Robust estimation of the Hurst parameter and selection of an onset scaling

Juhyun Park and Cheolwoo Park

Lancaster University and University of Georgia

Abstract: We consider the problem of estimating the Hurst parameter for long-range dependent processes using wavelets. Wavelet techniques have shown to effectively exploit the asymptotic linear relationship that forms the basis of constructing an estimator. However, it has been noticed that the commonly adopted standard wavelet estimator is vulnerable to various non-stationary phenomena that increasingly occur in practice and thus leads to unreliable results. In this paper, we propose a new wavelet method for estimating the Hurst parameter that is robust to non-stationarities such as peaks, valleys, and trends. We point out that the new estimator arises as a simple alternative to the standard estimator and does not require an additional correction term, which is subject to distributional assumptions. Additionally, we address the issue of selecting scales for the wavelet estimator, which is critical to properly exploit the asymptotic relationship. We propose a new method based on standard regression diagnostic tools, which is easy to implement and useful to provide informative goodness-of-fit measures. Several simulated examples are used for illustration and comparison. The proposed method is also applied to the estimation of the Hurst parameter of Internet traffic packet counts data.

Key words and phrases: Hurst parameter, Long-range dependence, Non-stationarities, Robustness, Wavelet spectrum.

1. Introduction

The Internet has brought major changes to the work places, and even the lifestyles, of many people. It also provides a rich source for research problems at several levels of interest to engineers, computer scientists, statisticians and probabilists. The Internet is often compared to the telephone network since there are interesting parallels between the two. Both are gigantic networks transporting large amounts of information between very diverse locations. Both are a concate-

nation of many pieces of equipment. But there are some important differences, which seriously affects traffic modeling.

An important statistical difference between the telephone network and the Internet comes in the distribution of the length of connections. The exponential distribution has provided a useful model for the telephone network. But it has been shown in a number of places, see e.g. Paxon and Floyd (1995), Crovella and Bestavros (1996) and Hernández-Campos et al. (2004), that the exponential distribution is very inappropriate for durations of Internet connections since some of them are very short (in milliseconds) and some are very long (in hours). Models for aggregated traffic are quite different from standard queueing theory when the distribution of the lengths are heavy tailed. Appropriate levels of heavy tails can induce long-range dependence as shown by the above authors.

Because of the widely accepted long-range dependent self-similar properties of network traffic, Hurst parameter estimation provides a natural approach to studying such models. Many approaches for estimating the Hurst parameter have been proposed including the aggregated variance (Beran, 1994), the local Whittle (Robinson, 1995), and the wavelet (Abry and Veitch, 1998) methods. Among various approaches, the wavelet method has attracted the interest owing to its robustness to non-stationarity and decorrelation property. Park et al. (2007b) thoroughly compared the three Hurst parameter estimators by using simulated, synthetic and real Internet traffic data sets. It reveals a number of important challenges which one faces when estimating the long-range dependence parameter in Internet data traffic traces. Stoev et al. (2005) explored some of these challenges in more detail by using the wavelet spectrum method. While the wavelet method is reliable in practice and quite robust with respect to smooth polynomial trends in the data, it can mislead the practitioner. For example, a traffic trace with a number of deterministic shifts in the mean rate results in a steep wavelet spectrum which leads to overestimating the Hurst parameter. We will come back to this issue in Section 4.

As an illustration we introduce a time series of packet counts (the numbers of packets arriving in consecutive 1 millisecond intervals) coming into the University of North Carolina, Chapel Hill (UNC) from outside. Figure 1 displays a packet count time series measured at the main internet link of UNC on April 13,

