A self-normalized confidence interval for the mean of a class of nonstationary processes

BY ZHIBIAO ZHAO

Department of Statistics, Penn State University, University Park, Pennsylvania 16802, U.S.A.

zuz13@stat.psu.edu

SUMMARY

We construct an asymptotic confidence interval for the mean of a class of nonstationary processes with constant mean and time-varying variances. Due to the large number of unknown parameters, traditional approaches based on consistent estimation of the limiting variance of sample mean through moving block or non-overlapping block methods are not applicable. Under a block-wise asymptotically equal cumulative variance assumption, we propose a self-normalized confidence interval that is robust against the nonstationarity and dependence structure of the data. We also apply the same idea to construct an asymptotic confidence interval for the mean difference of nonstationary processes with piecewise constant means. The proposed methods are illustrated through simulations and an application to global temperature series.

Some key words: Confidence interval; Global temperature; Invariance principle; Nonstationary process; Self-normalization; Time-varying variance.

1. INTRODUCTION

For independent and identically distributed observations, classical methods for constructing confidence intervals include asymptotic theory, the jackknife and Efron’s bootstrap; see DiCiccio & Efron (1996) for a review. There has been considerable interest in extending these results to dependent data. See, for example, Schreiber & Andrews (1984) for extensions to autoregressive moving average processes, Künsch (1989), Politis & Romano (1994) and Lahiri (2003) for extensions to stationary processes, Gonçalves & White (2002) for extensions to near-epoch dependent arrays and McElroy & Politis (2002) for discussion of linear processes with heavy-tailed innovations. To understand the difficulties raised by the dependence, let \{e_i\} be a stationary time series with zero mean and autocovariance function \( \gamma_k = \text{cov}(e_i, e_{i+k}) \) such that \( \sum_{k=0}^{\infty} |\gamma_k| < \infty \). For sample mean \( \bar{X} \) from \( X_i = \mu + e_i \) \((i = 1, \ldots, n)\),

\[
\text{var}\left( \sqrt{n}(\bar{X} - \mu) \right) = \gamma_0 + \frac{2}{n} \sum_{1 < i < j < n} \gamma_{j-i} \rightarrow \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k = \tau^2, \quad n \rightarrow \infty.
\]

The limiting variance \( \tau^2 \) is called the long-run variance, and popular choices for estimating it include the subseries approach (Carlstein, 1986), batch means (Glynn & Whitt, 1991), the prewhitening approach (Andrews & Monahan, 1992) and the block jackknife and bootstrap (Künsch, 1989; Politis & Romano, 1994; Lahiri, 2003). The stationarity assumption plays an important role in the aforementioned works as it asserts that the dependence structure does not change over time.
Our primary goal is to construct an asymptotic confidence interval for $\mu$ in the model

$$X_i = \mu + \sigma_i e_i \quad (i = 1, \ldots, n),$$

where $\sigma_i$ are unknown time-dependent constants. Then $X_i$ oscillates around the constant mean $\mu$ whereas its variance changes over time. Model (2) has two appealing features. First, we do not assume any specific time series models for $e_i$, and such a nonparametric setting provides a flexible framework for data analysis while avoiding model misspecification. Second, we allow $\sigma_i$ to change over time and hence introduce nonstationarity. In contrast, the majority of the aforementioned works are special cases of (2) with a constant variance. In §2·1, we impose two main conditions on $e_i$ and $\sigma_i$. First, we assume that the partial sum process of $e_i$ satisfies a strong invariance principle. Second, in contrast to the classical constant variance assumption, we propose a block-wise asymptotically equal cumulative variance condition for $\sigma_i$.

It is more challenging to deal with the time-varying variances case (2). For sample mean $\bar{X}$,

$$\text{var}\{\sqrt{n}(\bar{X} - \mu)\} = \frac{\gamma_0}{n} \sum_{i=1}^{n} \sigma_i^2 + \frac{2}{n} \sum_{1 \leq i < j \leq n} \sigma_i \sigma_j \gamma_{j-i} = \tau_n^2.$$ (3)

It is generally difficult to estimate $\tau_n^2$ without further restrictive assumptions on $\sigma_i$ and the dependence structure of $e_i$, since the number of unknown parameters $\sigma_i$ and $\gamma_k$ is even larger than the number of observations. Due to the nonstationarity, existing stationarity-based approaches are not applicable. Under the conditions in §2·1, we propose in §2·2 a self-normalization based confidence interval for $\mu$ that enjoys some appealing features. First, it does not involve estimating the variance $\tau_n^2$ in (3); instead the variance is cancelled out through self-normalization. Second, while classical methods require that both the block size and number of blocks go to infinity (Lahiri, 2003), our method works for a fixed number of blocks, which is a desirable feature for small sample size problems. Third, it is shown that the proposed methods are robust against the nonstationarity and dependence structure of the underlying model.

