Biometrika (2009), **96**, 4, *pp*. 847–860
doi: 10.1093/biomet/asp050

Advance Access publication 3 October 2009
Printed in Great Britain

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.

jan.hannig@unc.edu

AND THOMAS C. M. LEE

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong

tlee@sta.cuhk.edu.hk

SUMMARY

We apply Fisher's fiducial idea to wavelet regression, first developing a general methodology for handling model selection problems within the fiducial framework. We propose fiducial-based methods for wavelet curve estimation and the construction of pointwise confidence intervals. We show that these confidence intervals have asymptotically correct coverage. Simulations demonstrate that they possess promising empirical properties.

Some key words: Bayesian wavelet prior; Generalized fiducial inference; Pointwise confidence interval; Statistical model selection; Tree constraint; Wavelet regression.

1. INTRODUCTION

Fiducial inference was introduced by Fisher (1930) in an attempt to overcome what he saw as a serious deficiency in the Bayesian approach to inference: the use of a prior distribution even when no information was available. Fiducial inference created some controversy once it was realized that it could yield procedures that were not exact in the frequentist sense and did not possess other properties claimed by Fisher (Lindley, 1958; Zabell, 1992).

Fraser (1968) attempted to develop a rigorous framework for making fiducial inferences; Dawid & Stone (1982) provided further insight by studying situations where they yield exact confidence statements; Barnard (1995) proposed a view of fiducial distributions based on the pivotal approach that seems to remove some of the problems reported in earlier literature, and Dempster (2008) discussed the closely related Dempster–Shafer calculus. Nevertheless fiducial inference has failed to secure a place in mainstream statistics.

Tsui & Weerahandi (1989) proposed a new approach for hypothesis testing using the concept of generalized P-values, later extended to the construction of generalized confidence intervals (Weerahandi, 1993). Hannig et al. (2006) established a direct connection between fiducial and generalized confidence intervals, and proved their asymptotic frequentist correctness. These ideas were unified for parametric problems in E et al. (2008) and Hannig (2009). This unification is termed generalized fiducial inference and has been found to have good theoretical and empirical properties.

In this article we generalize the fiducial idea to statistical inference for wavelet regression. Our emphasis is on constructing pointwise confidence intervals, issues of estimation are also addressed. The development of a fiducial-based wavelet inference method raises two main challenges: incorporating a statistical model selection procedure into the fiducial framework; and applying the fiducial idea to nonparametric curve estimation problems.

The first challenge is addressed under a general methodology that integrates the fiducial idea with common statistical model selection methods in a development analogous to the extension of maximum likelihood to penalized maximum likelihood. With this new methodology, the second challenge is resolved by determining the structural equation that relates the wavelet coefficients and the noisy observations.

Wavelet curve estimation has been extensively studied in the literature. Aside from the universal thresholding of Donoho & Johnstone (1994), other methods of estimation have been proposed based on Stein's unbiased risk estimate (Donoho & Johnstone, 1995), crossvalidation (Nason, 1996), the Akaike information criterion (Hurvich & Tsai, 1998), the minimum description length principle (Antoniadis et al., 1997; Lee, 2002), the false discovery rate (Abramovich & Benjamini, 1996; Tadesse et al., 2005) and the fact that large wavelet coefficients tend to cluster together (Fryzlewicz, 2007). Bayesian and empirical Bayes approaches to wavelet curve estimation have been proposed in Chipman et al. (1997), Abramovich et al. (1998), Clyde & George (2000) and Johnstone & Silverman (2005). Wavelet denoising methods based on more sophisticated wavelet transformations have been studied in Downie & Silverman (1998) and Barber & Nason (2004).

Pointwise confidence intervals were obtained in Barber et al. (2002) by approximating, via cumulant matching, the posterior distribution of the Bayesian wavelet coefficient estimates from Abramovich et al. (1998). Semadeni et al. (2004) and Davison & Mastropietro (2009) improved this approximation using saddlepoint methods.

Additional material and technical details for this study can be found in a supplementary document obtainable from the authors.

2. BACKGROUND

2·1. *Introduction*

Suppose that observed equispaced data points $y_1, \ldots, y_n$ satisfy the model

$$y_i = g_i + \epsilon_i, \tag{1}$$

where $g = (g_1, \ldots, g_n)^{\mathrm{T}}$ is the true but unknown regression function, the errors ϵ_i are independent normal random variables with mean 0 and variance σ^2, and $n = 2^{J+1}$ is an integer power of 2. The proposed methods for estimating and constructing confidence intervals for g are developed from the wavelet thresholding method of Lee (2002) and the generalized fiducial approach of Hannig (2009).

Write $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ and denote the discrete wavelet transform matrix by H (Donoho & Johnstone, 1994). The true wavelet coefficients $d = (d_1, \ldots, d_n)^{\mathrm{T}}$ are given by $d = Hg$. We shall use upper-case letters, such as Z_i, to denote random variables, and lower-case letters, such as z_i, to denote observed values, unless it is clear from the context that a particular symbol is a random variable or an observed value.

2·2. *Wavelet regression*

Most wavelet regression methods consist of three steps. The first step is to apply a forward wavelet transform to the data y and obtain the empirical wavelet coefficients $w = (w_1, \ldots, w_n)^{\mathrm{T}} = Hy$. The second step is to apply a shrinkage operation to w to obtain an estimate $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_n)^{\mathrm{T}}$ of the true wavelet coefficients d. Lastly, the regression estimate $\hat{g} = (\hat{g}_1, \ldots, \hat{g}_n)^{\mathrm{T}}$ for g is computed via the inverse discrete wavelet transform: $\hat{g} = H^{\mathrm{T}}\hat{d}$. The second step of wavelet shrinkage is important because this is when statistical estimation is performed. Examples of such wavelet shrinkage methods can be found in the references given in § 1. Other factors that affect the quality of $\hat{g}$ include the choice of the wavelet, the type of wavelet transform and the primary resolution level.

