10-28-2003

Block Bootstrap Estimation of the Distribution of Cumulative Outdoor Degradation

Victor Chan
Western Washington University

Soumendra N. Lahiri
Iowa State University

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Block Bootstrap Estimation of the Distribution of Cumulative Outdoor Degradation

Abstract

An interesting prediction problem involving degradation of materials exposed to outdoor environments (weathering) is the estimation of the distribution of future cumulative degradation using small- to moderate-size degradation data sets. This distribution, which is assumed to arise as a result of the uncertainty/variability in the weather, can be expressed mathematically as the distribution of the sum of a periodic dependent time series, and is approximately normal by the Central Limit Theorem. The estimation of this distribution is thus equivalent to estimating the mean and the variance of the distribution. In this paper, we propose a block-bootstrap-based approach for the estimation and a novel technique to estimate the variance of the distribution. An example involving the degradation of a solar reflector material is provided. We also present the results of a simulation study to show the efficacy of the proposed estimators. A procedure for constructing an approximate confidence interval for the probability of failure is also given.

Keywords

periodic dependent time series, normal distribution, Central Limit Theorem

Disciplines

Statistics and Probability

Comments

This preprint was published in *Technometrics* 46 (2004): 215–224, doi:10.1198/004017004000000266.

Block Bootstrap Estimation of the Distribution of Cumulative Outdoor Degradation

Victor Chan
Department of Mathematics
Western Washington University
Bellingham, WA 98225

Soumendra N. Lahiri
Department of Statistics
Iowa State University
Ames, IA 50011

William Q. Meeker
Department of Statistics
Iowa State University
Ames, IA 50011

October 28, 2003

Abstract

An interesting prediction problem involving degradation of materials exposed to outdoor environments (weathering) is the estimation of the distribution of future cumulative degradation using small- to moderate-size degradation data sets. This distribution, which is assumed to arise as a result of the uncertainty/variability in the weather, can be expressed mathematically as the distribution of the sum of a periodic dependent time series, and is approximately normal by the Central Limit Theorem. The estimation of this distribution is thus equivalent to estimating the mean and the variance of the distribution. In this paper, we propose a block-bootstrap-based approach for the estimation and a novel technique to estimate the variance of the distribution. An example involving the degradation of a solar reflector material is provided. We also present the results of a simulation study to show the efficacy of the proposed estimators. A procedure for constructing an approximate confidence interval for the probability of failure is also given.

Key words: periodic dependent time series, normal distribution, Central Limit Theorem

1 Introduction

Reliability of products or materials exposed continuously to the outdoor environment is of great interest to, for example, the manufacturers of paints and coatings. One major problem in this area is to predict the amount of cumulative degradation of a particular paint or coating formulation that will occur in r years under a variable environment, with the assumption of a deterministic degradation behavior if the actual environmental conditions were known or fixed (as, for example, in a laboratory). An answer to the prediction problem is provided by the distribution of cumulative degradation in r years, where the distribution arises as a result of the uncertainty/variability in the environment (we assume the variability in the environment dominates the material or unit-to-unit variability, as is true in some applications). Such a distribution can be used, for example, to estimate the probability of failure of the material in r years, where failure is defined as the event in which the cumulative degradation exceeds some prespecified level.

One approach to estimate the future degradation distribution of cumulative degradation is by means of parametric modeling of the degradation data, which can be either observed or computed using a physical/chemical model of degradation and weather data. The degradation data are in the form of a time series of daily accumulated amount of degradation. One such approach is discussed in Chan and Meeker (2001). However, a nonparametric approach to estimate the distribution may be more desirable due to its flexibility. Because it does not require finding adequate parametric time series models to fit the observations, it is often able to handle more complicated types of data.

This paper presents a block-bootstrap-based method to estimate the distribution of cumulative degradation in r years, using a time series consisting of daily degradation observations. Such a time series exhibits not only correlation from day to day, but also periodicity as a result of seasonality in the daily weather. Therefore, the method must be able to handle *periodic* dependent data. Despite the volume of literature on the use of block bootstrap for time series data, we know of no work that deals with periodic dependent data.

As discussed in Section 2.1 below, the distribution of cumulative degradation is approximately normal/Gaussian. Because of this, the block bootstrap method essentially estimates the mean and the variance of the distribution. We investigate the statistical properties of the bootstrap estimators of the mean and the variance in Section 4. Based on these results and the periodic structure of the time series, we develop a new and improved estimator of the variance by making use of multiple bootstrap variance estimators. In Section 5, we describe the algorithm for finding the estimator and provide the details behind its formulation. In Section 6, we present the results of a simulation study to show the efficacy of the proposed estimator. We also describe, in Section 7, a construction scheme for an approximate confidence interval for the probability of failure, along with the results of a simulation study to determine the coverage probabilities. An advantage of our approach is that the results and the methodology developed here are applicable to nonstationary periodic time series data without any specific parametric model assumptions. Furthermore, because of the linearity of the sums, the estimators and confidence intervals proposed here can be computed without any data resampling, and hence, are not computationally demanding.

We conclude this section with a brief literature review. There is a fairly large amount of work on block bootstrap methods applicable to correlated data. Block bootstrap methods for time series data have been introduced by Hall (1985), Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992, 1994), among others. Review and general overview of the block bootstrap can be found in Leger et al. (1992), in Chapter 8 of Davison and Hinckley (1999), and in Bühlmann (2002). Park et al. (1999) and (2001) discuss threshold bootstrap and threshold jackknife, which provide other approaches to resampling serially correlated data. For a book-length treatment of the different bootstrap methods for dependent data, see Lahiri (2003). Recent work (including physical and chemical studies) related to the degradation of organic coatings under weathering conditions can be found in Bauer and Martin (1999, 2002).

2 Background

2.1 Distribution of Cumulative Degradation in r Years

Let W_i denote the degradation accumulated in day i. Due to the seasonal changes in the intensity of the weathering agents such as temperature and ultraviolet radiation, the time series of daily degradation W_i exhibits a periodicity of $\rho = 365$ days, ignoring leap years for simplicity. The problem of estimating the cumulative degradation in r years is equivalent to estimating the distribution of $\sum_{i=1}^{r\rho} W_i$, i.e., estimating the distribution of the sum of *periodic dependent* time series in r periods. By the Central Limit Theorem for dependent processes, $\sum_{i=1}^{r\rho} W_i$ has approximately a normal distribution when r is large. Therefore, estimating the distribution of the sum essentially reduces to estimating two parameters, namely the mean $\mathrm{E}(\sum_{i=1}^{r\rho} W_i)$ and the variance $\mathrm{Var}(\sum_{i=1}^{r\rho} W_i)$.