Our second goal is to construct an asymptotic confidence interval for the difference in means over two time periods. We assume that, for some $\mu_1$ and $\mu_2$,

$$X_i = \begin{cases} 
\mu_1 + \sigma_i e_i & (i = 1, \ldots, n_1), \\
\mu_2 + \sigma_i e_i & (i = n_1' + 1, \ldots, n_1' + n_2),
\end{cases}$$ (4)

where $1, \ldots, n_1$ and $n_1' + 1, \ldots, n_1' + n_2$ with $n_1' \geq n_1$ are two non-overlapping periods with means $\mu_1$ and $\mu_2$, respectively. Let $\Delta = \mu_1 - \mu_2$ be the mean difference of interest. As argued above, it would be difficult to apply traditional approaches to construct a confidence interval for $\Delta$ as it involves nonstationarity and dependence during two time periods. In §2·3, we extend the above self-normalization method to construct a confidence interval for $\Delta$. In §3·3, we apply the proposed methods to the monthly global temperature series.

2. Main results

2·1. Conditions on $e_i$ and $\sigma_i$ in (2)

First we impose some conditions on $e_i$ and $\sigma_i$ in (2). Assume that $e_i$ has the representation

$$e_i = G(\ldots, e_{i-1}, e_i),$$

where...
where $\varepsilon_i \ (i \in \mathbb{Z})$, are independent and identically distributed innovations, and $G$ is a measurable function such that $e_i$ is a well-defined random variable. For a random variable $e$ write $\|e\|_4 = (E(e^4))^{1/4}$. Let $\{e_i\}_{i \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$. Assume $E(e_0^4) < \infty$ and write

$$
\sum_{i=1}^{\infty} i \|e_i - e_i^*\|_4 < \infty, \quad e_i^* = G(\varepsilon, \varepsilon, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i). \tag{5}
$$

Corollary 4 in Wu (2007) establishes a strong invariance principle for $\{e_i\}_{1 \leq i \leq n}$ which states that, on a richer probability space, there exists a standard Brownian motion $\{B_t\}_{t \geq 0}$ such that

$$
\max_{1 \leq i \leq n} |S_i - \tau B_i| = O(n^{1/4} \log(n)), \quad S_i = \sum_{j=1}^{i} e_j, \tag{6}
$$

and $\tau^2 = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k$ is the long-run variance of $\{e_i\}$. Here and hereafter, $a_n = O(b_n)$ means that $|a_n|/b_n$ is bounded almost surely, and $a_n = o(b_n)$ means that $a_n/b_n \to 0$ almost surely. In the examples below we verify (5) for linear and nonlinear processes.

**Example 1 (Linear process).** Let $\varepsilon_i \ (i \in \mathbb{Z})$, be independent and identically distributed random variables with $E(e_i^4) < \infty$ and $E(e_0) = 0$. For $\{a_i\}_{i \geq 0}$ satisfying $\sum_{i=0}^{\infty} a_i^2 < \infty$, define the stationary process $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$. Then $\|e_i - e_i^\ast\|_4 \leq O(a_i)$ and (5) holds if $\sum_{i=1}^{\infty} i |a_i| < \infty$.

**Example 2 (Nonlinear process).** Let $\varepsilon_i \ (i \in \mathbb{Z})$, be independent and identically distributed random variables. Define recursively $e_i = R(\varepsilon_{i-1}, \varepsilon_i)$, where $R(\cdot, \cdot)$ is a measurable random map. Assume that there exists $x_0 \in \mathbb{R}$ such that $\|R(x_0, \varepsilon_0)\|_4 < \infty$ and $\rho = \sup_{x+x'} \|R(x, \varepsilon_0) - R(x', \varepsilon_0)\|_4/|x - x'| < 1$. By Wu & Shao (2004), $\|e_i - e_i^\ast\|_4 = O(\rho^i)$ and hence (5) holds. Special examples include the threshold autoregressive model, the autoregressive model with conditional heteroscedasticity, the exponential autoregressive model and the nonparametric autoregressive model.