The method of Lee (2002) is essentially a hard-thresholding method; i.e. for all i, $\hat{d}_i$ is either set to 0 or w_i. The number of nonzero $\hat{d}_i$s is selected by the minimum description length principle developed by Rissanen (2007). For introductory material on this principle, see Hansen & Yu (2001) and Lee (2001).

A tree constraint is also applied to these nonzero $\hat{d}_i$s. To describe this, we temporarily switch to the double-indexing scheme of Donoho & Johnstone (1994) to label the estimated wavelet coefficients $\hat{d} = \{(\hat{d}_{-1,0}), (\hat{d}_{0,0}), \ldots, (\hat{d}_{J,0}, \ldots, \hat{d}_{J,2^J-1})\}^{\mathrm{T}}$. With the exception of the first element, the indexing scheme is $\hat{d}_{j,k}$ $(j = 0, \ldots, J;\ k = 0, \ldots, 2^j - 1)$, where j and k are referred to as the resolution and location indices, respectively. In Lee (2002) the first two elements of $\hat{d}$ are never thresholded. For the remaining elements, $\hat{d}_{j+1,2k}$ and $\hat{d}_{j+1,2k+1}$ are called the children of $\hat{d}_{j,k}$ if $j = 0, \ldots, J - 1$. The tree constraint requires that if $\hat{d}_{j,k}$ is set to 0, then all of its children are set to 0 and is an extreme executioner of the intuition that wavelet coefficients at finer resolution levels have a higher chance of being set to zero. Numerical experience suggests that the tree constraint is essential to the success of the proposed methods. Under it, Lee (2002) proposes estimating g, which is equivalent to d, with the minimizer of $0{\cdot}5\, m \log_2 n + 0{\cdot}5\, n \log_2(\|y - \hat{g}\|^2/n)$, where m is the number of nonzero elements in $\hat{d}$. This is effectively the same as placing a penalty of $0{\cdot}5 \log_2 n$ in the base 2, loglikelihood scale.

2·3. *Generalized fiducial inference*

Generalized fiducial inference begins by expressing the relationship between the data X and the parameters θ as

$$X = G(\theta, U), \tag{2}$$

where $G(\cdot, \cdot)$ is termed the structural equation, and U is the random component of the structural equation whose distribution is completely known. For example, for simple linear regression one could write (2) as $Y_i = b_0 + b_1 x_i + \sigma Z_i \quad (i = 1, \ldots, n)$, where Z_i are independent $N(0, 1)$ random variables, x_i are known constants and $X = (Y_1, \ldots, Y_n)^{\mathrm{T}}, \theta = (b_0, b_1, \sigma)^{\mathrm{T}}$ and $U = (Z_1, \ldots, Z_n)^{\mathrm{T}}$.

Suppose that the structural relation (2) can be inverted and the inverse $G^{-1}(\cdot, \cdot)$ always exists. That is, for any observed x and for any arbitrary u, θ is obtained as $\theta = G^{-1}(x, u)$. As the distribution of U is completely known, one can always generate from it a random sample $\tilde{u}_1, \ldots, \tilde{u}_M$. This random sample of U is then transformed into a random sample for $\{\theta\colon \tilde{\theta}_1 = G^{-1}(x, \tilde{u}_1), \ldots, \tilde{\theta}_M = G^{-1}(x, \tilde{u}_M)\}$, which is called the fiducial sample. Estimates and confidence intervals for θ can be obtained from $\tilde{\theta}_1, \ldots, \tilde{\theta}_M$.

For many problems, including linear regression, the inverse $G^{-1}(\cdot, \cdot)$ may not exist. This can happen in two situations: for any particular u, either there is more than one θ satisfying (2), or there is no θ satisfying (2).

The first situation does not happen with positive probability for the model studied because the structural equation used in this paper can be viewed as a collection of sets of linear equations, each with at most one solution. Moreover, the probability that two different sets of equations in our collection have a solution at the same time is zero.

For the second situation, Hannig (2009) suggests removing the offending values of u from the sample space and then renormalizing the probabilities. Specifically, we generate u conditional on the event that the inverse $G^{-1}(\cdot,\cdot)$ exists. The rationale for this choice is that we know our data x were generated with some θ_0 and u_0, which implies there is at least one solution θ_0 satisfying (2) when the true u_0 is considered. Therefore, we restrict our attention to only those us for which $G^{-1}(\cdot,\cdot)$ exists.

In the particular case of the linear regression listed above, one could select the first three equations and solve for (b_0, b_1, σ^2) in terms of z_1, z_2, z_3. The inverse $G^{-1}(\cdot,\cdot)$ will exist if the vector z is such that the value of (b_0, b_1, σ^2) does not depend on which three equations were selected. To ensure this, one could plug the solutions for (b_0, b_1, σ^2), expressed in terms of z_1, z_2, z_3 into the remaining $n-3$ equations and obtain the following set of $n-3$ conditions:

$$\begin{aligned} 0 = {} & \frac{x_k\{y_3(z_1-z_2)+y_1(z_2-z_3)+y_2(z_3-z_1)\}}{x_3(z_1-z_2)+x_1(z_2-z_3)+x_2(z_3-z_1)} \\ & + \frac{x_2\{y_k(z_1-z_3)+y_1(z_3-z_k)+y_3(z_k-z_1)\}}{x_3(z_1-z_2)+x_1(z_2-z_3)+x_2(z_3-z_1)} \\ & + \frac{x_3\{y_k(z_2-z_1)+y_2(z_1-z_k)+y_1(z_k-z_2)\}}{x_3(z_1-z_2)+x_1(z_2-z_3)+x_2(z_3-z_1)} \\ & + \frac{x_1\{y_k(z_3-z_2)+y_3(z_2-z_k)+y_2(z_k-z_3)\}}{x_3(z_1-z_2)+x_1(z_2-z_3)+x_2(z_3-z_1)} \quad (k=4,\ldots,n). \end{aligned} \tag{3}$$

The inverse $G^{-1}(\cdot,\cdot)$ exists for any vector z satisfying (3).