If a large amount of data were available (i.e., many seasonal periods), then estimating the two parameters would be straightforward. Suppose, for simplicity, we have data for m periods in the form of a time series, where m is some large positive integer. Let $\{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{m\rho}\}$ denote the observed time series. (In this paper, we make a distinction between observed daily degradation, represented by $\mathcal{W}_i$, and future daily degradation, denoted by W_i, even though the two random variables have the same distribution.) In the context of degradation, the time series represents daily observations of degradation for m years. Let $T_k = \sum_{i=(k-1)\rho+1}^{k\rho} \mathcal{W}_i$ be the sum for the kth period, i.e., total amount of degradation accumulated in the kth year. If m is large, then one straightforward way to estimate the mean $\mathrm{E}(\sum_{i=1}^{r\rho} W_i)$ and variance $\mathrm{Var}(\sum_{i=1}^{r\rho} W_i)$

is by simply using the sample mean and sample variance of $T_1, T_2, \ldots, T_m$, (i.e., $\sum_{k=1}^{m} T_k/m$ and $\sum_{k=1}^{m}(T_k - \overline{T})^2/(m-1)$) multiplied by r, assuming independence of $T_1, T_2, \ldots, T_m$ up to a negligible error (which can be assumed here). For example, to estimate the mean and variance of the distribution of cumulative degradation in one year, use the sample mean and sample variance of annual degradation data.

However, long series of observations or realizations in reliability studies are rare, and this is especially true of degradation observations. For example, at this point there is less than ten years' worth of daily degradation data that have been observed or computed indirectly using a model and weather data for paints and coatings exposed to the outdoor environment. For small- to moderate-size data sets, i.e., for m *not* large, the sample variance does *not* provide a good estimate of the variance because of its *large* sampling variability. As such, a different method would have to be used to estimate this quantity with greater precision. In Section 5, we propose a method based on block bootstrap that provides a more precise estimate of the variance.

2.2 Properties of Periodic Dependent Time Series

Before describing the block bootstrap approach, we first make a formal definition of W_i and give a real-world example.

Let $\mathbf{Z}$ and $\mathbf{N}$ denote the set of all integers and the set of positive integers, respectively. Let $\{W_i : i \in \mathbf{Z}\}$ be a time series of random variables with the following properties:

1. W_i is periodic with period $\rho \in \mathbf{N}$, so that
$$W_i \overset{\mathrm{d}}{=} W_{i+\rho} \quad \text{for all} \quad i,$$
where $\overset{\mathrm{d}}{=}$ denotes equality in distribution.
2. W_i is δ-dependent, i.e., W_i and W_{i+j} are independent if $|j| > \delta > 0$.
3. Nonnegative covariance between any two W_i's, i.e., $\mathrm{Cov}(W_i, W_{i+j}) \geq 0$ for $|j| \leq \delta$.
4. δ is much smaller than period ρ, i.e., $\delta \ll \rho$.

Our interest is in estimating the distribution of $\sum_{i=1}^{r\rho} W_i$, where $r \in \mathbf{N}$.

2.3 Example

Predicted Daily Degradation of Solar Reflector Material at Boulder, CO from November 1995 to August 2000

An example of a time series with the above-mentioned properties is the predicted daily degradation (PDD) of solar reflector material. Solar reflector materials degrade over time from continuous exposure to weather elements such as ultraviolet (UV) radiation. The degradation is characterized by loss in performance, where performance is measured in terms of percentage of light reflected from the surface of the material. The predicted daily degradation (degradation accumulated within a 24-hour period) of such materials exposed to the outdoor environment is computed by integrating the degradation rate model of the material with respect to time using actual weather data; the instantaneous degradation rate depends on weather variables temperature, UV radiation, and humidity. See Chan and Meeker (2001) for more details. The development and parameter estimation of the degradation rate model and the data for the solar reflector material are given in Jorgensen et al. (1996).

Figure 1 shows the PDD of a solar reflector material called ECP-300A corresponding to the weather at Boulder, Colorado from November 1995 to August 2000. The period in this case is $\rho = 365$, ignoring leap years for simplicity. Because the daily degradation is induced by the outdoor weather and it is reasonable to assume noncorrelation in the weather patterns between any two different days that are far apart (say, 10 days), the δ-dependence condition is satisfied. The assumption is also supported by autocorrelation analysis of the daily degradation data.

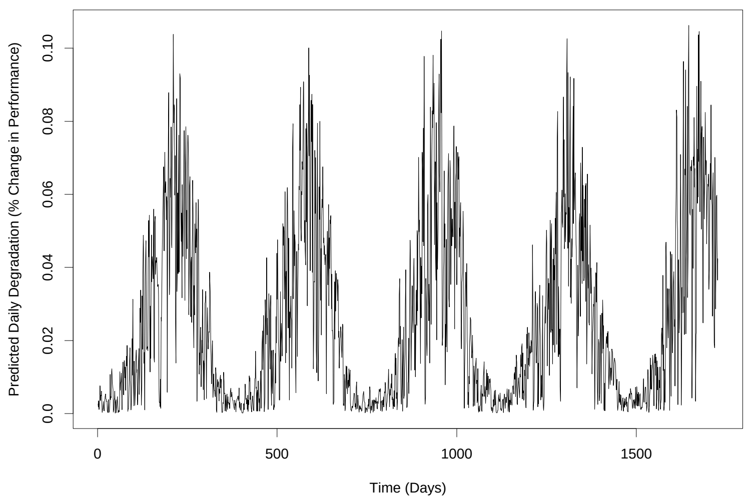

Figure 1: Predicted daily degradation in Boulder, CO. from November 1995 to August 2000 for a solar reflector material.

Using W_i to denote the daily degradation for day i, a problem of interest to reliability statisticians and engineers is the estimation of the distribution of cumulative degradation of the solar reflector material in r years, $\sum_{i=1}^{365r} W_i$.

3 Block Bootstrap with Periodic Data

The basic idea is to partition the period into small time intervals, and divide the periodic data into blocks that correspond temporally to each of the intervals. Independent sampling with replacement from the blocks from each interval is then used to form a new time series of bootstrap data. For example, in the context of degradation of materials subjected to weathering, the year consisting of 365 days is partitioned into intervals of, say, 10 days, so that the days from January 1 through January 10 will form the first interval, January 11 through January 20 the second interval and so on (the last interval will consist of only 5 days). The time series data of daily degradation observations are broken up into blocks that fit the corresponding intervals. Thus, a data set of length 5 years starting from January 1 of the first year through December 31 of the fifth year will result in 5 blocks of observations for each of the intervals. A block-bootstrapped time series of length, say three years, is formed by sampling three times with replacement from the 5 blocks for each of the intervals in chronological order so that the sampled blocks, when placed end-to-end, will form three years of continuous bootstrapped observations.

The following discussion describes the idea within a mathematical framework. For ease of presentation, we choose a different notation for the observed realization of W_i, which we denote as $\mathcal{W}_i$, to distinguish it from W_i whose cumulative-sum distribution we wish to estimate. Let $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{m\rho}$ be the observed time series of length $m \times \rho$, i.e., consisting of m periods, where

$$\mathcal{W}_i, \mathcal{W}_{\rho+i}, \mathcal{W}_{2\rho+i}, \ldots \stackrel{\mathrm{d}}{=} W_i$$

and are independent, because of the δ-dependence condition. If the length of the observed time series is not a multiple of period ρ, then the blocking scheme outlined below will still apply with a minor modification. This will be described at the end of the outlined scheme.

