The Frequency Domain Bootstrap

D. Izyumin, E. Shvarts & N. Ulle

Spring 2014

Introduction

Bootstrapping a Time Series

Since its introduction over three decades ago, the bootstrap has become a popular and widely applicable tool of statistical inference. At times, bootstrapping can provide valuable insight into the distributions of statistics, which would be unattainable by any analytical procedures. Over the years, numerous modifications and extensions of the original algorithm have made it possible to implement bootstrapping procedures in a variety of settings.

In the time series framework, observations are generally not independent and the dependence structure is a key object of interest. Moreover, there is usually an absence of repeated measurements, as only one realization of the process is observed. Since the bootstrap is based on the resampling of i.i.d objects, these two complications make the extension of the bootstrap to time series analysis a challenge.

The difficulties above may be overcome either by resampling whole blocks of observations (moving block bootstrap), or by first somehow transforming the data into i.i.d. objects and then resampling those. Examples of such i.i.d. transformations of the data are residuals and innovations in the time domain, or periodogram ordinates in the frequency domain. Dahlhaus and Janas outline a bootstrap procedure based on the resampling of Studentized tapered periodogram ordinates, and show that under reasonable assumptions the procedure has very favorable properties when dealing with an important class of ratio statistics [3].

Spectral Means and Ratio Statistics

Consider a real-valued, stationary, zero-mean time series $\{X_t\}_{t\in\mathbb{Z}}$, of which realizations $t = 1, \cdots, T$ are observed. The usual estimate of the spectral density function $f(\alpha)$ is the smoothed periodogram constructed from the observed data. For additional smoothing, the tapered periodogram, denoted by $I_T(\alpha)$, will be used instead. A brief motivation of tapering can be found in the *Empirical Examples* section, and Dahlhaus and Janas provide a detailed description of the procedure in their publication.

For any $\varphi = (\varphi^{(1)}, \cdots, \varphi^{(d)})$, where each $\varphi^{(i)}$ of bounded variation, the *spectral mean* is defined as

$$A(\varphi, f) = \int_0^\pi \varphi(\alpha) f(\alpha) d\alpha,$$

and the canonical estimate is given by

$$A(\varphi, I_T) = \int_0^\pi \varphi(\alpha) I_T(\alpha) d\alpha$$

Now consider the normalized spectral density $g(\alpha) = \frac{f(\alpha)}{F(\pi)}$, where F is the spectral distribution function. A natural estimate of this quantity is the normalized tapered periodogram, $J_T(\alpha) = \frac{I_T(\alpha)}{\hat{F}_T(\pi)}$, where $\hat{F}_T(\alpha)$ is the estimate of the spectral distribution function obtained using the tapered periodogram. The *normalized spectral mean* is obtained when g is plugged into $A(\varphi, \cdot)$ in place of f,

$$A(\varphi, g) = \frac{\int_0^\pi \varphi(\alpha) f(\alpha) d\alpha}{\int_0^\pi f(\alpha) d\alpha} = \frac{A(\varphi, f)}{A(1, f)},$$

and its estimate is obtained by similarly replacing I_T by J_T,

$$A(\varphi, J_T) = \frac{\int_0^\pi \varphi(\alpha) I_T(\alpha) d\alpha}{\int_0^\pi I_T(\alpha) d\alpha} = \frac{A(\varphi, I_T)}{A(1, I_T)}$$

Since the estimate can be expressed as a ratio of two spectral mean estimates, its is known as a *ratio statistic*.

Examples

The sample means and ratio statistics are both notable classes of statistics. Since it is only required that $\varphi(\alpha)$ is bounded in variation, a great number of functions may be chosen for φ, resulting in many different statistics and estimates. Below are several important estimates that can be expressed as either $A(\varphi, I_T)$ or $A(\varphi, J_T)$ with proper choice of φ:

$\varphi(\alpha)$	$A(\varphi, I_T)$	$A(\varphi, J_T)$
$\cos(\alpha u)$	Autocovariance $\hat{\gamma}(u), u \in \mathbb{Z}$	Autocorrelation $\hat{\rho}_T(u), u \in \mathbb{Z}$
$\mathbb{1}_{[0,\,\lambda]}(\alpha)$	Spectral Distribution Function $\hat{F}_T(\lambda)$	Normalized SDF $\frac{\hat{F}_T(\lambda)}{\hat{F}_T(\pi)}$

Table 1: Useful spectral means and ratio statistics.

