Supplementary document for

“Generalized Fiducial inference for wavelet regression”

By Jan Hannig

Department of Statistics and Operations Research

The University of North Carolina at Chapel Hill

Chapel Hill, North Carolina 27599-3260, U.S.A.

hannig@email.unc.edu

Thomas C. M. Lee

Department of Statistics, The Chinese University of Hong Kong

Shatin, Hong Kong

tlee@sta.cuhk.edu.hk

February 2, 2009

This document contains supplementary material for the article “Generalized Fiducial Inference for Wavelet Regression” written by Hannig and Lee (henceforth denoted as HL). It consists of four sections. The first section presents a full derivation of the generalized fiducial density (12) of HL. The second section describes a practical Markov chain Monte Carlo algorithm for simulating fiducial samples from this generalized fiducial density. The third section gives the proof for Theorem 1 of HL, while the issue of constructing confidence sets is provided in the last section.

1. Derivation of the generalized fiducial density

The structural equation incorporating the additional random variables P is

$$(W, \Psi) = G(d, t, \sigma^2, I, Z, P)$$

where I is collection of indexes,

$$W_i = \begin{cases} d_i + \sigma Z_i & \text{if } i \in I, \\ \sigma Z_i & \text{if } i \in I^{\complement}, \end{cases} \qquad \Psi_k = \begin{cases} P_k & \text{if } k \in I, \\ t_k + P_k & \text{if } k \in I^{\complement}, \end{cases}$$

$Z = (z_1, \ldots, z_n)$ are independent $N(0,1)$ random variables and $P = (P_1, \ldots, P_n)$ random variables with continuous density satisfying $f_P(0) = q$ independent of Z and each other. We delay the discussion of the choice of q until the end of this section as it will control the amount of penalty introduced by P. We denote the observed values of W by w and assume that the observed values of Ψ are $(0, \ldots, 0)$.

Notice that if I is known and fixed the penalizing random variables Ψ do not have any influence on the distribution of the observed data W due to independence. However, the values of Ψ will influence the conditional distribution in (9) of HL.

Now we compute the set inverse of the structural equation. The element $(d, t, \sigma^2, I) \in Q(w, 0, z, p)$ if and only if

$$\sigma^2 = \left(\frac{w_j}{z_j}\right)^2 \text{ for randomly chosen } j \in I^{\complement},$$

$$d_i = 0,\ t_i = 0 - p_i \text{ for } i \in I^{\complement}, \quad \text{and} \quad d_k = w_k - \frac{w_j}{z_j} z_k,\ t_k = 0 \text{ for } k \in I,$$

provided

$$w_i = \frac{w_j}{z_j} z_i \text{ for all } i \neq j,\ i \in I^{\complement} \quad \text{and} \quad p_k = 0 \text{ for all } k \in I.$$

Thus, if there is no such I for which z and p would satisfy the last two equalities then $Q(w, 0, z, p) = \emptyset$.

Consider now fixed I and $j \in I^{\complement}$ and denote $m(I)$ as the number of elements in I. Let us compute the joint density of $S = (w_j/Z_j)^2$, $D_k = w_k - w_j Z_k / Z_j$, $k \in I$, and $R_i = w_j Z_i / Z_j$, $i \in I^{\complement}, i \neq j$, together with $T_i = 0 - P_i$ for $i \in I^{\complement}$ and P_k, $k \in I$. A routine calculation shows that this joint density evaluated at $S = h$, $D_k = d_k$, $P_k = 0$, $k \in I$, $R_i = w_i$, $i \in I^{\complement}, i \neq j$, and $T_i = t_i$, $i \in I^{\complement}$ is

$$f(h, d, t, I) = |w_j| e^{-\left\{\frac{(d_0 - w_0)^2}{2h} + \frac{\sum_{k\in I}(d_k - w_k)^2}{2h} + \frac{\sum_{i \in I^{\complement}} w_i^2}{2h}\right\}} \left\{2(2\pi)^{\frac{n}{2}} h^{\frac{n}{2}+1}\right\}^{-1} q^{m(I)} \prod_{i \in I^{\complement}} f_p(-t_i).$$