The block bootstrap scheme consists of the following steps:

1. Partition the period into J intervals, each of length l. If ρ is not divisible by l, let every interval, except for the last one, be of size l. For simplicity, however, we will assume that $J \times l = \rho$.

2. Let $Y_1^*, Y_2^*, \ldots, Y_J^*$ be random variables defined as

$$\begin{aligned}
Y_1^* &\sim \text{Unif}\left\{\sum_{i=1}^{l} \mathcal{W}_i,\ \sum_{i=\rho+1}^{\rho+l} \mathcal{W}_i,\quad \ldots,\quad \sum_{i=(m-1)\rho+1}^{(m-1)\rho+l} \mathcal{W}_i\right\}, \\
Y_2^* &\sim \text{Unif}\left\{\sum_{i=l+1}^{2l} \mathcal{W}_i,\ \sum_{i=\rho+l+1}^{\rho+2l} \mathcal{W}_i,\quad \ldots,\quad \sum_{i=(m-1)\rho+l+1}^{(m-1)\rho+2l} \mathcal{W}_i\right\}, \\
\vdots \quad & \quad \vdots \\
Y_J^* &\sim \text{Unif}\left\{\sum_{i=(J-1)l+1}^{\rho} \mathcal{W}_i,\ \sum_{i=\rho+(J-1)l+1}^{2\rho} \mathcal{W}_i,\quad \ldots,\quad \sum_{i=(m-1)\rho+(J-1)l+1}^{m\rho} \mathcal{W}_i\right\},
\end{aligned}$$

where $Y \sim \text{Unif}\{\xi_1, \xi_2, \ldots, \xi_n\}$ implies sampling such that $\Pr(Y = \xi_i) = 1/n$, i.e., the distribution of Y puts probability masses uniformly among its finite set of values. In the context of block bootstrap, the outcomes of random variables $Y_1^*, Y_2^*, \ldots, Y_J^*$ could be obtained as follows: resample with replacement from the set of blocks of length l in the observed time series that matches the partitioned interval within the period, and sum the elements in the sampled block. For example, the blocks $\{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_l\}, \{\mathcal{W}_{\rho+1}, \mathcal{W}_{\rho+2}, \ldots, \mathcal{W}_{\rho+l}\}, \ldots, \{\mathcal{W}_{(m-1)\rho+1}, \mathcal{W}_{(m-1)\rho+2}, \ldots, \mathcal{W}_{(m-1)\rho+l}\}$ from the observed time series correspond to the *first* interval of length l within the period. One block from this set of m blocks is selected at random and its elements summed up, yielding an outcome of Y_1^*. This is done in the same way for the sets of blocks corresponding to the other intervals.

3. Use the distribution of $\sum_{j=1}^{rJ} Y_j^*$ to estimate the distribution of $\sum_{i=1}^{r\rho} W_i$, where the Y_j^*'s are *independent* and

$$Y_{J+j}^*, Y_{2J+j}^*, Y_{3J+j}^*, \ldots \overset{\mathrm{d}}{=} Y_j^*, \qquad 1 \leq j \leq J. \tag{1}$$

If the length of the observed time series is not a multiple of ρ, most, if not, all of the $\mathcal{W}_i$'s within the incomplete period can still be incorporated into the appropriate blocking variables Y_j^* as defined in step 2 above, with the consequence that some of the blocking variables Y_j^* (corresponding to the lower end of the period) may have one fewer element from which to resample in comparison with the other Y_j^*'s. This does not affect in any way the block bootstrap procedure.

For simplicity of presentation and mathematical convenience, we will focus only on time series observations with length of exact multiples of the period ρ, i.e., on $\{\mathcal{W}_i :\ i = 1, 2, \ldots, m\rho\}$. For practical estimation, m will have to be at least 4. For m smaller than 4, parametric models requiring strong assumptions on the joint distribution of W_i will have to be used.

4 Block Bootstrap Estimators of Mean and Variance

As mentioned in Section 2.1, the distribution of the sum of a periodic dependent time series is approximately normal. Using the block bootstrap approach with a specified block length l to estimate the distribution, the distribution of the bootstrapped sum is also approximately normal, by the Central Limit Theorem. That is, when r is large,

$$\sum_{j=1}^{rJ} Y_j^* \dot{\sim} \ N\Big(r\,\mathrm{E}(\sum_{j=1}^{J} Y_j^*),\ r\,\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)\Big), \tag{2}$$

because of the independence of Y_j^* for all j. Here $\sum_{j=1}^{J} Y_j^*$ represents the bootstrapped sum for one period, and note that Y_j^* and J have implicit dependence on block length l.

Because of the normality of both the actual and the bootstrapped distributions, the block bootstrap approach essentially estimates the mean, $\mathrm{E}(\sum_{i=1}^{r\rho} W_i)$, and the variance, $\mathrm{Var}(\sum_{i=1}^{r\rho} W_i)$, using the estimators $r\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$ and $r\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$. Without loss of generality, we can ignore the factor r in the estimators, and will do so for the rest of discussion in this paper.

Insofar as estimating the distribution of $\sum_{i=1}^{r\rho} W_i$ using the block bootstrap approach, the questions of interest regarding the two estimators

- $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$
- $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$

are how good these estimators are, and how they vary with the choice of l. The answers will be given in the next two sections.

Note that although these two bootstrap estimators can be obtained from generating bootstrap samples, an easier way is to compute them directly from the data. This is possible because the estimators are the mean and variance of a sum of independent random variables Y_j^* with simple discrete distributions. The expressions for these two estimators are given in the Appendix.

4.1 Bootstrap Estimator of the Mean

For a data set consisting of exactly m periods of observations, the block bootstrap estimator of the mean of the sum for one period, $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$, remains the same for any choice of block length $l \geq 1$. This is because the expectation is a linear operator and therefore is not affected by the way a sum is partitioned.

The properties of the estimator $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$ are as follows:

1. It is an unbiased estimator of $\mathrm{E}(\sum_{i=1}^{\rho} W_i)$.

2. It is equal to $(1/m)\sum_{k=1}^{m} T_k$, the sample mean of the periodic sums, where, as defined in Section 2.1, $T_k = \sum_{i=(k-1)\rho+1}^{k\rho} \mathcal{W}_i$ is the sum of the observations within period k. Thus, in our application $T_1, T_2, \ldots, T_m$ represent the annual cumulative degradation for m years, and the estimator is identical to the average of these m values.

3. Its variance is approximately equal to $(1/m)\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$, the true variance divided by the number of periods of the time series data.

These properties are proved in the Appendix.

4.2 Bootstrap Estimator of the Variance

Unlike the estimator of the mean $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$, the variance estimator $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ is strongly affected by the choice of block length l. This raises the question of how to choose l so that the estimator gives a "good" estimate of $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$. It turns out that the estimator $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ is biased regardless of the block length l. As in the case of a *stationary* time series, as the block length l increases, the bias of the bootstrap variance estimator decreases while its sampling variability increases. Thus, there exists an optimal value of the block length that minimizes the mean square error asymptotically. The problem of choosing this optimal block size empirically is, however, difficult. The existing methods, like the Hall, Horowitz and Jing (1995) method, based on subsampling, do not work here because of the nonstationarity of the time series and also because of the relatively small number of complete data periods or cycles available. This prompted us to consider an alternative method of estimating the variance for periodic time series data that is less sensitive to the choice of the block length. In the next section, we present an alternative method of estimating the variance $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$. The resulting estimator is approximately unbiased and approximately consistent.