Whittle Estimators

Consider the problem of selecting a parameter θ from a parameter family $\Theta \in \mathbb{R}$ with the corresponding family of spectral densities $\mathcal{F} = \{f_\theta : \theta \in \Theta\}, \Theta \in \mathbb{R}$. The Whittle estimate of the parameter, $\hat{\theta}$, is obtained by minimizing Whittle's likelihood function,

$$\mathcal{L}_T(\theta) = \frac{1}{2\pi} \int_0^\pi \left[\log f_\theta(\alpha) + \frac{I_T(\alpha)}{f_\theta(\alpha)} \right] d\alpha.$$

The likelihood $\mathcal{L}_T(\theta)$ can be interpreted as the distance between the (tapered) periodogram I_T and the candidate spectral density f_θ. In some settings Whittle estimates have very favorable properties and may be preferred over methods relying on the traditional likelihood [6].

Notably, a Whittle estimator can be expressed as a spectral mean $A(\varphi, I_T)$ with $\varphi(\alpha) = \nabla \frac{1}{f_\theta}(\alpha)$. Using this fact, Dahlhaus and Janas extend their ideas to Whittle estimates, and show that their bootstrap algorithm has favorable properties when working with Whittle estimates as well.

The Bootstrap Procedure

Resampling the Periodogram

We are interested in approximating the distribution of ratio statistic estimates, and so need an estimator for the spectral mean $A(\varphi, J_T)$ which we can compute via resampling. To this end, choose $n = \lfloor T/2 \rfloor$, and consider n equi-spaced samples of the periodogram, $I_j := I_T(\frac{j}{T/2}\pi)$ for $j = 1, \ldots, n$. By resampling from (a carefully normalized version of) these values, we can obtain an appropriate discretization of the spectral mean. Letting $\{I_j^*\}$ be the n resampled values, and $\varphi_j := \varphi(\frac{j}{T/2}\pi)$, consider

$$B(\varphi, I_T^*) := \frac{\pi}{n} \sum_{j=1}^{n} \varphi_j I_j^* \quad .$$

Our interest is in the behavior of the approximation error $A(\varphi, I_T) - A(\varphi, f)$, and so we consider the corresponding quantity $B(\varphi, I_T^*) - B(\varphi, \hat{f})$, where $\hat{f}$ is a suitable estimate for the spectral density (e.g., a consistent estimator formed from the periodogram values, perhaps via kernel smoothing). For the former, the (untapered) periodogram estimate obeys a central limit theorem (proved, e.g., in [4]), such that $\sqrt{T}\big(A(\varphi, I_T) - A(\varphi, f)\big)$ is asymptotically normal with variance equal to $2\pi \int \varphi^2 f^2 + (\kappa_4/\sigma^4) \left(\int \varphi f\right)^2$. Here κ_4 and σ^2 refer to the fourth cumulant and variance of the innovations in the underlying presumed linear process. Additionally, in [3], it is stated that $\sqrt{T}\big(B(\varphi, I_T^*) - B(\varphi, \hat{f})\big)$ is also asymptotically normal, with variance proportional to $2\pi \int \varphi^2 f^2$.

Hence, the latter statistic will only be useful for a bootstrap procedure if the second term in the former variance vanishes. Conveniently, this is fulfilled for example by Whittle estimates, and more generally by all ratio statistics. So, we will give an algorithm for resampling the periodogram which explains the 'careful normalization' mentioned above. In particular, the quantities which we resample are Studentized, because we expect them to asymptotically vary as independent exponential variables. [2][Theorem 5.2.6]

Periodogram Bootstrap Algorithm (of [3])

1. Obtain the periodogram sample values $\{I_j\}$ for $j = 1, \ldots, n$, possibly applying a taper to the data.
2. Obtain a consistent estimate $\hat{f}$ of the spectral density f, using the $\{I_j\}$. Kernel estimates are typical.
3. Obtain Studentized periodogram samples $\hat{\epsilon}_j := I_j/\hat{f}_j$.
4. Rescale the $\hat{\epsilon}_j$'s by their mean, so that $\tilde{\epsilon}_j := \hat{\epsilon}_j/\{(1/n)\sum_j \hat{\epsilon}_j\}$. This avoids unnecessary bias at the resampling stage.
5. Sample the empirical distribution of $\{\tilde{\epsilon}_j\}$ to obtain a bootstrap sample $\{\epsilon_j^*\}$.
6. Define bootstrap periodogram values via $I_j^* := \hat{f}_j \epsilon_j^*$.