After observing the (x_i, y_i)s, one could obtain a fiducial sample of (b_0, b_1, σ^2) by generating the vector Z as the independent $N(0, 1)$ variabless conditional on (3), and solving for the parameters. A simple calculation shows that the corresponding density for generating the fiducial sample is

$$r(b_0, b_1, \sigma^2) = C^{-1} \exp\left\{-\frac{\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2}{2\sigma^2}\right\} \sigma^{-n-2},$$

where C is a normalizing constant. For this particular example the generalized fiducial distribution equals the Bayesian posterior computed using the Jeffreys prior (Box & Tiao, 1973).

For a more rigorous description of the above generalized fiducial idea, we define the set inverse of G as

$$Q(x,u) = \{\theta : x = G(\theta, u)\}. \tag{4}$$

Let $\tilde{U}$ be a random variable with distribution U given $Q(x,U) \neq \emptyset$ and denote its realization $\tilde{u}$. Define $\tilde{G}^{-1}(x,\tilde{u})$ as one of the elements of $Q(x,\tilde{u})$ chosen according to some possibly random rule. We have $\theta = \tilde{G}^{-1}(x,\tilde{U})$ and through it a probability density function $r(\theta)$ for θ is defined. We call $r(\theta)$ the generalized fiducial density from which a fiducial sample for θ is generated.

It is often more computationally convenient to compute the generalized fiducial distribution directly as the conditional distribution of

$$\tilde{G}^{-1}(x,U) \mid Q(x,U) \neq \emptyset. \tag{5}$$

See Hannig (2009) for a detailed discussion of the issues surrounding the conditioning and choice of the elements in Q for parametric problems.

3. Main results

3·1. *The structural equation*

Due to (1) and the orthonormality of the discrete wavelet transform matrix H, a model for the empirical wavelet coefficients is $w = d + \sigma z$ with z being a n-dimensional vector of independent $N(0, 1)$ random variables. This model is the structural equation (2). It is straightforward to see $X = \{w\}$ and $U = \{z\}$. Initially, the model parameter θ appears to be composed of the n-dimensional d and the scalar σ^2. The assumption of sparsity, however, implies that most of the entries in the vector d are zero. Hence, for notational convenience, we record the indices of those d that are not zero. This information is stored in a set $I \subset \{1, \ldots, n\}$: d_k is not zero if $k \in I$. Including this index set $\theta = (d, \sigma^2, I)$. Following Lee (2002), we assume that I satisfies the tree condition introduced in § 2·2. We also assume that the complement $I^{\mathcal{C}}$ has at least $(1 - p_0)n$ elements, for $0 < p_0 < 1$; i.e. a nontrivial proportion of coefficients is thresholded. Using this notation, the structural equation is

$$w_k = \begin{cases} d_k + \sigma z_k & (k \in I), \\ \sigma z_k & (k \in I^{\mathcal{C}}). \end{cases} \tag{6}$$

Inverting (6), we see that the set inverse of G defined in (4) contains the element $(d, \sigma^2, I) \in Q(w, z)$ if and only if there is a fixed $j \in I^{\mathcal{C}}$ such that d, σ^2 and z satisfy the following relations: $\sigma^2 = w_j^2/z_j^2$, $d_i = 0$ $(i \in I^{\mathcal{C}})$, $d_k = w_k - w_j/(z_j z_k)$ $(k \in I)$ and

$$w_i = \frac{w_j}{z_j} z_i \quad (i \neq j, i \in I^{\mathcal{C}}). \tag{7}$$

The conditions on z introduced by (7) define a manifold on which z must lie in order for I to be feasible. We denote this manifold by $\mathcal{M}_I$. Consequently, $Q(w, z) \neq \emptyset$ if and only if $z \in \bigcup_{I \in \mathcal{I}} \mathcal{M}_I$, where $\mathcal{I}$ contains all subsets of $\{1, \ldots, n\}$ satisfying the tree constraint introduced in § 2·2 and having at most $p_0 n$ elements. Finally, $\tilde{G}^{-1}(\cdot, \cdot)$ is easily defined by picking any one of the elements in $Q(\cdot, \cdot)$.

3·2. *Generalized fiducial inference with model selection*

The next step is to calculate the generalized fiducial density $r(\theta)$ for our wavelet regression problem. However, any $r(\theta)$ resulting from a direct application of (5) to the structural equation (6) will lead to overfitting. This is similar to the case when the least-squares principle will favour models with more parameters. Below we address issues with overfitting and propose a general mechanism to resolve it.

Denote by $m(I)$ the number of elements in I and notice that $\mathcal{M}_I$ is determined by $n - m(I) - 1$ equations; i.e. it is a $\{m(I) + 1\}$-dimensional manifold in $\mathbb{R}^n$. This introduces the following problem when computing the conditional distribution of Z given $Q(w, Z) \neq \emptyset$: only conditions determined by the fewest possible number of equations could have conditional probability greater than 0. We illustrate this phenomenon with the following simple example. Let $V = (V_1, V_2, V_3)^{\mathrm{T}}$ be a jointly continuous random vector. Consider v_1 such that the marginal density $f_{V_1}(v_1) > 0$ and is continuous at v_1. Set $\mathcal{I}(dv) = (-dv, dv)/2$ and calculate for a Borel measurable set $\mathcal{A}$

