Resampling time series with seasonal components

Dimitris N. Politis
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
politis@euclid.ucsd.edu

Abstract

In the case of time series with a seasonal component, the well-known block bootstrap procedure is not directly applicable. We propose a modification of the block bootstrap that successfully addresses the issue of seasonalities, and show some of its properties.

1 Introduction

Consider observations of the form $X_1, \ldots, X_N$ arising from a time series $\{X_t, t \in \mathbf{Z}\}$. An interesting class of time series models involves the case where a seasonality effect is present, and is typically modeled by the equation

$$X_t = \mu_t + Y_t \quad \text{and} \quad \mu_t = \mu_{t-d} \quad \text{for all} \quad t \in \mathbf{Z}; \tag{1}$$

in the above, d is an integer denoting the period of the deterministic (but unknown) function μ_t, and $\{Y_t, t \in \mathbf{Z}\}$ is a (strictly) stationary sequence with mean zero. Usually, the period d is known (or obvious from the data), and in many cases corresponds to a daily, weekly or annual periodicity—hence the name "seasonality"; see e.g. Brockwell and Davis (1991).

Except in the rather special case where μ_t is a constant, model (1) is not stationary; hence, the Block Bootstrap (BB) method of Künsch (1989) is not directly applicable, and the same is true for the different BB variations; see e.g. Lahiri (1999) and the references therein.

In the present article, we suggest a way to fix this problem, and provide a resampling algorithm for time series with seasonal components such as model (1). The algorithm may be given the name: "Seasonal Block Bootstrap" (SBB) to distinguish it from the standard BB; however, the SBB is nothing other than a version of the BB with blocks whose size and starting points are restricted to be integer multiples of the period d.

2 Seasonal Block Bootstrap

For simplicity, let us assume that the sample size N is an integer multiple of the period d, i.e., $N = nd$ for some integer n; however, defining $n = [N/d]$, where $[\cdot]$ denotes the integer part, would lead to an identical algorithm effectively ignoring a part of a cycle at the end of the data series. The Seasonal Block Bootstrap (SBB) algorithm is defined below.

- First chose a positive integer $b(< n)$, and let the $i_0, i_1, \ldots, i_{k-1}$ be drawn i.i.d. with distribution uniform on the set $\{1, 2, \ldots, n-b+1\}$; here we may take $k = [n/b]$, although different choices for k are also possible. The SBB procedure constructs a bootstrap pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$, where $l = kbd$, as follows:
  For $m = 0, 1, \ldots, k-1$, and $j = 1, 2, \ldots, bd$, let $X^*_{mbd+j} := X_{i_m d+j-1}$.

The above procedure defines a probability measure (conditional on the data $X_1, \ldots, X_N$) that will be denoted P^*; expectation and variance with respect to P^* are denoted E^* and Var^* respectively. As claimed in the Introduction, the SBB is a version of the BB with blocks of size bd, and starting points $i_0 d, i_1 d, \ldots, i_{k-1} d$ that are integer multiples of the period d. We next give some applications of SBB's validity for inference under model (1).

2.1 Estimating the seasonal component and overall mean

An interesting issue may be estimation of the seasonal component μ_i, $i = 1, \ldots, d$, and the overall mean $\bar{\mu} = d^{-1} \sum_{i=1}^{d} \mu_i$. Point estimation is straightforward by means of averages of the 'sampled' series; in other words, define

$$\hat{\mu}_i = n^{-1} \sum_{j=0}^{n-1} X_{i+jd} \quad \text{and} \quad \hat{\bar{\mu}} = d^{-1} \sum_{i=1}^{d} \hat{\mu}_i. \tag{2}$$

The SBB is useful in getting interval estimates for μ_i and μ by means of successfully approximating the distribution of $\hat{\mu}_i$ and $\hat{\bar{\mu}}$; to do this, from each bootstrap pseudo-series $X_1^*, X_2^*, \ldots, X_l^*$ we construct bootstrap versions of $\hat{\mu}_i$ and $\hat{\bar{\mu}}$ by:

$$\hat{\mu}_i^* = (kb)^{-1} \sum_{j=0}^{kb-1} X_{i+jd}^* \quad \text{and} \quad \hat{\bar{\mu}}^* = d^{-1} \sum_{i=1}^{d} \hat{\mu}_i^*. \tag{3}$$

Denote the vectors $\mu = (\mu_1, \ldots, \mu_d)$, and $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_d)$. Let $\mathcal{L}(\sqrt{n}(\hat{\mu} - \mu))$ denote the probability law of $\sqrt{n}(\hat{\mu} - \mu)$, and $\mathcal{L}^*(\sqrt{kb}(\hat{\mu}^* - E^*\hat{\mu}^*))$ its bootstrap counterpart; similarly, define $\mathcal{L}(\sqrt{n}(\hat{\bar{\mu}} - \bar{\mu}))$, and $\mathcal{L}^*(\sqrt{kb}(\hat{\bar{\mu}}^* - E^*\hat{\bar{\mu}}^*))$. Finally, let $d_0(\cdot, \cdot)$ be a metric metrizing weak convergence, and $\alpha_Y(k)$ be the Rosenblatt strong mixing coefficients for series $\{Y_t\}$. If n is large, accurate point estimation is possible, and a central limit theorem and its bootstrap analog hold true under the assumptions of our Theorem 2.1.

