

# Local Polynomial Wavelet Regression with Missing at Random

**Alsaedi M. Altaher**

School of Mathematical Sciences  
Universiti Sains Malaysia  
11800 Penang, Malaysia  
assaedi76@yahoo.com

**Mohd Tahir Ismail**

School of Mathematical Sciences  
Universiti Sains Malaysia  
11800 Penang, Malaysia  
mtahir@usm.my

## **Abstract**

This paper considers the problem of missing at random for local polynomial wavelet regression (LPWR). Such missing observations can pose serious problems, affect the bias and finally misleading conclusions might be drawn. We introduce a new imputation method for LPWR; based on a combination of bootstrap and local polynomial prediction model. The proposed method uses an initial-bootstrap estimate which will be iterated through a local polynomial regression model until convergence. Practical implementations show that the proposed method is easy to implement, fast and achieves high performance even for data with non-Gaussian noises. A simulation study is conducted to assess the numerical achievement of the proposed method with comparison to other existing ones.

**Mathematics Subject Classification:** 65T60, 62F40, 68U20

**Keywords:** local polynomial regression, wavelet, missing data, simulation

## **1 Introduction**

Wavelet regression has been a power tool in data analysis and curve estimation over the past two decades and is still an expanding area of ongoing research.

It begins by the seminal paper of [4], where the concept of wavelet regression has been introduced to the statistical literature. The main target is to recover an unknown function say  $f$ ; based on sampled data  $y_1, y_2, \dots, y_n$  that are contaminated with additive noise.

$$y_i = f\left(\frac{i}{n}\right) + \varepsilon_i \quad ; i = 1, 2, \dots, n = 2^J; \quad J = 1, 2, \dots \quad (1)$$

Only a few general assumptions are made. The function  $f$  is assumed to be squared integral on the interval  $[0, 1]$ . The sequence  $\{\varepsilon_i\}_{i=1}^n$  is independent and identically distributed normally with mean zero and constant variance,  $\sigma^2$ . local polynomial wavelet regression fitting (LPWR, [13]) is usually considered as an improvement boundary adjustment in wavelet regression. Practical implementations have proved that LPWR performs much better than polynomial wavelet regression introduced by [14]. The LPWR has been using enormously, simply because the automatic boundary correction involved in (LPWR) offers high estimation quality without paying extra attention to determine what kind of boundary rules that should be used when applying classical wavelet regression. The question now: what if some observations in the response variable are missing (say; missing at random)?. The (LPWR) is designed for complete data sets and therefore, if some observations are missing, LPWR cannot immediately be applied due to two main restrictions: (1) the observations are equally spaced, and (2) the sample size, say,  $n$  is dyadic, i.e.,  $2^J$  for some integer  $J$ . Hence, it necessitates finding a proper way to handle the missing data in local polynomial wavelet regression. So far and to the best of the author's knowledge no investigations address the (LPWR) in the presence of missing data. However, there are some methods and suggestions to fill in the missing data in order to apply wavelet transform as a first step toward wavelet estimation.

A common recommendation in literature suggests to "pad with zeroes", that is, adding zeroes to increase the sample size up to the next largest size ( $n = 2^j$ ) [18]. Problem may arise when adding zeros at the end or beginning of the data set; that may count as a change point when there is none. Chatterjee [2] suggested using the median of the response variable to replace the missing values. If a large patch of observations is missing, then extra values are added by generating data randomly from normal distribution with  $\mu = \text{median}$  and small variance i.e. ( $\sigma^2 = 0.01$ ). Kim et al. [8] proposed a wavelet denoising method in the presence of missing data. This approach is based on a coupling of wavelet shrinkage and hierarchical or (h-likelihood).

Our contribution is mainly to study (LPWR) more intensively with respect to missing data. As a first step, one might achieve some simulations in order to discuss differences in performance among the most popular imputation methods such as: mean-median substitutions, non-parametric bootstrap and EM algorithm. With this in mind, we introduce a simple, an efficient method

based on a combination of bootstrap with local polynomial prediction model. Our proposed method uses an initial estimate obtained from nonparametric bootstrap method. This initial filling will be enhanced iteratively through a local polynomial regression model until convergence. The algorithm of our method is *simple*, *fast* and *easy* to implement with a high accuracy compared to traditional methods.

