Statistics and Computing **14**: 337–341, 2004

Automatic polynomial wavelet regression

THOMAS C. M. LEE* and HEE-SEOK OH†

*Department of Statistics, Colorado State University, CO 80523-1877, USA
tlee@stat.colostate.edu
†Department of Statistics, Seoul National University, Seoul, 151-742, Korea

Received July 2003 and accepted May 2004

In Oh, Naveau and Lee (2001) a simple method is proposed for reducing the bias at the boundaries for wavelet thresholding regression. The idea is to model the regression function as a sum of wavelet basis functions and a low-order polynomial. The latter is expected to account for the boundary problem. Practical implementation of this method requires the choice of the order of the low-order polynomial, as well as the wavelet thresholding value. This paper proposes two automatic methods for making such choices. Finite sample performances of these two methods are evaluated via numerical experiments.

Keywords: boundary adjustment, Bayesian Information Criterion, empirical Bayes, polynomial wavelet regression, Stein's unbiased risk estimation, wavelet thresholding

1. Introduction

Suppose a set of noisy data satisfying the following is observed: $y_i = f(i/n) + \epsilon_i, i = 1, \ldots, n = 2^J$, where the unknown regression function f is assumed to be square integrable on the interval [0, 1], and the errors ϵ_i's are independent and identical zero-mean normal random variables. The goal is, given the y_i's, to estimate f using wavelet techniques.

When performing wavelet regression, it is customary to impose on f some boundary assumptions, such as periodicity or symmetry. However, such assumptions may not always be reasonable. To overcome this problem, it is suggested by Oh, Naveau and Lee (2001) to decompose f as the sum of a set of wavelet basis functions, f_W, plus a low-order polynomial, f_P. That is, $f = f_W + f_P$. The hope is that, once f_P is removed from f, the remaining portion f_W can be well estimated using wavelet regression with say periodic boundary assumption. In practice this approach requires the choosing of the polynomial order for f_P and the wavelet thresholding value for f_W. The main contribution of this article is the proposals of two automatic methods for selecting such values.

The rest of this article is organized as follows. Background material is provided in Section 2. Section 3 presents the two proposed methods and Section 4 reports simulation results. Conclusion is offered in Section 5.

2. Background: Polynomial wavelet regression

Let ϕ and ψ be a father and a mother wavelet respectively. Any square integrable function f admits the following expansion (e.g., Daubechies 1992, p. 130):

$$f(x) = \sum_{k=-\infty}^{\infty} c_{0,k}\phi_k(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(x), \qquad (1)$$

where $\phi_k(x) = 2^{1/2}\phi(2x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$. Here the scaling and detail coefficients are respectively equal to $c_{0,k} = \int_{-\infty}^{\infty} f(x)\phi_k(x)\,dx$ and $d_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)\,dx$. Equation (1) suggests the following classical nonlinear wavelet regression estimator:

$$\hat{f}_W(x) = \sum_{k=1}^{2^J-1} \hat{c}_{0,k}\phi_k(x) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \hat{d}^S_{j,k}\psi_{j,k}(x), \qquad (2)$$

where $\hat{c}_{0,k} = \sum_i y_i\phi_k(i/n)$ and $\hat{d}_{j,k} = \sum_i y_i\psi_{j,k}(i/n)$ are respectively the empirical scaling and detail coefficients, and $\hat{d}^S_{j,k} = \mathrm{sgn}(\hat{d}_{j,k})\max(0, |\hat{d}_{j,k}| - \lambda)$ denotes the soft-thresholded wavelet coefficients with thresholding value λ. Sometimes the soft-thresholded coefficients $\hat{d}^S_{j,k}$ are replaced by the hard-thresholded coefficients $\hat{d}^H_{j,k} = \hat{d}_{j,k}I_{\{|\hat{d}_{j,k}|>\lambda\}}$ (see, for examples, Donoho and Johnstone (1994, 1995).

