# Undercoverage of Wavelet-based Bootstrap Confidence Intervals for a Mean Difference

*Yuqing Jia*

College of Management, Henan University of Technology,Zhengzhou, 450007, China

yuqingjia@126.com

**Abstract:** The decorrelating property of the discrete wavelet transformation (DWT) appears to be a valuable feature that avoids estimating the correlation structure in the original data space by bootstrap resampling the DWT. Data analysts have been using these so-called wavestrap methods in recent years without theoretical or empirical proof of their validity. However, our simulation studies show that these wavestraps yield undercoverage of parameters of interest for a mean difference. Thus, the wavestrap method is not preferred in obtaining resamples related to mean structure and should be used with caution. The reasons for these undercoverages are also discussed in this paper.

**Keywords:** discrete wavelet transformation (DWT), inverse discrete wavelet transform (IDWT), bootstrap

## 1. INTRODUCTION

The wavelet transform has emerged as a powerful mathematical tool for decomposing a function f(t) in terms of its time and frequency components. The Fourier transform is another tool for obtaining the frequency components of a function of time. The wavelet transform is superior to the Fourier transform when one wants the localization in time as well as the frequency localization for nonstationary signals or events. Wavelet transforms have been drawing much attention from statisticians since Donoho and his coauthors introduced wavelet denoising and shrinkage (Donoho,1993; Donoho and Johnstone, 1994; Donoho, 1995; Donoho et al.,1995).

The discrete wavelet transform (DWT) is discussed in Section 2. The DWT can be computationally efficient using Mallat's pyramid algorithm and the algorithm can be easily generalized to higher dimensional DWT's (Mallat, 1989a, Mallat, 1989b). One desirable property of the DWT is the decorrelating property of the wavelet coefficients (Frazier et al., 1991; Zhang and Walter, 1994). This decorrelating property in the two dimensional setting provides the basic support for the enhanced False Discovery Rate which reduces the number of local hypotheses in wavelet domain and proves to have more power than classical False Discovery Rate (Shen et al., 2002).

Resampling techniques including bootstrap and permutation test are discussed in Section 3. Bootstrap methods work well and preserve good second-order properties when the data are uncorrelated (Efron, 1993). The classical bootstrap methods fail when data are correlated. The decorrelating property of wavelet coefficients suggests that valid resamples from correlated data can be obtained by transforming the original data to the wavelet domain and resampling decorrelated wavelet coefficients. Bullmore et al. (2001) proposed this new resampling method based on independent permutations of wavelet coefficients of the observed time series. Their simulations show that the resampled time series have very similar autocorrelation functions to the original. Breakspear et al. (2004) extended the wavelet-based resampling method to identify dynamic interactions between brain regions from fMRI data. In a recent paper, Whitcher (2004) proposed that the bootstrap resamples of the approximately uncorrelated wavelet coefficients (wavestrap) can give a proper bootstrap distribution of the statistic based on the inverse DWT (IDWT).

However, wavelet-based resampling is not completely sound. Simulations in Section 4 show that confidence intervals for means based on the wavelet resamples have serious undercoverage, even in the special case of white noise sequences.

## 2. DISCRETE WAVELET TRANSFORM

### 2.1 Simple Wavelets

0.2 in Wavelets are small waves. The wavelet functions usually have finite support unlike sine or cosine functions oscillating consistently for all arguments from-$\infty$ to $+\infty$. A wavelet function $\psi(\cdot)$ defined on the real line satisfies two conditions

(1) $\psi(\cdot)$ integrates to zero

$$\int_{-\infty}^{+\infty} \psi(x)\mathrm{dx} = 0 \qquad (1)$$

(2) $\psi^2(\cdot)$ is integrable

$$\int_{-\infty}^{+\infty} \psi^2(x)\mathrm{d}x < \infty \qquad (2)$$

Hence, it follows from Equation (2) that the wavelet function cannot have nonzero values on an infinite interval. Equation (1) suggests that a wavelet fluctuates above and below the real line and dies out at least asymptotically. Along with the constraint of Equation (2), a graph of any wavelet function will look like a small wave on a finite interval.