In (2), $\sigma_i^2$ is the variance at time $i$. For $j \leq j'$ define the cumulative variance over $j, \ldots, j'$ as

$$
V(j, j') = \sum_{i=j}^{j'} \sigma_i^2. \tag{7}
$$

**Definition 1.** For simplicity let $n = km$ for $m \in \mathbb{N}$ and a fixed $k \in \mathbb{N}$. We say that $\sigma_i, i = 1, \ldots, n = km$, satisfy the $k$-block asymptotically equal cumulative variance condition if

$$
\frac{V\{ (j-1)m + 1, jm \}}{V(km)} \to \frac{1}{k}, \quad m \to \infty. \tag{8}
$$

Condition (8) states that the data can be divided into $k$ equal blocks with asymptotically equal cumulative variances, which extends the assumption of constant variance to piecewise asymptotically equal cumulative variance. In particular, (8) holds if

$$
V(1, m) = V(m+1, 2m) = \cdots = V\{ (k-1)m + 1, km \}. \tag{9}
$$

A special case is $\sigma_i = \sigma$ for a constant $\sigma$. Condition (8) is more flexible than (9) by allowing for some deviations among $V\{ (j-1)m + 1, jm \} \ (j = 1, \ldots, k)$. If the data-generating mechanism has a periodic feature with period $m$, then (8) holds. For instance, for monthly temperature series,
due to seasonal effects, the variations within one year may not be the same for different months. However, across different years, it is reasonable to expect that the variations would be roughly the same due to the periodic feature of temperature series, so $m$ can be taken as multiples of 12.

2.2. Confidence intervals through self-normalization

In this section, we propose a self-normalization based asymptotic confidence interval for $\mu$ with observations $X_1, \ldots, X_n$ from (2). In (2), define the variation of $\sigma_1, \ldots, \sigma_n$ as

$$V_n = |\sigma_n| + \sum_{i=1}^{n-1} |\sigma_{i+1} - \sigma_i|.$$ 

Intuitively, $V_n$ measures the oscillation of $\sigma_1, \ldots, \sigma_n$. If $\sigma_i = \sigma$ for a constant $\sigma$, then $V_n = |\sigma|$. If $\sigma_i = s(i/n)$ for some function $s(t)$, $0 \leq t \leq 1$, then $V_n$ is bounded from above by the total variation of $s(\cdot)$, and in particular $V_n$ has a finite bound if $s(\cdot)$ is Lipschitz continuous.

Assume that (6) holds. For the sample mean $\bar{X}$, by the summation by parts formula,

$$n(\bar{X} - \mu) = \sum_{i=1}^{n} \sigma_i(S_i - S_{i-1}) = \sigma_n S_n + \sum_{i=1}^{n-1} (\sigma_i - \sigma_{i+1}) S_i = \sigma_n \tau B_n + \sum_{i=1}^{n-1} (\sigma_i - \sigma_{i+1}) \tau B_i + O(V_n n^{1/4} \log(n))$$

$$= \tau \sum_{i=1}^{n} \sigma_i(B_i - B_{i-1}) + O(V_n n^{1/4} \log(n)). \quad (10)$$

Since the increments $B_i - B_{i-1}$ ($i \in \mathbb{N}$), are independent standard normal random variables, a confidence interval for $\mu$ would be constructed from (10) if $\tau$ and $\sigma_i$ can be estimated. Unfortunately, the number of unknown parameters is even larger than the number of observations, and therefore it is difficult if not impossible to construct a confidence interval through (10).