$$\begin{aligned}
&\mathrm{pr}(V \in \mathcal{A} \mid V_1 = v_1 \text{ or } V_2 = v_2, V_3 = v_3) \\
&\quad = \lim_{dv \to 0} \frac{\mathrm{pr}[\{V \in \mathcal{A}, V_1 \in v_1 + \mathcal{I}(dv)\} \cup \{V \in \mathcal{A}, V_2 \in v_2 + \mathcal{I}(dv), V_3 \in v_3 + \mathcal{I}(dv)\}]}{\mathrm{pr}[\{V_1 \in v_1 + \mathcal{I}(dv)\} \cup \{V_2 \in v_2 + \mathcal{I}(dv), V_3 \in v_3 + \mathcal{I}(dv)\}]} \\
&\quad = \lim_{dv \to 0} \frac{\mathrm{pr}\{V \in \mathcal{A}, V_1 \in v_1 + \mathcal{I}(dv)\}}{\mathrm{pr}\{V_1 \in v_1 + \mathcal{I}(dv)\}} \\
&\quad = \mathrm{pr}(V \in \mathcal{A} \mid V_1 = v_1),
\end{aligned}$$

where the second equality follows from the fact that, as $dv \to 0$,

$$\mathrm{pr}\,\{V \in \mathcal{A}, V_2 \in v_2 + \mathcal{I}(dv), V_3 \in v_3 + \mathcal{I}(dv)\} = o\,[\mathrm{pr}\,\{V_1 \in v_1 + \mathcal{I}(dv)\}]\,.$$

Hence, when considered together, the higher-dimensional condition has no effect on the conditional distribution. This demonstrates that the generalized fiducial distribution will, with probability 1, pick models with minimal thresholding.

The propensity of statistical procedures to favour larger models over smaller models is overcome by the inclusion of a penalty term encoding our preference for parsimony. We will follow this approach and introduce a penalty, on the loglikelihood scale, of $0{\cdot}5 \log n$ for each nonthresholded wavelet coefficient, cf. Lee (2002).

A penalty can be incorporated into the generalized fiducial distribution by introducing additional synthetic random variables and equations into the structural equation. This modification will not change the sampling distribution of the data, but it will allow a penalty term to be included. In particular, we add to the structural equation (6)

$$\psi_k = \begin{cases} p_k & (k \in I), \\ t_k + p_k & (k \in I^{\mathcal{C}}), \end{cases} \tag{8}$$

where $t = (t_1, \ldots, t_n)$ are unknown parameters and $p = (p_1, \ldots, p_n)$ are independent random variables with density $f_p(p)$, independent of the wavelet coefficients w. Considering equations (6) and (8) together introduces a symmetry into the problem: if d_k is set to zero then t_k may take any value and vice versa.

In order to derive a generalized fiducial distribution for the enlarged model, we need observed values for $\psi = (\psi_1, \ldots, \psi_n)$. Since these values were artificially introduced and never actually observed, we assume that $\psi = (0, \ldots, 0)$. We also assume that the density $f_p(p)$ is continuous and $f_p(0) = n^{-1/2}$; this last value is chosen to match the minimum description length penalty $0{\cdot}5 \log n$. Further comments about this are provided in § 1 of the supplementary document.

Due to independence, the presence of the random variables ψ does not alter the sampling distribution of w; however, it does alter the conditional distribution used in the definition of the generalized fiducial distribution. In particular, inverting the structural equation (8), we see that the set inverse of G defined in (4) contains the element $(t, I) \in Q(w, z)$ if and only if t and p satisfy the following relations: $t_i = -p_i$ $(i \in I^{\mathcal{C}})$, $t_k = 0$ $(k \in I)$ and

$$p_k = 0 \quad (k \in I). \tag{9}$$

As before, this last set of equations (9) introduces conditions that p must satisfy in order for I to be feasible. Combining the conditions on z from (7) with the conditions on p from (9), the total number of conditions on the joint distribution of (z, p) is $n - 1$ regardless of the choice of I. This alleviates the problem of overfitting described at the beginning of this subsection. See § 1 of the supplementary document for further technical details.

3·3. *Generalized fiducial distribution for wavelet regression*

Appendix 1 outlines the development of the idea presented in the previous subsections, and shows that the generalized fiducial distribution function is obtained as an integral of the following generalized density:

$$r(\sigma^2, d, I) = C^{-1}2^{-1}(\pi)^{-(n+1)/2}\sigma^{-n-2}\frac{\sum_{j\in I^C}|w_j|}{n-m(I)} \times \exp\left[-\frac{m(I)\log n}{2} - \frac{\{\sum_{k\in I}(d_k - w_k)^2 + \sum_{i\in I^C} w_i^2\}}{2\sigma^2}\right]\prod_{i\in I^C}\delta_0(d_i), \quad (10)$$

where C is a normalizing constant and $\delta_0(s)$ is the Dirac function; i.e. $\int_{\mathcal{A}} \delta_0(s)\, ds = 1$ if $0 \in \mathcal{A}$ and 0 otherwise.

The constant C in (10) cannot be computed in a closed form so a sample from $r(\sigma^2, d, I)$ must be simulated using Markov chain Monte Carlo techniques. Details of the algorithm used may be found in § 2 of the supplementary document.

The generalized fiducial distribution (10) is defined in the wavelet domain. The inverse wavelet transform is used to define a generalized fiducial distribution on the function domain. In the next section we demonstrate how this distribution is used for statistical inference.