The simplest example of a wavelet function is the Haar wavelet. The Haar wavelet is named in honor of Alfred Haar who developed the Haar function in 1909. It is defined as:

$$\psi^H(x) = \begin{cases} 1/\sqrt{2}, & -1 < x \le 0 \\ -1/\sqrt{2}, & 0 < x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

The Harr wavelet make an excellent choice for illustration purpose due to its simplicity can be seen in Figure 1.

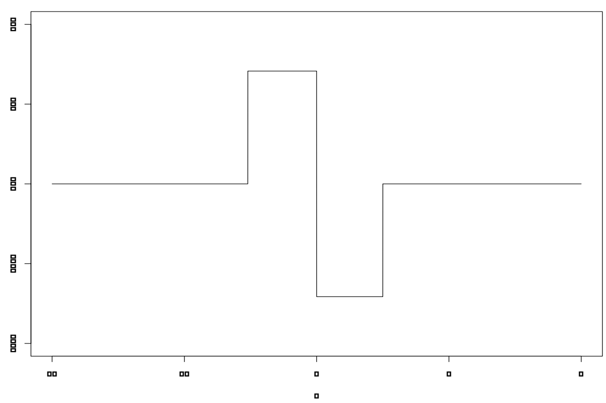

**Figure 1. Haar wavelet function: the haar wavelet is not continuous**

Among all the continuous wavelet functions, the probably most popular family of wavelets are Daubechies wavelets, which are named after Ingrid Daubechies who introduced it in great mathematical details (Daubechies, 1992). Daubechies wavelets are compactly supported as shown in Figure 2. Daubechies wavelets also have different vanishing moments, where a wavelet $\psi(x)$ with $n$ vanishing moments is defined to satisfy

$$\int_{-\infty}^{+\infty} x^i \psi(x) \mathrm{d}x = 0, \quad \text{for} \quad i = 0,1,...,n-1$$

A Daubechies wavelet with $n$ vanishing moments is of $n$ the-order (Daubechies, 1992), and denoted by D($n$) in this dissertation. The Haar wavelet is a special member of this family with zero vanishing moment. Daubechies wavelets of order two and four, denoted by D(2) and D(4), are plotted separately in Figure 2. A important advantage of Daubechies wavelets over other wavelet functions is that the Daubechies wavelet is the most compactly supported wavelet with a certain number of vanishing moments (Daubechies, 1988). For example, D(2) and D(4) have the minimal local support of all wavelet functions with two and four vanishing moments respectively; D(2) is supported over four time point and D(4) is supported over eight time points as given in Figure 2.

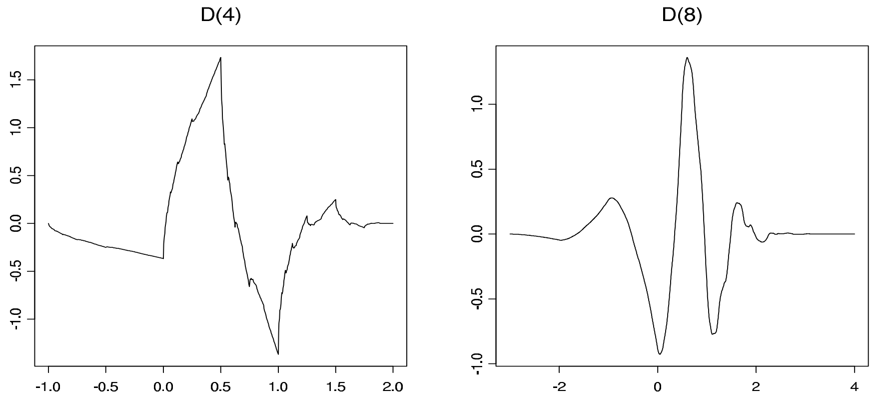


**Figure 2. Daubechies wavelet functions: Daubechies wavelet of the second-order, D(2), is plotted in the left panel; the fourth-order Daubechies wavelet, D(4), is in the right panel**