Since $d_i = 0$ for $i \in I^{\complement}$, this density can be used to compute for any set $A \subset \mathbb{R}^{n+1}$

$$A_{I,j}(dx) = P\left\{(S,D) \in A; R_i \in \left(w_k - \frac{dx}{2}, w_k + \frac{dx}{2}\right), i \in I^{\complement}, i \neq j; P_k \in \left(\frac{-dx}{2}, \frac{dx}{2}\right), k \in I\right\}$$
$$= \frac{dx^{n-1}q^{m(I)}}{2(2\pi)^{\frac{n}{2}}} \iint_A e^{-\left\{\frac{\sum_{k\in I}(d_k-w_k)^2}{2h} + \frac{\sum_{i\in I^{\complement}} w_i^2}{2h}\right\}} |w_j| h^{-(\frac{n}{2}+1)} \prod_{i\in I^{\complement}} \delta_0(d_i)\, dh\, dd + o(dx^{n-1}), \tag{1}$$

where δ_0 is the Dirac function. Notice that the t_i, $i \in I^{\complement}$ were integrated out.

Thus following (9) of HL and using the fact that j is picked at random from $I^{\complement}$, the conditional distribution in the definition of fiducial probability (9) of HL can be understood as

$$P\{(S,D) \in A, I\} = \lim_{dx\to 0} \frac{\sum_{j\in I^{\complement}} A_{I,j}(dx)}{n - m(I)} \left\{\sum_{I'\in\mathcal{I}} \frac{\sum_{j\in I'^{\complement}} A_{I',j}(dx)}{n-m(I')} + o(dx^{n-1})\right\}^{-1}, \tag{2}$$

where $\mathcal{I}$ contains all subsets of $\{1,\dots,n\}$ satisfying the tree constraint described in § 2·1 of HL and having at most $p_0 n$ elements; see § 3·1 of HL. We remark that in principle $Q(w, Z, P)$ can contain more than one element. Therefore the numerator of (2) had to be computed by inclusion and exclusion. This leads to the extra $o(dx^{n-1})$ as the intersection terms in the inclusion and exclusion formula are of a lower order than dx^{n-1}. This is due to the fact that more than the usual $n-1$ conditioning equations would have to be satisfied in order to have z and p that are compatible with more than one I.

From (1) and (2) we conclude that the generalized fiducial density is

$$r(h,d,I) = C^{-1} q^{m(I)} \frac{\sum_{j\in I^{\complement}} |w_j|}{n-m(I)} e^{-\left\{\frac{\sum_{k\in I}(d_k-w_k)^2}{2h} + \frac{\sum_{i\in I^{\complement}} w_i^2}{2h}\right\}} \left\{2(2\pi)^{\frac{n}{2}} h^{\frac{n}{2}+1}\right\}^{-1} \prod_{i\in I^{\complement}} \delta_0(d_i), \tag{3}$$

where

$$\begin{aligned} C &= \sum_{I'\in\mathcal{I}} \frac{q^{m(I')}}{2(2\pi)^{\frac{n}{2}}} \iint \frac{\sum_{j\in I'^{\complement}} |w_j|}{n-m(I')} e^{-\left\{\frac{\sum_{k\in I'}(d_k-w_k)^2}{2h} + \frac{\sum_{i\in I'^{\complement}} w_i^2}{2h}\right\}} h^{-(\frac{n}{2}+1)} \prod_{i\in I'^{\complement}} \delta_0(d_i)\, dh\, dd \\ &= \sum_{I'\in\mathcal{I}} q^{m(I')} \Gamma\left\{\frac{n-m(I')}{2}\right\} \frac{\sum_{j\in I'^{\complement}} |w_j|}{n-m(I')} \left\{ 2\pi^{\frac{n-m(I')}{2}} \left(\sum_{i\in I'^{\complement}} w_i^2\right)^{\frac{n-m(I')}{2}} \right\}^{-1}. \end{aligned}$$