To reduce the boundary effects present in $\hat{f}_W(x)$, the following so-called *polynomial wavelet regression* estimator was proposed by Oh, Naveau and Lee (2001):

$$\hat{f}_{PW}(x) = \hat{f}_P(x) + \hat{f}_W(x) = \sum_{l=0}^{d} \hat{\alpha}_l x^l + \sum_{k=1}^{2^J-1} \hat{c}_{0,k}\phi_k(x) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \hat{d}^S_{j,k}\psi_{j,k}(x), \quad (3)$$

where $\hat{f}_P(x) = \sum_{l=0}^{d} \hat{\alpha}_l x^l$ is a polynomial estimator of degree d. Thus, the use of $\hat{f}_{PW}(x)$ requires the choosing of d as well as the thresholding value λ. With appropriately chosen d and λ, it is demonstrated in Oh, Naveau and Lee (2001), both analytically and empirically, that $\hat{f}_{PW}(x)$ is superior to $\hat{f}_W(x)$. The goal of this article is to propose two automatic methods for choosing both d and λ. Notice that no such automatic methods are proposed by Oh, Naveau and Lee (2001). Also notice that the best choice of λ for the $\hat{f}_W$ in (3) may be different from the best λ for the $\hat{f}_W$ in (2).

It is desirable to maintain the orthogonality between the set of polynomial basis $\{x, \ldots, x^d\}$ and the wavelet basis. This means that the equations $\int x^l\psi(x)\,dx = \int x^l\phi(x)\,dx = 0$ have to be satisfied for $l = 1, \ldots, d$. Wavelets with such properties were constructed by Daubechies (1992) and named coiflets. Hence, the use of a coiflet with at least $d+1$ vanishing moments in (3) implies that the polynomial regression term is orthogonal to the wavelet regression term. Due to this orthogonality property, Oh, Naveau and Lee (2001) suggest estimating the parameters in (3) by first regressing the observations $\{y_i\}$ on the set $\{x, \ldots, x^d\}$ for fixed d and then applying wavelet regression to the residuals of the polynomial regression.

3. The proposed selection methods

This section describes two automatic methods for selecting d and λ.

3.1. *Stein's unbiased risk estimation*

The first method is a two-step procedure for choosing the values of d and λ that aim to minimize the risk between f and $\hat{f}_{PW}$, defined as $E\|f - \hat{f}_{PW}\|^2$. First a criterion similar to Mallows' C_p (Mallows 1973) is used to choose the d that aims to minimize the risk between f and $\hat{f}_P$, where $\hat{f}_P$ is the polynomial estimator computed by regressing y_i's on $x, \ldots, x^d$. Then the SURE wavelet regression procedure of Donoho and Johnstone (1995) is applied to choose the λ that aims to minimize the risk between $f - \hat{f}_P$ and $\hat{f}_W$, where $\hat{f}_W$ is obtained by applying ordinary wavelet regression to the polynomial residuals $y_i - \hat{f}_P(i/n)$.

For the selection of d, we suggest using the maximizer of $r(d)$

$$r(d) = \sum_{l=0}^{d} \hat{\alpha}_l^2 - \frac{2\hat{\sigma}^2 d}{n}, \quad d = 0, 1, \ldots,$$

where $\hat{\sigma}^2$ is any consistent estimator of σ^2. In this article, the simplest MAD is used. It can be shown that maximizing $r(d)$ is equivalent to minimizing an approximately unbiased estimator of the risk between f and $\hat{f}_P$.

Once d and hence $\hat{f}_P$ is obtained, the SURE procedure of Donoho and Johnstone (1995) is applied to the residuals $y_i - \hat{f}_P(i/n)$ to obtain $\hat{f}_W$. Then the final polynomial wavelet regression estimate is obtained as $\hat{f}_{PW} = \hat{f}_P + \hat{f}_W$.

3.2. *Bayesian approach*

The second method adopts a Bayesian framework for choosing d and λ. Similar to before, it is a two-step procedure. First the Bayesian Information Criterion (BIC, Schwarz 1978) is employed to select d. Then the empirical Bayes thresholding method of Johnstone and Silverman (2003) is applied to the polynomial residuals.

For the selection of d, we choose the d that minimizes the following criterion:

$$\mathrm{BIC}(d) = n \log\left[\frac{1}{n}\left\{\sum_{i=1}^{n} y_i - \hat{f}_P\left(\frac{i}{n}\right)\right\}^2\right] + d\,\log n.$$

It can be shown that choosing the model with the minimum BIC value is approximately equivalent to choosing the model with the largest posterior probability (e.g., Haste, Tibshirani and Friedman 2002).

Once d is selected, the empirical Bayes wavelet thresholding procedure of Johnstone and Silverman (2003) is applied to the residuals $y_i - \hat{f}_P(i/n)$ to obtain $\hat{f}_W$. The final estimate $\hat{f}_{PW}$ is then obtained as $\hat{f}_P + \hat{f}_W$.

3.3. *Other approaches*

We have also studied other approaches for choosing d and λ. These include cross-validation (Nason 1996) and the minimum description length principle (e.g., see Lee 2002 and references given therein). However, the practical performances of these approaches are inferior to the Bayesian approach, and hence we omit their descriptions.