The rest of this paper is organized as follows: In section 2, we briefly review the so-called local polynomial wavelet regression (LPWR). Section 3 is devoted for model and Missing Mechanism. In section 4 we give a brief description to some missing data imputation methods followed by our proposed method in section 5. Section 6 examines the practical performance of six imputation methods via a simulation study. Conclusion is drawn in section 7.

## 2 Review on Local Polynomial Wavelet Regression

This section describes local polynomial shrinkage (Regression) proposed by [13]. Let  $\phi$  and  $\psi$  be father (scaling) and mother (dilation) wavelet functions respectively. The function estimator of  $f$  can be written in terms of wavelets as:

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

Where:  $\phi_k(x) = 2^{1/2} \phi(2x - k)$ ,  $\psi_{j,k}(x) = 2^{j/2} \psi(2x - k)$ ,  $\hat{c}_{0,k} = \sum y_i \phi_k(x_i)$  and  $\hat{d}_{j,k}^s$  denotes the soft threshold coefficient.

Given a wavelet coefficient,  $d_{j,k}$  and threshold value,  $\lambda$  the soft threshold rule,  $\hat{d}_{j,k}^s$  of the wavelet coefficient can be defined as:

$$\hat{d}_{j,k}^s = \text{sgn}(\hat{d}_{j,k}) ( |\hat{d}_{j,k}| - \lambda ) I_{(|\hat{d}_{j,k}| > \lambda)}$$

Here  $I$  refers to the usual indicator function. The Hybrid local polynomial wavelet regression method was introduced by [13] as an improvement boundary adjustment in wavelet regression. Instead of using the global polynomial fit as in [14], it was proposed using a local polynomial fit,  $\hat{f}_{LP}$ . Therefore, the Hybrid Local Polynomial wavelet Shrinkage estimator,  $\hat{f}_H(x)$  can be written as:

$$\hat{f}_H(x) = \hat{f}_{LP}(x) + \hat{f}_W(x)$$

As shown in [13],  $\hat{f}_H(x)$  is computed through an iterative algorithm inspired by the back-fitting algorithm of [6]. The following steps summarize the key points to find the final Hybrid Local polynomial wavelet regression fitting  $\hat{f}_H$

1. Select an initial estimate  $\hat{f}_0$  for  $f$  and let  $\hat{f}_H = \hat{f}_0$ .
2. For  $j = 1, 2, \dots$  iterate the following steps:
  - Apply wavelet regression to the residuals  $y_i - \hat{f}_H$  and obtain  $\hat{f}_W^j$ .
  - Estimate  $\hat{f}_{LP}^j$  by fitting local polynomial regression to  $y_i - \hat{f}_W^j$ .
3. Stop if  $\hat{f}_H = \hat{f}_{LP}^j + \hat{f}_W^j$  converges.

The initial estimate,  $\hat{f}_0$  can be found by using Friedman's smoother (known as *supsmu*, available in R), while smoothing parameter is selected by cross validation or direct Plug-in criterion (available at *KernSmooth* package in R). EbayesThresh procedure [7] is used for wavelet estimation at the second step.

### 3 Model and Missing Mechanism

Consider again the non-parametric regression model (1). It is obvious that the basic inference begins by considering the random sample  $(x_i, y_i)$  which can be rewritten as  $(x_i, y_i, \delta_i)$  in presence of missing data. Where: the design point  $x_i$  is assumed to be completely observed;  $\delta_i = 0$  if the response variable  $y_i$  is missing. Otherwise,  $\delta_i = 1$  for  $i = 1, 2, \dots, n$ . Three common different missing mechanisms are in literature: MCAR (missing completely at random), MAR (missing at random) and MNAR (missing not at random). However, a more realistic assumption widely used in literature is the so-called MAR (missing at random) assumption see, [15],[1] and [10]. The MAR assumption would usually require that there exists a chance mechanism, denoted by  $p(y)$  such that

$$p(\delta = 1/y, x) = p(\delta = 1/x) = p(x)$$

In other words, it means that  $\delta$  and  $y$  are conditionally independent given  $x$ .

## 4 Some Classical Missing Data Imputations

### 4.1 Mean–Median Imputation

The first way to address the issue of missing data is to replace these missing values by the annual mean of the response variable. Since the annual mean is easily affected by presence of outliers, the median can be used as an alternative. In case of a large patch of observations is missing, Chatterjee [2] suggested adding extra values by generating data randomly from normal distribution with mean ( $\mu = \text{median}$ ) and small variance i.e. ( $\sigma^2 = 0.01$ ). Another kind of mean substitution is stochastic mean imputation which also known as zero order regression (ZOR+). It can be used to fill in the missing observations

$(y_j)$ . The stochastic mean is simply the annual mean,  $\bar{y}$  plus random error from normal distribution:

$$y_j = \bar{y} + N(0, \sigma^2)$$

Where  $\sigma^2$  is about the variance of the residuals of the regression model,[12].