Finally, we discuss the choice of q. The use of the penalizing random variables P led to the term $q^{m(I)}$ in (3), which, after taking logarithm, corresponds to an additive penalty term of $m(I)\log q$ to the log likelihood. In this article we choose $q = n^{-1/2}$. This leads to, on the log likelihood scale, the penalty of $-0{\cdot}5m(I)\log n$ matching the MDL penalty. Then equation (14) of HL now follows immediately. We also remark that other choices could be reasonable. For example the choice of $q = e^{-2}$ would lead to the AIC. penalty.

2. A Markov chain Monte Carlo algorithm for simulating fiducial samples

The generalized fiducial density $r(d, \sigma^2, I)$ in (14) of HL contains a normalizing constant C that is not computable in a closed form. Moreover, the number of choices for I in (14) of HL is exponentially increasing, which makes numerical integration infeasible. Due to these facts we used an Markov chain Monte Carlo procedure to generate a sample of $\tilde{d}_1, \ldots, \tilde{d}_M$ from the generalized fiducial distribution $r(d, \sigma^2, I)$. This procedure consists of two major steps.

Step 1: generation of I. Use the Metropolis–Hastings algorithm to generate a realization $\tilde{I}$ of I. To specify this algorithm, we need the probability mass function $p(I)$ of I, which from (14) of HL is calculated as

$$r(I) = C^{-1} e^{-\frac{m(I)\log n}{2}} 2^{-\frac{1}{2}} \pi^{-\frac{n-m(I)}{2}} \left(\sum_{i \in I^{\complement}} w_i^2 \right)^{-\frac{n-m(I)}{2}} \Gamma\left\{ \frac{n-m(I)}{2} \right\} \frac{\sum_{i\in I^{\complement}} |w_i|}{n-m(I)}. \tag{4}$$

We also need the proposal distribution, which is defined as follows. Suppose that our chain is in state I and we generate a proposed state I^p. In order to maintain the tree constraint we use the following rules. First, with equal probability, locate a candidate non-thresholded coefficient node that satisfies exactly one of the following four possibilities:

(a) a node with a right non-thresholded child (and the left child is thresholded),

(b) a node with a left non-thresholded child (and the right child is thresholded),

(c) a node at the highest resolution (i.e. lowest level in the tree),

(d) and a node that is not at the highest resolution and have no children (i.e. both children are thresholded).

If the selected candidate satisfies case (a) (case (b)), its left (right) child will be turned non-thresholded. If the selected candidate is from case (c), this candidate will be made thresholded. For case (d), with equal probability one of the following three actions will be taken: (i) the left child is turned non-thresholded, (ii) the right child is turned non-thresholded, (iii) remove this node. Thus at the end we are left with a proposed I^p that has either one more or one less non-thresholded coefficient than the original state I. Once the proposal is generated, we then follow the usual Metropolis-Hastings algorithm to accept or reject the proposal.

Step 2: generation of σ^2 and d. Given $\tilde{I}$ and (14) of HL, it is straightforward to generate an observation $\{\tilde{\sigma}^2, \tilde{d}\}$ from $r(\sigma^2, d|\tilde{I})$. In particular $\tilde{\sigma}^2$ can be generated as

$$\sigma^2 = \sum_{i \in \tilde{I}^{\complement}} w_i^2 / \chi^2_{n-m(\tilde{I})}$$

where $\chi^2_{n-m(\tilde{I})}$ has the chi-squared distribution with $n - m(\tilde{I})$ degrees of freedom. Then given $\tilde{I}$ and σ^2, we can generate $\tilde{d}_i$ from

$$\tilde{d}_i \mid \tilde{I}, \sigma^2 \sim \begin{cases} N(w_i, \sigma^2) & \text{if } i \in \tilde{I}, \\ 0 & \text{if } i \in \tilde{I}^{\complement}. \end{cases}$$