Note that it's also possible to skip steps 1–5 by sampling directly from an exponential distribution with mean 1. This alternate procedure has the same convergence properties as the one outlined above, but may also avoid biases introduced by the Studentization, and is therefore preferable.

Assumptions

In order to estimate the distribution of the approximation error $A(\varphi, I_T) - A(\varphi, f)$ by that of $B(\varphi, I_T^*) - B(\varphi, \hat{f})$, the time series and desired spectral mean must be sufficiently nice; we describe some conditions and their relationship with the result.

- First, $\{X_t : t \in \mathbb{Z}\}$ must be a real-valued linear process: $X_t = \sum_{u \in \mathbb{Z}} a_u \xi_{t-u}$.

  The $\{\xi_t\}$ satisfy $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$, $\mathbb{E}\xi_1^8 < \infty$, $\mathbb{E}\xi_t^3 = 0$. The first two conditions are necessary for weak stationarity. Finite eighth moments are necessary for the bootstrap distribution to converge to the exponential distribution. [3][Corollary 1] Having vanishing third moments satisfies a technical constraint guaranteeing a good approximation of the relevant covariance matrix. [3][Remark 5] Without this, we cannot attain the claimed convergence rate.

  The $\{a_t\}$ vanish exponentially; this is a requirement imposed by the use of Edgeworth expansions in the proof.

- The estimate $\hat{f}$ of f is uniformly strongly consistent, such that

$$\sup_{\alpha \in [0,\pi]} |\hat{f}(\alpha) - f(\alpha)| \to 0 \text{ almost surely.}$$

  This is necessary for showing convergence in distribution. Further, the spectral density f itself is everywhere-positive and bounded away from zero; in particular, there are no vanishing Fourier modes.

- φ is a d-dimensional vector of bounded-variation functions $\varphi^{(r)} : [0, \pi] \to \mathbb{R}$, each with even periodic extension to all of $\mathbb{R}$. This broad definition allows for a rich choice of spectral means for estimating. The use of Edgeworth expansions further imposes that the Fourier coefficients $\{\hat{\varphi}(\omega)\}$ vanish exponentially. While this would technically omit some important classes of statistics, this requirement is not sharp, and the method's validity beyond the theoretical guarantee is suggested by empirical evidence.

- The taper function has several technical limitations placed on it; the Tukey-Hanning tapers are given in [3] as an example of an acceptable choice. The innovations $\{\xi_t\}$ and φ also must satisfy some technical conditions for the use of Edgeworth expansions.

Convergence Rate

Under the necessary assumptions, estimating the distribution of the standardized spectral density by resampling from the standardized periodogram is consistent, and the convergence rate is asymptotically better than that achieved by the normal approximation.

To be precise, let $A(\varphi, g)$ be the ratio statistic corresponding to the time series with spectral density f, J_T be the corresponding standardized periodogram, J_T^* be the bootstrapped periodogram, and $\hat{g}$ be an estimator of the normalized spectral density, defined according to $\hat{g}_j := \hat{f}_j / \{(\pi/n) \sum_k \hat{f}_k\}$. Denote V_T as the covariance matrix of $\sqrt{T} A(\varphi, J_T)$. Define D_T according to $D_T^2 = V_T^{-1}$, and define $\hat{D}_T$ analogously via $\sqrt{T} B(\varphi, J_T^*)$. Let P^* denote the conditional probability given the data. Then, we have the result [3][Theorem 1]:

Theorem: Suppose the necessary assumptions hold. Then for almost all samples $\{I_j\}$,

$$\sup_C \left| \mathbb{P}\big(\sqrt{T} D_T (A(\varphi, J_T) - A(\varphi, g)) \in C\big) - \mathbb{P}^*\big(\sqrt{T} \hat{D}_T (B(\varphi, J_T^*) - B(\varphi, \hat{g})) \in C\big) \right| = o(T^{-1/2}) \quad .$$

The sup is taken over convex measurable subsets of $\mathbb{R}^d$.