## 4.2 Non-Parametric Bootstrap Imputation

Bootstrap method is a powerful tool for statistical inference when the standard approximations are not available or unreliable. It was first introduced as in [3] and since then, much research in more complicated sittings has followed. Here, we state a few steps based on bootstrap to bear on missing data problem in the response variable of local polynomial wavelet regression.

1. Replace all the missing value by the estimated mean,  $\bar{y}$ .
2. Draw  $B$  independent bootstrap samples and compute the mean for each sample say;  $\bar{y}_i$ .
3. Compute the overall bootstrap mean and the variance such as:

$$\bar{y}_b = \left( \sum_{i=1}^B \bar{y}_i \right) / B, \quad s_b^2 = \frac{1}{B-1} \sum_{i=1}^B (\bar{y}_i - \bar{y}_b)^2$$

4. Generate  $m$  data point randomly from normal distribution with  $(\mu = \bar{y}_b; \sigma^2 = s_b^2)$  to replace the  $m$ -missing observations. This is how to obtain pseudo observations in a simple and a quick way.

## 4.3 EM Algorithm

Another modern way to address the missing issue is by obtaining maximum likelihood estimators (MLE), and one of the most common is called the Expectation-Maximization algorithm, abbreviated as the EM algorithm. The EM algorithm is an iterative procedure that finds the MLE of the parameter vector by repeating the following steps:

1. The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates.
2. The maximization M-step: Given complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step.

These two steps have to be iterated until the iterations converge. An excellent discussion of the EM algorithm and its solution is provided by [9].

## 5 Bootstrap-Iterative Polynomial Prediction

The basic idea of iterative polynomial prediction is based on replacing the missing values with their corresponding predicted values from fitting polynomial model plus random error. (Calculations are based on deleting the pairs that involve any missing). This procedure was used by [5] for incomplete longitudinal data. However, this idea has been modified for missing data in polynomial wavelet regression. The main steps of this procedure can be illustrated as follows:

1. Instead of deleting the pairs that involve any missing as in literature, it is suggested using the entire non-parametric bootstrap process illustrated in 4.2 to fill in the missing data.
2. Find the ordinary least squares fit of quadratic polynomial. As shown in [5], choose the  $2^{nd}$  order model since it is the lowest degree polynomial that admits non-zero coefficients.
3. Find the predicted values ( $\hat{y}_j^c; j = 1, 2, \dots, m$ ) according to the complete data set. Where:  $\hat{y}_i^c = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$  and then calculate the mean squared error ( $mse$ ).
4. Replace the missing values as  $y_j = \hat{y}_j^c + N(\mu = 0, \sigma^2 = mse)$ . Here  $mse$  refers to the mean squared error from the  $2^{nd}$  order polynomial model.
5. Iterate the last two steps for  $r$  times (say  $r = 100$ ) or until convergence. Finally construct the complete data  $Y = (y_1, y_2, \dots, y_n)$  where

$$y_i = \left( \sum_{j=1}^r y_{ij} \right) / r$$

## 6 Bootstrap-Iterative local Polynomial Prediction

This method is similar to the previous one in the sense that it is a two-stage imputation method. It starts by initial estimate for missing data from the above non-parametric bootstrap method. This initial estimate will be iterated through a local polynomial regression model until convergence. The final estimate of each missing value will be the estimate at last iteration plus random error from normal distribution. Details are below:

1. At first, instead of losing some information by deleting the pairs involved any missing (as in literature); we suggest using non-parametric bootstrap method illustrated in section 4.2 to fill in the missing data.

2. Once the response variable being completely filled, Find the predicted values,  $\hat{f}$  using local polynomial regression and restore the values which were missing to be replaced with their current corresponded predicted values.

$$\hat{f}(x; p; h) = e_1'(X'_{p,x} W_x X_{p,x})^{-1} (X'_{p,x} W_x X_{p,x})$$

Here  $p$  refers to the order of local polynomial regression while  $h$  is smoothing parameter.  $W_x$  is diagonal matrix of weights with:  $W_i \equiv K_h(x_j - x) \equiv \{K[(x_j - x)/h]\}$ ,  $e_1$  is the  $(p + 1) \times 1$  vector having 1 in the first entry and zeros elsewhere.  $K$  is a kernel function. See [17] for details.

