BOOTSTRAPPING KERNEL SPECTRAL DENSITY ESTIMATES WITH KERNEL BANDWIDTH ESTIMATION

Abdelhak M Zoubir

Curtin University of Technology
GPO Box U1987, Perth WA 6845, Australia
e-mail: zoubir@ieee.org

ABSTRACT

We address the problem of confidence interval estimation of spectral densities using the bootstrap. Of special interest is the choice of the kernel global bandwidth. First, we investigate resampling based techniques for the choice of the bandwidth. We then address the question of whether the accuracy of the distributional bootstrap estimation is influenced by using the resample version, rather than the sample version of an empirical bandwidth. Aligned with recent results on non-parametric probability density estimation, we found that varying an empirical bandwidth across resamples is largely unnecessary and thus, the computational burden is greatly reduced while maintaining estimation accuracy.

1. INTRODUCTION

Non-parametric spectral density estimation has been extensively investigated in the signal processing as well as the statistical literature. However, assessing accuracy when little is known about the statistical distributional properties of the signal or when only a small number of observations is available and large sample theory does not apply has found only little coverage. The issue is of great importance because many signals encountered in real-life applications are non-Gaussian and/or non-stationary. Non-stationarity can be relaxed by limiting the number of observations to a small one and assuming (quasi) stationarity within the observation interval. This leads to the problem of spectral resolution and thus, the choice of the kernel bandwidth becomes fundamental. Although this has been largely investigated [1, 2], most results are asymptotic and inapplicable when the signal is non-Gaussian.

The objective in this paper is to propose an automatic choice of bandwidth in kernel spectral density estimation using the bootstrap [3, 4]. Further, we propose the construction of confidence intervals under the above conditions. We address the question of whether the accuracy of the confidence intervals found using the bootstrap are influenced by using the resample version, rather than the sample version of an empirical bandwidth. The approach we use in the latter has been motivated by the recent work by Hall *et al.* [5] in the context of kernel *probability* density estimation.

The paper is organised as follows. In Section 2, we present the data model and define the problem. In Section 3, we discuss confidence interval estimation for spectral densities, using a residual based method, rather than a block method. In Section 4, we address the problem of bandwidth selection and highlight its importance in view of accuracy of distributional properties of kernel spectral densities in Section 5, before we conclude.

2. DATA MODEL

Let $X_1, \ldots, X_T$ be observations from a real-valued, discrete-time, zero-mean, stationary signal $X_t,\ t \in \mathbb{Z}$, whose variance is finite and has spectral density

$$C_{XX}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \mathsf{E} X_0\, X_{|\tau|} e^{-j\omega\tau}\,, \quad -\infty < \omega < \infty\,.$$

Denote the periodogram by

$$I_{XX}(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t e^{-j\omega t} \right|^2, \quad -\pi \le \omega \le \pi\,.$$

A kernel spectral density estimate is given by

$$\hat{C}_{XX}(\omega;\, h) = \frac{1}{Th} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_{XX}(\omega_k)\,, \qquad (1)$$

where the kernel $K(\cdot)$ is a known symmetric, non-negative, real-valued function, h is its bandwidth and N denotes the largest integer less than or equal to $T/2$. The discrete frequencies ω_k are given by $\omega_k = 2\,\pi\,k/T$, $-N \le k \le N$.

The performance of $\hat{C}_{XX}(\omega; h)$ depends on the kernel bandwidth h; this dependence is stressed in the notation in Eq. (1). The objective in this work is to assess the accuracy of $\hat{C}_{XX}(\omega;\, h)$ when the distribution of $X_t,\ t \in \mathbb{Z}$, is unknown and T is small. Further, we wish to optimise the parameter h and investigate its effects on the distributional approximation of $\hat{C}_{XX}(\omega;\, h)$.