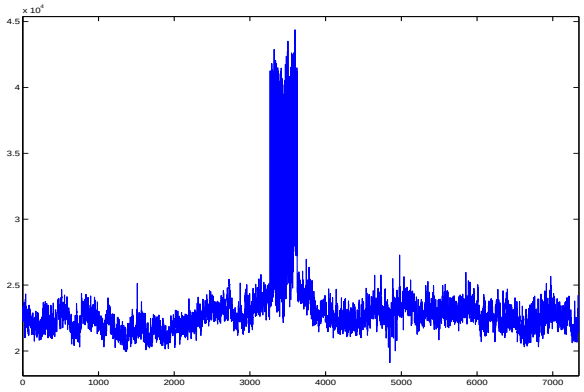
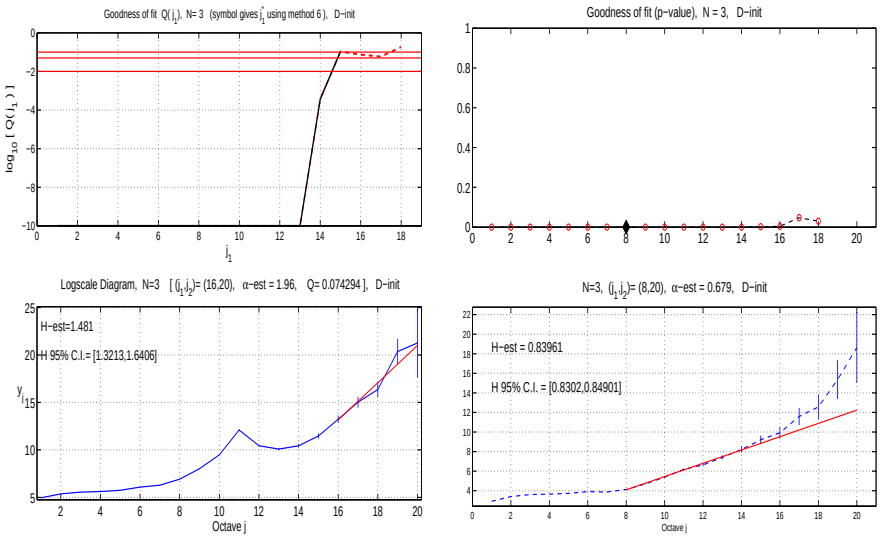


Figure 1: Sat1300: packet count time series of aggregated traffic at 1 second.

Saturday, from 1 p.m. to 3 p.m., 2002 (Sat1300). They were originally measured every 1 millisecond (7.2 million data points) but aggregated by a factor of 1000 (that is at 1 second) for a better display of trends. The time series plot shows a huge spike for about 6 minutes in the middle.

Figure 2 (a) shows the wavelet spectrum and the estimated Hurst parameter of the Sat1300 time series using the Abry and Veitch's wavelet method. The detail of the method is given in Section 2.2.1. Briefly the bottom panel plots the $\log_2$ of the (estimated) variance of the wavelet coefficients at a scale (or octave) value against $j = \log_2(\text{scale})$ (blue solid line). For processes that are long-range dependent, the wavelet spectrum will exhibit a region in which there is an approximately linear relationship with positive slope at the right (coarser scale) side. One can estimate the Hurst parameter, H, along with confidence intervals on the estimate by applying a weighted least squares (H=(slope+1)/2) to a particular range of scales chosen from the top panel. In this case, the chosen range is $16 \leq j \leq 20$. The spectrum roughly forms a linear line, which exhibits long-range dependence. However, $\hat{H} = 1.48$ (the estimated slope is overlaid), which cannot happen in theory for a stationary process and it suggests that the time series contains a non-stationary segment(s). Note that there is a bump at $j = 11$. Park et al. (2004) verified that the high-frequency behavior inside the big spike shown in Figure 1 causes this bump, which is a reflection of scaling behavior. We will revisit this issue in Section 4.



(a) Abry and Veitch method (b) Proposed method

Figure 2: Wavelet spectra and the Hurst parameter estimates by (a) Abry and Veitch and (b) the proposed methods

As Stoev et al. (2005) pointed out, the wavelet spectrum can serve as a diagnostic tool in this case since the unusual shape of the spectrum reveals the local non-stationary behavior in the original time series. If the segment of the time series where the spike occurs is taken out and the remaining parts are concatenated, then $\hat{H}$ is around 0.84, which is consistent with the Hurst parameter estimates obtained from other UNC data sets. This implies that the time series can be decomposed into a stationary long-range dependent process with $H = 0.84$ plus a local non-stationary behavior. The Hurst exponent of interest in this case is $H = 0.84$ but the 6-minute long spike dramatically changes the global Hurst parameter. While the bump is an indication of a non-stationary behavior, it affects the method to select the range of the scale j differently, which causes the estimation of H unreliable. In other words, the selected range of the scale is narrow ($16 \leq j \leq 20$) due to the bump, which makes $\hat{H}$ higher and its confidence interval wider.