4. Inference for wavelet regression

4·1. *Point estimation*

Let $l(\cdot, \theta)$ be a loss function and $r_X(\theta)$ be the corresponding density of the generalized fiducial distribution for parameter θ. We show the subscript X to emphasize its dependence on the observed data. Define the fiducial risk of a point estimator T as

$$\int_\Theta l(T, \theta) r_X(\theta)\, d\theta. \quad (11)$$

The statistic T that minimizes this fiducial risk can be taken as a point estimator of θ. If $l(\cdot, \cdot)$ is the squared loss, the estimator that minimizes (11) is the mean of the generalized fiducial distribution (Casella & Berger, 2002, p. 353), so we propose to use the mean of the generalized fiducial distribution as our point estimator. Due to the linearity of the inverse wavelet transformation, the estimator is the same whether we first take the mean of the generalized fiducial distribution and then apply the inverse transform, or vice versa. In practice, estimation is achieved by first simulating a fiducial sample $\tilde{d}_1, \ldots, \tilde{d}_M$ for d from (10) using the method described in § 2 of the supplementary document, and then computing the estimate as the inverse wavelet transform of the elementwise average of this sample.

4·2. *Confidence intervals*

The marginal generalized fiducial distribution at any location i allows the construction of an approximate pointwise confidence interval for the unknown g_i. We generate a fiducial sample $\tilde{d}_1, \ldots, \tilde{d}_M$ from (10) and apply the inverse wavelet transform to obtain a random sample for the unknown function. We then use this sample at location i: the approximate pointwise $(1-\alpha)100\%$ confidence interval is defined as the $\alpha/2$ and $1-\alpha/2$ quantiles of this sample. This procedure is repeated for $i = 1, \ldots, n$ to obtain a pointwise confidence band for g. As reported in § 6·2, the pointwise confidence intervals generated using this method have favourable frequentist properties as measured by repeated sample performance.

Curvewise confidence sets can also be constructed using the fiducial sample $\tilde{d}_1, \ldots, \tilde{d}_M$. See § 4 of the supplementary document for details. For related work on curvewise confidence intervals for wavelet regression, see Genovese & Wasserman (2005) and Cai & Low (2006).

5. Theoretical properties

In this section we establish basic asymptotic properties of the generalized fiducial distribution defined by (10). Assume that the true regression function $g(x)$ in (1) is observed equidistantly over a compact interval with $n = 2^{J+1}$ observations.

Theorem 1. *Suppose that $g(x)$ is a finite linear combination of the wavelet basis functions. Then, as $J \to \infty$, the generalized fiducial distribution defined by* (10) *converges weakly to the distribution concentrated on $g(x)$ for almost all sequences of the observed data. Additionally, if $g(x)$ is not zero, then the pointwise confidence intervals of* § 4·2 *have asymptotically correct coverage.*

The proof of this theorem is outlined in Appendix 2.

This theorem guarantees good large sample properties of the proposed confidence intervals for the case when the unknown function is a finite linear combination of wavelet basis functions. Other, more general, functions $g(x)$ can be often well approximated by such finite linear approximations. In particular, if the wavelet coefficients of $g(x)$ decay fast enough, then one can use ideas outlined in Donoho & Johnstone (1995) to show the consistency of the generalized fiducial distribution; however, the resulting confidence intervals might be conservative. This is because the estimator of the noise variance used implicitly in the generalized fiducial distribution could be too large due to contamination by small wavelet coefficients.

The proofs outlined in Appendix 2 clearly reveal the crucial importance of the tree constraint. Specifically, a careful examination of the proof of Lemma A1 shows that the generalized fiducial distribution is asymptotically consistent even if the penalty used $\exp\{c\, m(I) \log n\}$ were modified from the current $c = 0{\cdot}5$ to any $c > 0$. Without the tree constraint, the generalized fiducial distribution is inconsistent for any $c < 2$.

6. Simulation results

6·1. *Point estimation*

The experimental settings adopted for point estimation are essentially the same as those in Semadeni et al. (2004). The five test functions were Block, Bumps, Doppler and Heavisine of Donoho & Johnstone (1994), and the piecewise polynomial function Ppoly of Nason & Silverman (1994). Five hundred noisy datasets were generated for each test function with a sample size of $n = 1024$ and signal-to-noise ratio $\|g\|/\sigma = 4$. For each noisy dataset five wavelet regression methods were applied to estimate the true function. The five regression methods are the generalized fiducial-based method described in § 4·1, the complex wavelet thresholding method of Barber & Nason (2004), the empirical Bayes method of Johnstone & Silverman (2005), the Bayesian method of Abramovich et al. (1998) and the minimum description length with tree constraint method of Lee (2002). The methods of Barber & Nason (2004) and Johnstone & Silverman (2005) are two of the best wavelet thresholding methods and were chosen for benchmark comparison. The methods of Abramovich et al. (1998) and Lee (2002) were selected because they form the basis for the wavelet pointwise confidence interval methods to be compared in the next subsection.

Table 1. *Averaged mean square errors, multiplied by 1000*

	FMEAN	CTHRESH	EBAYES	BTHRESH	MDLTREE
Blocks	19·0	16·0	20·0	29·0	33·0
Bumps	18·0	13·0	18·0	26·0	69·0
Doppler	10·0	8·2	10·0	9·5	35·0
Heavisine	7·0	4·0	5·2	4·7	4·5
Ppoly	5·8	2·7	6·6	5·0	4·0

FMEAN, the proposed method from § 4·1; CTHRESH, the method of Barber & Nason (2004); EBAYES, the method of Johnstone & Silverman (2005); BTHRESH, the method of Abramovich et al. (1998); MDLTREE, the method of Lee (2002). The largest standard error multiplied by 1000 is 0·06.