5 Block-Length Regression Estimator of the Variance

5.1 Motivation

We first show why the estimator $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ is biased. For this purpose and also for reference later on, we introduce the variables

$$Y_j = \sum_{i=(j-1)l+1}^{\min(jl,\rho)} W_i,\ j = 1, 2, \ldots, J,$$

where $J = \lceil \rho/l \rceil$ is the smallest integer not less than ρ/l. Thus, Y_j is the sum of the time series over interval j. Note that if $l \geq \delta$, where δ refers to the δ-dependence as defined in Section

2.2, Y_j and Y_k are independent whenever $|j-k|>1$. The index for Y_j can be extended to all positive integers by imposing the periodic property

$$Y_{kJ+j} \stackrel{\mathrm{d}}{=} Y_j \qquad 1 \leq j \leq J, \quad k \in \mathbf{N}.$$

Note that Y_j^* is obtained by an equiprobable sampling of the realizations of the random variable Y_j.

In the Appendix, we show that $\mathrm{E}[\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)] = (1-1/m)\sum_{j=1}^{J}\mathrm{Var}(Y_j)$. Now because $\sum_{j=1}^{J}\mathrm{Var}(Y_j)$ is equal to $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$ minus some positive correlation terms (except when $J=1$, in which case the two quantities are equal), $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ is also a biased estimator of $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$.

To present our proposed method to obtain an estimate of the variance $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$ that is approximately unbiased and approximately consistent, we first give a result.

Result 1 *Under the conditions for W_i given in Section 2.2,*

$$\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - \sum_{j=1}^{J}\mathrm{Var}(Y_j) = 2\sum_{j=1}^{J-1}\sum_{i=jl-\delta+1}^{jl}\sum_{k=jl-i+1}^{\delta}\mathrm{Cov}(W_i, W_{i+k}) \tag{3}$$

for $J>1$ and $\delta \leq l \leq \rho/2$. Assuming that the average of the covariance terms in the sum on the right-hand side is approximately the same for all l over some range $[l_1, l_2]$ where $\delta \leq l_1 < l_2 \leq \rho/2$, then

$$\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - \sum_{j=1}^{J}\mathrm{Var}(Y_j) \quad \dot{\propto} \quad (\rho/l) - 1$$

over that range.

Proof Given in the Appendix. □

The right side of Equation 3 in Result 1 expresses the sum of covariance terms between pairs of variables across independent non-overlapping blocks, constrained by the δ-dependence condition. The assumption that the average of the covariance terms is approximately the same over some range of block length l greater than δ is reasonable for actual data on intuitive grounds.

Based on Result 1 and the fact that $\mathrm{E}[\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)] = (1-1/m)\sum_{j=1}^{J}\mathrm{Var}(Y_j)$, we conclude that for l in the range for which the assumption in Result 1 is true,

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)\right] &\approx (1-\frac{1}{m})\left[\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - c(\frac{\rho}{l}-1)\right] \\
&\approx (1-\frac{1}{m})\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - c(1-\frac{1}{m})(\frac{\rho}{l}-1), \qquad (4)
\end{aligned}$$

where c is some constant. The estimator of the variance can therefore be rewritten as

$$\text{Var}(\sum_{j=1}^{J} Y_j^*) \approx (1 - \frac{1}{m})\text{Var}(\sum_{i=1}^{\rho} W_i) - c(1 - \frac{1}{m})(\frac{\rho}{l} - 1) + \epsilon_m(l), \tag{5}$$

where $\epsilon_m(l)$ is an error term that depends on both the amount of data (reflected in m, the number of periods of observation) and the block length l, and has mean 0.

Equation 5 shows that the block bootstrap estimator of the variance is approximately a linear function of $\rho/l-1$ plus a zero-mean error term, with an intercept at $(1-1/m)\text{Var}(\sum_{i=1}^{\rho} W_i)$. This suggests that one can estimate $(1 - 1/m)\text{Var}(\sum_{i=1}^{\rho} W_i)$ by fitting a simple linear regression of the form $\text{Var}(\sum_{j=1}^{J} Y_j^*) = \hat{\beta}_0 + \hat{\beta}_1(\rho/l - 1)$, where $\rho/l - 1$ is the independent variable; $[m/(m-1)]\hat{\beta}_0$ would provide an approximately unbiased estimator of $\text{Var}(\sum_{i=1}^{\rho} W_i)$. The estimator is also approximately consistent because $\text{Var}(\sum_{j=1}^{J} Y_j^*)/\sum_{j=1}^{J}\text{Var}(Y_j) \xrightarrow{p} 1$ as the length of data $m \to \infty$. We shall call the estimator the block-length regression (BLR) estimator of the variance.

5.2 Algorithm for finding the BLR estimator

Our proposed method to obtain an approximately unbiased and approximately consistent estimator of $\text{Var}(\sum_{i=1}^{\rho} W_i)$ can be summarized as follows:

1. Obtain the block bootstrap estimator $\text{Var}(\sum_{j=1}^{J} Y_j^*)$ for block lengths of, say, $l = \delta, \delta + 1, \ldots, 3\delta$, or for a range of block lengths for which the assumption in Result 1 is satisfied.
2. Fit a simple linear regression with $[m/(m-1)]\text{Var}(\sum_{j=1}^{J} Y_j^*)$ as the response and $(\rho/l-1)$ as the independent variable, over the range of l.
3. The regression estimate of the intercept $\hat{\beta}_0$ provides an estimate of $\text{Var}(\sum_{i=1}^{\rho} W_i)$, and hence $\hat{\beta}_0$ is our proposed BLR variance estimator.

5.3 Example

As an illustration, we will use the data for daily degradation of solar reflector material at Boulder, Colorado described in Example 2.3.

Table 1 provides the values of the bootstrap mean estimator $\text{E}(\sum_{j=1}^{J} Y_j^*)$ and the bootstrap variance estimator $\text{Var}(\sum_{j=1}^{J} Y_j^*)$ as a function of block length l (up to $l = 30$) based on the solar reflector daily degradation data described in Example 2.3. The amount of data is $m \approx 5$ because the length of the time series is close to 5 years. The values of $\text{E}(\sum_{j=1}^{J} Y_j^*)$ are approximately constant across the block sizes; they would all be the same if the length of data were exactly 5 years. The estimate of $\text{E}(\sum_{i=1}^{\rho} W_i)$, the mean of the distribution of the total cumulative degradation for one year, based on our block-bootstrap approach is, therefore, about 9.056.

