02/06

Automatic Polynomial Wavelet Regression

Thomas C. M. Lee* and Hee-Seok Oh†

May 9, 2002

Abstract

In Oh, Naveau & 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 three automatic methods for making such choices. Finite sample performances of these three methods are evaluated via numerical experiments.

Keywords: boundary adjustment, cross–validation, minimum description length principle, polynomial wavelet regression, Stein's unbiased risk estimation, wavelet thresholding

*Corresponding Author. Department of Statistics, Colorado State University and Geophysical Statistics Project, National Center for Atmospheric Research. Postal: Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA. Phone: (970) 491-2185. Fax: (970) 491-7895. Email: tlee@stat.colostate.edu.

†Geophysical Statistics Project, National Center for Atmospheric Research. Postal: Geophysical Statistics Project, National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307-3000, USA. Phone: (303) 497-1340. Fax: (303) 497-1333. Email: heeseok@ucar.edu.

1 Introduction

Suppose a set of noisy data satisfying the following is observed:

$$y_i = f(\frac{i}{n}) + \epsilon_i, \quad i = 1, \ldots, n = 2^J,$$

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 et al. (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 three automatic methods for selecting such values. These three methods are based on three different statistical model selection principles. They are cross–validation, the minimum description length principle (Rissanen 1989), and Stein's unbiased risk estimation (Stein 1981). To the best of the authors' knowledge, no such automatic methods have been previously discussed in the literature.

The rest of this article is organized as follows. Background material is provided in Section 2. Section 3 presents the three 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:

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

where $\phi_k(x) = 2^{1/2}\phi(2x-k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^jx-k)$. 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:

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

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 $\hat{d}^H_{j,k} = \hat{d}_{j,k}I_{\{|\hat{d}_{j,k}|>\lambda\}}$.

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

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

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 et al. (2001), both analytically and empirically, that $\hat{f}_{PW}(x)$ is superior to $\hat{f}_W(x)$. The goal of this paper is to propose and compare the use of three different automatic methods for choosing both d and λ. Notice that no such automatic methods are proposed by Oh et al. (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 $\{x, ..., x^d\}$ 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 $l = 1, ..., d$. 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 et al. (2001) suggest estimating the parameters in (3) by first regressing the observations $\{y_i\}$ on the set $\{x, ..., x^d\}$ for fixed d and then applying wavelet regression to the residuals of the polynomial regression.

3 The Proposed Automatic Selection Methods

This section proposes three automatic methods for choosing d and λ. These three methods are based on cross–validation, the minimum description length principle, and Stein's unbiased risk estimation.

3.1 Cross–Validation

This section presents a two–step cross–validation (CV) procedure for choosing the values of d and λ. The first step of the procedure constructs a CV score for choosing d, while the second step of the procedure applies the method of Nason (1996) to choose λ.

Let $\hat{f}_P$ be the polynomial estimator of degree d computed by regressing y_i's on $x, \ldots, x^d$. Also let $\hat{f}_{P,-i}$ be a polynomial estimator of degree d obtained similarly, but without using the ith data point y_i. Note that the dependence of both $\hat{f}_P$ and $\hat{f}_{P,-i}$ on d is suppressed from the notation. The first step of the procedure is to choose the d that minimizes the following CV score:

$$\mathrm{CV}(d) = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \hat{f}_{P,-i}(\frac{i}{n})\right\}^2.$$

It is straightforward to show that

$$y_i - \hat{f}_{P,-i}(\frac{i}{n}) = \frac{y_i - \hat{f}_P(\frac{i}{n})}{1 - H_{ii}},$$

where H_{ii} is the ith diagonal element of the hat matrix (e.g., see Weisberg 1985, pp. 109). Thus, the above CV score can be expressed as

$$\mathrm{CV}(d) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i - \hat{f}_P(\frac{i}{n})}{1 - H_{ii}}\right\}^2.$$

The first step of the procedure chooses d as the minimizer of $\mathrm{CV}(d)$.

Once d is selected by minimizing $\mathrm{CV}(d)$, one can then apply wavelet regression, with an appropriate value of λ, to the residuals of the polynomial regression. To choose such a λ, we suggest using the procedure developed by Nason (1996). This procedure is an approximation to the traditional leave–one–out CV. For details, see Nason (1996).