3. BOOTSTRAPPING SPECTRAL DENSITIES

There exist several methods for approximating the distribution of spectral density estimates and confidence intervals for $C_{XX}(\omega)$ based on an estimate $\hat{C}_{XX}(\omega; h)$. One approach uses the blocks of blocks bootstrap, which was developed for weakly dependent and stationary observations [6]. Alternate methods for bootstrapping dependent data [7] may also be used. The difficulty with these techniques, however, is the choice of parameters that define a block. Throughout the paper we will use a residual based method for bootstrapping spectral density estimates, see for example [8]. We explore the fact that the spectral density is a scale parameter of the asymptotic distribution of $I_{XX}(\omega)$. Thus, we interpret the spectral estimation problem as an approximate multiplicative regression problem, where

$$I_{XX}(\omega_k) = C_{XX}(\omega_k) \cdot \varepsilon_k\,, \quad k = 1, \ldots, N\,.$$

The real-valued residuals ε_k, $k = 1, \ldots, N$ are, for large N, approximately independent and identically distributed [1]. We will assume this asymptotic result holds and resample from $\varepsilon_1, \ldots, \varepsilon_N$. For this, we will also need an initial estimate $\hat{C}_{XX}(\omega; h_i)$, say, of the form given in Eq. (1) because the true spectral density $C_{XX}(\omega)$ is unknown. Its bandwidth h_i is arbitrary and may not coincide with h. In the resampling step, another kernel spectral density estimate $\hat{C}_{XX}(\omega; g)$ is used to approximate the distribution of $\hat{C}_{XX}(\omega; h)$. Its form is equivalent to Eq. (1) and its "resampling" bandwidth g may not be the same as h_i or h. A discussion on the spectral density bandwidth will be given in Section 4. In Table 1, we give the single steps of the bootstrap procedure based on residuals. We note that similarly to additive regression where we centre

Table 1. The Bootstrap procedure.

Step 0. *Data Collection*. Collect the data $X_1, \ldots, X_T$ and detrend it by substracting the sample mean.

Step 1. *Initial Estimate*. Choose an $h_i > 0$ and compute

$$\hat{C}_{XX}(\omega;\, h_i) = \frac{1}{Th_i} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h_i}\right) I_{XX}(\omega_k)\,.$$

Step 2. *Compute Residuals*. Calculate the residuals

$$\hat{\varepsilon}_k = \frac{I_{XX}(\omega_k)}{\hat{C}_{XX}(\omega_k; h_i)}, \qquad k = 1, \ldots, N\,.$$

Step 3. *Rescaling*. Rescale the empirical residuals to

$$\tilde{\varepsilon}_k = \frac{\hat{\varepsilon}_k}{\hat{\varepsilon}_\bullet}, \quad k = 1, \ldots, N, \quad \hat{\varepsilon}_\bullet = \frac{1}{N} \sum_{k=1}^{N} \hat{\varepsilon}_k\,.$$

Step 4. *Resampling*. Draw independent bootstrap residuals $\tilde{\varepsilon}_1^*, \ldots, \tilde{\varepsilon}_N^*$ from the empirical distribution of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 5. *Bootstrap Estimates*. With a bandwidth g, find

$$\begin{aligned} I_{XX}^*(\omega_k) &= I_{XX}^*(-\omega_k) = \hat{C}_{XX}(\omega_k; g)\, \tilde{\varepsilon}_k^*\,, \\ \hat{C}_{XX}^*(\omega; h, g) &= \frac{1}{T\,h} \sum_{k=-N}^{N} K\left(\frac{\omega - \omega_k}{h}\right) I_{XX}^*(\omega_k)\,. \end{aligned}$$

Step 6. *Confidence Bands Estimation*. Repeat **Steps 4-5** a large number of times and find c_U^* (and proceed similarly for c_L^*) such that

$$\mathsf{Pr}_*\left(\sqrt{T\,h}\,\frac{\hat{C}_{XX}^*(\omega; h, g) - \hat{C}_{XX}(\omega; g)}{\hat{C}_{XX}(\omega; g)} \le c_U^*\right) = \alpha\,,$$

that is $\{1 + c_U^*(T\,h)^{-1/2}\}^{-1} \hat{C}_{XX}(\omega; h)$ is the upper bound of an $(1 - 2\alpha)\%$-confidence interval for $C_{XX}(\omega)$.

residuals to avoid an additional bias [3], the re-scaling of the residuals in **Step 3** has a similar effect in multiplicative regression. We will set in **Step 5** $I_{XX}^*(0) = 0$, the periodogram at $\omega_k = 0$ of a mean corrected sample.