3. For  $j = 1, 2, \dots$  ; iterate the previous step till convergence with respect to mean squared error.
4. Upon convergence, refresh the missing value estimates, that is to be replaced with their current corresponded predicted values( at last iteration); plus random error from normal distribution with zero mean and variance equal to  $mse^{(j)}$ . Here  $mse^{(j)}$  denotes mean squared error from local polynomial regression at last iteration.

## 7 Simulation

In this section, we have used the codes of the R statistical package to carry out a simulation study to assess the achievement of the proposed method in addition to some existing ones:

1. Mean substitution.
2. Median substitution.
3. Non parametric bootstrap imputation.
4. EM Imputation.
5. Bootstrap-iterative polynomial prediction.
6. Bootstrap-iterative local polynomial prediction.

Throughout the whole simulation mother wavelet  $N = 10$  was used in every wavelet transform with soft thresholding rule. EbayesThresh as in [7] was used. The median absolute deviation of the wavelet coefficients at the finest level was used to find variances for EbayesThresh. Laplace density and the median of posterior were used for EbayesThresh. For bootstrap local polynomial imputation, we used: the second order local polynomial fitting with standard

Gaussian density function as a kernel function. The smoothing parameter was selected by direct plug-in methodology as described in [16]. Altogether four test functions were used. They are listed in Table 1 and are depicted in Figure 1. Each function has some abrupt changing features such as discontinuities or sharp bumps. It is reasonable to assume that functions 1, 2 and 4 denote non-periodic case while Function 3 as a periodic. We also used three kinds of noise structure (normal, noise from t-distribution with three degrees of freedom and mixture normal). We test four different missing percentages: 15%, 25%, 35%, and 45%. All missing locations are randomly selected. Two levels of signal to noise ratio ( $snr$ ) were used:  $snr = 5$  and 10. We consider two different sample sizes:  $n = 512$  and 128. For every combination of test function, noise structure,  $snr$ , sample size and missing data percentage, 100 samples (with missing at random) were generated. For each generated data set, we applied the above six imputation methods to fill in the missing data. Then, we applied LPWR to obtain an estimate( $\hat{f}$ ) for each test function. The mean squared error ( $mse$ ) was computed as a numerical measure for assessing the quality of  $\hat{f}$ .

$$mse(\hat{f}) = \frac{1}{n} \sum_{i=1}^n [f(i/n) - \hat{f}(i/n)]^2$$

Figures 2-13 display the box plot of the values of  $mse(\hat{f})$  for all test functions listed in Table 1 for ( $n = 128$ ,  $snr = 5$ ). Other results are similar and hence, omitted. Having examined the results; In terms of the mean squared error criterion, the following empirical remarks were observed:

1. In general, for all test functions our proposed method (Bootstrap iterative local polynomial imputation) performs much better than the other five methods regardless of noise structure, periodic assumption or the missing data percentage.
2. The mean and median imputation methods and non-parametric bootstrap seem to give almost similar results except for test function 2 where the median imputation shows the largest value for the mean squared error.
3. The EM imputation behaves badly for test functions 3-4. In contrast, it seems to be quite competitive with mean, median and bootstrap for Test function 2 and competitive with bootstrap-iterative polynomial method for test function 1. This clearly can be seen as the missing data percentage increases.



## 8 Conclusion

In this paper the problem of missing at random is considered for local polynomial wavelet regression. A new method called (Bootstrap-iterative local polynomial imputation.) has been introduced. The practical performance is evaluated through a simulation experiment. Results revealed that the proposed method seems to be more suitable to fill in the missing data for local polynomial wavelet regression than: mean-median substitutions, non-parametric bootstrap and EM algorithm and bootstrap iterative polynomial.