To sum up, a polynomial estimator $\hat{f}_P$ is first obtained by regressing y_i's on $x_1, \ldots, x_d$, where d is chosen by minimizing $\mathrm{CV}(d)$. Then the CV wavelet regression procedure of Nason (1996) is applied to the residuals $y_i - \hat{f}_P(\frac{i}{n})$ to obtain $\hat{f}_W$. The final estimate $\hat{f}_{PW}$ of f is then computed as $\hat{f}_P + \hat{f}_W$.

3.2 The Minimum Description Length Principle

Another automatic model selection method, the minimum description length (MDL) principle of Rissanen (1989), has also been applied to wavelet regression with *hard–thresholding*;

see Lee (2002), Moulin (1996) and Saito (1994). We first review the use of MDL for ordinary wavelet regression (i.e., without polynomial adjustment). Then we demonstrate how MDL can be applied for choosing both d and λ.

In general the MDL principle *defines* the best fitting model as the one that produces the shortest code length of the data. Here the code length of an object can be treated as the amount of memory space that is required to store the object. One way to apply the principle is to split the code length for a set of data into two parts: (i) a fitted model plus (ii) the data "conditioned on" the fitted model; i.e., the residuals. For ordinary wavelet regression without polynomial adjustment, the data are $y_1, \ldots, y_n$, a fitted model is $\hat{f}_W$, and the residuals are $y_i - \hat{f}_W(\frac{i}{n})$, $i = 1, \ldots, n$. Notice that any $\hat{f}_W$ can be uniquely specified by a set of non–zero hard–thresholded values $\hat{d}^H_{j,k}$'s.

If $L(z)$ denotes the code length of object z, one has the following decomposition

$$\begin{aligned} L(\text{“data”}) &= L(\text{“fitted model } \hat{f}_W\text{”}) + L(\text{“residuals”}) \\ &= L(\text{“non–zero } \hat{d}^H_{j,k}\text{'s”}) + L(\text{“residuals”}), \end{aligned}$$

and MDL defines the best fitting $\hat{f}_W$ as the minimizer of $L(\text{“data”})$. If a *tree constraint* is imposed on those non–zero $\hat{d}^H_{j,k}$'s (see below), the following expression for $L(\text{“data”})$ is derived by Lee (2002):

$$\mathrm{MDL}(\lambda) = \frac{m}{2}\log_2 n + \frac{n}{2}\log_2 \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \hat{f}_W(\frac{i}{n})\right\}^2, \qquad (4)$$

where m is the number of non–zero $\hat{d}^H_{j,k}$'s. It is suggested by Lee (2002) to select $\hat{f}_W$, or λ, as the minimizer of $\mathrm{MDL}(\lambda)$. Also, a tree–growing algorithm is proposed by Lee (2002) for the practical minimization of $\mathrm{MDL}(\lambda)$.

The above–mentioned tree constraint can be described as follows. For $0 \leq j < J$, if $\hat{d}^H_{j,k}$ is thresholded to be zero, then both $\hat{d}^H_{j+1,2k}$ and $\hat{d}^H_{j+1,2k+1}$ are automatically set to zero, regardless of their magnitudes. The idea behind this constraint is that, it follows the intuition that wavelet coefficients at a finer level (or resolution) should have a higher chance of being thresholded when comparing to those coarser level wavelet coefficients at the same relative location.

The criterion MDL(λ) can be straightforwardly extended for the choosing of both d and λ: one just has to add the code length for encoding the estimated polynomial coefficients $\hat{\alpha}_0, \ldots, \hat{\alpha}_d$. Then the expression for L("data") admits the form

$$\begin{aligned} L(\text{``data''}) &= L(\text{``polynomial coefficients''}) + L(\text{``fitted model } \hat{f}_W\text{''}) + L(\text{``residuals''}) \\ &= L(\hat{\alpha}_0, \ldots, \hat{\alpha}_d) + L(\text{``non--zero } \hat{d}^H_{j,k}\text{'s''}) + L(\text{``residuals''}). \end{aligned}$$

The term $L(\hat{\alpha}_0, \ldots, \hat{\alpha}_d)$ can be approximated by using the following result of Rissanen (1989, pp 55–56): if a parameter estimate is estimated from N observations, then its code length is asymptotically $\frac{1}{2}\log_2 N$. Since each of the $\hat{\alpha}_l$'s is estimated from n data points, $L(\hat{\alpha}_0, \ldots, \hat{\alpha}_d) = \frac{d+1}{2}\log_2 n$. Therefore we have

$$\text{MDL}(d, \lambda) = \frac{d+1+m}{2}\log_2 n + \frac{n}{2}\log_2 \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \hat{f}_{PW}(\frac{i}{n})\right\}^2. \tag{5}$$

Thus our MDL choice of (d, λ) is defined as the pair that minimizes MDL(d, λ).