Under some regularity conditions, one can show that the Mallows distance between the pivotal quantity

$$\sqrt{Th}\{\hat{C}_{XX}(\omega; h) - C_{XX}(\omega)\}/C_{XX}(\omega)$$

and its bootstrap approximation,

$$\sqrt{Th}\{\hat{C}_{XX}^*(\omega; h, g) - \hat{C}_{XX}(\omega; g)\}/\hat{C}_{XX}(\omega; g)\,,$$

approaches zero in probability [8]. We conclude this section with an example.

Consider the autoregressive (AR) processes of order 5, X_t and Y_t defined as $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + N_t$ and $Y_t = Y_{t-1} - 0.7Y_{t-2} - 0.4Y_{t-3} + 0.6Y_{t-4} - 0.5Y_{t-5} + U_t$, where N_t are independently standard normally distributed variates and U_t are independently uniformly distributed variates on the interval $[-2.5, 2.5)$. Let $T = 256$ and consider the estimation of the spectral density of X_t at frequencies $\omega_k = 2\pi k/256$ for $k = 41, 42, 43, 66, 67, 68, 83, 84, 85$, which correspond to two peaks and the trough between both peaks (see Figure 1). We ran the algorithm of Table 1 with a Bartlett-Priestley kernel [2] for several values of h. A good global bandwidth was found to take a value somewhere around $h = 0.1$ (see Section 4). The confidence interval approximation results are presented in Table 2 and are based on 1000 replications. For the sake of comparison, we also include the χ^2 approximation, in which the $100\alpha\%$ confidence interval for the spectral density is approximated by

$$\frac{2\lfloor hT \rfloor \hat{C}_{XX}(\omega; h)}{\chi^2_{2\lfloor hT \rfloor}\left(\frac{1+\alpha}{2}\right)} < C_{XX}(\omega) < \frac{2\lfloor hT \rfloor \hat{C}_{XX}(\omega; h)}{\chi^2_{2\lfloor hT \rfloor}\left(\frac{1-\alpha}{2}\right)},$$

where $\chi^2_\nu(\alpha)$ is such that $\mathsf{Pr}(\chi^2_\nu < \chi^2_\nu(\alpha)) = \alpha$ [1]. A typical result of confidence bands of the spectral density of X_t is shown in Figure 1 along with the true density. We proceed similarly in the

Table 2. Performance of the residual based bootstrap method and the χ^2 method for a Gaussian AR(5) process.

	Residual Method				χ^2 Method			
ω_k	Lo.	Cov.	Up.	Length	Lo.	Cov.	Up.	Length
1.00	0	918	82	5.78	21	902	77	5.39
1.03	0	906	94	5.86	31	894	75	5.49
1.05	0	914	86	5.89	18	893	89	5.51
1.61	1	982	17	1.79	60	900	40	1.57
1.64	1	972	27	1.78	52	911	37	1.56
1.66	2	972	26	1.76	66	896	38	1.56
2.03	2	916	82	2.59	31	898	71	2.37
2.06	0	922	78	2.60	26	905	69	2.40
2.08	0	928	72	2.60	26	915	59	2.40

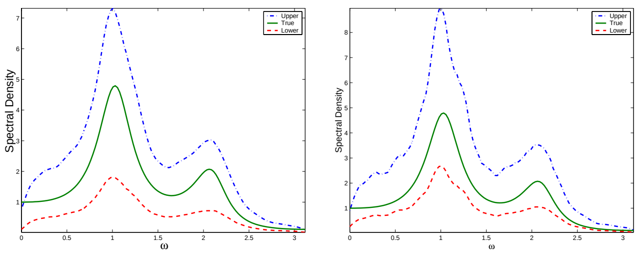

Fig. 1. 95% confidence bands for the spectral density of X_t using the residual based method (left) and the χ^2 approximation (right).

non-Gaussian case and consider the frequencies $\omega_k = 2\pi k/256$

Table 3. Performance of the residual based bootstrap method and the χ^2 method for an AR(5) process driven by non-Gaussian noise.