To attenuate this problem, we assume that (8) holds. Divide the data into $k$ equal blocks each with $m$ observations and denote by $\bar{X}(j) = m^{-1} \sum_{i=(j-1)m+1}^{jm} X_i$ the sample mean for block $j = 1, \ldots, k$. By the argument in (10),

$$m(\bar{X}(j) - \mu) = \tau \sum_{i=(j-1)m+1}^{jm} \sigma_i(B_i - B_{i-1}) + O(V_n n^{1/4} \log(n)). \quad (11)$$

Assume that

$$\frac{\sum_{i=1}^{n} \sigma_i^2 n^{1/2} \log^2(n)}{\sum_{i=1}^{n} \sigma_i^2} \to 0, \quad n \to \infty. \quad (12)$$

Therefore, by (8),

$$\frac{m(\bar{X}(j) - \mu)}{\sqrt{V(1,km)}} = \frac{\tau}{\sqrt{V(1,km)}} \sum_{i=(j-1)m+1}^{jm} \sigma_i(B_i - B_{i-1}) + o(1) \quad (j = 1, \ldots, k), \quad (13)$$

are asymptotically independent and identically distributed normal random variables.
DEFINITION 2. Let \( H(z_1, \ldots, z_{k-1}) \) be a \((k-1)\)-variate function. We say that \( H \) is of order one if \( H(\lambda z_1, \ldots, \lambda z_{k-1}) = |\lambda| H(z_1, \ldots, z_{k-1}) \) for all \( \lambda, z_1, \ldots, z_{k-1} \in \mathbb{R} \).

Let \( H \) be a continuous \((k-1)\)-variate function of order one. By (13) and the continuous mapping theorem, as \( m \to \infty \), we have the convergence in distribution

\[
\frac{\sum_{j=1}^{k} (\bar{X}(j) - \mu)}{H(\bar{X}(1) - \bar{X}(2), \ldots, \bar{X}(k-1) - \bar{X}(k))} \to \frac{\sum_{j=1}^{k} W_j}{H(W_1 - W_2, \ldots, W_{k-1} - W_k)} = \xi_k, \tag{14}
\]

where \( W_1, \ldots, W_k \) are independent standard normal random variables. Under condition (8), the unknown quantities \( \sigma^2 \sum_{j=(j-1)m+1}^{jm} \sigma_i^2 \) cancel out in the numerator and denominator of (14), thus resulting in a pivot. Theorem 1 follows from (14).

**Theorem 1.** In (2), let the errors \( e_i \) satisfy (5). Assume that (8) holds with \( n = km \) for a fixed integer \( k \). Further assume that (12) holds. Denote by \( q_k(1-\alpha/2) \) the 100(1 - \( \alpha/2 \))% quantile of \( \xi_k \) in (14). Then a 100(1 - \( \alpha \))% asymptotic confidence interval for \( \mu \) is

\[
\frac{1}{k} \sum_{j=1}^{k} \bar{X}(j) \pm \frac{q_k(1-\alpha/2)}{k} H(\bar{X}(1) - \bar{X}(2), \ldots, \bar{X}(k-1) - \bar{X}(k)). \tag{15}
\]

Theorem 1 provides a general self-normalized asymptotic confidence interval. Examples of \( H \) include, for instances, \( H(z_1, \ldots, z_{k-1}) = (\sum_{j=1}^{k-1} z_j^2)^{1/2}, \sum_{j=1}^{k-1} |z_j| \) or median\(1 \leq j \leq k-1 |z_j| \). For simplicity, we choose \( H \) so that (14) becomes the convergence in distribution of

\[
\frac{\sum_{j=1}^{k} (\bar{X}(j) - \mu) / \sqrt{k}}{\sqrt{\sum_{j=1}^{k} (\bar{X}(j) - \bar{X})^2 / (k-1)}} \to \frac{\sum_{j=1}^{k} W_j / \sqrt{k}}{\sqrt{\sum_{j=1}^{k} (W_j - \bar{W})^2 / (k-1)}} = t_{k-1}, \tag{16}
\]

where \( \bar{X} = \sum_{j=1}^{k} \bar{X}(j)/k, \bar{W} = \sum_{j=1}^{k} W_j/k, t_{k-1} \) is Student’s \( t \)-statistic with \( k - 1 \) degrees of freedom and the equality holds in distribution. Consequently, (15) becomes

\[
\frac{1}{k} \sum_{j=1}^{k} \bar{X}(j) \pm t_{k-1} \left( 1 - \frac{\alpha}{2} \right) \left[ \sum_{j=1}^{k} (\bar{X}(j) - \bar{X})^2 / (k(k-1)) \right]^{1/2},
\]

where \( t_{k-1}(1 - \alpha/2) \) is the 100(1 - \( \alpha/2 \))% quantile of \( t_{k-1} \).