Table 1: The four test functions used in simulation

Test Function	Formula
1	fg1 of Donoho and Johnstone [3]
2	$4x(1 - \sin x)$ $x \in [0, 0.88] \cup [0.93, 1]$ $1.10$ $x \in [0.88, 0.93]$
3	Dopper of Donoho and Johnstone [3]
4	Piecewise polynomial function [11]

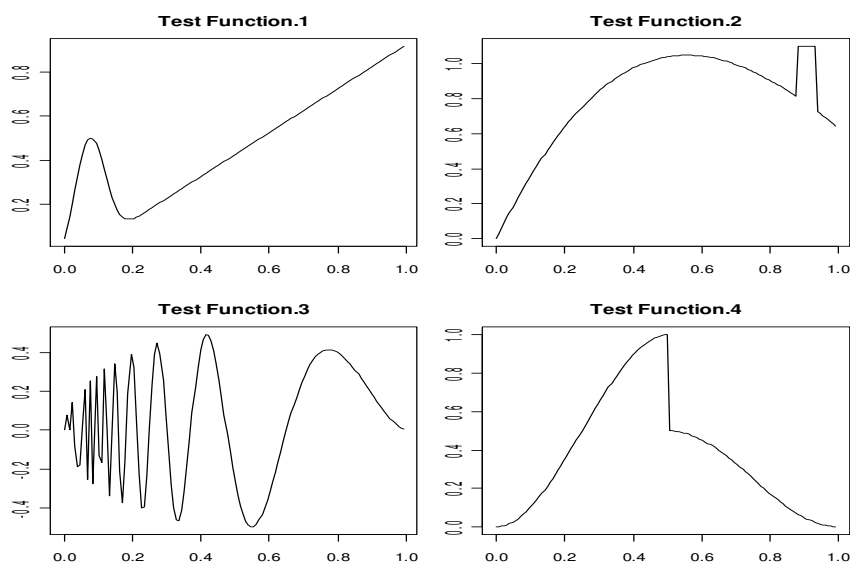


Figure 1: The four test functions used in simulation.

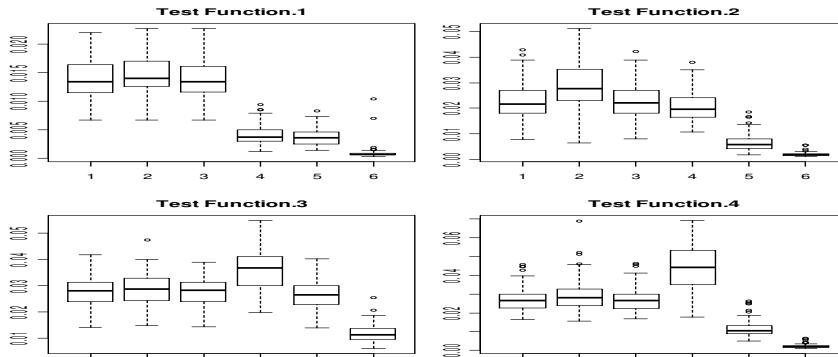


Figure 2: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 15% missing, Gaussian noise.

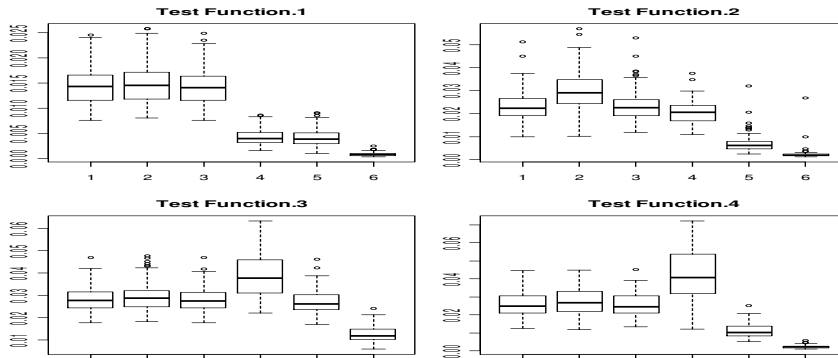


Figure 3: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 25% missing, Gaussian noise.

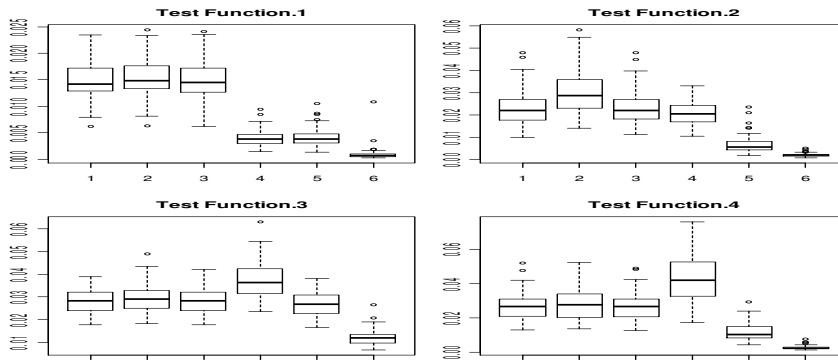


Figure 4: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 35% missing, Gaussian noise.