In practice MDL(d, λ) is minimized in the following manner. First, for a given d, a polynomial estimator $\hat{f}_P$ is obtained by regressing y_i's on $x, \ldots, x^d$. Then the MDL wavelet procedure of Lee (2002) (i.e., aims to minimize MDL(λ)) is applied to the polynomial residuals $y_i - \hat{f}_P(\frac{i}{n})$ to obtain $\hat{f}_W$, and hence $\hat{f}_{PW} = \hat{f}_P + \hat{f}_W$. This process is repeated for different values of d, and the $\hat{f}_{PW}$ that gives the smallest value of $\mathrm{MDL}(d, \lambda)$ is taken as the final estimate.

3.3 Stein's Unbiased Risk Estimation

In this section we attempt to choose the pair of d and λ that minimizes the risk between f and $\hat{f}_{PW}$, defined as $E\|f - \hat{f}_{PW}\|^2$. A two–step procedure similar to the CV procedure described in Section 3.1 is used. 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 & 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 $y_i - \hat{f}_P(\frac{i}{n})$.

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

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

where $\hat{\sigma}^2$ is any consistent estimator of σ^2. For $\hat{\sigma}^2$, we can use the estimators of Gasser, Sroka & Jennen-Steinmetz (1986) or Hall, Kay & Titterington (1990). 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 $\hat{f}_P$. This sort of scheme has been used by other authors in various nonparametric regression problems; e.g., see Rice (1984) and Eubank & Hart (1992).

Once d and hence $\hat{f}_P$ is obtained, the SURE procedure of Donoho & Johnstone (1995) is applied to the residuals $y_i - \hat{f}_P(\frac{i}{n})$ to obtain $\hat{f}_W$. Then the final estimate is obtained as

$\hat{f}_{PW} = \hat{f}_P + \hat{f}_W$.

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. pcv: the cross–validation procedure described in Section 3.1;
2. pmdl: the MDL procedure described in Section 3.2;
3. psure: the SURE procedure described in Section 3.3; and
4. osure: the original SURE procedure developed by Donoho & Johnstone (1995); 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 $\{x, ..., x^d\}$ and the wavelet basis is preserved.

Altogether 8 test functions were used. They are listed in Table 1 and are displayed in Figure 1. Test Functions 1 and 2 are the classical wavelet testing functions *Blocks* and *Doppler* advocated by Donoho & Johnstone (1994). Test Functions 3 to 6 are constructed by adding either a linear or quadratic trend to these two functions. Lastly Test Functions 7 and 8 are two simple functions that have been used by other authors (e.g., Fan & Gijbels

1995). Notice that for Test Functions 1 and 2, it is reasonable to assume periodic boundary condition, while for Test Functions 3 to 8 boundary adjustment is strongly preferred.

The signal–to–noise ratio (snr) is defined as: snr $= \|f\|/\sigma$ (the same as Donoho & 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 $\mathrm{MSE}(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}\{f(\frac{i}{n}) - \hat{f}(\frac{i}{n})\}^2$. Paired Wilcoxon tests were also applied to test if the difference between the median $\mathrm{MSE}(\hat{f})$ 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 (see Wand 2000).

Boxplots of $\mathrm{MSE}(\hat{f})$ values and Wilcoxon test rankings for low and high snrs are similar, and hence are omitted. However, averaged Wilcoxon rankings for each test function are reported in Table 3. These averages were taken over all combinations of snrs and sample sizes.

For the purpose of visual evaluation, Figure 3 displays some typical estimated regression functions obtained by the above four wavelet regression methods. From these plots it is obvious that osure suffers from boundary effects.

4.2 Results

From the simulation results, the following empirical observations can be made.

1. No method performed uniformly the best.

2. psure and pcv gave nearly identical results.

3. psure and pcv always outperformed osure, including Test Functions 1 and 2, which do not require boundary adjustment.

4. psure and pcv gave the best results for functions contain abrupt changing structures (Test Functions 1 to 6).

5. pmdl performed the best for functions with simple structures (Test Functions 7 and 8).

