>
</tr>
<tr>
<td>.975</td>
<td>.0016</td>
<td>( .9753, .9775, .9743, .9728, .9803 )</td>
<td>( .9805, .9748, .9553, .9553, .9816 )</td>
</tr>
</tbody>
</table>

See note to Table 1.

Table 3: \( \phi(t) = \sqrt{t} \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \pm )</th>
<th>( \rho = .5 )</th>
<th>( \rho = .9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>( \pm )</td>
<td>( \sqrt{t} )</td>
<td>( \sqrt{t} )</td>
</tr>
<tr>
<td>( .025 )</td>
<td>.0016</td>
<td>( .0392, .0130, .0046, .0046, .0392 )</td>
<td>( .0444, .0155, .0040, .0040, .0444 )</td>
</tr>
<tr>
<td>( .050 )</td>
<td>.0022</td>
<td>( .0695, .0309, .0151, .0146, .0695 )</td>
<td>( .0758, .0328, .0138, .0138, .0758 )</td>
</tr>
<tr>
<td>( .100 )</td>
<td>.0030</td>
<td>( .1259, .0770, .0456, .0456, .1259 )</td>
<td>( .1292, .0731, .0422, .0422, .1292 )</td>
</tr>
<tr>
<td>( .200 )</td>
<td>.0040</td>
<td>( .2270, .1759, .1354, .1346, .2270 )</td>
<td>( .2291, .1742, .1371, .1371, .2291 )</td>
</tr>
<tr>
<td>( .250 )</td>
<td>.0043</td>
<td>( .2709, .2284, .1909, .1892, .2716 )</td>
<td>( .2771, .2209, .1907, .1907, .2780 )</td>
</tr>
<tr>
<td>( .300 )</td>
<td>.0046</td>
<td>( .3193, .2840, .2496, .2478, .3219 )</td>
<td>( .3288, .2819, .2484, .2484, .3288 )</td>
</tr>
<tr>
<td>( .400 )</td>
<td>.0049</td>
<td>( .4176, .3925, .3772, .3759, .4211 )</td>
<td>( .4263, .3940, .3723, .3723, .4271 )</td>
</tr>
<tr>
<td>( .500 )</td>
<td>.0050</td>
<td>( .5088, .5037, .5060, .5021, .5194 )</td>
<td>( .5216, .5146, .5039, .5028, .5217 )</td>
</tr>
<tr>
<td>( .600 )</td>
<td>.0049</td>
<td>( .6071, .6154, .6375, .6050, .6375 )</td>
<td>( .6111, .6261, .6296, .6111, .6343 )</td>
</tr>
<tr>
<td>( .700 )</td>
<td>.0046</td>
<td>( .6979, .7291, .7581, .6979, .7620 )</td>
<td>( .7048, .7386, .7633, .7048, .7633 )</td>
</tr>
<tr>
<td>( .750 )</td>
<td>.0043</td>
<td>( .7478, .7831, .8180, .7478, .8180 )</td>
<td>( .7532, .7890, .8198, .7532, .8198 )</td>
</tr>
<tr>
<td>( .800 )</td>
<td>.0040</td>
<td>( .7966, .8334, .8730, .7966, .8734 )</td>
<td>( .8052, .8421, .8717, .8052, .8717 )</td>
</tr>
<tr>
<td>( .900 )</td>
<td>.0030</td>
<td>( .8881, .9312, .9595, .8881, .9599 )</td>
<td>( .9029, .9360, .9584, .9029, .9584 )</td>
</tr>
<tr>
<td>( .950 )</td>
<td>.0022</td>
<td>( .9424, .9741, .9869, .9424, .9875 )</td>
<td>( .9480, .9759, .9839, .9480, .9850 )</td>
</tr>
<tr>
<td>( .975 )</td>
<td>.0016</td>
<td>( .9686, .9897, .9966, .9686, .9970 )</td>
<td>( .9728, .9909, .9948, .9728, .9948 )</td>
</tr>
</tbody>
</table>