This motivates us to develop a robust Hurst parameter estimation method that resists the effect of non-stationary behavior such as peaks, valleys, and

trends. Figure 2 (b) shows the proposed wavelet spectrum and the estimate of H of the Sat1300 times series. The spectrum shows no bump and $\hat{H} = 0.84$ which is consistent with the estimate when we exclude the non-stationary segment from the time series. In addition, the range of the scale chosen from the top panel is from $j = 8$ to 20, which makes the confidence interval narrower. Thus, it clearly shows that the proposed method is robust to the spike in the middle and produce the stable estimate of H.

We utilize a robust estimation for the finite variance case. We provide a brief justification similar to Veitch and Abry (1999), showing that the robust regression model arises as a natural alternative to the standard regression model. The same regression model has been studied independently for the infinite variance case. See Stoev et al. (2002), Stoev and Taqqu (2003), and Stoev and Taqqu (2005). Therefore, it can be argued that the idea developed under the finite variance assumption naturally extends to the infinite variance case and that the method is not limited to finite assumption in the end.

We also extend the idea of the estimation to the problem of selection of an onset scaling by formulating it as a model selection problem. Since the linear relationship in a wavelet spectrum is asymptotic in nature, the restriction of scales into proper subsets would result in a *better* estimate. The practical implication is that one needs to detect a scaling phenomenon for a given data. This involves the selection of the range based on the observation, where the asymptotic property can be reasonably assumed to be true. Veitch et al. (2003) addressed the issue by proposing a model selection based on a series of test statistics. With an aid of visualization of a goodness-of-fit measure, the onset scale can be selected automatically or interactively. However, for examples with non-standard processes such as the Sat1300 time series, this goodness-of-fit measure tends to show instability. Moreover, the measure is meaningful only for selection purposes and the number itself is not interpretable (for example, refer to the Q statistic (0.07) in Figure 2 (a)).

We reformulate the problem in hypothesis testing framework and propose an improved goodness-of-fit measure using p-values, which are easy to understand and which reflect the underlying behavior informatively.

The remainder of the paper is structured as follows. In Section 2, we define

our robust wavelet estimator and make a comparison to the standard wavelet estimator. The issue of selection of scales is discussed in Section 3. Some simulations studies are given in Section 4, followed by real data examples in Section 5. We conclude in Section 6.

2. Hurst parameter estimation

This section introduces the proposed robust Hurst parameter estimator in Section 2.1 and compares it with the Abry and Veitch's estimator in Section 2.2.

2.1 Robust wavelet estimation

We consider an estimator constructed through the discrete wavelet transform. Let $\psi(t)$ be a square integrable function with $M \in \mathcal{Z}$ zero moments, $M \geq 1$, so that

$$\int_{\mathcal{R}} t^m \psi(t)\, dt = 0\,, \qquad \text{for all } m = 0, \ldots, M-1\,. \tag{2.1}$$

Consider a family of functions

$$\{\psi_{j,k} = 2^{-j/2}\psi(2^{-j}t - k), j, k \in \mathcal{Z}\}$$

obtained by dyadic dilations and translations of ψ, which forms a basis of multiresolution analysis. For a second order stationary stochastic process $X = \{X(t)\}$, the discrete wavelet transform is defined as

$$D(j,k) = \int_{\mathcal{R}} X(t)\psi_{j,k}(t)\, dt\,, \qquad j, k \in \mathcal{Z}\,.$$

Suppose that $\{X(t), t \in \mathcal{R}\}$ is a self-similar process, with self-similarity parameter H. Then for fixed $j \in \mathcal{Z}$,

$$D(j,k) \stackrel{d}{=} 2^{j(H+1/2)} D(0,k)$$

as a process in $k \in \mathcal{Z}$ (Abry et al., 2003). Thus, we have

$$\begin{aligned} E[\log_2 D(j,k)^2] &= E[\log_2 (2^{j(H+1/2)} D(0,k))^2] \\ &= j(2H+1) + E[\log_2 D(0,k)^2]\,. \end{aligned}$$

This also suggests that the Hurst parameter H can be estimated by a linear regression model using a sample mean estimator for the left hand side against the scale parameter j.