Throughout all of our numerical experiments the number of burn-in samples was 10,000. A total of 1,000 samples (i.e. $\tilde{d}_1, \ldots, \tilde{d}_{1000}$) were collected for every consecutive 200 samples after the burn-in period. When $n = 1024$, our implementation completes this procedure in about 30 seconds on a Intel(R) Core(TM)2 1.86GHz machine.

3. Proofs

In this section we develop our theoretical results that lead to Theorem 1. To set up the proper asymptotics, consider a sequence of rooted binary trees B_J (each with $n = 2^{J+1}$

elements) and label its nodes in a sequential matter with 1 through 2^{J+1}, level by level, left to right. Each subset of B_J can be identified by a set I of these indexes. Denote by $\mathcal{I}_J$ the collection of all subtrees of B_J satisfying the conditions introduced in § 3·1; i.e. $I \in \mathcal{I}_J$ if I represents a tree having at most $(1 - p_0)n$ elements.

At each node we observe a value w_i, $i = 1, \ldots, n$, the wavelet coefficient. The marginal fiducial probability $r(I)$ of each $I \in \mathcal{I}_J$ is given by (4). Recall that $r(I)$ depends on the observed values of the w_i. Thus we will want to show that the fiducial distribution converges weakly for almost all sequences of observed sets of wavelet coefficients $\{w_i\}_{i=1}^{2^{J+1}}$, $J = 1, 2 \ldots$

We will now prove two lemmas establishing consistency of the generalized fiducial distribution. We first turn our attention to investigating what happens if the truth is a zero function.

Lemma 1. *Let w_i, $i = 1, \ldots, n$ be independent $N(0, \sigma^2)$ and $I_0 = \emptyset$ be the empty tree (corresponding to all coefficents being thresholded). Then $r(I_0) \to 1$ almost surely.*

Proof. Let us denote $x_{(1)} > x_{(2)} > \cdots > x_{(n)}$ the order statistics of $w_1^2, \ldots, w_n^2$. It is well known that (Embrechts et al., 1997, Theorem 3.5.1)

$$x_{(1)} \leq 2\log(n) + 2\log\log n \tag{5}$$

eventually almost surely. Additionally, it follows from Davis & Resnick (1984), Theorem 5.1, that if $0 < \alpha < 1$ and $m = [n^\alpha]$ then

$$\sum_{i=1}^{m} x_{(i)} < 2(1-\alpha)m\log(n) \tag{6}$$

eventually almost surely.

Take $I \in \mathcal{I}_J$ and write $m = m(I)$ the number of elements of I. It follows that

$$\begin{aligned}\frac{r(I)}{r(I_0)} \leq & e^{-\frac{m}{2}\log n}(2\pi)^{\frac{m}{2}} \frac{n}{n-m} \frac{\sum_{i\in I^\complement}|w_i|}{\sum_{i=1}^n |w_i|} \left(\frac{\sum_{i\in I^\complement} w_i^2}{n-m}\right)^{\frac{m}{2}} \frac{\left(\frac{n-m}{2}\right)^{\frac{m}{2}}\Gamma\left(\frac{n-m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{\sum_{i=1}^n w_i^2}{\sum_{i\in I^\complement} w_i^2}\right)^{\frac{n}{2}} \\ \leq & e^{-\frac{1}{2}m\log n + C_1 m}\left(\frac{\sum_{i=1}^n w_i^2}{\sum_{i\in I^\complement} w_i^2}\right)^{\frac{n}{2}},\end{aligned}$$

where C_1 is a constant depending on p_0 and σ^2. Further compute

$$\left(\frac{\sum_{i=1}^n w_i^2}{\sum_{i\in I^\complement} w_i^2}\right)^{\frac{n}{2}} \le \exp\left\{\frac{n}{2}\log\left(1+\frac{\sum_{i\in I} w_i^2}{\sum_{i\in I^\complement} w_i^2}\right)\right\}$$
$$\le \exp\left(\frac{n}{2}\frac{\sum_{i\in I} w_i^2}{\sum_{i\in I^\complement} w_i^2}\right) \le \exp\left(\frac{n}{2}\frac{\sum_{i=1}^m x_{(i)}}{\sum_{i=m+1}^n x_{(i)}}\right).$$