	Residual Method				χ^2 Method			
ω_k	Lo.	Cov.	Up.	Length	Lo.	Cov.	Up.	Length
0.63	0	916	84	35.45	17	894	89	35.06
0.66	0	885	115	35.31	11	858	131	35.05
0.68	0	882	118	35.50	11	889	100	35.43
1.27	0	876	124	35.81	5	836	159	35.64
1.30	0	886	114	36.16	9	864	127	35.54
1.32	0	891	109	36.12	27	888	85	35.08
2.33	6	983	11	0.32	82	895	23	0.28
2.35	5	986	9	0.31	83	887	30	0.26
2.42	3	983	14	0.30	69	906	25	0.24

for $k = 26, 27, 28, 52, 53, 54, 95, 96, 97$, which correspond to two peaks and the minimum of the true spectral density (see Figure 2). The confidence interval results are presented in Table 3.

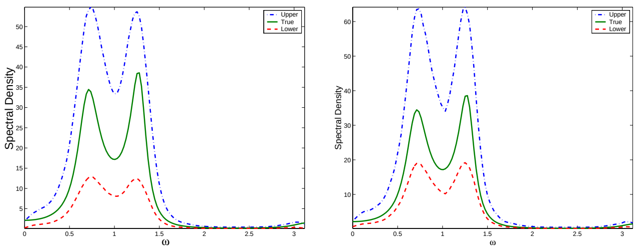

Fig. 2. 95% confidence bands for the spectral density of Y_t using the residual based method (left) and the χ^2 approximation (right).

4. BANDWIDTH OPTIMISATION

Appropriate selection of h is essential to achieving good estimates. Several criteria such as the integrated mean squared error have been proposed for choosing h in various context, such as kernel probability density functions. One widely used measure of the local performance of spectral density estimates is the mean squared percentage error [2], defined as

$$\eta(\omega; h) = \mathsf{E}\{(\hat{C}_{XX}(\omega; h) - C_{XX}(\omega))/C_{XX}(\omega)\}^2.$$

Because $C_{XX}(\omega)$ is unknown, one cannot calculate the minimising h_0. Under some assumptions [1, 2] on X_t and on the kernel K, $\eta(\omega; h)$ is minimised asymptotically if h is chosen to vanish at a rate $T^{-1/5}$. Using the bootstrap we can locally optimise the bandwidth h. We estimate $\eta(\omega; h)$ by its bootstrap counterpart,

$$\eta^*(\omega; h) = \mathsf{E}_*\{(\hat{C}^*_{XX}(\omega; h, g) - \hat{C}_{XX}(\omega; g))/\hat{C}_{XX}(\omega; g)\}^2,$$

and we choose the optimal bandwidth $\hat{h}^*_0$ such that

$$\eta^*(\omega; \hat{h}^*_0) = \min_{h \in B_T} \eta^*(\omega; h),$$

where B_T is an interval of bandwidths which shrinks to zero at the optimal rate $T^{-1/5}$, i.e., $B_T = [aT^{-1/5}, bT^{-1/5}]$, where $0 < a < b < \infty$ are suitable constants. One can avoid resampling in the minimisation if one explores the asymptotic independence of the bootstrap residuals $\tilde{\varepsilon}^*_k$ and $\mathsf{E}_*\tilde{\varepsilon}^*_k = 1$, $k = 1, \ldots, N$. Franke and Härdle [8] show that in this case a closed form expression exists for $\eta^*(\omega; h)$, which can be minimised with respect to h. The minimising $\hat{h}^*_0$ is (under some regularity conditions) consistent in that $T^{1/5}(\hat{h}^*_0 - h_0)$ converges to zero and the ratio $\eta^*(\omega; \hat{h}^*_0)/\eta(\omega; h_0)$ tends to 1 in probability as $T \to \infty$.

We focus our attention to the problem of global bandwidth optimisation and use the averaged mean squared percentage error

$$\bar{\eta}(h) = \frac{1}{N}\sum_{k=1}^{N} \eta(\omega_k; h),$$

which is to be minimised with respect to h. We do not resort to the approximation for $\eta^*(\omega_k; h)$ given in [8] and implement the bootstrap approach for finding $\hat{h}^*_0$. For simplicity, we will make no distinction between h and g (see Table 1) and proceed as in Table 4, where **Steps 0-3** are identical to the ones given in Table 1. The choice of the range of h used for the optimisation depends

Table 4. The Bootstrap procedure for bandwidth optimisation.