The values of the bootstrap variance estimator $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ are calculated for block lengths $l = 1, 2, 3, \ldots, 60$ (so that $J = 365, 183, 122, \ldots, 6$), and are plotted in Figure 2. The first 30 values are given in Table 1. Figure 2 shows that $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ increases rapidly with block size l and then levels off. This is due to the fact that the larger the block length l, the fewer the (positive) covariance terms being excluded in the estimate of the one-year variance, and hence the increase. The variability of $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ also increases with l, as seen by the greater fluctuations in the block lengths $l > 30$.

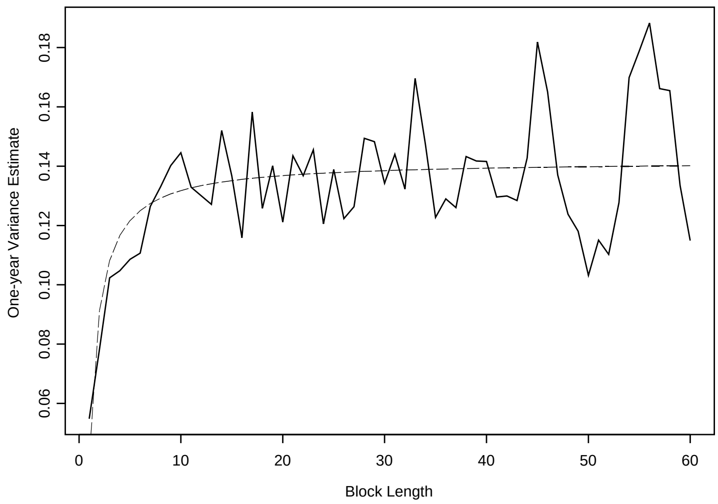

Figure 2: Plot of the one-year variance estimate $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ versus block size l for daily degradation of solar reflector material at Boulder CO. The dashed line corresponds to the fitted regression model.

Following the algorithm outlined in Section 5.2, we fit a simple linear regression to the data points $(\frac{5}{4}\mathrm{Var}(\sum_{j=1}^{J} Y_j^*), (365/l) - 1)$, and used the estimated value of the intercept as an estimate of the variance of the sum of cumulative degradation for one year. The dependence structure of daily degradation is assumed to have a lag of at most 10 days, i.e., the δ-dependence is such that $\delta = 10$. This is reasonable because presumably the weather on any one day will not affect the outcome of the weather at the same location ten days later, and this independence carries over to daily degradation because the weather is responsible for degradation. The range of block lengths used in the regression is from $l = 10$ through 40. The estimate of the variance of the sum of cumulative degradation for one year $\mathrm{Var}(\sum_{i=1}^{365} W_i)$ using the BLR estimator is 0.174. Other choices of range of block lengths yield estimates of the variance that differ by at most a few percent; for example, the range $l = 10, \ldots, 60$ gives 0.178, and even the full range

Table 1: Block bootstrap estimates of the mean $E(\sum_{j=1}^{J} Y_j^*)$ and the variance $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ as a function of block size for daily degradation of solar reflector material at Boulder CO.

Block Size	Mean	Variance
1	9.0560	0.0548
2	9.0560	0.0783
3	9.0558	0.1023
4	9.0560	0.1047
5	9.0560	0.1086
6	9.0559	0.1107
7	9.0566	0.1264
8	9.0558	0.1331
9	9.0562	0.1402
10	9.0559	0.1446
11	9.0572	0.1329
12	9.0547	0.1300
13	9.0563	0.1271
14	9.0552	0.1520
15	9.0549	0.1367
16	9.0554	0.1159
17	9.0567	0.1582
18	9.0562	0.1258
19	9.0728	0.1402
20	9.0559	0.1212
21	9.0565	0.1435
22	9.0542	0.1368
23	9.0613	0.1455
24	9.0541	0.1206
25	9.0658	0.1389
26	9.0563	0.1223
27	9.0484	0.1263
28	9.0728	0.1494
29	9.0620	0.1483
30	9.0541	0.1343

$l = 1, \ldots, 60$ gives 0.177.

The dashed line in Figure 2 corresponds to the fitted values of the variance from the simple linear regression. The fitted values for $l = 1, 2, \ldots, 9$ are extrapolations from the regression estimates.

6 Simulation Study

This section uses simulation to check the properties of our bootstrap-based estimators. In particular we compare the efficiency of our proposed BLR estimator of the variance of cumulative degradation with the sample variance based on annual cumulative degradation. The model used to generate simulated data is the time series model developed by Chan and Meeker (2001) to fit the predicted daily degradation data for the solar reflector material.

The model corresponding to Boulder, Colorado yields an annual cumulative degradation that has approximately a normal distribution with a mean of 9.054 and a variance of 0.2073. We considered three cases of data length: $m = 5, 10, 20$ years. For each case, 2000 time series were generated from the model. For each time series, we calculated the estimate of the mean, the estimate of variance (using our proposed method) and the sample variance for the cumulative degradation in one year, obtaining 2000 values of each of the three statistics for each m. The range of block lengths used in computing the BLR variance estimate was between 10 and 40, inclusive. The mean and variance of each statistic were computed and are presented in Table 2.

Table 2: Results of 2000 simulations from model based on Boulder data

	Bootstrap Estimator of $E(\sum_{i=1}^{365} W_i)$		BLR Estimator of $Var(\sum_{i=1}^{365} W_i)$		Sample Variance	
m	Mean	Variance	Mean	Variance	Mean	Variance
5	9.051	0.04409	0.2054	0.003847	0.2074	0.02234
10	9.055	0.02007	0.2073	0.001755	0.2096	0.01025
20	9.055	0.01049	0.2069	0.000818	0.2081	0.00460
	True value = 9.054		True value = 0.2073		True value = 0.2073	

The table shows the following:

- The means of the three estimators are close to the true values being estimated. In particular, the mean of our proposed BLR estimator of the variance is close to the true variance (0.2073) in each case, validating the approximate unbiasedness of our proposed estimator. Also the means of the BLR variance estimator are no larger than the true value, i.e., biased low, a typical characteristic of bootstrap variance estimation for positively correlated data.

- The variances of the estimators decrease with m, the number of years' worth of data, as expected. Based on the simulations, the variance of the BLR variance estimator appears to decrease with a factor of approximately $1/m$.
- The variance of the estimator of the mean $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$ is approximately equal to the true variance (0.2073) divided by the length of data m in years, as predicted by theory.
- Comparing the variance of the proposed BLR estimator of the variance with that of the sample variance, we see that the BLR estimator is much more efficient than the sample variance, with a relative efficiency (the variance ratio between two estimators) of close to 6 in each case.

The results of the simulation provide strong support for the block bootstrap approach to estimating the distribution of cumulative degradation. In particular, the efficiency of the BLR variance estimator translates to a more precise estimate of the variance.

As a further illustration, we present the results of another simulation study based on the model for daily degradation of solar reflector material in Fort Peck, Montana in Table 3. As in the previous study, 2000 simulations were generated for each $m = 5, 10, 20$. The true mean and variance of the cumulative degradation for one year are 6.018 and 0.1634. The relative efficiency of our BLR estimator of the variance with respect to the sample variance is about 4 in each case.

