

The Frequency Domain Bootstrap

D. Izyumin, E. Shvarts, & N. Ulle

Spring 2014

Introduction

The Bootstrap Procedure

Empirical Examples

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 [3]. Following the procedure of Dahlhaus and Janas, a 10% Tukey-Hanning taper (see [2]) 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

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

where $j = 0, \dots, \lfloor T/2 \rfloor$. 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 $U(-\sqrt{3}, \sqrt{3})$ 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.

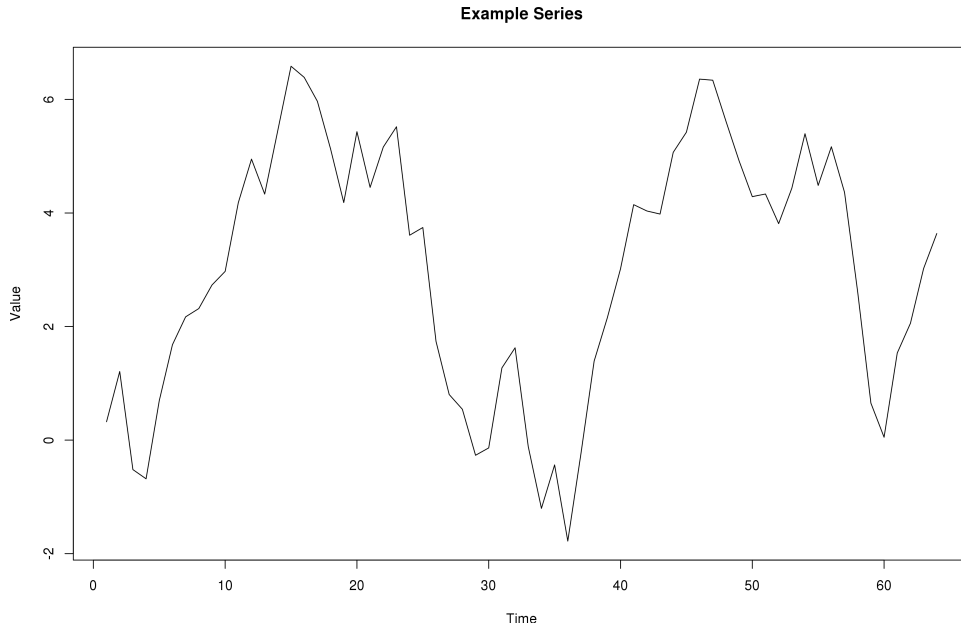


Figure 1: A realization of the AR(1) process in Example 1.

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 $T \rightarrow \infty$,

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

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.

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 ° 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 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 1. For reference, these autocorrelations were also estimated with R's `acf()` function. Using the smaller bandwidth, 0.05, gives slightly larger estimates, although all of