4. Simulation results

4.1. *Setup*

This section investigates the relative practical performances of four wavelet regression methods. The four wavelet regression methods tested were:

1. psure: the polynomial SURE procedure described in Section 3.1;
2. osure: the original SURE procedure developed by Donoho and Johnstone (1995); i.e., no polynomial boundary treatment is present.

Table 1. *Formulae of the test functions. All have the same domain* $x \in [0, 1]$

Test function	Formula
1	*blocks* of Donoho and Johnstone (1994)
2	*doppler* of Donoho and Johnstone (1994)
3	$7x + blocks$
4	$x + doppler$
5	$7(x - 0.8)^2 + blocks$
6	$3(x - 0.6)^2 + doppler$

3. pebayes: the polynomial empirical Bayes procedure described in Section 3.2; and
4. oebayes: the original empirical Bayes procedure developed by Johnstone and Silverman (2003); i.e., no polynomial boundary treatment is present.

Throughout the whole simulation a maximal value of $d = 3$ was used for $\hat{f}_P$, while a coiflet with 5 vanishing moments and the periodic boundary assumption were used for $\hat{f}_W$. Note that the orthogonality between the polynomial basis $\{x, \ldots, x^d\}$ and the wavelet basis is preserved.

Altogether 6 test functions were used. They are listed in Table 1 and are displayed in Fig. 1. Test Functions 1 and 2 are the classical wavelet testing functions *Blocks* and *Doppler* advocated by Donoho and Johnstone (1994). Test Functions 3 to 6 are constructed by adding either a linear or quadratic trend to these two functions. Notice that for Test Functions 1 and 2, it is reasonable to assume periodic boundary conditions, while for Test Functions 3 to 6 boundary adjustment is strongly preferred.

The signal-to-noise ratio (snr) is defined as: snr $= \|\boldsymbol{f}\|/\sigma$ (the same as Donoho and Johnstone 1994), and three levels were used: high snr $= 7$, medium snr $= 5$ and low snr $= 3$. Also, three different sample sizes were used: $n = 512$, 1024 and 2048.

For each combination of test function, snr and n, 100 sets of noisy observations were simulated. For each simulated data set, the above four wavelet regression methods were applied to estimate the test function. Figure 2 displays, for those cases associated with medium snr, boxplots of the values of the mean-squared-errors (MSE) for all estimated regression functions. Here MSE of a $\hat{f}$ is defined as $\text{MSE}(\hat{f}) = n^{-1} \sum_{i=1}^{n} \{f(i/n) - \hat{f}(i/n)\}^2$. Paired Wilcoxon tests were also applied to test if the difference between the median $\text{MSE}(\hat{f})$ values of two wavelet methods is significant or not. The significance level used was 1.25%, and the relative rankings, with 1 being the best, are listed in Table 2. Ranking the methods in this manner provides an indicator of the relative merits of the methods (e.g.,