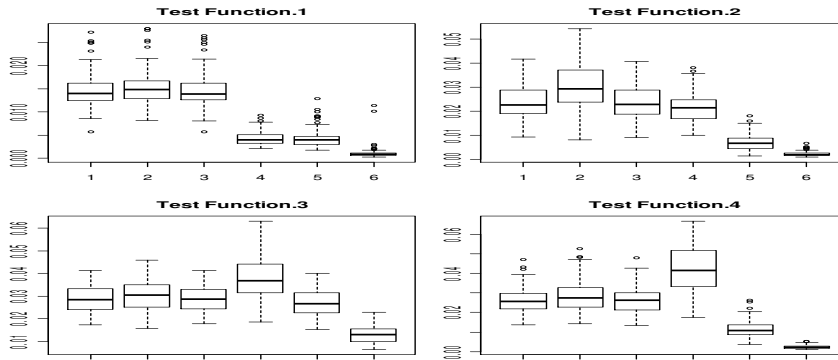


Figure 5: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 45% missing, Gaussian noise.

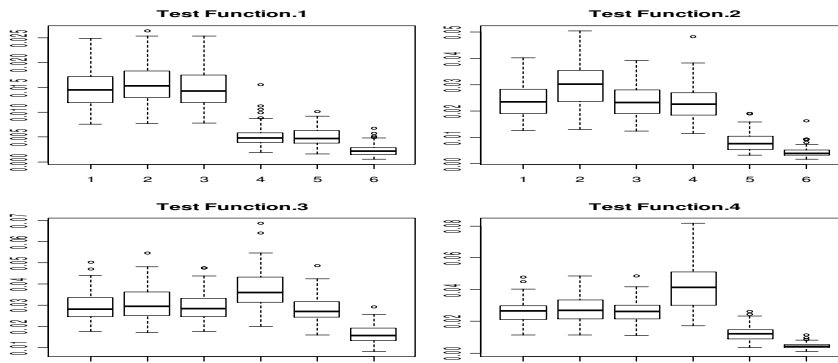


Figure 6: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 15% missing, mixture Gaussian noise.

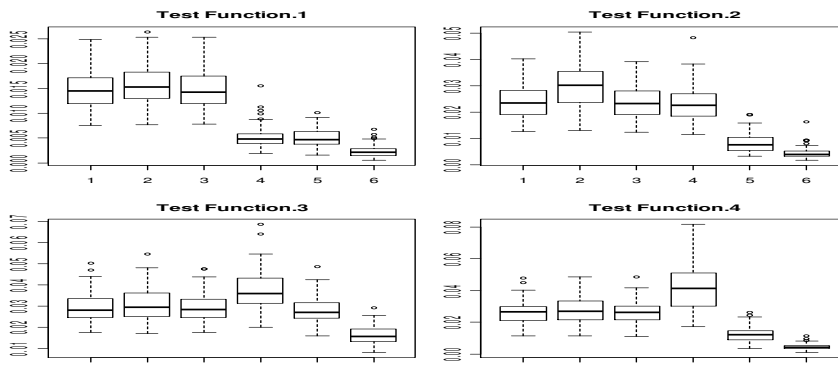


Figure 7: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 25% missing, mixture Gaussian noise.

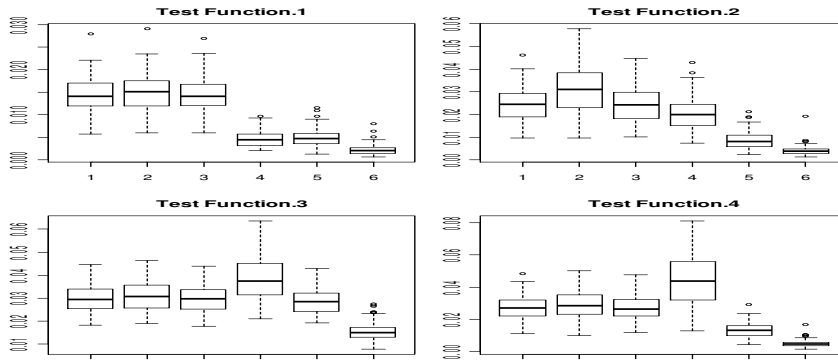


Figure 8: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 35% missing, mixture Gaussian noise.