By Theorem 1, the constructed asymptotic confidence interval (15) does not depend on the unknown nonstationarity or dependence structure. In contrast to traditional approaches that require consistent estimation of the complicated variance \( \tau_n^2 \) in (3), the proposed approach cancels out \( \tau_n^2 \) through a self-normalization procedure under condition (8). Also, unlike the traditional approaches that require both the block length and number of blocks go to infinity, the proposed method works for a fixed \( k \), which is a desirable feature for small sample problems.

There is no universal rule to choose \( m \) in (8), and one may get a reasonable value either from experience or the nature of the data-generating mechanism. For example, for monthly data, a reasonable choice of \( m \) would be 12; for daily data, one may take \( m = 365 \) to be the number of observations within one year. Clearly, if (8) holds for a particular \( m \), then it also holds for any multiples of \( m \). Thus, it is desirable to use a relatively large \( m \) in (8) to reduce the possibility of a misspecified \( m \). In our simulation in §3-1, we also examine the effect of misspecifications of \( m \). In §3-2, we propose a simulation scheme to examine the empirical performance for a given dataset and a testing procedure to check condition (8).
2.3. Confidence intervals for mean difference

We extend the idea in §2.2 to construct an asymptotic confidence interval for the mean difference $\Delta = \mu_1 - \mu_2$ based on observations from (4). Assume that (8) holds with $m$. For simplicity, let $n_1 = k_1 m$, $n'_1 = k'_1 m$ and $n_2 = k_2 m$ for some integers $k_1 \leq k'_1 < k_2$. Let $X(j)$ be defined as in §2.2. By the same argument in (11), $\bar{X}(1) - \mu_1, \ldots, \bar{X}(k_1) - \mu_1, \bar{X}(k'_1 + 1) - \mu_2, \ldots, \bar{X}(k'_1 + k_2) - \mu_2$ are asymptotically independent and identically distributed normal random variables. Therefore, we can use the two-sample $t$-statistic to construct an asymptotic confidence interval for $\Delta$.

**Theorem 2.** Let $X_i$ be from (4) with $n_1 = k_1 m$, $n'_1 = k'_1 m$ and $n_2 = k_2 m$ for some integers $k_1 \leq k'_1 < k_2$. Assume that the same conditions in Theorem 1 hold. Denote by $t_{k_1+k_2-2}(1 - \alpha/2)$ the $100(1 - \alpha/2)$% quantile of $t$-distribution with $k_1 + k_2 - 2$ degrees of freedom. Then a $100(1 - \alpha)$% asymptotic confidence interval for $\Delta = \mu_1 - \mu_2$ is

$$
\bar{X} - \bar{X} \pm t_{k_1+k_2-2} \left(1 - \frac{\alpha}{2}\right) \Omega(k_1, k'_1, k_2),
$$

where $\bar{X} = \sum_{j=1}^{k_1} X(j)/k_1$, $\bar{X} = \sum_{j=k'_1+1}^{k'_1+k_2} X(j)/k_2$ and

$$
\Omega(k_1, k'_1, k_2) = \left\{ \frac{k_1 + k_2}{k_1 k_2 (k_1 + k_2 - 2)} \right\}^{1/2} \left[ \sum_{j=1}^{k_1} (\bar{X} - \bar{X})^2 + \sum_{j=k'_1+1}^{k'_1+k_2} (\bar{X} - \bar{X})^2 \right]^{1/2}.
$$

3. Finite sample performance and an application

3.1. Empirical coverage probabilities based on simulated models

We compare the finite sample performance of the proposed confidence interval to that of existing methods. Most existing methods deal with stationary processes, and we briefly introduce the idea using the setting in (1). Using asymptotic normality, we can construct confidence intervals by inserting in a consistent estimate $\hat{\tau}^2$ of the long-run variance $\tau^2$ in (1). Two popular choices of $\hat{\tau}^2$ are the moving block and non-overlapping block methods; see Lahiri (2003). Alternatively, one can also construct confidence intervals using the bootstrap distribution of $\sqrt{n}(\bar{X} - \mu)/\hat{\tau}$ with the aforementioned two estimates $\hat{\tau}^2$. To implement the above methods, one needs a block length parameter, denoted by $m$, such that $m \to \infty$ and $n/m \to \infty$. See Lahiri (2003, Ch. 2 and 3). Finally, we also consider the naive confidence interval constructed by ignoring the dependence and nonstationarity so that observations are assumed to be independent and identically distributed.