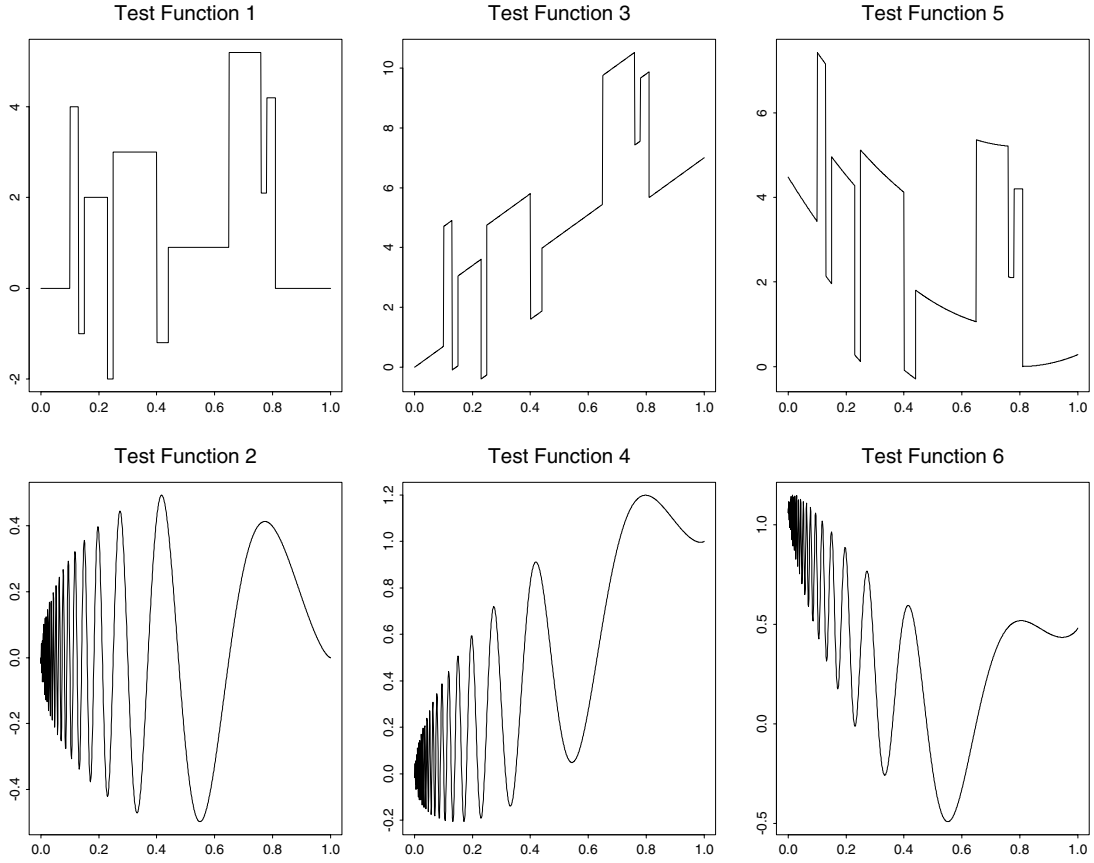

Fig. 1. *Plots of test functions used in the simulation*

Table 2. *Pairwise Wilcoxon rankings, for medium snr, for the four wavelet regression procedures tested*

	$n = 512$				$n = 1024$				$n = 2048$			
Test function	psure	osure	pebayes	oebayes	psure	osure	pebayes	oebayes	psure	osure	pebayes	oebayes
1	3	4	1.5	1.5	3	4	1.5	1.5	3	4	1.5	1.5
2	3	4	2	1	3	4	2	1	3	4	2	1
3	3	4	1	2	3	4	1	2	3	4	1	2
4	3	4	1	2	3	4	1	2	3	4	1	2
5	3	4	1	2	3	4	1	2	3	4	1	2
6	3	4	1	2	3	4	1	2	3	4	1	2