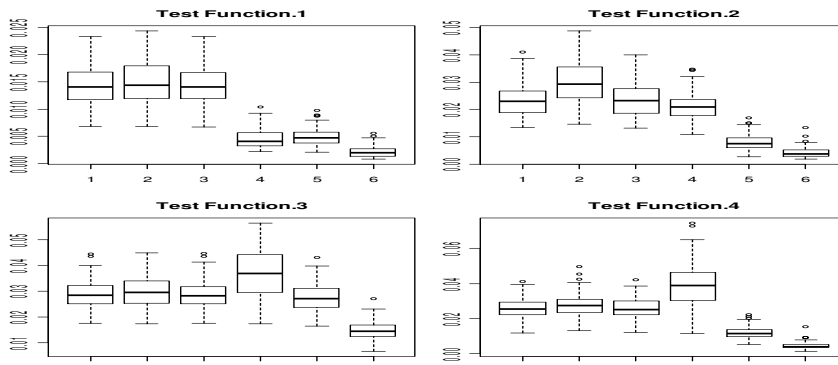


Figure 9: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 45% missing, mixture Gaussian noise.

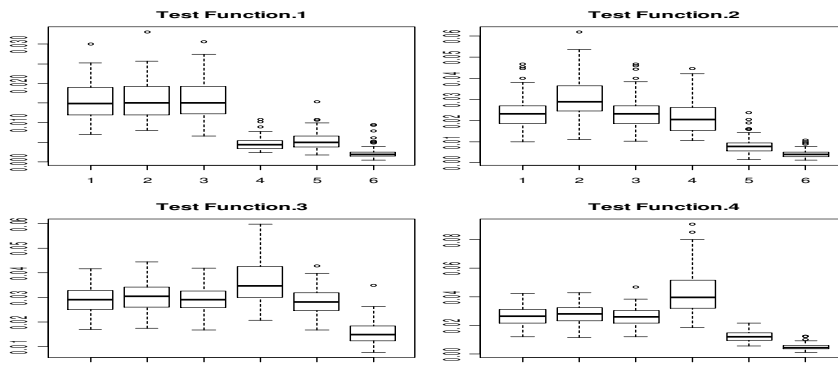


Figure 10: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 15% missing, noise from  $t(3)$ .

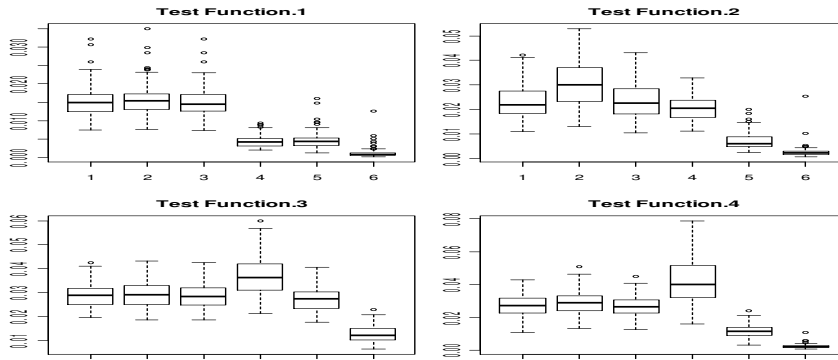


Figure 11: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 25% missing, noise from  $t(3)$ .

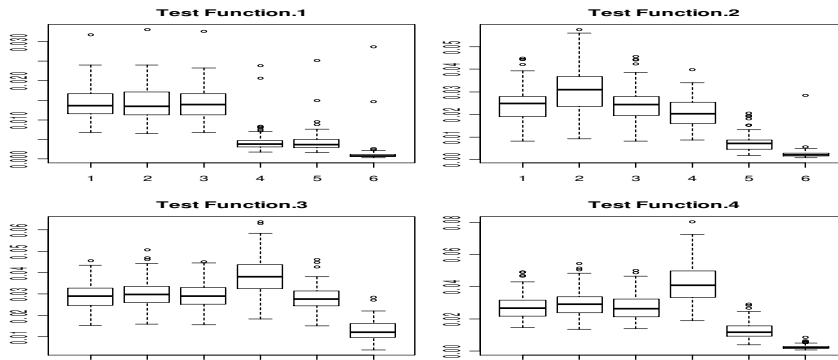


Figure 12: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 35% missing, noise from  $t(3)$ .