Suppose that $\{D(j,k) : k = 1, \ldots, n_j\}, j = 1, \ldots, J$ are wavelet coefficients from the process with a length of 2^J. Here n_j are the number of wavelet coefficients at scale j. Define

$$Y_j = \frac{1}{n_j} \sum_{k=1}^{n_j} \log_2 D(j,k)^2.$$

Because n_j varies with j, it is natural to use a weighted least squares approach with weights proportional to n_j. An estimator of H can be constructed using a weighted linear regression as

$$\hat{H} = \frac{1}{2} \sum_j w_j Y_j - \frac{1}{2},$$

where $\sum_j w_j = 0$ and $\sum_j j w_j = 1$. Note that although the estimator is written in terms of second order statistic of $D(j,k)$, because of logarithm, it only requires the existence of $E[\log_2 |D(j,k)|]$.

To understand the behavior of the estimator, we may need more assumptions about the sequence $\{Y_j : j \in \mathcal{Z}\}$. For example if the sequence $\{D(j,k) : k \in \mathcal{Z}\}$ is stationary, we would have

$$Y_j \stackrel{a.s.}{\to} (2H+1)j + E[\log_2 D(0,k)^2],$$

as $j \to \infty$. Then, the estimator is consistent. A weaker assumption that warrants consistency is that the logarithm of the sequence, $\{\log_2 D(j,k)^2 : k \in \mathcal{Z}\}$, is stationary and this is where *robustness* stems from. See also Stoev et al. (2002).

Because of similarity in behavior of wavelet coefficients to long-range dependent processes, the same idea applies to long-range dependent processes such as Fractional Gaussian Noise (FGN) or Fractional Auto-Regressive Integrated Moving Average (FARIMA). For example, the cumulative sum processes of FGN recovers Fractional Brownian Motion (FBM) that satisfies self-similar property. We formulate the problem for self-similar processes because the argument is more transparent with self-similar processes in several aspects. When the process has an infinite variance, the same idea of self-similarity can be easily extended, as in Stoev et al. (2002), Stoev and Taqqu (2003), and Stoev and Taqqu (2005).

In the next section, we go back to second order stationary long-range dependent processes and make some comparison to the standard wavelet estimator. In

particular, the robust estimator can be seen as a simplified alternative derivation that justified the standard estimator.

2.2 Comparison under long-range dependent processes

2.2.1 Standard regression model

We briefly review the standard wavelet estimator defined in Abry and Veitch (1998), and Veitch and Abry (1999) under second order stationary long-range dependent processes. We are mainly interested in the relationship between the robust estimator and the standard estimator arising from linear regression models.

For a long-range dependent process $X(t)$, it has been shown that, as $j \to \infty$,

$$\mathrm{E}[D^2(j,\cdot)] \sim 2^{j\gamma} C\,, \quad 0 < \gamma < 1\,,$$

where C is the constant defined in Veitch and Abry (1999). The last relationship suggests that the long-range dependent parameter γ (or $H = (\gamma+1)/2$) can be estimated from

$$\log_2\left(\mathrm{E}[D^2(j,\cdot)]\right) = j\gamma + \text{constant}\,, \qquad \text{as } j \to \infty.$$

This linear relationship popularizes wavelet-based techniques for estimating γ (or H). The idea is to replace the expected value $\mathrm{E}[D^2(j,\cdot)]$ by the corresponding sample quantity calculated at each scale j,

$$\mu_j = \frac{1}{n_j}\sum_{k=1}^{n_j} D(j,k)^2,$$

where n_j is the number of wavelet coefficients at scale j.