Table 2. *Averaged empirical coverage rates, in percentage*

	Blocks	Bumps	Doppler	Heavisine	Ppoly
			Nominal rate = 90		
FBAND	89·1	89·7	90·5	91·7	95·2
SBAND	91·3	89·5	89·5	84·8	92·2
DM	91·4	90·5	90·6	87·7	92·6
			Nominal rate = 95		
FBAND	94·5	94·9	95·4	96·3	98·1
SBAND	95·6	94·8	94·7	93·8	96·9
DM	95·6	95·3	95·2	94·8	97·0
			Nominal rate = 99		
FBAND	98·8	98·9	99·1	99·4	99·7
SBAND	98·9	99·0	98·7	98·4	99·4
DM	98·9	99·0	98·7	98·6	99·4

FBAND, the proposed method from § 4·2; SBAND, the method of Semadeni et al. (2004) with normal prior; DM, the method of Davison & Mastropietro (2009) with Laplace prior. The largest standard error is 1·0%.

The mean square error for each function estimate $\hat{g}$ is $\|g - \hat{g}\|^2/n$; these values are tabulated in Table 1. The complex wavelet method of Barber & Nason (2004) gave the best performance for all test functions. For the more complex test functions Blocks, Bumps and Doppler, both the proposed method from § 4·1 and the empirical Bayes method of Johnstone & Silverman (2005) produced the next best performance, while the method of Lee (2002) came last. For the remaining two test functions, Heavisine and Ppoly, the method of Lee (2002) performed the second best while the methods of § 4·1 and Johnstone & Silverman (2005) performed the worst.

6·2. *Pointwise confidence intervals*

This section compares the empirical results of the method proposed in § 4·2 for constructing pointwise confidence intervals with two other methods: the saddlepoint method of Semadeni et al. (2004) with a normal prior, and the improved saddlepoint method proposed by Davison & Mastropietro (2009) with a Laplace prior.

The experimental settings are the same as those in Davison & Mastropietro (2009). One thousand noisy datasets were generated for each of the same five test functions from § 6·1 with $n = 512$ and signal-to-noise ratio 4. The method described in § 4·2 was then applied to each of these datasets to obtain the 90%, 95% and 99% pointwise confidence intervals. The empirical coverage rate for each test function and each sample time, i.e. the percentage of confidence intervals covering g_i $(i = 1, \ldots, n)$, was calculated and averaged over all sample times. These averaged empirical coverage rates and those from the methods of Semadeni et al. (2004) and Davison & Mastropietro (2009), reported in Table 2 of Davison & Mastropietro (2009), are shown in Table 2.

The method of Semadeni et al. (2004) was selected for comparison because it seems to be the best published method for constructing pointwise wavelet regression confidence intervals. Table 2 suggests that both the proposed method and the method of Davison & Mastropietro (2009) are superior to the method of Semadeni et al. (2004), and gave similar performances.

Plots of the empirical coverage rates for the proposed method suggest that the empirical and nominal coverage rates differ the most near the discontinuities of the test functions. These plots are omitted due to space limitations but can be obtained from the authors.

ACKNOWLEDGEMENT

The authors are grateful to two reviewers and the editor for their extremely helpful comments, and to Jessi Cisewski and Damian Wandler for proofreading the paper. The work of Hannig was supported in part by the U.S. National Science Foundation. The work of Lee was supported in part by the Hong Kong Research Grants Council and the National Science Foundation. Hannig and Lee are also affiliated with Colorado State University.

APPENDIX 1

Derivation of the generalized fiducial density (10)

We compute the set inverse of the structural equation given by (6) and (8). The element $(d, t, \sigma^2, I) \in Q(w, 0, z, p)$ if and only if the equations in and immediately above (7) and (9) are satisfied.

Consider fixed I and $j \in I^{\mathcal{C}}$ and denote $m(I)$ as the number of elements in I. We 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 $R_i = w_j Z_i/Z_j$ $(i \in I^{\mathcal{C}}, i \neq j)$, together with $T_i = 0 - P_i$ $(i \in I^{\mathcal{C}})$ and P_k $(k \in I)$. A routine calculation shows that for any set $\mathcal{A} \subset \mathbb{R}^{n+1}$,

$$\mathcal{A}_{I,j}(dx) = \text{pr}\left\{(S, D) \in \mathcal{A}; R_i \in \left(w_k - \frac{dx}{2}, w_k + \frac{dx}{2}\right), i \in I^{\mathcal{C}}, i \neq j; P_k \in \left(\frac{-dx}{2}, \frac{dx}{2}\right), k \in I\right\}$$
$$= \frac{dx^{n-1} q^{m(I)}}{2(2\pi)^{\frac{n}{2}}} \int\int_{\mathcal{A}} e^{-(2h)^{-1}\left\{\sum_{k\in I}(d_k - w_k)^2 + \sum_{i\in I^{\mathcal{C}}} w_i^2\right\}} |w_j| h^{-(\frac{n}{2}+1)} \prod_{i\in I^{\mathcal{C}}} \delta_0(d_i)\, dh\, dd + o(dx^{n-1}), \quad \text{(A1)}$$

where δ_0 is the Dirac function.

Following (5) and using the fact that j is picked at random from $I^{\mathcal{C}}$, the conditional distribution in the definition of fiducial probability (5) is

$$\text{pr}\{(S, D) \in \mathcal{A}, I\} = \lim_{dx\to 0} \frac{\sum_{j\in I^{\mathcal{C}}} \mathcal{A}_{I,j}(dx)}{n - m(I)} \left\{\sum_{I'\in\mathcal{I}} \frac{\sum_{j\in I'^{\mathcal{C}}} \mathcal{A}_{I',j}(dx)}{n - m(I')} + o(dx^{n-1})\right\}^{-1}, \quad \text{(A2)}$$

where $\mathcal{I}$ contains all subsets of $\{1, \ldots, n\}$ satisfying the tree constraint described in § 2·2 and having at most $p_0 n$ elements; see § 3·1.