From these observations two important lessons could be learnt. First, it seems to be preferable to incorporate the polynomial basis to the wavelet regression problem (Point 3). Secondly, if the true regression function is suspected to be simple in structure, use pmdl; otherwise, use pcv or psure (Points 4 and 5).

5 Conclusion

In this article the problem of automatic polynomial wavelet regression was considered. Three 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.

References

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Vol. 1, Philadelphia: Society for Industrial and Applied Mathematics.

Donoho, D. L. & Johnstone, I. M. (1994), 'Ideal spatial adaptation by wavelet shrinkage', *Biometrika* **81**, 425–455.

Donoho, D. L. & Johnstone, I. M. (1995), 'Adapting to unknown smoothness via wavelet shrinkage', *Journal of the American Statistical Association* **90**, 1200–1224.

Eubank, R. L. & Hart, J. D. (1992), 'Testing goodness-of-fit in regression via order selection criteria', *The Annals of Statistics* **20**, 1412–1425.

Fan, J. & Gijbels, I. (1995), 'Data–driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation', *Journal of the Royal Statistical Society Series B* **57**, 371–394.

Gasser, T., Sroka, L. & Jennen-Steinmetz, C. (1986), 'Residual variance and residual pattern in nonlinear regression', *Biometrika* **73**, 625–633.

Hall, P., Kay, J. W. & Titterington, D. M. (1990), 'Asymptotically optimal difference-based estimation of variance in nonparametric regression', *Biometrika* **77**, 521–528.

Lee, T. C. M. (2002), 'Tree–based wavelet regression for correlated data using the minimum description length principle', *Australian & New Zealand Journal of Statistics* 44, 23–39.

Mallows, C. L. (1973), 'Some comments on C_p', *Technometrics* **15**, 661–675.

Moulin, P. (1996), Signal estimation using adapted tree–structured bases and the MDL principle, *in* 'Proceedings of the IEEE Signal Processing Symposium on Time–Frequency and Time–Scale Analysis', Paris, pp. 141–143.

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. & Lee, G. (2001), 'Polynomial boundary treatment for wavelet regression', *Biometrika* **88**, 291–298.

Rice, J. A. (1984), 'Bandwidth choice for nonparametric regression', *The Annals of Statistics* **12**, 1215–1230.

Rissanen, J. (1989), *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore.

Saito, N. (1994), Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion, *in* E. Foufoula-Georgiou & P. Kumar, eds, 'Wavelets in Geophysics', New York: Academic Press, pp. 299–324.

Stein, C. M. (1981), 'Estimation of the mean of a multivariate normal distribution', *The Annals of Statistics* **9**, 1135–1151.

Wand, M. P. (2000), 'A comparison of regression spline smoothing procedures', *Computational Statistics* **15**, 443–462.

Weisberg, S. (1985), *Applied Linear Regression*, second edn, John Wiley & Sons: New York.

Test Function	Formula
1	*blocks* of Donoho & Johnstone (1994)
2	*doppler* of Donoho & Johnstone (1994)
3	$7x + blocks$
4	$x + doppler$
5	$7(x-0.8)^2 + blocks$
6	$3(x-0.6)^2 + doppler$
7	$(4x-2) + \exp\{-16(4x-2)^2\}$
8	$(4x-2) + \sin(8x-4) + 2\exp\{-16(4x-2)^2\}$

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

Test Function	$n = 512$				$n = 1024$				$n = 2048$			
	pcv	pmdl	psure	osure	pcv	pmdl	psure	osure	pcv	pmdl	psure	osure
1	1.5	4	1.5	3	1.5	3.5	1.5	3.5	1.5	4	1.5	3
2	1.5	3	1.5	4	1.5	3	1.5	4	1.5	4	1.5	3
3	1.5	3	1.5	4	1.5	3	1.5	4	1.5	3	1.5	4
4	1.5	3	1.5	4	2	2	2	4	1.5	4	1.5	3
5	1.5	3.5	1.5	3.5	1.5	3.5	1.5	3.5	1.5	4	1.5	3
6	1.5	3	1.5	4	1.5	3	1.5	4	1.5	4	1.5	3
7	2.5	1	2.5	4	2.5	1	2.5	4	2.5	1	2.5	4
8	3.5	1	3.5	2	3.5	1	3.5	2	2.5	1	2.5	4

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

Test Function	pcv	pmdl	psure	osure
1	1.50	3.89	1.50	3.11
2	1.56	3.33	1.56	3.56
3	1.72	3.11	1.72	3.44
4	1.56	3.11	1.56	3.78
5	1.50	3.72	1.50	3.28
6	1.50	3.33	1.50	3.67
7	2.33	1.33	2.33	4.00
8	2.72	1.56	2.72	3.00
overall	1.80	2.92	1.80	3.48

Table 3: Averaged pairwise Wilcoxon rankings for the four wavelet regression procedures tested. The averages were taken over all combinations of snrs and sample sizes.