Empirical Examples

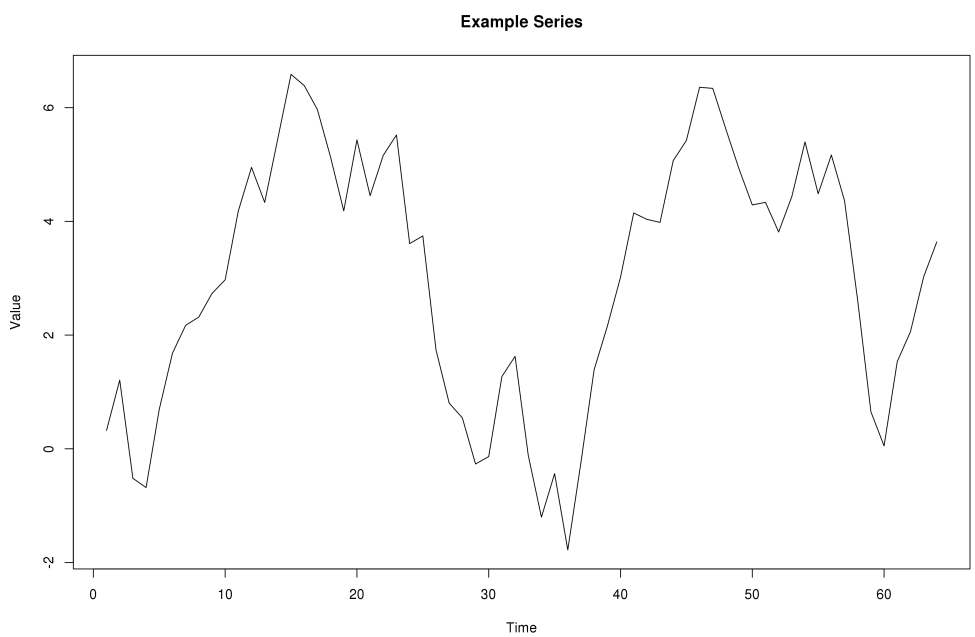

Figure 1: A realization of the AR(1) process in Example 1.

Practical Details

A random process can only be observed over a finite span of time. Consequently, the sample endpoints will typically interrupt any oscillations at two distinct points in their cycle. The discrete Fourier transform treats the sample as periodic, and the resulting discontinuity at the endpoints generates noise, or *leakage*, in the periodogram. Smoothing the tails of the sample towards zero, a process known as *tapering*, eliminates the discontinuity and can help to alleviate leakage [7]. Following the procedure of Dahlhaus and Janas, a 10% Tukey-Hanning taper (see [5]) was used for all examples presented here. Small changes to the taper proportion had little, if any, effect on the smoothed spectral density estimate $\hat{f}$. In practice, other tapers can

be used with the frequency domain bootstrap. These were not investigated, to avoid distraction from our focus.

Although tapering reduces noise in the periodogram, it isn't sufficient to make the periodogram a consistent estimator for the spectral density. This requires further smoothing. In the examples, kernel smoothing is used, with the Epanechnikov kernel

$$K(x) = \frac{3}{4}\pi\left[1 - \left(\frac{x}{\pi}\right)^2\right].$$

Smoothing has a tendency to reduce or eliminate modes of the periodogram, and care must be taken to preserve important details. To this end, it's suggested to smooth the log-periodogram rather than smoothing the periodogram directly. Further bias correction is possible in light of the distribution of the periodogram ordinates. With kernel weights $w_j(x)$ defined in the standard way, the smoothed, bias-corrected spectral density estimate is given by

$$\hat{f}(x) = \exp\left[\sum_j w_j(x) \ln I_j - \ln \Gamma\big(1 + w_j(x)\big)\right],$$

where $j = 0, \ldots, \lfloor T/2 \rfloor$. The bandwidth is also essential, and should be selected to offer a good compromise between noise reduction and loss of detail.