From (A1) and (A2), we conclude that the generalized fiducial density is

$$r(h, d, I) = C^{-1} q^{m(I)} \frac{\sum_{j\in I^{\mathcal{C}}} |w_j|}{n - m(I)} e^{-(2h)^{-1}\left\{\sum_{k\in I}(d_k - w_k)^2 + \sum_{i\in I^{\mathcal{C}}} w_i^2\right\}} \left\{2(2\pi)^{\frac{n}{2}} h^{\frac{n}{2}+1}\right\}^{-1} \prod_{i\in I^{\mathcal{C}}} \delta_0(d_i). \quad \text{(A3)}$$

The use of the penalizing random variables P led to the term $q^{m(I)}$ in (A3) which, after taking a logarithm, corresponds to an additive penalty term of $m(I)\log q$ to the loglikelihood. In this article we choose $q = n^{-1/2}$. This leads to, on the loglikelihood scale, the penalty of $-0{\cdot}5\, m(I)\log n$ matching the minimum description length penalty, and the equation (10) follows immediately. There are other reasonable choices of q, such as $q = e^{-2}$ leading to the Akaike information criterion penalty.

Appendix 2

Proofs

We outline our theoretical results that lead to Theorem 1. A more detailed treatment can be found in § 3 of the supplementary document. 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 manner 1 through 2^{J+1}, level by level, left to right. Each subset of B_J is 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 of the wavelet coefficient w_i $(i = 1, \ldots, n)$. The marginal fiducial probability $r(I)$ for each $I \in \mathcal{I}_J$ is

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

Recall that $r(I)$ depends on the observed values of the w_i. We will show that the fiducial distribution converges weakly for almost all sequences of observed sets of wavelet coefficients $\{w_i\}_{i=1}^{2^{J+1}}$ $(J = 1, 2, \ldots)$.

Lemma A1. *Let w_i, $i = 1, \ldots, n$ be independent $N(0, \sigma^2)$ and $I_0 = \emptyset$ be the empty tree corresponding to all coefficients being thresholded. Then $r(I_0) \to 1$ almost surely.*

Proof. Let us denote by $x_{(1)} > \cdots > x_{(n)}$ the order statistics of $w_1^2, \ldots, w_n^2$. It is well known that (Embrechts et al., 1997, Theorem 3·5·1) $x_{(1)} \leqslant 2\log(n) + 2\log\log n$ eventually almost surely. Additionally, it follows from Davis & Resnick (1984), Theorem 5·1, that if $0 < \alpha < 1$ and $m = \lfloor n^\alpha \rfloor$, where $\lfloor x \rfloor$ is the integer part of x, then

$$\sum_{i=1}^{m} x_{(i)} < 2(1 - \alpha) m \log(n) \tag{A4}$$

eventually almost surely.

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

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

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

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

Fix $\epsilon_1 > 0$. By (A4) and the strong law of large numbers, there are $q > 0$ and $\upsilon_1 > 0$ such that, if $n^{1/2+\epsilon_1} < m \leqslant qn$, then

$$\exp\left(\frac{n}{2} \frac{\sum_{i=1}^{m} x_{(i)}}{\sum_{i=m+1}^{n} x_{(i)}} \right) \leqslant \exp\left\{ \left(\frac{1}{2} - \upsilon_1 \right) m \log n \right\}$$

and consequently,

$$\frac{r(I)}{r(I_0)} \leqslant e^{-v_1 m \log n + C_1 m}. \tag{A5}$$

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

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

and consequently,

$$\frac{r(I)}{r(I_0)} \leqslant e^{-\frac{1}{2} m \log n + C_1 m + C_2 n}. \tag{A6}$$

The case $0 < m \leqslant n^{1/2+\epsilon_1}$ is more complicated. Unlike in the previous two cases, the ratio $r(I)/r(I_0)$ could grow unbounded. However, due to the tree constraint, the probability of such an event is small. 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

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

By the strong law of large numbers, there is $v_2 > 0$ such that

$$\frac{r(I)}{r(I_0)} \leqslant e^{-v_2 m \log n + C_1 m}. \tag{A7}$$

Let A_n be the event that all trees of size less than $n^{1/2+\epsilon_1}$ contain at most one-third of its $w_i^2 > x_{(n^{3/4+\epsilon_2})}$. Since the number of binary trees of size m is the Catalan number $C_m \leqslant 4^m$ (Stanley, 1999), equations (A5)–(A7) imply that on A_n,

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

where $r_n \to 0$.

The rest of the proof follows from the Borel–Cantelli lemma. Fix a tree of size m. The probability that the arrangement of the w_is is such that the fixed tree contains at most $1/3$ of $w_i^2 > x_{(n^{3/4+\epsilon_2})}$ is less than

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

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

$$\text{pr}\left(A_n^{\mathcal{C}}\right) \leqslant \sum_{m=1}^{\lfloor n^{1/2+\epsilon_1}\rfloor} 4^m e^{-v_3 m \log n} \leqslant \frac{4 - (4n^{-v_3})^{n^{1/2+\epsilon_1}}}{n^{v_3} - 4}.$$

This implies that $\sum_{J=1}^{\infty} \text{pr}(A_{2^{J+1}}^{\mathcal{C}}) < \infty$. □

Lemma A1 implies that for almost all sequences of the observed data, the marginal generalized fiducial distribution of d converges weakly to 0 almost surely.

If some of the wavelet coefficients are not zero, the discrete wavelet transform guarantees that they increase as $n^{1/2}$ (Donoho et al., 1995), leading to the following lemma. The proof is similar to the proof of Lemma A1 and is omitted.

LEMMA A2. *Assume that* $w_i = d_i n^{1/2} + \sigma z_i$, *where* z_i *are independent* $N(0, 1)$*, and that there is a fixed, finite, nonempty 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 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 A1 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 $N(d_i n)^{1/2}, \sigma^2$ and the assumptions of Lemma A2 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 intervals follows directly from Lemmas A1, A2 and results in Hannig (2009). □

REFERENCES

ABRAMOVICH, F. & BENJAMINI, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comp. Statist. Data Anal.* **22**, 351–61.