**2.2 Discrete wavelet transform**

Wavelet coefficients are derived from wavelet functions. For the Haar wavelet, its wavelet coefficients are $h_1 = 1/\sqrt{2}$ and $h_2 = -1/\sqrt{2}$. Its scaling coefficients are $l_1 = l_2 = 1/\sqrt{2}$ through the quadrature mirror filters that has the form of $l_m = (-1)^m h_{M-m}$, for $m = 1,2,...,M$ with $M$ =2 in this case. The $l_m$ s are also called low pass filters and $h_m$ s are called high pass filters, because the $l_m$ s decompose a signal into a low frequency portion, and $h_m$ s into a high frequency portion. The wavelet coefficients for $D(2)$ are defined to be $h_1 = \frac{1-\sqrt{3}}{4\sqrt{2}}, h_2 = \frac{-3+\sqrt{3}}{4\sqrt{2}}, h_3 = \frac{3+\sqrt{3}}{4\sqrt{2}}, h_4 = \frac{-1-\sqrt{3}}{4\sqrt{2}},$ and the corresponding scaling coefficients are defined to be $l_1 = -h_4, l_2 = h_3, l_3 = -h_2, l_4 = h_1$.

Given a time series {$\mathbf{y}(t): t = 0,1,...,N-1$, where $N$ is a power of 2}, the first level of coefficients from the one-dimensional discrete wavelet transform (DWT) of $\mathbf{y}(t)$ is expressed as

$$s_{t,1} = \sum_{m=0}^{m=M} l_m y(m+2t) \Rightarrow \mathbf{s}_1 = L\mathbf{y}(t)$$

$$d_{t,1} = \sum_{m=0}^{m=M} h_m y(2t+1-m) \Rightarrow \mathbf{d}_1 = H\mathbf{y}(t), \tag{3}$$

These filters have similar properties to $L$ and $H$, and they perform the IDWT which begins from $\mathbf{w}$ and works backwards recursively to obtain the original data $\mathbf{s}_0 = \mathbf{y}(t)$. This remarkably fast pyramid algorithm algorithm for decomposition and reconstruction was developed by Mallat (1989) from the theory of quadrature mirror filters. It provides fast, requiring $O(N)$ operations which is even faster than the Fast Fourier Transform (FFT) which is $O(Nlog_2N)$.

**3. WAVELET-BASED RESAMPLING**

Several authors (Frazier et al.,1991; Zhang and Walter,1994) have discussed the wavelet decorrelating property. This property is the basis for wavelet-based resampling techniques in many recent articles. The basic idea of wavelet decorrelating is that after the correlated time series is transformed into the wavelet domain via the DWT, the detail coefficients $\mathbf{d}_j$ s are approximately uncorrelated within each level and between levels. The decorrelated detail coefficients are then resampled within levels and transformed back to the time domain via the IDWT to obtain a resample of original time series.

Let $\mathbf{y}(t)$ be a time series of *N* points (*N* is a power of 2). The wavelet-based resampling algorithm is as follows:

Choose a suitable orthogonal wavelet basis.

Compute the DWT of $\mathbf{y}(t)$, i.e., DWT($\mathbf{y}$)= $\mathbf{w}$.

Apply the ordinary "naive" bootstrap (resampling with replacement) or permutation (resampling without replacement) to all the wavelet detail coefficients $\mathbf{d}_j$ within each level $j$ to obtain the resampled coefficients $\mathbf{d}_j^*$. Leave the scaling coefficients $\mathbf{s}_J$ unchanged.

Apply the inverse discrete wavelet transform to obtain the resampled time series $\mathbf{y}^*$.

Calculate the statistic of interest $\hat{\theta}^*$ on $\mathbf{y}^*$.

Repeat the above procedures $B$ times to estimate the sampling distribution of $\hat{\theta}^*$.

When a time series is not very long with usually 128 or 256 time points, Bullmore et al. (2001) recommend adopting the 4th-order Daubechies wavelet, D(4), to reduce the typical boundary effect of wavelets. D(2) and D(4) are used as the wavelet bases in the simulation study.