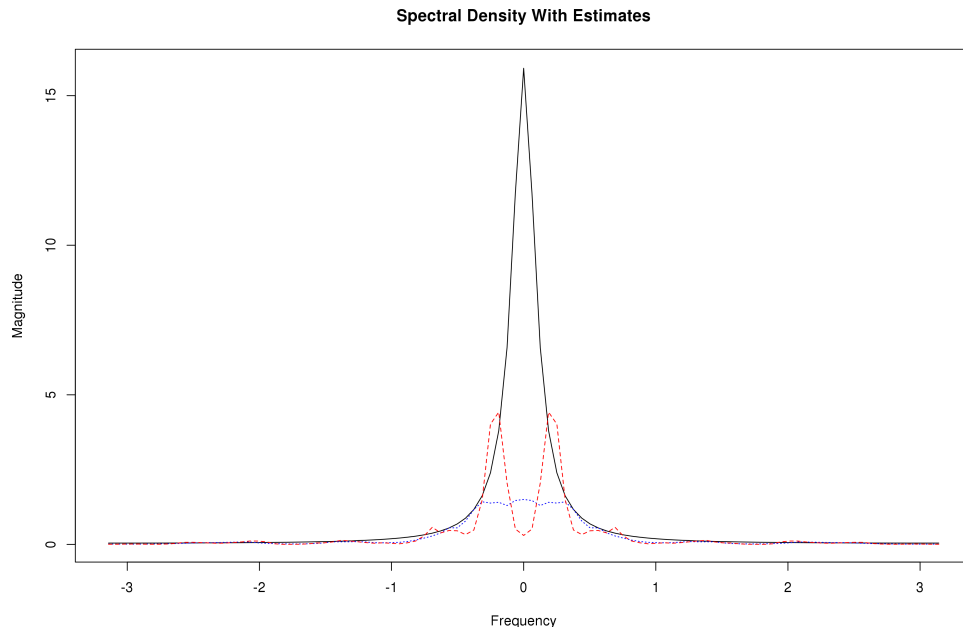


Figure 2: Example 1 spectral density (solid, black) and spectral density estimates at bandwidth 0.1 (dotted, blue) and 0.05 (dashed, red).

	Lag-1	Lag-2
0.01	(0.547, 0.618)	(0.236, 0.357)
0.05	(0.524, 0.580)	(0.206, 0.291)
<code>acf()</code>	0.774	0.630

Table 1: Autocorrelation estimates for the Melbourne temperatures time series, using two different bandwidths and the `acf()` function.

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.

Conclusion

References

References

- [1] Daily minimum temperatures in melbourne, australia, 1981-1990. Time Series Data Library.
- [2] 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.
- [3] Robert H Shumway, David S Stoffer, and David S Stoffer. *Time series analysis and its applications*, volume 3. Springer New York, 2000.

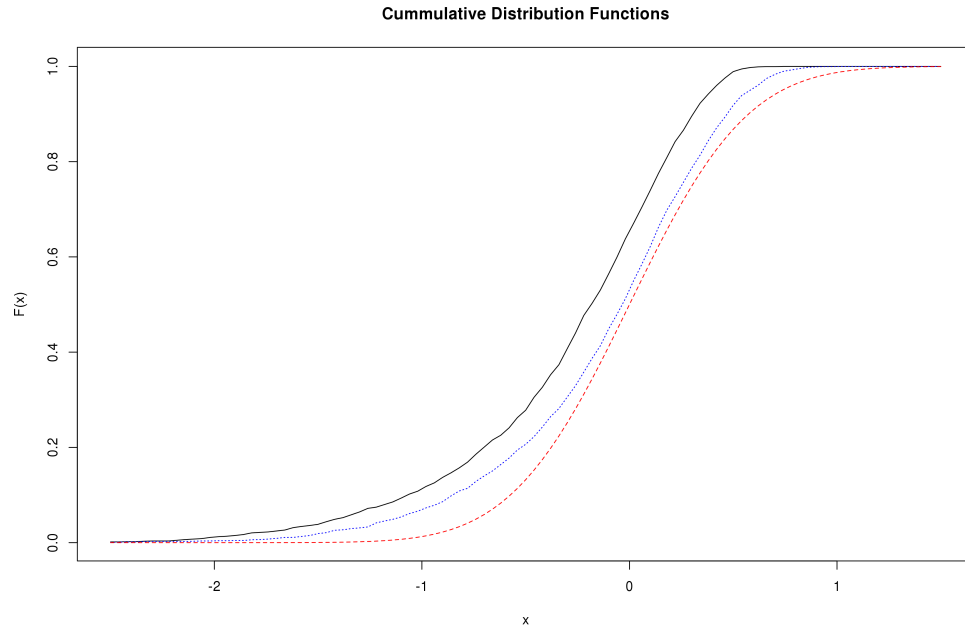


Figure 3: Example 1 CDFs for the true (solid, black), bootstrapped (dotted, blue), and asymptotic (dashed, red) error distributions.