Table 3: Results of 2000 simulations from model based on Ft. Peck data

	Bootstrap Estimator of $\mathrm{E}(\sum_{i=1}^{365} W_i)$		BLR Estimator of $\mathrm{Var}(\sum_{i=1}^{365} W_i)$		Sample Variance	
m	Mean	Variance	Mean	Variance	Mean	Variance
5	6.021	0.03382	0.1618	0.003184	0.1641	0.01376
10	6.017	0.01617	0.1620	0.001444	0.1649	0.00579
20	6.014	0.00803	0.1611	0.000684	0.1638	0.00273
	True value = 6.018		True value = 0.1634		True value = 0.1634	

7 Approximate Confidence Interval for Probability of Failure

Failure is defined as the event in which the cumulative degradation reaches a specified failure level. In other words, failure is said to occur when $\sum_{i=1}^{r\rho} W_i > D_f$, where D_f is the failure level. In applications such as reliability related to degradation of materials or products, the probability of failure in r years of such materials would be a quantity of interest.

Because the distribution of the cumulative degradation of materials subject to outdoor weathering in r years is approximately normal, the uncertainty in the exact values of the mean and the variance of the normal distribution (which stems from the uncertainty/variability in the weather) is responsible for the uncertainty in the probability of failure. Therefore, if a confidence interval for the mean and a good estimate of the variance can be obtained, we can also construct a confidence interval for the probability of failure.

The estimator $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$ of the mean level of cumulative degradation in a year is essentially the sum of a large number of (correlated) random variables (see Result 4), and so by the Central Limit Theorem, it has approximately a normal distribution. The mean of this distribution is $\mathrm{E}(\sum_{i=1}^{\rho} W_i)$, and the variance is approximately $(1/m)\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$ (see Result 2). Because we can get a reasonable estimate of the variance using our proposed approach, we can therefore construct an approximate confidence interval for $\mathrm{E}(\sum_{i=1}^{\rho} W_i)$.

The formula for an approximate $100(1-\alpha)\%$ confidence interval for $\mathrm{E}(\sum_{i=1}^{r\rho} W_i)$, the mean of the cumulative degradation in r years, is given by

$$r\mathrm{E}(\sum_{j=1}^{J} Y_j^*) \pm z_{\alpha/2}\sqrt{\frac{r^2}{m}\mathrm{Var}(\sum_{i=1}^{\rho} W_i)},$$

where $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$ is unknown. The BLR variance estimator provides a good estimate of this unknown variance.

The formula for the confidence interval for the probability of failure $\Pr(\sum_{i=1}^{r\rho} W_i > D_f)$ is given by

$$\left[1-\Phi\left(\frac{D_f - lb}{\sqrt{r\mathrm{Var}(\sum_{i=1}^{\rho} W_i)}}\right), \quad 1-\Phi\left(\frac{D_f - ub}{\sqrt{r\mathrm{Var}(\sum_{i=1}^{\rho} W_i)}}\right)\right], \tag{6}$$

where lb and ub are the lower and upper bounds of the $100(1-\alpha)\%$ confidence interval for $\mathrm{E}(\sum_{i=1}^{r\rho} W_i)$. As in the previous confidence interval formula, the BLR estimate of the variance can be used to replace the unknown variance $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$.

To assess the accuracy of the confidence interval for the probability of failure, we made use of the 2000 simulated time series for each of the three different data lengths $m = 5, 10, 20$ years from the model for Boulder, Colorado discussed in the last section. For each simulated time series, we computed the confidence interval using Equation 6 for the probability of failure in 5 years based on the failure level at $D_f = 46$. The coverage probability (the probability that the confidence interval constructed from a specified procedure will capture the quantity being estimated, which in this case is the probability of failure) of the confidence interval for the true probability of failure in 5 years (which is 0.2367) was computed for each of the different combinations of confidence level and data length. The results are given in Table 4.

It is apparent that the approximate confidence interval construction gives a slightly lower (i.e., anticonservative) confidence level than nominal in every case. Also as the size of data reflected in m increases from 5 years to 10 years, the coverage probabilities improve. There appears to be no significant change from $m = 10$ years to $m = 20$.

Table 4: Coverage Probabilities using 2000 Simulations

	Confidence Level		
m	95%	90%	80%
5	0.930	0.870	0.781
10	0.945	0.894	0.789
20	0.944	0.891	0.787

8 Concluding Remarks

In this paper we introduce a block-bootstrap approach to estimate the cumulative degradation of materials subjected to outdoor weathering. We also propose a method to obtain an approximately unbiased and approximately consistent estimator (which we call the BLR estimator) of the variance of the cumulative degradation distribution, and we have shown that it is more efficient than the sample variance for small to moderate-size data sets. Two advantages of the method over traditional block bootstrap procedures are that generating bootstrap samples from data is not required, and that the optimal block length l does not have to be determined. Although the method proposed in the paper was motivated by the degradation of materials exposed to outdoor weathering, other problems involving estimating the distribution of the sum of periodic dependent time series satisfying the conditions listed in Section 2.2 are amenable to the proposed methodology.

The block-bootstrap approach and the method to estimate the variance of the cumulative degradation proposed in this paper suggest a number of interesting problems for further research. Some of these are:

1. Determine the theoretically best bounds of block length l to obtain the most efficient BLR estimator of the variance using our proposed procedure, assuming that the assumptions in Result 1 are met. In the paper, we used the range of l between 10 and 40, and found that, at least for the degradation of solar reflector material, the range of l is not a major issue because changing the range does not alter the results by much. However, simulations suggest that the classic trade-off between the variance and the bias of the estimator comes into play when the range of l is varied. Minimization of mean square error (MSE) of the estimator for different bounds of l may be called for.

2. An understanding of how the error term $\epsilon_m(l)$ in Equation 5 varies with data size m and block length l will help determine the properties of this error term, such as its variance.

3. The block bootstrap variance estimator $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ increases with block length l, as shown in Figure 2. Hence, by the generalized Gauss-Markov theorem, the generalized least squares method yields a better estimate of the variance than the proposed BLR variance estimator. Implementing the generalized least squares, however, requires knowl-

edge of the covariance structure of the error term $\epsilon_m(l)$. The challenge here is to estimate the covariance matrix of $\epsilon_m(l)$.

4. Our method assumes a trend-free periodic dependent time series. For other applications where there is a trend or drift, our method would require some modifications.

Acknowledgements We would like thank the editor, the associate editor and an anonymous referee for reviewing our manuscript and for their valuable comments and suggestions. This research was partially supported by NSF grant number DMS 0072571, and by 2002 Summer Research Grant, Western Washington University.