## 4. SIMULATIONS

### Simulations on Mean Difference

0.2inIn one simulation study, the parameter of interest is the mean difference $\mu_d$ between the first half $y_1$ and the second half mean $y_2$ of the time series $\mathbf{y}(t)$. Here, the time series is chosen to be 128 white noise points for simplicity. The fourth-order Daubechies wavelet basis is used and $B = 499$. Thus, the algorithm in Section 3 yields the resampling distribution of the mean differences from 499 resamples for each white noise realization. The $100\alpha/2$ th percentile, $\hat{\theta}^*_{((B+1)\alpha/2)}$ and $100(1-\alpha/2)$ th percentile, $\hat{\theta}^*_{((B+1)(1-\alpha/2))}$ are calculated from this resampling distribution. The "basic" confidence interval is obtained using the limits (Sec 5.2.1, Davison and Hinkley, 1997)

$$CI_{\alpha/2} = 2\hat{\theta} - \hat{\theta}^*_{((B+1)(1-\alpha/2))} \quad \text{and} \quad CI_{1-\alpha/2} = 2\hat{\theta} - \hat{\theta}^*_{((B+1)\alpha/2)}.$$

The studentized bootstrap confidence limits (Sec 5.2.1, Davison and Hinkley, 1997) is also evaluated here. The variance estimate $v$ of the mean difference has a closed formula. Thus $v$ and $\hat{\theta}$ for the original data as well as $v^*$ and $\theta^*$ for the resampled data can be calculated without any difficulty. For each resample the bootstrap version of $N(0,1)$ approximation is $z^* = (\theta^* - \theta)/\sqrt{v}$. The $B = 499$ resampled values of $z^*$ give the $100\alpha/2$ th percentile, $z^*_{((B+1)\alpha/2)}$ and $100(1-\alpha/2)$ th percentile, $z^*_{((B+1)(1-\alpha/2))}$. Thus the studentized confidence limits are

$$CI_\alpha/2 = \theta - \sqrt{v}z^*_{((B+1)(1-\alpha/2))}, \quad \text{and} \quad CI_{1-\alpha/2} = \theta - \sqrt{v}z^*_{((B+1)\alpha/2)}$$

Davison and Hinkley (1997) prefer this studentized limits to the previous basic method, and both methods are implemented here. The simulation is repeated 1000 times, and 1000 confidence intervals are obtained, therefore the coverage of the confidence interval can be estimated from the number of confidence intervals that cover the true mean difference zero.

In the simulation $\alpha = 0.10$, and therefore around 90% of the calculated confidence intervals should contain zero ($SE \approx 0.9\%$). Table 1 is the summary of the results for wavestrapping with either basic or studentized confidence limits as well as naive bootstrap method. Wavetrapping methods give coverage of about 60%, which is far below the nominal level 90%, while the naive bootstrap gives 89.8%, close to the nominal level. The naive bootstrap only works well when the observations are independent and identically distributed such as in this special case of white noise.

**Table 1**

| | Basic | Studentized | Naive Bootstrap |
|---|---|---|---|
| Bootstrap | 52.40% | 56.10% | 89.80% |
| Permutation | 58.10% | 62.60% | NA |

Table 1. Simulation results for wavelet boostrappings and naive bootstrap methods. The nominal confidence level is 90%. The column headings are the confidence-limit methods. The first column lists the resampling methods for wavelet coefficients in wavestrapping. The wavelet basis is the fourth-order Daubechies wavelet.

The reason for the serious under-coverage of wavestrapping is that the wavelet transform preserves the mean structure in the scaling coefficients. During the wavestrapping procedure, the scaling coefficients remain unchanged every time the detail coefficients are resampled. Therefore, the variability of the mean structure will not be sufficient in wavestrapping. In the above simulations, once a sequence of white noise $\mathbf{y}(t)$ is simulated, the overall mean $\bar{\hat{\mathbf{y}}}$ for every wavestrapping resample $\hat{\mathbf{y}}(t)$ remains the same as that of the original data $\bar{\mathbf{y}}$. Therefore, there is perfect negative correlation (-1.0) between the first half mean $\bar{y}_1$ and the second half mean $\bar{y}_2$ in the resample $\hat{\mathbf{y}}(t)$: $\bar{y}_1 = \bar{\mathbf{y}} - \bar{y}_2$, and the mean differences from each resample will have smaller standard deviation than they should ($\sqrt{\frac{1}{64}+\frac{1}{64}} \approx .17$). These are shown in Figure 4. The histograms in Figure 4 also show that the permutation-based wavestrapping performs better than the bootstrap-based wavestrapping.