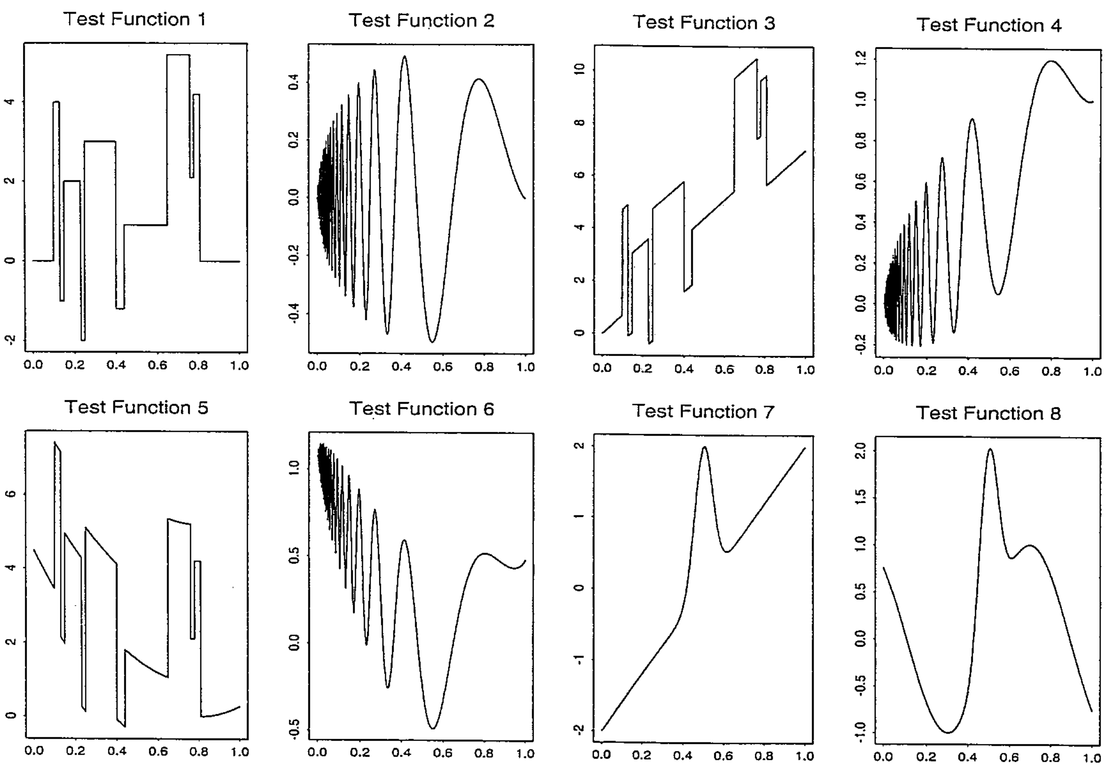

Figure 1: Plots of test functions used in the simulation.