Fix $\epsilon_1 > 0$. By (6) and the Strong Law of Large Numbers there are $q > 0$ and $v_1 > 0$ such that, if $n^{1/2+\epsilon_1} < m \le qn$, then

$$\exp\left(\frac{n}{2}\frac{\sum_{i=1}^m x_{(i)}}{\sum_{i=m+1}^n x_{(i)}}\right) \le \exp\left\{\left(\frac{1}{2}-v_1\right)m\log n\right\}$$

and consequently

$$\frac{r(I)}{r(I_0)} \le e^{-v_1 m\log n + C_1 m}. \tag{7}$$

Similarly, if $qn < m \le p_0 n$, then by strong law of large numbers there is $C_2 < \infty$ such that

$$\left(\frac{\sum_{i=1}^n w_i^2}{\sum_{i\in I^\complement} w_i^2}\right)^{\frac{n}{2}} \le \left(\frac{\sum_{i=1}^n x_{(i)}}{\sum_{i=p_0 n}^n x_{(i)}}\right)^{\frac{n}{2}} \le e^{C_2 n}$$

and consequently

$$\frac{r(I)}{r(I_0)} \le e^{-\frac{1}{2}m\log n + C_1 m + C_2 n}. \tag{8}$$

The case $0 < m \le n^{1/2+\epsilon_1}$ is more complicated. Notice that, unlike in the previous two cases, the ratio $r(I)/r(I_0)$ could grow unbounded. We will show however that, thank to the tree constraint, the probability of such an event is small. To this end, fix a small $\epsilon_2 > 0$ and consider I such that at most 1/3 of its $w_i^2 > x_{(n^{3/4+\epsilon_2})}$. Then

$$\frac{1}{2}\sum_{i\in I} w_i^2 \le \frac{m}{3}(\log n + \log\log n) + \frac{2m}{3}\left(\frac{1}{4}-\epsilon_2\right)\log n \le \left(\frac{1}{2}-\frac{2\epsilon_2}{3}\right)m\log n + m\log\log n.$$

Thus by Strong Law of Large Numbers there is $v_2 > 0$ such that

$$\frac{r(I)}{r(I_0)} \le e^{-v_2 m\log n + C_1 m}. \tag{9}$$

Let A_n be the event that all trees of size less then $n^{1/2+\epsilon_1}$ contain at at most 1/3 of its $w_i^2 > x_{(n^{3/4+\epsilon_2})}$. Since the number of binary trees of size m is the Catalan number $C_m \le 4^m$, Stanley (1999), equations (7), (8), and (9) imply that on A_n

$$\begin{aligned}\sum_{I\in\mathcal{I}}\frac{r(I)}{r(I_0)} \leq & 1 + \sum_{m=1}^{[n^{1/2+\epsilon_!}]} 4^m e^{-v_2 m\log n + C_1 m} \\ & + \sum_{m=[n^{1/2+\epsilon_1}]+1}^{[qn]} 4^m e^{-v_1 m \log n + C_1 m} + \sum_{m=[qn]+1}^{[p_0 n]} 4^m e^{-\frac{1}{2}m\log n + C_1 m + C_2 n} \\ \leq & 1 + r_n,\end{aligned}$$

where $r_n \to 0$.