See note to Table 1.
Table 4: The Boundary Corrected Estimator

<table>
<thead>
<tr>
<th>$\phi = t^2 (t - \frac{1}{2}) + \sqrt{t}$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>$n = \infty$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
<td>$\rho = .9$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
<td>$\rho = .9$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
</tr>
<tr>
<td>0.5</td>
<td>.284</td>
<td>.248</td>
<td>.304</td>
<td>.238</td>
<td>.268</td>
<td>.293</td>
<td>.257</td>
<td>.260</td>
<td>.279</td>
</tr>
<tr>
<td>1.0</td>
<td>.532</td>
<td>.478</td>
<td>.558</td>
<td>.477</td>
<td>.506</td>
<td>.548</td>
<td>.511</td>
<td>.503</td>
<td>.533</td>
</tr>
<tr>
<td>1.5</td>
<td>.728</td>
<td>.682</td>
<td>.765</td>
<td>.706</td>
<td>.712</td>
<td>.747</td>
<td>.732</td>
<td>.704</td>
<td>.732</td>
</tr>
<tr>
<td>2.0</td>
<td>.866</td>
<td>.834</td>
<td>.898</td>
<td>.870</td>
<td>.857</td>
<td>.884</td>
<td>.886</td>
<td>.854</td>
<td>.868</td>
</tr>
<tr>
<td>2.5</td>
<td>.945</td>
<td>.927</td>
<td>.966</td>
<td>.960</td>
<td>.942</td>
<td>.954</td>
<td>.962</td>
<td>.941</td>
<td>.951</td>
</tr>
<tr>
<td>3.0</td>
<td>.978</td>
<td>.976</td>
<td>.992</td>
<td>.992</td>
<td>.983</td>
<td>.986</td>
<td>.992</td>
<td>.981</td>
<td>.985</td>
</tr>
<tr>
<td>MSE</td>
<td>1.790</td>
<td>2.022</td>
<td>1.501</td>
<td>1.772</td>
<td>1.842</td>
<td>1.630</td>
<td>1.658</td>
<td>1.866</td>
<td>1.718</td>
</tr>
</tbody>
</table>

Table 5: The Penalized Estimator

<table>
<thead>
<tr>
<th>$\phi = t^2 (t - \frac{1}{2}) + \sqrt{t}$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
<th>$\rho = 0$</th>
<th>$\rho = .5$</th>
<th>$\rho = .9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>$n = \infty$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
<td>$\rho = .9$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
<td>$\rho = .9$</td>
<td>$\rho = 0$</td>
<td>$\rho = .5$</td>
</tr>
<tr>
<td>0.5</td>
<td>.265</td>
<td>.267</td>
<td>.238</td>
<td>.141</td>
<td>.274</td>
<td>.257</td>
<td>.175</td>
<td>.266</td>
<td>.242</td>
</tr>
<tr>
<td>1.0</td>
<td>.511</td>
<td>.529</td>
<td>.477</td>
<td>.364</td>
<td>.521</td>
<td>.511</td>
<td>.413</td>
<td>.507</td>
<td>.481</td>
</tr>
<tr>
<td>1.5</td>
<td>.706</td>
<td>.739</td>
<td>.706</td>
<td>.648</td>
<td>.737</td>
<td>.732</td>
<td>.685</td>
<td>.713</td>
<td>.706</td>
</tr>
<tr>
<td>2.0</td>
<td>.853</td>
<td>.885</td>
<td>.870</td>
<td>.869</td>
<td>.886</td>
<td>.886</td>
<td>.876</td>
<td>.863</td>
<td>.861</td>
</tr>
<tr>
<td>2.5</td>
<td>.940</td>
<td>.959</td>
<td>.95</td>
<td>.968</td>
<td>.960</td>
<td>.963</td>
<td>.970</td>
<td>.945</td>
<td>.951</td>
</tr>
<tr>
<td>3.0</td>
<td>.989</td>
<td>.989</td>
<td>.992</td>
<td>.996</td>
<td>.990</td>
<td>.992</td>
<td>.995</td>
<td>.984</td>
<td>.987</td>
</tr>
<tr>
<td>MSE</td>
<td>1.923</td>
<td>1.648</td>
<td>1.772</td>
<td>1.983</td>
<td>1.647</td>
<td>1.658</td>
<td>1.853</td>
<td>1.820</td>
<td>1.829</td>
</tr>
</tbody>
</table>

**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.

The empirical distribution function of $\Xi_n/\kappa$ 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 $t_0 = .75$ and .9 and also for $t_0 = .6$ 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 $\sigma$. 

19
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 $\lambda_n$ must be specified. These may be chosen to minimize a criterion, and we use the criterion mean square error below. By (20) and (21), 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 (24) and (23). From simulations these are minimized by $b^* \approx .31$ and $c^* \approx .61$. So,

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

are suggested as asymptotically optimal choices in (23) and (24). These depend on parameters, but can be estimated as explained in the last part of Section 4. From the same simulations, the minimized values are $E(T_{b,n}^2) \approx .179$ and $E(U_{c,n}^2) \approx 1.92$ 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 with the values for $n = 150$ well, especially for $\rho = .9$. However, the distributions of the absolute values when $n = 150$ are much closer to the asymptotic values; and 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 $\rho = .9$, the penalized estimator has the larger MSE. The smoothing parameter for the boundary corrected estimator is easier to interpret, however: $\hat{\mu}_{b,n} = \hat{\mu}_{m,n}$ is the least squares estimator of $\mu_{m,n}$. 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 (25) yields $\dot{\phi}'_1 = 2.1567$. Using this value together with $\hat{\sigma} = .1248$ 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 $\hat{\mu}_{b,n} = .4700$. The choices of $t_n$ and $\epsilon_n$ in (25) 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 $\epsilon_n$ was changed to .2. With this data, the boundary corrected estimator is not doing much correcting. This is appropriate, since there is no evidence of spiking. □
References


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)