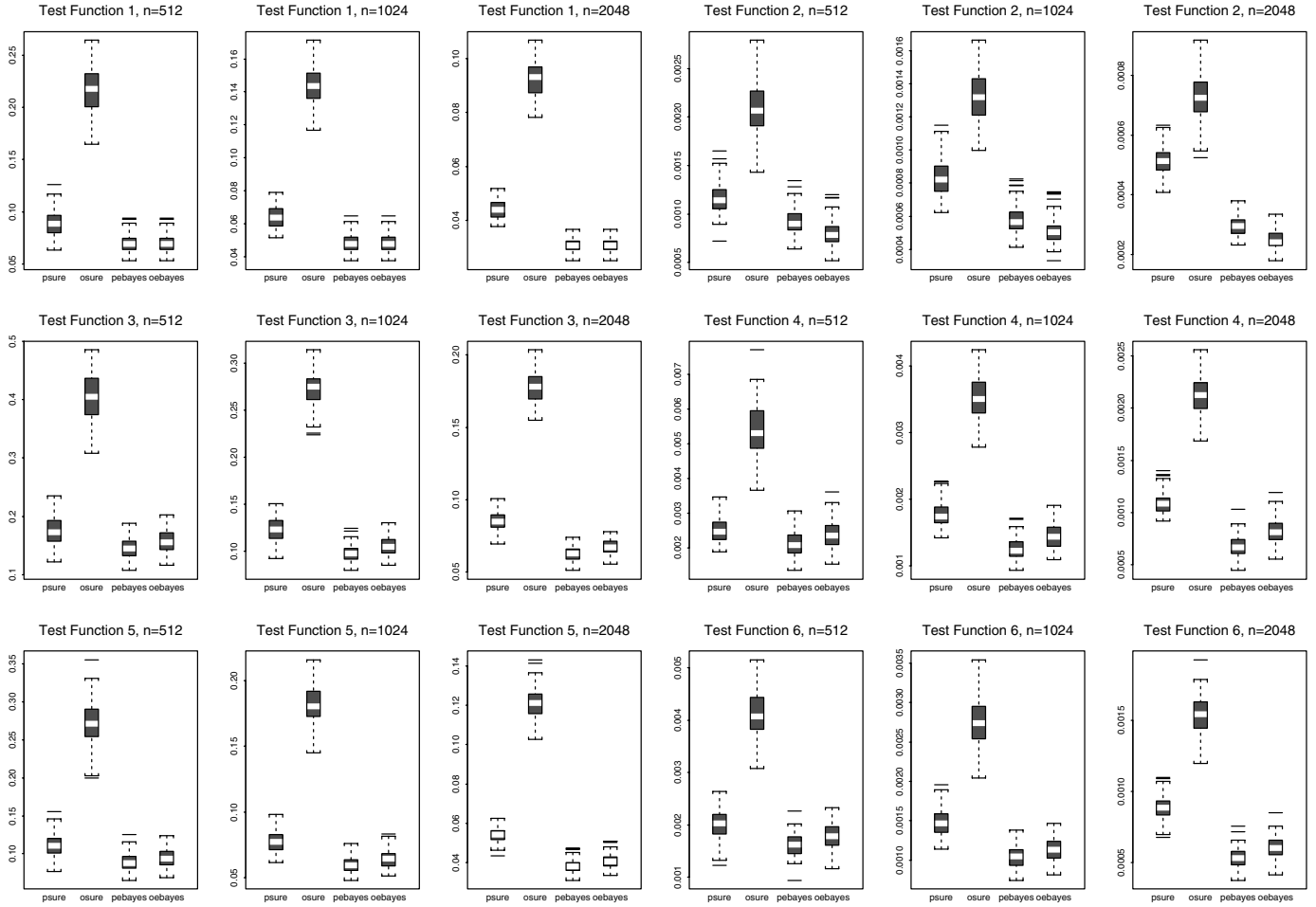

Fig. 2. *Boxplots of MSE($\hat{f}$) values, medium snr*

see Wand 2000). Boxplots of MSE($\hat{f}$) values and Wilcoxon test rankings for low and high snrs are similar, and hence are omitted.

4.2. *Results*

The grand averaged Wilcoxon rankings taken over all combinations of snrs and sample sizes are 3.01, 3.98, 1.25 and 1.76, respectively, for psure, osure, pebayes and oebayes. We have also computed similar grand averaged rankings for the MSE of those observations that are in the boundary region $[0, 0.05] \cup [0.95, 1]$ of the test functions. These rankings are, in the same order as before, 2.26, 3.58, 1.76 and 2.40.

From the simulation results, the following empirical observations can be made: (i) No method performed uniformly the best, (ii) pebayes seems to be superior to oebayes, while psure is superior to osure, and, (iii) psure seems to perform better than oebayes at the boundary. Overall these simulation results seem to suggest that it is preferable to incorporate the polynomial basis to the wavelet regression problem.

5. Conclusion

In this article the problem of automatic polynomial wavelet regression was considered. Two automatic methods were proposed for choosing the free parameters involved. Results from a simulation study show that automatic polynomial wavelet regression is a promising alternative to ordinary wavelet regression.

Acknowledgments

The authors would like to thank the referee and the associate editor for their constructive comments which lead to the development of the Bayesian method and a more succinct version of the paper. This work was partially supported by the National Science Foundation under Grant No. 0203901 and the Natural Sciences and Engineering Research Council of Canada.

References

Daubechies I. 1992. Ten Lectures on Wavelets. SIAM, Philadelphia.

Donoho D.L. and Johnstone I.M. 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81: 425–455.

Donoho D.L. and Johnstone I.M. 1995. Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association 90: 1200–1224.

Haste T., Tibshirani R., and Friedman J. 2002. The Elements of Statistical Learning. Springer, New York.

Johnstone I.M. and Silverman B.W. 2003. Empirical Bayes selection of wavelet thresholds. Unpublished manuscript.

Lee T.C.M. 2002. Tree-based wavelet regression for correlated data using the minimum description length principle. Australian and New Zealand Journal of Statistics 44: 23–39.

Mallows C.L. 1973. Some comments on C_p. Technometrics 15: 661–675.

Nason G.P. 1996. Wavelet shrinkage using cross-validation. Journal of the Royal Statistical Society Series B 58: 463–479.

Oh H.-S., Naveau P., and Lee G. 2001. Polynomial boundary treatment for wavelet regression. Biometrika 88: 291–298.

Schwarz G. 1978. Estimating the dimension of a model. The Annals of Statistics 6: 461–464.

Wand M.P. 2000. A comparison of regression spline smoothing procedures. Computational Statistics 15: 443–462.