9 Appendix

Proof of Result 1

For simplicity of exposition, let $Jl = \rho$. We first note that

$$\mathrm{Var}(\sum_{i=1}^{\rho} W_i) = \sum_{i=1}^{\rho} \mathrm{Var}(W_i) + 2\sum_{i=1}^{\rho-1}\sum_{k=1}^{\rho-i} \mathrm{Cov}(W_i, W_{i+k}).$$

As for the sum of the other variance,

$$\begin{aligned}
\sum_{j=1}^{J} \mathrm{Var}(Y_j) &= \sum_{j=1}^{J} \mathrm{Var}\left(\sum_{(j-1)l+1}^{jl} W_i\right) \\
&= \sum_{j=1}^{J}\left[\sum_{i=(j-1)l+1}^{jl} \mathrm{Var}(W_i) \quad + \quad 2 \sum_{(j-1)l+1\le i<i'\le jl} \mathrm{Cov}(W_i, W_{i'})\right] \\
&= \sum_{i=1}^{\rho} \mathrm{Var}(W_i) + 2\sum_{j=1}^{J} \sum_{i=(j-1)l+1}^{jl-1} \sum_{k=1}^{jl-i} \mathrm{Cov}(W_i, W_{i+k}).
\end{aligned}$$

Thus, for $J > 1$

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - \sum_{j=1}^{J} \mathrm{Var}(Y_j) &= 2\sum_{i=1}^{\rho-1}\sum_{k=1}^{\rho-i} \mathrm{Cov}(W_i, W_{i+k}) - 2\sum_{j=1}^{J} \sum_{i=(j-1)l+1}^{jl-1} \sum_{k=1}^{jl-i} \mathrm{Cov}(W_i, W_{i+k}) \\
&= 2\sum_{j=1}^{J} \sum_{i=(j-1)l+1}^{jl-1} \sum_{k=1}^{\rho-i} \mathrm{Cov}(W_i, W_{i+k}) + 2\sum_{j=1}^{J-1}\sum_{k=1}^{\rho-i} \mathrm{Cov}(W_{jl}, W_{jl+k}) \\
&\quad -2\sum_{j=1}^{J} \sum_{i=(j-1)l+1}^{jl-1} \sum_{k=1}^{jl-i} \mathrm{Cov}(W_i, W_{i+k}) \\
&= 2\sum_{j=1}^{J-1} \sum_{i=(j-1)l+1}^{jl} \sum_{k=jl-i+1}^{\rho-i} \mathrm{Cov}(W_i, W_{i+k}).
\end{aligned}$$

For the range of l where $\delta \leq l \leq \rho/2$, the condition of δ-dependence implies that

$$\begin{aligned} \mathrm{Var}(\sum_{i=1}^{\rho} W_i) - \sum_{j=1}^{J} \mathrm{Var}(Y_j) &= 2\sum_{j=1}^{J-1} \sum_{i=jl-\delta+1}^{jl} \sum_{k=jl-i+1}^{\delta} \mathrm{Cov}(W_i, W_{i+k}) \\ &= \delta(\delta+1)(J-1)\overline{\mathrm{Cov}(W_i, W_{i'})}_J \end{aligned} \tag{7}$$

where $\overline{\mathrm{Cov}(W_i, W_{i'})}_J$ is the average of the covariance terms in (7).

By the assumption that $\overline{\mathrm{Cov}(W_i, W_{i'})}_J$ is approximately the same for all l in some range $[l_1, l_2]$ where $\delta \leq l_1 < l_2 \leq \rho/2$, then within this range of l

$$\mathrm{Var}(\sum_{i=1}^{\rho} W_i) - \sum_{j=1}^{J} \mathrm{Var}(Y_j) \approx C\delta(\delta+1)(J-1) = C\delta(\delta+1)(\rho/l-1),$$

for some constant C, giving the desired result. □

Result 2 *With data consisting of exactly m periods, the block bootstrap estimator of the mean* $\mathrm{E}(\sum_{j=1}^{J} Y_j^*)$

1. *is an unbiased estimator of* $\mathrm{E}(\sum_{i=1}^{\rho} W_i)$.
2. *is equal to the sample mean for periodic sum* $(1/m)\sum_{k=1}^{m} T_k$ *where, as defined in Section 2.1,* $T_k = \sum_{i=(k-1)\rho+1}^{k\rho} \mathcal{W}_i$ *is the sum of the observations within period k and the data are of size m periods.*
3. *has variance that is approximately equal to* $(1/m)\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$*, the true variance divided by the length of data in periods.*

Proof Without loss of generality, let $Jl = \rho$. Define $Y_{jk} = \sum_{i=(k-1)\rho+(j-1)l+1}^{(k-1)\rho+jl} \mathcal{W}_i$, i.e., the sum of the block of data corresponding to the jth interval and kth period. Then, as in Section 3, $Y_j^*,\ j = 1, \ldots, J$ have independent discrete distributions such that

$$Y_j^* \sim \mathrm{Unif}\,\{Y_{j1}, Y_{j2}, \ldots, Y_{jm}\}$$

1. First note that $\mathrm{E}(\mathrm{E}(Y_j^*)) = \mathrm{E}((1/m)\sum_{k=1}^{m} Y_{jk}) = (1/m)\sum_{k=1}^{m} \mathrm{E}(Y_{jk}) = (1/m)\sum_{k=1}^{m} \mathrm{E}(Y_j) =$

$\mathrm{E}(Y_j) = \mathrm{E}(\sum_{i=(j-1)l+1}^{jl} W_i)$ so that

$$\begin{aligned}
\mathrm{E}(\mathrm{E}(\sum_{j=1}^{J} Y_j^*)) &= \sum_{j=1}^{J} \mathrm{E}(\mathrm{E}(Y_j^*)) \\
&= \sum_{j=1}^{J} \mathrm{E}(\sum_{i=(j-1)l+1}^{jl} W_i) \\
&= \mathrm{E}(\sum_{j=1}^{J} \sum_{i=(j-1)l+1}^{jl} W_i) \\
&= \mathrm{E}(\sum_{i=1}^{\rho} W_i)
\end{aligned}$$

since $Jl = \rho$.

2. By interchanging the summations in the third step,

$$\begin{aligned}
\mathrm{E}(\sum_{j=1}^{J} Y_j^*) &= \sum_{j=1}^{J} \mathrm{E}(Y_j^*) \\
&= \sum_{j=1}^{J} \frac{1}{m} \sum_{k=1}^{m} Y_{jk} \\
&= \frac{1}{m} \sum_{k=1}^{m} \sum_{j=1}^{J} Y_{jk} \\
&= \frac{1}{m} \sum_{k=1}^{m} \sum_{j=1}^{J} \sum_{i=(k-1)\rho+(j-1)l+1}^{(k-1)\rho+jl} \mathcal{W}_i \\
&= \frac{1}{m} \sum_{k=1}^{m} T_k.
\end{aligned}$$

3. The approximation on the 3rd line of the following set of equations is justified because δ is much smaller than the period ρ and hence the small covariance between the sums of two adjacent periods can be ignored. Also the equality on the 3rd line holds because of

the periodicity of W_i.