In (2), without loss of generality let $\mu = 0$ so that $X_i = \sigma_i \epsilon_i$. For $\sigma_i$, consider the periodic function $\sigma_i = \cos(2\pi i/100)$. Clearly, (9) and (8) hold for $m = 25, 50, 75, \ldots$. To access the effect of misspecification of $m$ in (8), we try three choices $m = 15, 25, 30$. For $\epsilon_i$, first we consider the nonlinear threshold autoregressive model

$$
\text{Model I: } \epsilon_i = \eta_i - E(\eta_i), \quad \eta_i = \theta|\eta_{i-1}| + \sqrt{(1 - \theta^2)}\epsilon_i, \quad |\theta| < 1. \quad (17)
$$

Here, $\epsilon_i$ are independent standard normal random variables. By Example 2, (5) holds. By Andel et al. (1984), the stationary solution $\eta_i$ has the skew-normal distribution with $E(\eta_i) = \theta \sqrt{2/\pi}$. For $\theta$, we consider four choices $\theta = 0, 0.2, 0.5, 0.8$, representing independence, mild,
Table 1. Comparisons of empirical coverage percentages for models (17) and (18)

<table>
<thead>
<tr>
<th>Model (17)</th>
<th>m = 15</th>
<th>m = 25</th>
<th>m = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>0</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>SN</td>
<td>95.3</td>
<td>95.5</td>
<td>95.4</td>
</tr>
<tr>
<td>MB</td>
<td>90.9</td>
<td>91.3</td>
<td>90.4</td>
</tr>
<tr>
<td>NB</td>
<td>90.7</td>
<td>90.8</td>
<td>90.7</td>
</tr>
<tr>
<td>MBB</td>
<td>79.7</td>
<td>81.4</td>
<td>80.1</td>
</tr>
<tr>
<td>NBB</td>
<td>97.0</td>
<td>96.5</td>
<td>95.4</td>
</tr>
<tr>
<td>ID</td>
<td>94.9</td>
<td>94.5</td>
<td>89.5</td>
</tr>
</tbody>
</table>

Model (18) with standard normal innovations

| Model (18) with m = 25 and \( t \)-distributed innovations with \( k \) degrees of freedom |
|--------|--------|--------|--------|
| \( \beta \) | 2.5    | 2.01   | 1.01   | 0.7    | 2.5    | 2.01   | 1.01   | 0.7    | 2.5    | 2.01   | 1.01   | 0.7    |
| SN     | 94.6   | 94.7   | 94.4   | 95.0   | 94.8   | 94.5   | 95.6   | 96.8   | 95.2   | 96.4   | 97.4   | 98.0   |
| MB     | 89.6   | 89.0   | 87.4   | 88.9   | 86.5   | 86.4   | 86.6   | 89.2   | 85.0   | 84.7   | 85.4   | 87.8   |
| NB     | 89.2   | 89.1   | 88.7   | 90.5   | 85.6   | 86.3   | 88.2   | 91.0   | 84.4   | 84.5   | 87.2   | 90.1   |
| MBB    | 81.3   | 82.0   | 82.4   | 87.6   | 79.1   | 78.0   | 82.2   | 87.4   | 75.7   | 78.9   | 81.9   | 86.7   |
| NBB    | 97.0   | 95.7   | 96.2   | 97.0   | 97.1   | 96.9   | 97.5   | 98.1   | 98.8   | 97.7   | 98.3   | 98.3   |
| ID     | 86.8   | 81.7   | 58.0   | 51.3   | 86.8   | 81.6   | 57.6   | 51.2   | 86.4   | 81.0   | 57.9   | 51.3   |

\( \theta \), the proposed self-normalization based confidence interval; \( \text{MB} \) and \( \text{NB} \), confidence intervals via the asymptotic normality with the moving block and non-overlapping block estimates of \( \tau^2 \) in (1) when assuming stationarity; \( \text{MBB} \) and \( \text{NBB} \), the moving block and non-overlapping block bootstrap methods; ID, the naive method by ignoring the dependence and nonstationarity.