Veitch and Abry (1999) provided distributional justification for a linear regression approach under an ideal situation, by noting that under the above setting the $D(j,k)$ is a collection of zero mean random variables which are quasi-decorrelated. Hence, if we assume that $D(j,k)$s are independent and identically distributed Gaussian variables and that $D(j,\cdot)$ and $D(j',\cdot)$ are independent when $j \neq j'$, then

$$\mu_j \overset{d}{\sim} \frac{\sigma_j^2}{n_j}\chi^2(n_j)$$

where $\sigma_j^2 = 2^{j\gamma}C$ and $\chi^2(\nu)$ is a chi-square random variable with ν degrees of freedom. It follows that

$$\begin{aligned} \log_2(\mu_j) &\overset{d}{\sim} \log_2 \sigma_j^2 - \log_2(n_j) + \log_2 \chi^2(n_j) \\ &\overset{d}{\sim} j\gamma + \log_2(C) - \log_2(n_j) + \ln \chi^2(n_j)/\ln 2\,. \end{aligned} \tag{2.2}$$

From the following result,

$$\mathrm{E}[\ln \chi^2(\nu)] = \psi(\nu/2) + \ln 2\,, \qquad \mathrm{Var}[\ln \chi^2(\nu)] = \zeta(2, \nu/2)\,, \tag{2.3}$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the Psi function and $\zeta(z, \nu)$ is a generalized Riemann Zeta function (Gradshteyn and Ryzhik, 2000), it follows that

$$\mathrm{E}[\log_2(\mu_j)] = j\gamma + \log_2(C) + g_j\,, \qquad \mathrm{Var}[\log_2(\mu_j)] = \zeta(2, n_j/2)/(\ln 2)^2\,,$$

where

$$g_j = \psi(n_j/2)/\ln 2 - \log_2(n_j/2)\,. \tag{2.4}$$

Let $\tilde{Y}_j \equiv \log_2(\mu_j) - g_j$. Here g_j is a bias correction factor that compensates for the difference between $\mathrm{E}[\log_2(\mu_j)]$ and $\log_2(\mathrm{E}[d^2(j,\cdot)])$, to make $\tilde{Y}_j$ an asymptotically unbiased estimator of $\log_2(\mathrm{E}[d^2(j,\cdot)])$. The parameter is then estimated by applying a weighted least squares method based on the model

$$\tilde{Y}_j = j\gamma + \text{constant} + \tilde{\varepsilon}_j\,,$$

where $\tilde{\varepsilon}_j$ has mean 0 and variance $\zeta(2, n_j/2)/(\ln 2)^2$. Consequently, the Hurst parameter H can be obtained from the relationship $\gamma = 2H - 1$.

2.2.2 Robust regression model

As a motivation for the robust estimation, we begin with the same assumptions as above. Instead of directly focusing on the estimator μ_j, we may treat each individual coefficient $D(j,k)$ equally as a possible response. Then, from (2.2) with $n_j = 1$, we have

$$\begin{aligned} \log_2 D(j,k)^2 &\overset{d}{\sim} \log_2 \sigma_j^2 + \log_2 \chi^2(1) \\ &\overset{d}{\sim} j\gamma + \log_2(C) + \ln \chi^2(1)/\ln 2\,. \end{aligned} \tag{2.5}$$

Let $Y_{j,k} = \log_2 D(j,k)^2$. Then it can be shown from (2.3) that

$$E[Y_{j,k}] = j\gamma + \gamma_0\,, \qquad Var[Y_{j,k}] = \sigma^2\,,$$

where $\gamma_0 = \log_2 C + \psi(1/2)/\ln 2 + 1$ and $\sigma^2 = \zeta(2, 1/2)/(\ln 2)^2$. This leads to a simple linear regression model with constant variance σ^2. The least squares estimates are defined by

$$argmin_{\gamma,\gamma_0} \sum_{j=1}^{J} \sum_{k=1}^{n_j} (Y_{j,k} - j\gamma - \gamma_0)^2 .$$