The lemma will follow by Borel-Cantelli lemma provided that $\sum_{J=1}^{\infty} P(A^{\complement}_{2^{J+1}}) < \infty$, recall we are only interested in $n = 2^{J+1}$. To this end, fix a tree of size m. The probability that the arrangement of the w_i's is such that the fixed tree contains at most $1/3$ of $w_i^2 > x_{(n^{3/4+\epsilon_2})}$ is less then

$$p_{m,n} = \frac{\binom{[n^{3/4+\epsilon_2}]}{[m/3]}\binom{n}{[2m/3]}}{\binom{n}{[2m/3]}} \leq \frac{n^{(3/4+\epsilon_2)m/3} n^{2m/3}}{(n-m)^m} \leq \left(\frac{n}{n-m}\right)^m n^{-\left(\frac{1}{12}+\frac{\epsilon_2}{3}\right)m}.$$

By choosing ϵ_1 and ϵ_2 small enough, we can find $v_3 > 0$, so that for all n large enough and all $m < n^{1/2+\epsilon_1}$, $p_{m,n} \leq e^{-v_3 m \log n}$. From here

$$P(A^{\complement}_n) \leq \sum_{m=1}^{[n^{1/2+\epsilon_1}]} 4^m e^{-v_3 m \log n} \leq \frac{4 - (4n^{-v_3})^{n^{1/2+\epsilon_1}}}{n^{v_3} - 4}.$$

This implies that $\sum_{J=1}^{\infty} P(A^{\complement}_{2^{J+1}}) < \infty$ and the Lemma is proved. □

Notice that Lemma 1 implies that for almost all sequences of the observed data, the marginal generalized fiducial distribution of d converges weakly to 0 almost surely.

The discrete wavelet transform guarantees that, if some of the coefficients are not zero, they actually are increasing with n as $n^{1/2}$ (Donoho et al., 1995). This motivates the following lemma.

Lemma 2. *Let us assume that* $w_i = d_i n^{1/2} + \sigma z_i$*, where* z_i *are independent* $N(0,1)$*. Additionally assume that there is fixed, finite, non-empty tree* I_T *such that,* $d_i \neq 0$ *for all leaves of* I_T *and* $d_i = 0$ *for all* $i \notin I_T$*. Then* $r(I_T) \to 1$ *almost surely.*

Proof. Recall that the fiducial distribution is assumed to be supported by $\mathcal{I}$, the set of all trees having at most $p_0 n$ elements. Consider the set of all subtrees $\mathcal{J} = \{I_1, \ldots, I_K\}$ of the tree I_T that also have non-zero a_i on all their leaves. By definition, both the empty tree and I_T are members of $\mathcal{J}$. Moreover, for all $I_j \in \mathcal{J}$, define the set $\mathcal{I}_j$ of all supertrees of $I_j \in \mathcal{I}$ having exactly the same non-zero a_i as I_j. Notice that $\mathcal{I}_j,\ j = 1, \ldots K$ is a partition of $\mathcal{I}$ and therefore

$$\sum_{I \in \mathcal{I}} \frac{r(I)}{r(I_T)} = \sum_{j=1}^{K} \frac{r(I_j)}{r(I_T)} \sum_{I \in \mathcal{I}_j} \frac{r(I)}{r(I_j)}. \tag{10}$$

First notice that by the same argument as in the proof of Lemma 1, $\sum_{I \in \mathcal{I}_j} \frac{r(I)}{r(I_j)} \to 1$ for all $j = 1 \ldots, K$ almost surely. Second, consider the factor $r(I_j)/r(I_T)$. Denote the sizes of these trees m_j and m_T respectively and assume that $D = I_T \setminus I_j \neq \emptyset$ and $d = m_T - m_j > 0$. Compute