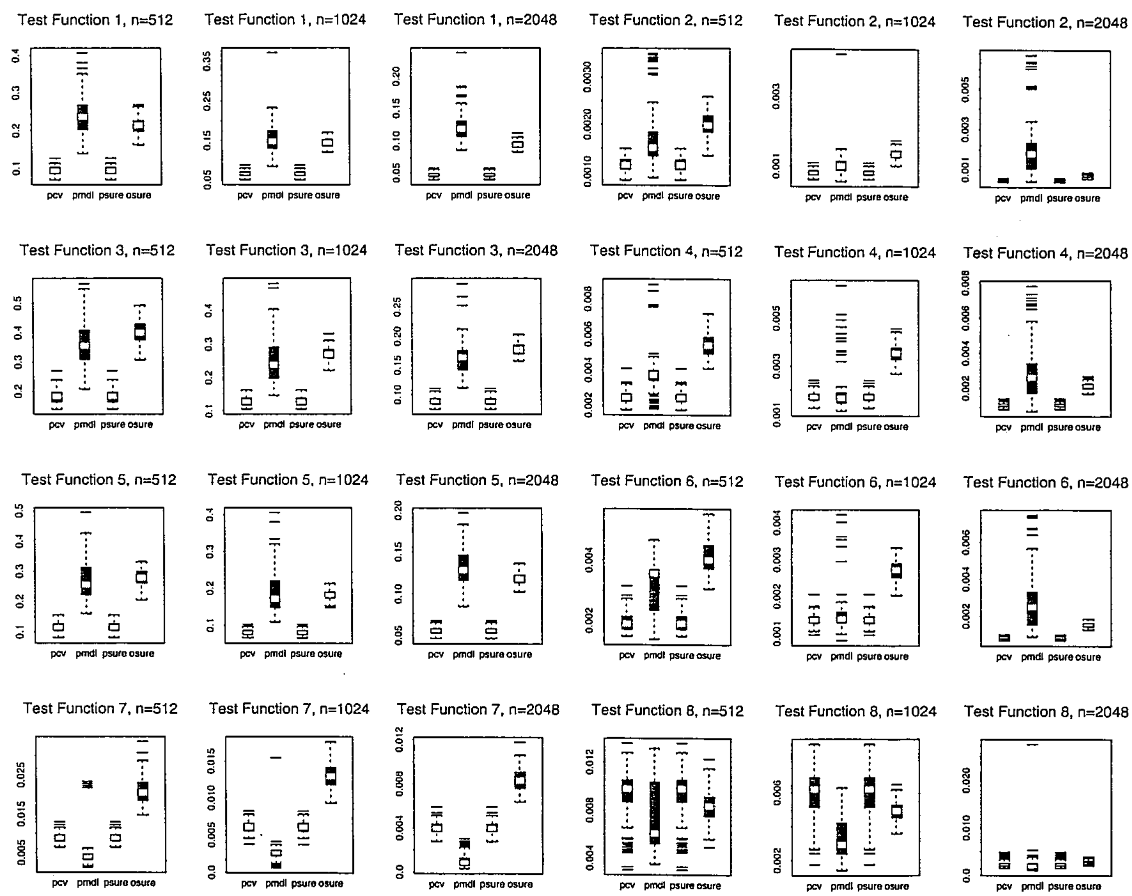

Figure 2: Boxplots of MSE($\hat{f}$) values, medium snr.

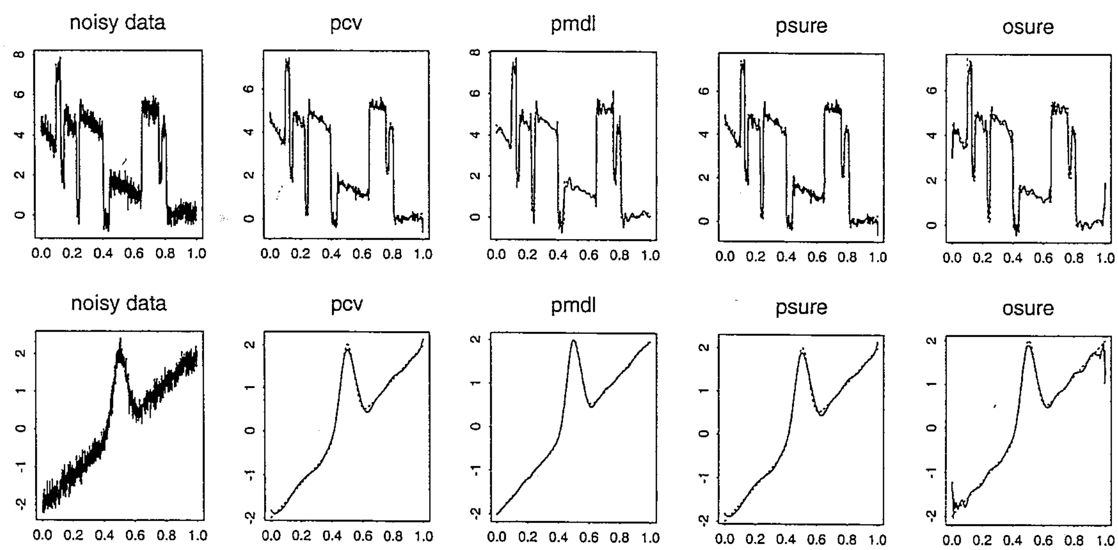

Figure 3: Plots of true (dotted lines) and estimated (solid lines) regression functions.