ABRAMOVICH, F., SAPATINAS, T. & SILVERMAN, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. R. Statist. Soc.* B **60**, 725–49.

ANTONIADIS, A., GIJBELS, I. & GREGOIRE, G. (1997). Model selection using wavelet decomposition and applications. *Biometrika* **84**, 751–63.

BARBER, S. & NASON, G. P. (2004). Real nonparametric regression using complex wavelets. *J. R. Statist. Soc.* B **66**, 927–39.

BARBER, S., NASON, G. P. & SILVERMAN, B. W. (2002). Posterior probability intervals for wavelet thresholding. *J. R. Statist. Soc.* B **64**, 189–205.

BARNARD, G. A. (1995). Pivotal models and the fiducial argument. *Int. Statist. Rev.* **63**, 309–23.

BOX, G. E. P. & TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.

CAI, T. & LOW, M. (2006). Adaptive confidence balls. *Ann. Statist.* **34**, 202–8.

CASELLA, G. & BERGER, R. L. (2002). *Statistical Inference,* 2nd ed. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

CHIPMAN, H. A., KOLACZYK, E. D. & MCCULLOCH, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *J. Am. Statist. Assoc.* **92**, 1413–21.

CLYDE, M. & GEORGE, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *J. R. Statist. Soc.* B **62**, 681–98.

DAVIS, R. & RESNICK, S. (1984). Tail estimates motivated by extreme value theory. *Ann. Statist.* **12**, 1467–87.

DAVISON, A. C. & MASTROPIETRO, D. (2009). Saddlepoint approximation for mixture models. *Biometrika.* **96**, 479–86.

DAWID, A. P. & STONE, M. (1982). The functional-model basis of fiducial inference (with discussion). *Ann. Statist.* **10**, 1054–74.

DEMPSTER, A. P. (2008). The Dempster-Shafer calculus for statisticians. *Int. J. Approx. Reason.* **48**, 365–77.

DONOHO, D. L. & JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–55.

DONOHO, D. L. & JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.* **90**, 1200–24.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. & PICARD, D. (1995). Wavelet shrinkage: Asymptopia? (with discussion). *J. R. Statist. Soc.* B **57**, 301–69.

DOWNIE, T. & SILVERMAN, B. (1998). The discrete multiple wavelet transform and thresholding methods. *IEEE Trans. Sig. Proces.* **46**, 2558–61.

E, L., HANNIG, J. & IYER, H. K. (2008). Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *J. Am. Statist. Assoc.* **103**, 854–65.

EMBRECHTS, P., KLÜPPELBERG, C. & MIKOSCH, T. (1997). *Modelling Extremal Events.* Applications of Mathematics (New York) 33. Berlin: Springer.

FISHER, R. A. (1930). Inverse probability. *Proc. Camb. Phil. Soc.* **xxvi**, 528–35.

FRASER, D. A. S. (1968). *The Structure of Inference*. New York: John Wiley & Sons.

FRYZLEWICZ, P. (2007). Bivariate hard thresholding in wavelet function estimation. *Statist. Sinica* **17**, 1457–81.

GENOVESE, C. R. & WASSERMAN, L. (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.* **33**, 698–729.

HANNIG, J. (2009). On generalized fiducial inference. *Statist. Sinica* **19**, 491–544.

HANNIG, J., IYER, H. K. & PATTERSON, P. (2006). Fiducial generalized confidence intervals. *J. Am. Statist. Assoc.* **101**, 254–69.

HANSEN, M. H. & YU, B. (2001). Model selection and the principle of minimum description length. *J. Am. Statist. Assoc.* **96**, 746–74.

HURVICH, C. M. & TSAI, C.-L. (1998). A crossvalidatory AIC for hard wavelet thresholding in spatially adaptive function estimation. *Biometrika* **85**, 701–10.

JOHNSTONE, I. M. & SILVERMAN, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.* **33**, 1700–52.

LEE, T. C. M. (2001). An introduction to coding theory and the two–part minimum description length principle. *Int. Statist. Rev.* **69**, 169–83.

LEE, T. C. M. (2002). Tree–based wavelet regression for correlated data using the minimum description length principle. *Aust. New Zeal. J. Statist.* **44**, 23–39.

LINDLEY, D. V. (1958). Fiducial distributions and Bayes' theorem. *J. R. Statist. Soc.* B **20**, 102–7.

NASON, G. P. (1996). Wavelet shrinkage using cross–validation. *J. R. Statist. Soc.* B **58**, 463–79.

NASON, G. P. & SILVERMAN, B. W. (1994). The discrete wavelet transform. *J. Comp. Graph. Statist.* **3**, 163–91.

RISSANEN, J. (2007). *Information and Complexity in Statistical Modeling*. Springer.

SEMADENI, C., DAVISON, A. C. & HINKLEY, D. V. (2004). Posterior probability intervals in Bayesian wavelet estimation. *Biometrika* **91**, 497–505.

STANLEY, R. P. (1999). *Enumerative Combinatorics.* vol. 2, Cambridge Studies in Advanced Mathematics 62. Cambridge: Cambridge University Press.

TADESSE, M. G., IBRAHIM, J. G., VANNUCCI, M. & GENTLEMAN, R. (2005). Wavelet thresholding with Bayesian false discovery rate control. *Biometrics* **61**, 25–35.

TSUI, K.-W. & WEERAHANDI, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *J. Am. Statist. Assoc.* **84**, 602–7.

WEERAHANDI, S. (1993). Generalized confidence intervals. *J. Am. Statist. Assoc.* **88**, 899–905.

ZABELL, S. L. (1992). R. A. Fisher and the fiducial argument. *Statist. Sci.* **7**, 369–87.

[*Received January* 2008. *Revised April* 2009]