Step 4. *Bandwidth Optimisation.* For $h_L < h < h_U$,

Step 4.1. *Resampling.* Draw independent bootstrap residuals $\tilde{\varepsilon}^*_1, \ldots, \tilde{\varepsilon}^*_N$ from the empirical distribution of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_N$.

Step 4.2. Find

$$I^*_{XX}(\omega_k) = I^*_{XX}(-\omega_k) = \hat{C}_{XX}(\omega_k; h)\,\tilde{\varepsilon}^*_k,$$

$$\hat{C}^*_{XX}(\omega_k; h) = \frac{1}{T\,h}\sum_{j=-N}^{N} K\left(\frac{\omega - \omega_j}{h}\right) I^*_{XX}(\omega_k).$$

Step 4.3. Calculate

$$\zeta^*(h) = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{\hat{C}^*_{XX}(\omega_k; h) - \hat{C}_{XX}(\omega_k; h)}{\hat{C}_{XX}(\omega_k; h)}\right\}^2$$

Step 4.4. Repeat Steps 4.1-4.3 a large number of times and compute $\mathsf{E}_*\zeta^*(h)$

Step 5. Find the minimising $\hat{h}^*_0$ for $\mathsf{E}_*\zeta^*(h)$

on *a priori* knowledge about the smoothness of $C_{XX}(\omega)$. In the absence of any information, the interval $[h_L, h_U]$ should be large enough to cover a broad range of bandwidths. However, there is a trade-off between spectral resolution and computational expense.

We use the same example as in Section 3 and run a bootstrap based bandwidth optimisation. We did run the algorithm of Table 4 in 0.01 increments for h. We used a total of 200 bootstrap runs in **Step 4**. The results are given in Figure 3 for both X_t and Y_t. They show that there is a distinct minimum for h, which in both cases is around 0.1. For $T = 128$ samples we found that $\hat{h}$ lies between 0.09 and 0.1, while for $T = 256$ it was found to be between 0.092 and 0.12. An increase of the number of bootstrap runs does not affect the results. If one is to approximate the distribution of the kernel spectral density estimate, including bandwidth estimation, the additional computational expense due to the optimisation of h is minimal.

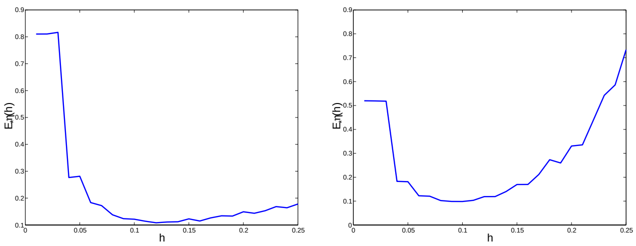

Fig. 3. 95% Average mean squared percentage error vs. kernel spectral bandwidth h for X_t (left) and Y_t (right).

5. ACCURACY ANALYSIS

The Bootstrap is a means of estimating the distributional properties of estimates using resampling. A bootstrap version of the population-sample relationship is constructed in that each component that depends on the population is replaced by its sample version, and each part that depends on the sample is replaced by its resample counterpart [3]. Many examples such as bootstrapping means and variances show that if one fails to adhere to this rule, the accuracy of the bootstrap approximation, which is of second order (i.e., the distributions are within $O_p(T^{-1})$ apart), quickly degrades. In the case of estimating $C_{XX}(\omega)$ using a kernel estimate, $\hat{C}_{XX}(\omega; h)$ is determined by a bandwidth h. For simplicity, let us assume that $g = h$. In Section 4 we discussed the estimation of h, which is calculated from the data $X_1, \ldots, X_T$. In that case, h is not fixed, but depends on the data, and thus, we denote it by $\hat{h}$. If in kernel spectral density estimation we require to assess the accuracy or determine the distribution of the estimate of $C_{XX}(\omega)$ along with the estimation of a pilot estimate $\hat{h}$ for h, as discussed in the previous section, the bootstrap dictates that the estimate $\hat{C}^*_{XX}(\omega; h)$ should be computed using $h = \hat{h}^*$ of $\hat{h}$. It is well-known that this reduces the order of error associated with the bootstrap approximation to a distribution.