$$\frac{r(I_j)}{r(I_T)} \leq e^{\frac{d}{2}\log n}(2\pi)^{-\frac{d}{2}} \frac{n - m_T}{n - m_j} \frac{\sum_{i \in I_j^{\complement}} |w_i|}{\sum_{i \in I_T^{\complement}} |w_i|} \left(\frac{\sum_{i \in I^{\complement}} w_i^2}{n} \right)^{-\frac{d}{2}} \frac{\left(\frac{n}{2}\right)^{-\frac{d}{2}} \Gamma\left(\frac{n - m_j}{2}\right)}{\Gamma\left(\frac{n - m_T}{2}\right)} \left(\frac{\sum_{i \in I_T^{\complement}} w_i^2}{\sum_{i \in I_j^{\complement}} w_i^2} \right)^{\frac{n}{2}},$$

and

$$\frac{\sum_{i \in I_T^{\complement}} w_i^2}{\sum_{i \in I_j^{\complement}} w_i^2} = \frac{n^{-1} \sum_{i \in I_T^{\complement}} w_i^2}{n^{-1} \sum_{i \in I_T^{\complement}} w_i^2 + n^{-1} \sum_{i \in D} w_i^2}.$$

Recall, that $n^{-1/2} w_i \to d_i$. Since D contains at least one non-zero d_i, we have by assumption of the lemma $n^{-1} \sum_{i \in D} w_i^2 \to r > 0$ almost surely. Thus by the Strong Law of Large Numbers there are constants $C < \infty$ and $0 < q < 1$ such that

$$\frac{r(I_j)}{r(I_T)} \leq C n^{\frac{d}{2}} q^n$$

eventually almost surely and (10) converges to $r(I_T)/r(I_T) = 1$. The statement of the lemma follows. ∎

Let us now consider the fiducial distribution under the assumptions of Lemma 2. Conditionally on selecting the right I_T, the fiducial distribution of σ^2 is $\sum_{i \in I_T^{\complement}} w_i^2 / U$, where $U \sim \chi^2_{n - m_T}$, which converges weakly to the true value of σ^2 almost surely. Moreover, it is well-known that the one-sided and two-sided confidence intervals based on this distribution are the classical exact confidence intervals for σ^2. Similarly, conditionally on selecting the right I_T, the fiducial distribution of d_i is $n^{-\frac{1}{2}}w_i + n^{-\frac{1}{2}}T_i \left\{\sum_{i\in I_T^{\complement}} w_i^2/(n-m_T)\right\}^{\frac{1}{2}}$. This again leads consistent generalized fiducial distribution and classical exact confidence intervals. Since, by Lemma 2, we pick the correct I_T with probability going to 1, the (unconditional) fiducial distribution is consistent and leads to asymptotically correct confidence intervals for each of the $d_i,\ i \in I_T$ and σ.

Moreover, write R_n for a random variable having the marginal generalized fiducial distribution of d. Define the projection $\pi(x) = (y_1, \ldots, y_n)$, where $y_i = x_i$ if $i \in I_T$ and 0 otherwise. Recall that, under the assumptions of Lemma 2, the true values $d = \pi(d)$. Also define $\Pi(x)$ as $\pi(x)$ with the zeroed elements omitted.

Consider a collection of functions $l_n(x)$ and a fixed function $h(y)$, such that $h(\Pi x) = l_n(x)$. Simple delta method calculations show that the distribution of $l_n(n^{-1/2}R_n)$, provided $\nabla h(\Pi d) \neq 0$, converges in distribution to $l_n(d)$ almost surely, and satisfies the conditions stated in Hannig (2009). Thus confidence intervals based on $l_n(n^{-1/2}R_n)$ are asymptotically correct confidence intervals for $l_n(d) = h(\Pi d)$. This leads to the following theorem.

Theorem 2. *Assume conditions of Lemma 1 or Lemma 2. Then the generalized fiducial distribution of d and σ^2 is consistent, i.e. the generalized fiducial distribution converges weakly to the distribution concentrating all its mass on the true value of d and σ^2 almost surely.*

Assume conditions of Lemma 2. Let l_n and $r(y)$ be functions introduced above. Then the fiducial distribution $l_n(n^{-1/2}R_n)$ leads to asymptotically correct confidence intervals for the function of true wavelet coefficients $l_n(d)$, provided the gradient $\nabla h(\Pi a) \neq 0$.