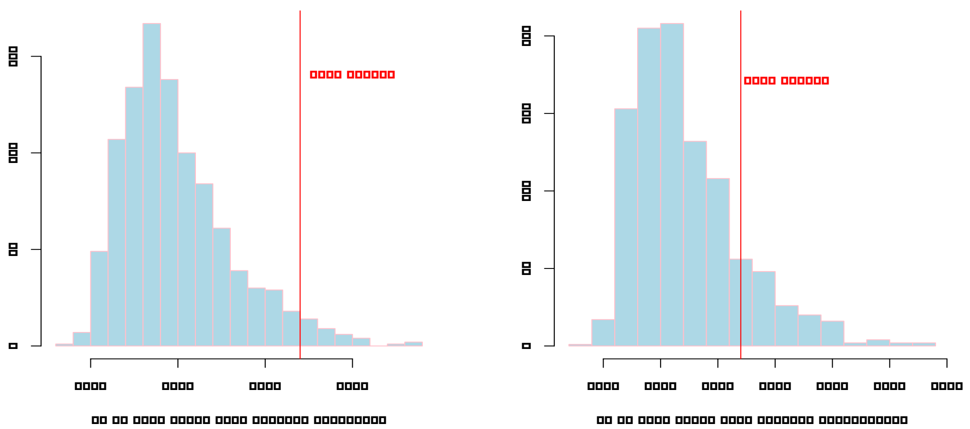

**Figure 3. Histograms of resampled standard deviations. The Left histogram is based on wavelet bootstrapping; the right one is for wavelet permutation**

**5. CONCLUSION**

The decorrelating property of DWT seems attractive, and bootstrapping wavelet coefficients of one-dimensional time series and higher dimensional images are tried by several researchers in neuroimaging and other research areas. This chapter has demonstrated that the wavelet-based bootstrap procedure is not appropriate when dealing with the inference on the mean structure. Therefore, the wavestrap procedures are not optimal nonparametric bootstrap methods on estimating mean structure of data and should be performed with careful consideration.

**REFERENCES**

[1]Breakspear, M., Brammer,M., Bullmore, E., Das, P. and Williams L. (2004). “Spatio-temporal Wavelet Resampling for Functional Neuroimaging Data.” Human Brain Mapping, 23(1), 1-25.

[2]Bullmore, E., Long, C, Suckling, J, Fadili, J, Calvert, G, Zelaya, F, Carpenter, T, Brammer, M. (2001). “Colored noise and computational inference in neurophysiological time series analysis: Resampling methods in time and wavelet domains.” Human Brain Mapping, 12, 61-78.

[3]Chui, C. (1997). Wavelets: A Mathematical Tool for Signal Analysis, SIAM Monographs on Mathematical Modeling and Computation. Philadelphia: Society for Industrial and Applied Mathematics.

[4]Daubechies, I. (1992). “Ten Lectures on Wavelets.” Volume 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.

[5]Donoho, D. (1993). “Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data.” Proceedings of Symposia in Applied Mathematics, 47, 173-205.

[6]Donoho, D. (1995). “De-noising by soft-thresholding.” IEEE Transactions on Information Theory, 41(3), 613-627.

[7]Donoho, D. and Johnstone, I. (1994). “Ideal spatial adaptation by wavelet shrinkage.” Biometrika, 81(3), 425-455.

[8]Whitcher, B.(2004). “Wavelet-Based Bootstrapping of Spatial Patterns on a Finite Lattice.” Computational Statistics and Data Analysis, To Appear.