Example 1

Consider estimation of the lag-1 autocorrelation $\rho(1)$, in the AR(1) model $X_t = 0.9X_{t-1} + Z_t$, using $T = 64$ observations. The innovations $\{Z_t\}$ are independent and identically distributed $\mathrm{U}(-\sqrt{3}, \sqrt{3})$ random variables, making them a mean 0, variance 1 white noise. In this setting, the conditions of Theorem 1 are satisfied. A realization of this process is shown in Figure 1.

A key feature of this model is that its spectral density exhibits a sharp mode at zero. This makes spectral density estimation challenging, because of the trade-offs inherent in smoothing the periodogram. Figure 2 shows that with a bandwidth of 0.1, much of the mode's peak is lost. On the other hand, reducing the bandwidth to 0.05 produces a misleading, bimodal estimate. The former was selected for use here.

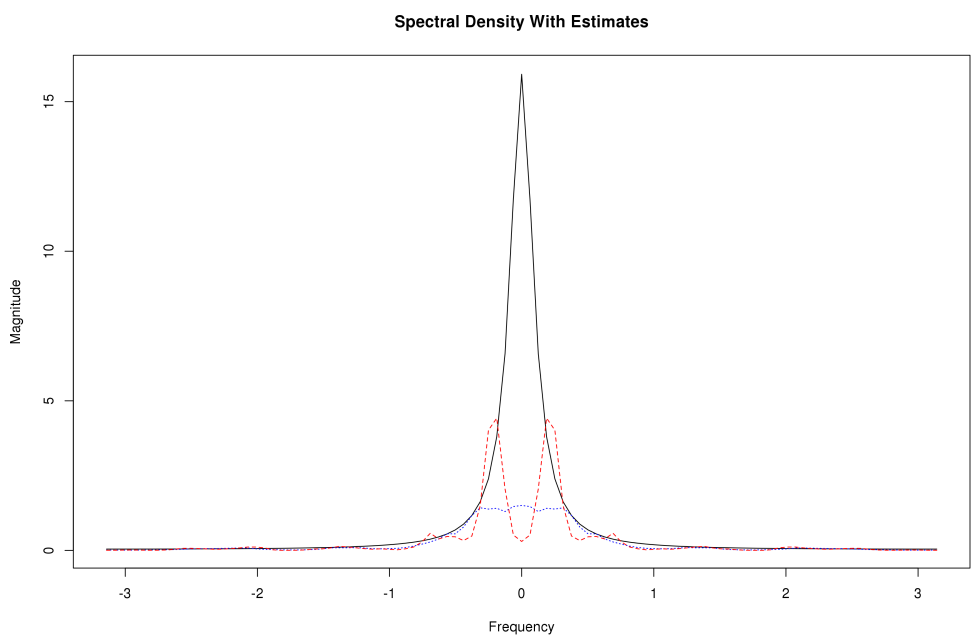

Figure 2: Example 1 spectral density (solid, black) and spectral density estimates at bandwidth 0.1 (dotted, blue) and 0.05 (dashed, red).

The estimate $\hat{\rho}(1)$ for the lag-1 autocorrelation $\rho(1)$ coincides with the Yule-Walker estimate of the AR parameter, so we have the standard result that as $T \to \infty$,

$$\sqrt{T}\left(\frac{\hat{\rho}(1) - 0.9}{c}\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, 1 - 0.9^2),$$

where c is a correction for the taper. This asymptotic error distribution is what would typically be used to construct confidence bounds on the estimate. The frequency domain bootstrap provides a bootstrapped error distribution as a competitor. In this case, 2000 estimates were bootstrapped. Figure 3 shows a plot of the cumulative distribution functions for both, as well as the true distribution (as approximated by repeated simulation). The figure confirms the result of Theorem 1: the bootstrapped distribution significantly outperforms the asymptotic distribution. This result is especially impressive because the spectral density estimate was not very good—smoothing masked the peak at zero. Repeated simulations show similarly positive results. Finally, we note that the true lag-1 autocorrelation 0.9 typically did fall within the 95% confidence intervals produced using the bootstrapped error distribution.