Theorem 2.1 *Assume that, for some* $\delta > 0$, $E|Y_t|^{2+\delta} < \infty$ *and* $\sum_{k=1}^{\infty} \alpha_Y^{\delta/(2+\delta)}(k) < \infty$. *If* $b \to \infty$ *as* $n \to \infty$ *but with* $b = o(n)$, *then*

$$d_0\left(\mathcal{L}(\sqrt{n}(\hat{\mu} - \mu)), \mathcal{L}^*(\sqrt{kb}(\hat{\mu}^* - E^*\hat{\mu}^*))\right) \xrightarrow{P} 0 \tag{4}$$

as well as

$$d_0\left(\mathcal{L}(\sqrt{n}(\hat{\bar{\mu}} - \bar{\mu})), \mathcal{L}^*(\sqrt{kb}(\hat{\bar{\mu}}^* - E^*\hat{\bar{\mu}}^*))\right) \xrightarrow{P} 0. \tag{5}$$

The proof of (4) relies on a bootstrap central limit theorem analogous to the one given in Radulovic (1996); eq. (5) follows from (4) by the continuous mapping theorem. Note that, under the stronger conditions $E|Y_t|^{6+\delta} < \infty$ and $\sum_{k=1}^{\infty} k^2 \alpha_Y^{\delta/(6+\delta)}(k) < \infty$, convergence of second moments will also hold in conjunction with eq. (4) and (5).

2.2 Least squares estimators

Suppose that eq. (1) holds with $\mu_t = \sum_{j=1}^{p} \beta_j m_j(t)$ for all $t \in \mathbf{Z}$, where $m_j(\cdot)$, $j = 1, \ldots, p$, are some *known* functions of period d, and β_j, $j = 1, \ldots, p$, are some unknown coefficients; for identifiability we need $p \leq d$ here. For example, $m_1(\cdot)$ may be the constant function, and $m_j(\cdot)$ for $j > 1$ may be some other functions of interest, e.g., trigonometric functions.

Let $\hat{\beta}_j$ denote the Least Squares (LS) estimator of β_j, i.e., the minimizer of the quantity: $\sum_{t=1}^{N}(X_t - \sum_{j=1}^{p} \beta_j m_j(t))^2$. Let $\beta = (\beta_1, \ldots, \beta_p)$ and $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$. Finally, let $\hat{\beta}^*$ be the LS estimator of β computed from an SBB bootstrap series $X_1^*, X_2^*, \ldots, X_l^*$. It is easy to calculate that $\hat{\beta}$ is a continuous (actually linear) function of $\hat{\mu}$. Thus, by the continuous mapping theorem, the following is an immediate corollary of Theorem 2.1.

Corollary 2.1 *Under the conditions of Theorem 2.1 we have*

$$d_0\left(\mathcal{L}(\sqrt{n}(\hat{\beta}-\beta)), \mathcal{L}^*(\sqrt{kb}(\hat{\beta}^*-E^*\hat{\beta}^*))\right) \xrightarrow{P} 0. \tag{6}$$

2.3 Concluding remark

A foreseeable competitor to the SBB under model (1) is a procedure involving BB resampling of the 'residuals' $\hat{Y}_t := X_t - \hat{\mu}_t$ as a basic step in creating a bootstrap series $X_1^*, \ldots, X_l^*$ by $X_t^* := \hat{\mu}_t + \hat{Y}_t^*$, where $\hat{Y}_1^*, \ldots, \hat{Y}_l^*$ is the BB residual pseudo-series. The SBB is certainly more direct and intuitive as compared to the 'residual' BB, but is it as efficient?

It is well-known that the degree of overlap among blocks to be bootstrapped plays a major role in efficiency: maximum overlap leads to maximum efficiency. The overlap proportion in the SBB is actually equal to $(b-1)/b$. For $b=1$, it corresponds to zero overlap (the so-called Carlstein blocking); however, even for b as low as 2 it moves to 50% overlap which has improved efficiency. More importanty, as $b \to \infty$, the degree of overlap tends to 100% which implies that the SBB enjoys an asymptotic relative efficiency of one as compared to 'residual' BB blocking with full overlap; see Lahiri (1999), or eq. (3.46) in Politis et al. (1999).

References

[1] Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods, 2nd ed.*, Springer: New York.

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

[3] Lahiri, S.N. (1999), Theoretical comparisons of block bootstrap methods, *Ann. Statist.*, 27, 386-404.

[4] Politis, D.N., Romano, J.P. and Wolf, M. (1999), *Subsampling*, Springer: New York.

[5] Radulovic, D. (1996), The bootstrap of the mean for strong mixing sequences under minimal conditions, *Statist. Prob. Letters*, 28, 65-72.