Recently Hall *et al.* [5] showed that for kernel probability density estimation replacing $\hat{h}$ by $\hat{h}^*$ when computing the bootstrap estimator does not necessarily improve the order of accuracy of confidence procedure based on the bootstrap. Unlike the percentile-t approach [4], it will not usually improve performance by an order of magnitude. The reason is mainly that standard bootstrap methods are unable to capture the bias of a curve estimate. Also, pivoting [9], which ensures the order of accuracy, is not reproduced.

Following [5], one can show that similar conclusions can be drawn for spectral densities. Indeed $\hat{C}^*_{XX}(\omega; \hat{h}) - \hat{C}_{XX}(\omega; \hat{h})$ as well as $\hat{C}^*_{XX}(\omega; \hat{h}^*) - \hat{C}_{XX}(\omega; \hat{h})$ are first order accurate. Extensive simulations have confirmed this result. In Figure 4, we show a typical result of confidence interval estimation when using the plug-in method, i.e., using $\hat{h}^*$ at each resampling step against using a pilot estimate $\hat{h}$. Here, we used the same parameters as in Section 3. From the figure, it can be seen that the results are in close agreement. The next step will be to compare these results with percentile-t methods, in which we studentise the statistics [4]. Findings will be presented elsewhere.

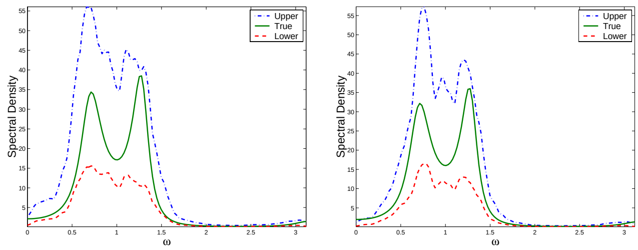

Fig. 4. 95% confidence bands for the spectral density of Y_t using the plug-in method (left) and using a pilot estimate $\hat{h}$ (right).

6. DISCUSSION

We have considered the problem of kernel spectral density estimation. First, we presented a bootstrap based method for confidence interval (distributional) estimation. The method is based on resampling residuals and thus does not require block optimisation. Then, we proposed a bootstrap method for the choice of the kernel global bandwidth. The method does not rely on asymptotic findings and results show good performance with little computational burden. We then have addressed the question of whether the accuracy of the distributional bootstrap estimation is influenced by using the resample version, rather than the sample version of an empirical bandwidth. Aligned with recent results on non-parametric probability density estimation it has been found that varying an empirical bandwidth across resamples is largely unnecessary and thus, the computational burden is greatly reduced while maintaining estimation accuracy.

7. REFERENCES

[1] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day San Francisco, 1981.

[2] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press London, 1981.

[3] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, 1993.

[4] A.M. Zoubir and B. Boashash, "The Bootstrap: Signal processing Applications," *IEEE Signal Processing Magazine*, vol. 15, pp. 56–76, 1998.

[5] P. Hall and K.-H. Kang, "Bootstrapping Nonparametric Density Estimators with Empirically Chosen Bandwidths," *The Annals of Statistics*, vol. 29, pp. 1443–1468, 2001.

[6] D. N. Politis, J. P. Romano, and T.-L. Lai, "Bootstrap Confidence Bands for Spectra and Cross-Spectra," *IEEE Transactions on Signal Processing*, vol. 40, pp. 1206–1215, 1992.

[7] D. Politis, J. Romano, and M. Wolf, *Subsampling*, Springer-Verlag, 1999.

[8] J. Franke and W. Härdle, "On Bootstrapping Kernel Estimates," *The Annals of Statistics*, vol. 20, pp. 121–145, 1992.

[9] A. M. Zoubir, "Multiple Bootstrap Tests and Their Application," in *Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Proc., ICASSP-94*, Adelaide, Australia, 1994, vol. VI, pp. 69–72.