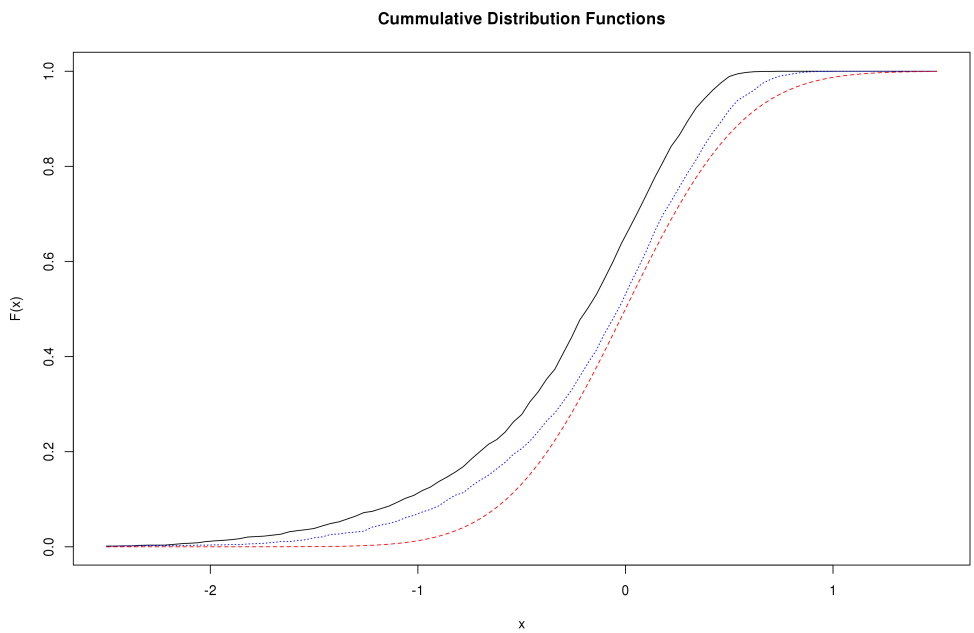

Figure 3: Example 1 CDFs for the true (solid, black), bootstrapped (dotted, blue), and asymptotic (dashed, red) error distributions.

Application

Now we turn our attention to an application of the frequency domain bootstrap. The Melbourne temperatures time series [1] records the daily minimum temperature (in $^\circ$ Celsius) in Melbourne, Australia from the beginning of 1981 to the end of 1990. Shown in Figure 4, this time series exhibits a strong oscillations with approximate period 365. It follows that there is a sharp peak in the spectral density in the vicinity of $2\pi/365 \approx 0$, somewhat similar to the first example. Again, this makes estimation of the spectral density function difficult, because the mode may be diminished by smoothing.

Estimates for the spectral density are shown in Figure 5. Bandwidths 0.01, 0.05, and 0.1 were considered. At the narrowest of these, the peak near zero appears to be well-preserved, but a great deal of noise is also preserved at higher frequencies. On the other hand, the widest bandwidth loses both the shape and height of the mode. Based on this, only the bandwidth 0.01 and 0.05 estimates were used in the bootstrap procedure.

The frequency domain bootstrap was used to compute confidence intervals for the lag-1 and lag-2 autocorrelations. The results are shown in Table 2. For reference, these autocorrelations were also estimated