Now we are ready to provide the proof for Theorem 1 of HL.

Proof of Theorem 1. If $g(x) = 0$, then the discrete wavelet transform coefficients w_i are independent $N(0, \sigma^2)$ and the assumptions of Lemma 1 are satisfied. If $g(x) = \sum_{i=1}^{\infty} d_i \xi_i(x)$, where ξ_i are the wavelet basis functions and only finitely many $d_i \neq 0$, then the discrete wavelet transform coefficients computed using the same basis functions are independent $N(n^{1/2}d_i, \sigma^2)$ and the assumptions of Lemma 2 are satisfied.

In any case, since the value of the function at a point is a linear combination of the wavelet coefficients, the result on consistency and asymptotic correctness of pointwise confidence interval follows directly from Theorem 2.

The result on curvewise confidence intervals then follows from Hannig (2009) by verifying that the shape of the curvewise confidence region conforms to the conditions stated there. The calculations are standard and hence are omitted. □

4. Curvewise confidence sets

Since we have a distribution on the space of functions, we can use it to define a curvewise approximate confidence sets. This can be done by considering a functional $\nu(\cdot)$ on the space of the wavelet coefficients and setting the curvewise confidence set based on quantiles $\nu(V_d)$, where V_d is a random variable having the generalized fiducial distribution of d as its distribution.

In practice this can be achieved with the following two steps. First for each member $\tilde{d}_i$ of a generated fiducial sample $\{\tilde{d}_1, \ldots, \tilde{d}_M\}$, calculate a functional $\nu(\tilde{d}_i)$. Then define an approximate $(1-\alpha)100\%$ curvewise confidence set for d as

$$c(d) = \{d : \nu(d) \leq (1-\alpha) \text{ quantile of } \nu(\tilde{d}_1), \ldots, \nu(\tilde{d}_M)\}.$$

An approximate curvewise confidence set can be obtained via inverse wavelet transform. In our numerical experiments we have tested the following choice of the functional (Cai & Low, 2006)

$$\nu(d^*) = \sum_{j=0}^{J} 2^{js} \left\{ \sum_{k=0}^{2^j-1} (d^*_{j,k} - \bar{d}_{j,k})^2 \right\}^{\frac{1}{2}} \Big|_{s=0.001},$$

where $d^*_{j,k}$'s and $\bar{d}_{j,k}$'s are the elements of d^* and $\bar{d} = (\tilde{d}_1 + \ldots + \tilde{d}_M)/M$, respectively. Notice that for clarity we have used the double indexing scheme (2) in HL. In below this

	nominal coverage		
test function	90	95	99
Blocks	88·5	95·2	99·1
Bumps	94·3	97·8	99·4
Doppler	86·5	92·8	97·5
Heavisine	85·1	91·4	97·3
Ppoly	85·2	92·5	98·6

Table 1: Curvewise empirical coverage rates, in percentage, for FBALL. The largest standard error is 1·1%.

method is referred as FBALL.

For illustrative purpose we applied FBALL to the same noisy data sets generated in § 6.2 of HL. That is, for each test function, we have constructed 1000 curvewise confidence sets $c(d)$. We then tested how many of these confidence sets actually cover the corresponding true function. The resulting empirical coverage rates are reported in Table 1. From these values one could see that FBALL produced barely adequate empirical results.

References

Cai, T. & Low, M. (2006). Adaptive confidence balls. *Ann. Statist.* **34**, 202–208.

Davis, R. & Resnick, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467–1487.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., & Picard, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). *J. R. Statist. Soc. B* **57**, 301–369.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin.

Hannig, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19**. To appear.

STANLEY, R. P. (1999). *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.