$$\begin{aligned}
\mathrm{Var}(\mathrm{E}(\sum_{j=1}^{J} Y_j^*)) &= \mathrm{Var}(\sum_{j=1}^{J} \mathrm{E}(Y_j^*)) = \mathrm{Var}(\sum_{j=1}^{J} \frac{1}{m} \sum_{k=1}^{m} Y_{jk}) \\
&= \mathrm{Var}(\frac{1}{m} \sum_{i=1}^{m\rho} \mathcal{W}_i) = \frac{1}{m^2}\mathrm{Var}(\sum_{i=1}^{m\rho} \mathcal{W}_i) = \frac{1}{m^2}\mathrm{Var}(\sum_{i=1}^{m\rho} W_i) \\
&\approx \frac{1}{m^2} \sum_{k=1}^{m} \mathrm{Var}(\sum_{i=(k-1)\rho+1}^{k\rho} W_i) = \frac{1}{m^2} \sum_{k=1}^{m} \mathrm{Var}(\sum_{i=1}^{\rho} W_i) \\
&= \frac{1}{m}\mathrm{Var}(\sum_{i=1}^{\rho} W_i)
\end{aligned}$$

$\square$

Result 3 *The block bootstrap variance estimator* $\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)$ *is a biased estimator of* $\mathrm{Var}(\sum_{i=1}^{\rho} W_i)$ *and of* $\sum_{j=1}^{J} \mathrm{Var}(Y_j)$*. To be more precise,* $\mathrm{E}[\mathrm{Var}(\sum_{j=1}^{J} Y_j^*)] = (1 - 1/m) \sum_{j=1}^{J} \mathrm{Var}(Y_j)$ *for all* $m > 1$*.*

Proof Since $Y_j^* \sim \mathrm{Unif}\,\{Y_{j1}, Y_{j2}, \dots, Y_{jm}\}$, we first note that

$$\begin{aligned}
\mathrm{Var}(Y_j^*) &= \mathrm{E}(Y_j^{*2}) - [\mathrm{E}(Y_j^*)]^2 \\
&= \frac{1}{m} \sum_{k=1}^{m} Y_{jk}^2 - (\frac{1}{m} \sum_{k=1}^{m} Y_{jk})^2 \\
&= \frac{1}{m} \sum_{k=1}^{m} (Y_{jk} - \overline{Y}_{j\cdot})^2
\end{aligned}$$

where $\overline{Y}_{j\cdot}$ is the average over index k. Observing that $Y_{j1}, Y_{j2}, \dots, Y_{jm}$ are independent for $J > 1$ (because of the δ-dependence condition and that $\delta \ll \rho$) and identically distributed (Y_{jk} is a realization of Y_j for all k), the last expression is essentially the same as the maximum likelihood estimator of the variance using a random sample under normality, i.e., $(1/n) \sum_{i=1}^{n} (x_i - \bar{x})^2)$. And it is well known that the mean of this estimator is biased, and is equal to $\mathrm{E}(\sum_{i=1}^{n} (x_i - \bar{x})^2/n) = ((n-1)/n)\sigma^2$.

Therefore,

$$\mathrm{E}(\mathrm{Var}(Y_j^*)) = (1 - \frac{1}{m})\mathrm{Var}(Y_j),$$

and it follows that

$$\begin{aligned}\mathrm{E}(\mathrm{Var}(\sum_{j=1}^{J}(Y_j^*))) &= \sum_{j=1}^{J}\mathrm{E}(\mathrm{Var}(Y_j^*)) \\ &= (1-\frac{1}{m})\sum_{j=1}^{J}\mathrm{Var}(Y_j).\end{aligned}$$

□

Result 4 *The block bootstrap estimators of the mean and variance can be expressed in terms of the data in the following way:*

$$\mathrm{E}(\sum_{j=1}^{J}Y_j^*) = \frac{1}{m}\sum_{j=1}^{J}\sum_{k=1}^{m}Y_{jk}$$

$$\mathrm{Var}(\sum_{j=1}^{J}Y_j^*) = \frac{1}{m}\sum_{j=1}^{J}\sum_{k=1}^{m}Y_{jk}^2 - \frac{1}{m^2}\sum_{j=1}^{J}(\sum_{k=1}^{m}Y_{jk})^2$$

where, as in Result 2, $Y_{jk} = \sum_{i=(k-1)\rho+(j-1)l+1}^{(k-1)\rho+jl}\mathcal{W}_i$ *is the sum of the block of data corresponding to the* j*th interval and* k*th period.*

Proof The expression for the mean is derived in the proof in part two of Result 2. The expression for the variance follows from the independence of Y_j^* and the results given in the proof of Result 3. □

References

Bauer, D. R., and Martin, J. W., Editors (1999), *Service Life Prediction of Organic Coatings, A Systems Approach*, American Chemical Society: Washington.

Bauer, D. R., and Martin, J. W., Editors (2002), *Service Life Prediction: Methodologies and Metrologies*, Bauer, D. R., and Martin, J. W., Editors, American Chemical Society: Washington.

Bühlmann, P. (2002), "Bootstraps for Time Series", *Statistical Science*, **17**, 52-72

Carlstein, E. (1986), "The use of subseries values for estimating the variance of a general statistics from a stationary sequence", *Ann. Statist.*, **14**, 1171-1179.

Chan, V. and Meeker, W. Q. (2001), "Estimation of degradation-based reliability in outdoor environments", Preprint Series, Dept. of Statistics, Iowa State University.

Davison, A. C. and Hinkley, D. V. (1997), *Bootstrap Methods and Their Applications*, Cambridge Univ. Press.

Hall, P. (1985), "Resampling a coverage pattern", *Stochastic Process Appl.*, **20**, 231-246.

Hall, P, Horowitz, J. L. , and Jing, B-Y. (1995), "On blocking rules for the bootstrap with dependent data", *Biometrika*, **82**, 561-574.

Jorgensen, G. J., Kim, H. M., and Wendelin, T. J. (1996), "Durability studies of solar reflector materials exposed to environmental stresses", *Durability Testing of Nonmetallic Materials, ASTM STP 1294*, Herling, R. J., Editor, American Society for Testing and Materials.

Künsch, H. R. (1989), "The jackknife and the bootstrap for general stationary observations", *Ann. Statist.*, **17**, 1217-1241.

Léger, C., Politis, D. N. and Romano, J. P. (1992), "Bootstrap technology and applications", *Technometrics*, **34**, 378-398.

Lahiri, S.N. (2003), *Resampling Methods for Dependent Data.* Springer, NY.

Liu, R.Y. and Singh, K. (1992), "Moving blocks jackknife and bootstrap capture weak dependence", *Exploring the Limits of Bootstrap* (Eds.– R. Lepage and L. Billard), Wiley, New York, 225-248.

Park, D. and Willemain, T. R. (1999), "The threshold bootstrap and threshold jackknife", *Computational Statistics and Data Analysis*, **31**, 187-202.

Park, D., Kim, Y., Shin, K. and Willemain, T. R. (2001),"Simulation output and analysis using the threshold bootstrap", *European Journal of Operational Research*, **134**, 17-28.

Politis, D. and Romano, J.P. (1992), "A circular block resampling procedure for stationary data", *Exploring the Limits of Bootstrap* (Eds.– R. Lepage and L. Billard), Wiley, New York, 263-270.

Politis, D. and Romano, J.P. (1994). "The Stationary Bootstrap", *J. Amer. Statist. Assoc.* **89**,1303-1313.