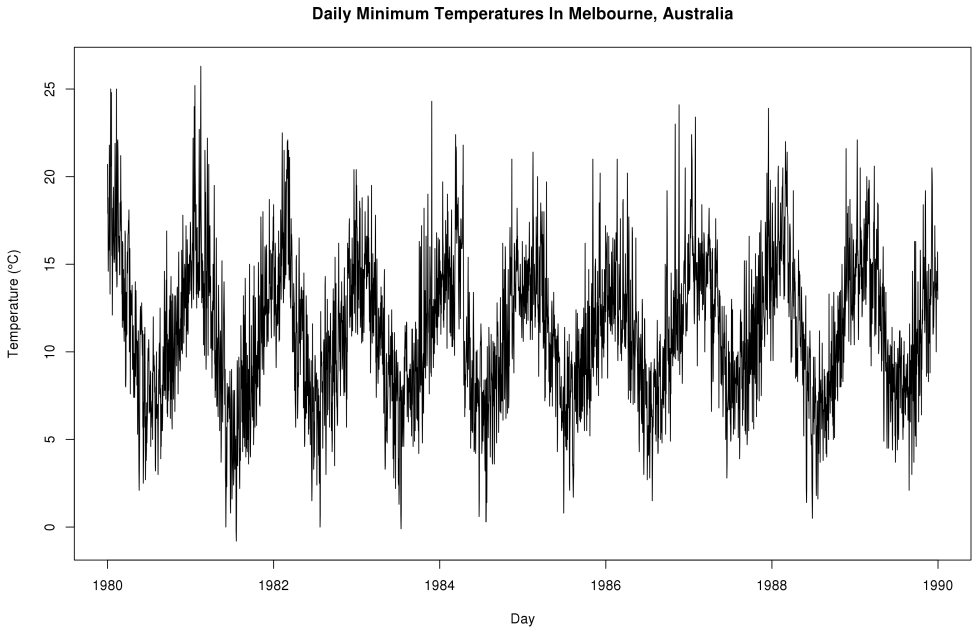

Figure 4: The Melbourne temperatures time series used in the application.

	Lag-1	Lag-2
0.01	$(0.547, 0.618)$	$(0.236, 0.357)$
0.05	$(0.524, 0.580)$	$(0.206, 0.291)$
`acf()`	0.774	0.630

Table 2: Autocorrelation estimates for the Melbourne temperatures time series, using two different bandwidths and the `acf()` function.

with R's `acf()` function. Using the smaller bandwidth, 0.05, gives slightly larger estimates, although all of the bootstrapped estimates are substantially smaller than the reference estimates. This behavior seems to indicate that there is a serious penalty for the attenuation of the mode, and the results of the frequency domain bootstrap should be viewed with scrutiny in applications where strong seasonality is present.

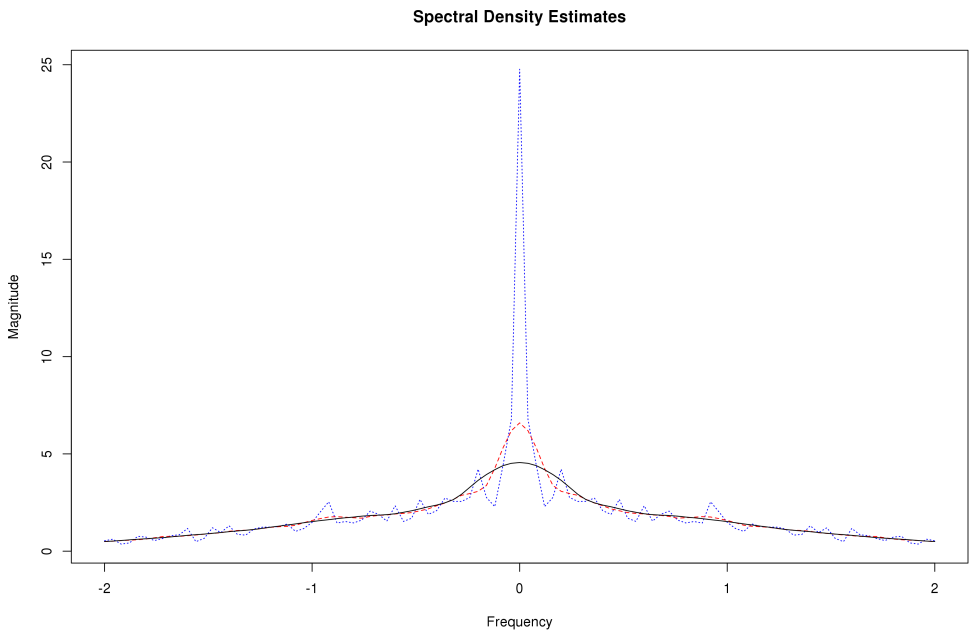

Figure 5: Spectral density estimates for the Melbourne temperature time series, with bandwidth 0.01 (dotted, blue), 0.05 (dashed, red), and 0.1 (solid, black).

References

[1] Daily minimum temperatures in melbourne, australia, 1981-1990. Time Series Data Library.

[2] David R Brillinger. *Time series: data analysis and theory*, volume 36. Siam, 2001.

[3] R Dahlhaus, D Janas, et al. A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics*, 24(5):1934–1963, 1996.

[4] Rainer Dahlhaus. Spectral analysis with tapered data. *Journal of Time Series Analysis*, 4(3):163–175, 1983.

[5] Rainer Dahlhaus. Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *The Annals of Statistics*, pages 808–841, 1988.

[6] Robert Fox and Murad S Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, pages 517–532, 1986.

[7] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer New York, 2000.