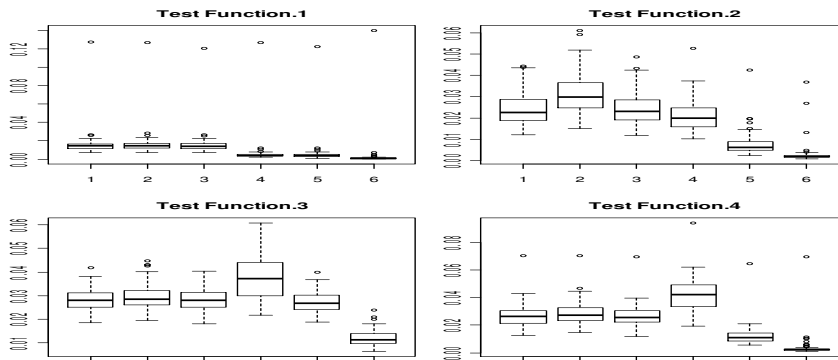


Figure 13: Box plot of  $mse(\hat{f})$  from simulation study with  $snr = 5$ ,  $n = 128$ , 45% missing, noise from  $t(3)$ .

## References

- [1] G. Boente, W. Gonzalez-Manteiga, and A. Perez-Gonzalez, *Robust non-parametric estimation with missing data*, Journal of Statistical Planning and Inference **139** (2009), no. 2, 571–592.
- [2] A. Chatterjee, *Detection of change points using wavelet analysis*, Ph.D. thesis, Department of Statistics, University of Wyoming, 2009.
- [3] D.L. Donoho and I.M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, Journal of the american statistical association (1995), 1200–1224.
- [4] D.L. Donoho and J.M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81** (1994), no. 3, 425–455.
- [5] P.J. Foster, A.M. Mami, and A.M. Bala, *On treatment of the multivariate missing data*, Probability and Statistics Group 13, School of Mathematics, The University of Manchester, 2009.
- [6] T.J. Hastie and R.J. Tibshirani, *Generalized additive models*, Chapman & Hall/CRC, 1990.
- [7] I.M. Johnstone and B.W. Silverman, *Empirical bayes selection of wavelet thresholds*, Annals of Statistics (2005), 1700–1752.
- [8] D. Kim, Y. Lee, and H.S. Oh, *Hierarchical-likelihood-based wavelet method for denoising signals with missing data*, Signal Processing Letters, IEEE **13** (2006), no. 6, 361–364.
- [9] G.J. McLachlan and T. Krishnan, *The em algorithm and extensions*, vol. 382, LibreDigital, 2008.
- [10] M. Mojirsheibani, *Nonparametric curve estimation with missing data: A general empirical process approach*, Journal of statistical planning and inference **137** (2007), no. 9, 2733–2758.
- [11] G.P. Nason and B.W. Silverman, *The discrete wavelet transform in  $s$* , Journal of Computational and Graphical Statistics (1994), 163–191.
- [12] T. Nittner, *Missing at random (mar) in nonparametric regression—a simulation experiment*, Statistical Methods and Applications **12** (2003), no. 2, 195–210.
- [13] H.S. Oh and T.C.M. Lee, *Hybrid local polynomial wavelet shrinkage: wavelet regression with automatic boundary adjustment*, Computational Statistics and Data Analysis **48** (2005), no. 4, 809–820.

- [14] H.S. Oh, P. Naveau, and G. Lee, *Polynomial boundary treatment for wavelet regression*, *Biometrika* **88** (2001), no. 1, 291–298.
- [15] D. B. Rubin and R. J. A. Little, *Statistical analysis with missing data*, J Wiley & Sons, 2002.
- [16] D. Ruppert, S.J. Sheather, and M.P. Wand, *An effective bandwidth selector for local least squares regression*, *Journal of the American Statistical Association* (1995), 1257–1270.
- [17] M.P. Wand and M.C. Jones, *Kernel smoothing*, vol. 60, Chapman & Hall/CRC, 1995.
- [18] Y. Wang, *Jump and sharp cusp detection by wavelets*, *Biometrika* **82** (1995), no. 2, 385–397.

**Received: September, 2011**