Intermediate and strong dependence, respectively. Next, we consider linear process

Model II: \( \epsilon_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}, \quad a_j = (j + 1)^{-\beta} / 10, \quad \beta > 1/2. \) (18)

By Example 1, (5) holds for \( \beta > 2 \). The cases \( \beta > 1 \) and \( \beta \in (0.5, 1] \) correspond to short-range and long-range dependence, respectively. We choose \( \beta = 0.7 \), 1.01, 2.01, 2.5. Here 0.7 and 1.01 are included to examine the performance when the dependence is so strong that (5) is violated. To access the effect of different distributions of the innovations \( \epsilon_i \), we use standard normal innovations and \( t \)-distributed innovations with \( k = 3, 4, 5 \) degrees of freedom.

In Table 1, we report the empirical coverage percentages, rounded to one decimal place, among 10 000 realizations of the aforementioned methods with nominal level 95% and sample size \( n = 150 \). The coverage probabilities for the two block bootstrap methods are based on 1000 realizations to save computation time. Clearly, the proposed self-normalization method has empirical coverage percentages close to 95% and outperforms other methods for almost all cases. Moreover, the proposed method works reasonably well even when \( m \) is misspecified. In contrast, the naive method when ignoring dependence and nonstationarity performs poorly when dependence is strong with \( \theta = 0.8 \) for Model I and \( \beta = 0.7 \) and 1.01 for Model II. The asymptotic normality based confidence intervals when assuming stationarity have coverage probabilities between 85%
and 90%. The moving block bootstrap method has coverage probabilities of about 80% whereas the non-overlapping block bootstrap is conservative. On the other hand, when the dependence is very strong with $\beta = 0.7$ in Model II, the proposal tends to be slightly conservative. To examine the performance at other nominal levels, we plot the empirical quantiles of 10,000 realizations of the simulated statistic in the left-hand side of (16) with $m = 25$ and $k = 6$ versus the theoretical Student’s $t$-quantiles with 5 degrees of freedom. The quantile-quantile plots show that the empirical quantiles are reasonably close to their theoretical counterparts, see the Supplementary Material.

3.2. Empirical coverage probabilities based on a data-generated process and condition (8)

Now we consider a data-generated process. To illustrate the idea, let $X_1, \ldots, X_n$ be observations with sample mean $\bar{X}$. Then $e_i^* = X_i - \bar{X}$ can be viewed as the residual noises. To examine the empirical coverage probabilities, we propose the following simulation scheme:

(A) Simulate independent and identically distributed random variables $\xi_1, \ldots, \xi_n$ such that $\xi_i$ takes $-1$ and 1 with probability 50% each. Clearly, $E(\xi_i) = 0$ and $\text{var}(\xi_i) = 1$.

(B) Compute the data-generated innovations as $\tilde{e}_i^* = \xi_i e_i^*$.

(C) Use $\tilde{e}_1^*, \ldots, \tilde{e}_n^*$ to construct the confidence interval, and check whether it covers zero.

(D) Repeat (A)–(C) 10,000 times and obtain the proportion of realizations that cover zero.

By adding the random sign $\xi_i$, the simulation scheme above generates innovations $\tilde{e}_i^*$ directly from the data, and hence preserves the nonstationarity of the original data. See Liu (1988) for related discussions. We apply the idea to the monthly temperature observations in §3.3 to examine the coverage probabilities for the two time periods 1871–1900 and 1951–1980. The result is summarized in Table 2. We see that the proposed method works reasonably well.

Recall $V(j, j')$ in (7). Let $e_i^*$ be defined as above. To examine condition (8), we use $\tilde{V}(j, j') = \sum_{l=j}^{j'} e_i^* e_l^*$ as a proxy of $V(j, j')$. Define

$$D_j = \tilde{V}(\{j - 1\}m + 1, jm) - \tilde{V}(1, km)/k \quad (j = 1, \ldots, k).$$

(19)

If (8) is true, then $D_j$ should be close to zero. Using the 95% criterion, we check whether $|D_j| > 1.96s$, where $s$ is the standard deviation of $D_1, \ldots, D_k$.