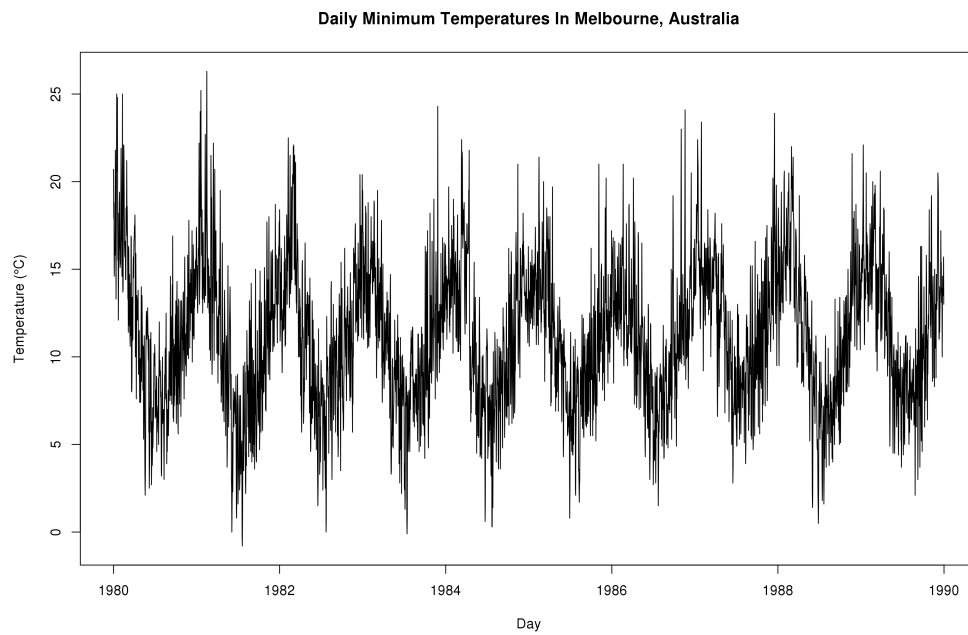


Figure 4: The Melbourne temperatures time series used in the application.

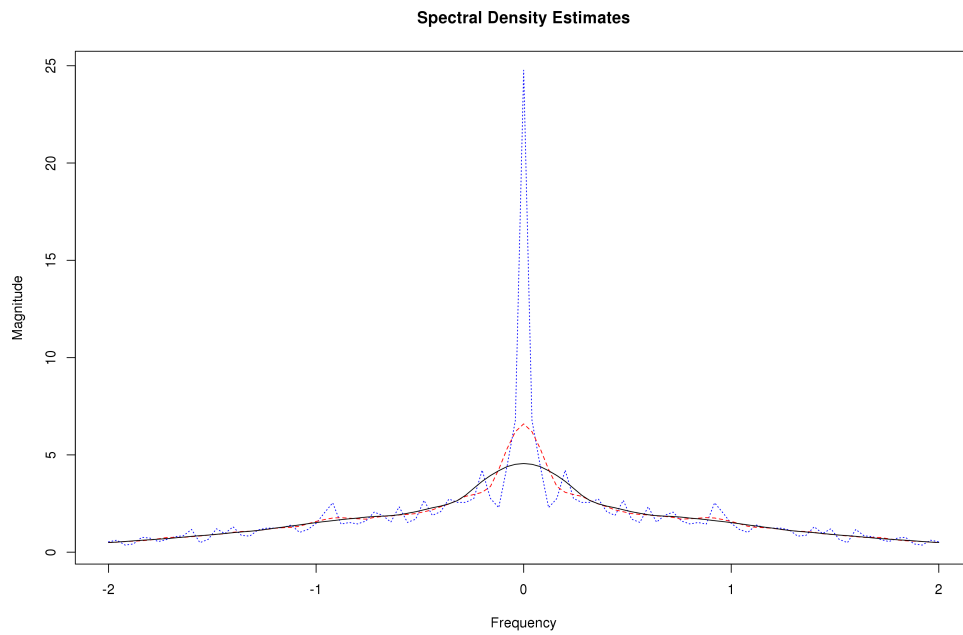


Figure 5: Spectral density estimates for the Melbourne temperature time series, with bandwidth 0.01 (dotted, blue), 0.05 (dashed, red), and 0.1 (solid, black).