3.3. An application to global temperature data


Table 2. Comparisons of empirical coverage percentages for data-generated innovations. The abbreviations in Table 1 are used

<table>
<thead>
<tr>
<th>$m$</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
<th>12</th>
<th>24</th>
<th>36</th>
<th>48</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td>60</td>
<td>12</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td>60</td>
<td>12</td>
</tr>
<tr>
<td>$m$</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td>60</td>
<td>12</td>
<td>24</td>
<td>36</td>
<td>48</td>
<td>60</td>
<td>12</td>
</tr>
</tbody>
</table>
Inference for a class of nonstationary processes

Fig. 1. Monthly global temperature anomalies in degrees Celsius during 1850–2008. The first two and the last two dashed vertical lines correspond to 1871–1900 and 1951–1980.

Table 3. 95% confidence intervals for \(\mu_1\), \(\mu_2\) and \(\Delta = \mu_1 - \mu_2\) in degrees Celsius

<table>
<thead>
<tr>
<th>(m)</th>
<th>(\mu_1)</th>
<th>(\mu_2)</th>
<th>(\Delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>([-0.36, -0.28])</td>
<td>([-0.16, -0.08])</td>
<td>([-0.25, -0.15])</td>
</tr>
<tr>
<td>24</td>
<td>([-0.37, -0.27])</td>
<td>([-0.17, -0.07])</td>
<td>([-0.27, -0.13])</td>
</tr>
<tr>
<td>36</td>
<td>([-0.38, -0.27])</td>
<td>([-0.17, -0.08])</td>
<td>([-0.27, -0.13])</td>
</tr>
<tr>
<td>48</td>
<td>([-0.39, -0.25])</td>
<td>([-0.19, -0.08])</td>
<td>([-0.27, -0.11])</td>
</tr>
<tr>
<td>60</td>
<td>([-0.40, -0.25])</td>
<td>([-0.16, -0.08])</td>
<td>([-0.27, -0.13])</td>
</tr>
</tbody>
</table>

The land portion consists of surface air temperature data from over 3000 station records after correcting for such non-climatic errors as station shifts and/or instrument changes. The marine data consist of sea surface temperature measurements from ships and buoys after correcting for different types of buckets used before 1942.

From the time series plot in Fig. 1, we see that while the temperature steadily increased during 1901–1950 and 1981–2008, the trend was rather flat during 1871–1900 and 1951–1980 and the temperature series were oscillating around certain levels; see the window between the first two and the last two dashed lines in Fig. 1. We use (4) to model the latter two time periods. The point estimates for \(\mu_1\), \(\mu_2\) and \(\Delta = \mu_1 - \mu_2\) are \(-0.32\), \(-0.12\) and \(-0.20\) degrees Celsius, respectively. We construct their 95% confidence intervals using different choices of \(m\). Due to the nature of monthly observations, we take \(m = 12, 24, \ldots, 60\) to be multiples of 12. To examine condition (8), we apply the idea in §3.2 to the two time periods 1871–1900 and 1951–1980 separately and in combination. We find that there are at most one or two \(D_{\text{s}}\)s in (19) outside 1.96 standard deviations. Thus, we conclude that there is no evidence to reject (8). The simulation study in §3.2 also suggests that the proposal works well.

As the summary results in Table 3 show, the confidence intervals are consistent across different choices of \(m\). It is interesting that the confidence intervals for \(\mu_2\) lie strictly to the right of that for \(\mu_1\). Thus, we conclude that the temperature during 1951–1980 is significantly higher than that during 1871–1900. Moreover, a 95% confidence interval for such an increase, negative \(\Delta\), in mean temperature is \([0.13, 0.27]°C\) with \(m = 36\). While our conclusion does not go beyond the claim of an increasing trend in previous studies (Wu et al., 2001; Wu & Zhao, 2007), the imposed assumptions are weaker and allow for unknown nonstationarity and dependence structure.
ACKNOWLEDGEMENT

I am grateful to the editor and two referees for their constructive comments. This work was supported by a National Institute on Drug Abuse grant.

SUPPLEMENTARY MATERIAL

Supplementary Material available at *Biometrika* online includes plots of the simulation in §3.1.

REFERENCES


[Received December 2009. Revised October 2010]