It is easy to check that this approach is equivalent to weighted least squares criterion as

$$argmin_{\gamma,\gamma_0} \sum_{j=1}^{J} n_j (\bar{Y}_{j\cdot} - j\gamma - \gamma_0)^2 ,$$

where $\bar{Y}_{j\cdot} = \frac{1}{n_j} \sum_{k=1}^{n_j} Y_{j,k}$. We will replace $\bar{Y}_{j\cdot}$ with Y_j from now on, so $Y_j = \frac{1}{n_j} \sum_{k=1}^{n_j} \log_2 \left(D(j,k)^2\right)$. Therefore, an equivalent formulation can be made as

$$Y_j = j\gamma + \gamma_0 + \varepsilon_j ,$$

where ε_j has mean 0 and variance σ^2/n_j, for which the weighted least squares method is used. Soltani et al. (2004) proposed $Y_j^* = (Y_j + Y_{n_j/2})/2$, which is shown to follow a Gumbel distribution for FBM processes but still suggested to use least squares approach for practical consideration.

While the Gaussian assumption of $D(j,k)$ does not guarantee a Gaussian distribution for the error term, the least squares approach in general is not sensitive to distributional assumption and the standard estimator is shown to be asymptotically unbiased and efficient. For more detailed analysis with correlated errors in the standard wavelet estimator, see Bardet et al. (2000). Some discussions on the comparison of these two regression models are given in Section 3.5.1.

Both estimators fall in a general class of linear estimators in linear regression models and thus statistical properties are similar. Below we summarize a well known property of least squares estimators as a reference.

Proposition 1 (Example 1, p.27, Ferguson, 1996) *Suppose that*

$$Y_j = \alpha + \beta z_j + \epsilon_j \qquad j = 1, \ldots,$$

where z_j's are known numbers that are not all equal and the ϵ_j's are i.i.d. random variables all with mean zero and share a common variance σ^2. Then, the least squares estimate, $\hat{\beta}_n$ is consistent provided that as $n \to \infty$,

(a) $\sum_{j=1}^{n}(z_j - \bar{z}_n)^2 \to \infty$

(b) $\max_{j\le n}(z_j - \bar{z}_n)^2 / \sum_{j=1}^{n}(z_j - \bar{z}_n)^2 \to 0$.

Moreover, $\hat{\beta}_n$ *is asymptotically normal with*

$$\sqrt{n}s_n(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \sigma^2)\,,$$

where $s_n^2 = \sum_{j=1}^{n}(z_j - \bar{z}_n)^2/n$.

3. Model selection

When these estimators are computed with a finite number of observations N, there is a more delicate issue than justification of distributional assumptions. Because of its asymptotic nature in approximation, performance of the estimator is heavily dependent on the choice of regime where the relationship is reasonably justifiable. In theory, Bardet et al. (2000) showed that the standard wavelet estimator $\hat{H}_{[}j_1, j_2]$ is consistent, as j_1 and $N/2^{j_2} \to \infty$, where $\hat{H}_{[}j_1, j_2]$ means that the estimator is constructed based on selected scales of $j_1, \ldots, j_2$. In practice, it has been observed that the choice of the onset parameter has a stronger influence on the estimation than on the distributional assumption (Abry et al., 2003). Although this issue has been rightly acknowledged, there are few discussions in the current practice beyond heuristically trimming scales at both ends. One exception is the work of Veitch et al. (2003), where they propose an automatic procedure based on sequential testing, assuming second order long-range dependent processes. It was motivated that the exact value of $\log_2 E[D(j,\cdot)]$ can be computed or well approximated by a sample statistic, closely related to the fact that the standard wavelet estimator is constructed based on the same quantity. However, when the robust estimator is used, it is not clear whether the same argument would apply or is necessary.

We develop a general approach borrowed from ideas of regression diagnostics. A usual aim of regression diagnostics is to examine deviations from assumed linear models through outliers and influential points. Improvements in estimation are made when those points are removed or downweighted. A similar story can be told with wavelet estimators. We want to exclude scales that pull them away from linearity and the magnitude of influence can be measured by various diagnostic

measures. Alternatively, we are expecting linearity to start to appear at a certain scale, which means there is a *change point*. Again the phenomena will be reflected in the estimation and some type of diagnostic measures will pick them up. There is huge literature on those topics, depending on possible scenarios. Our aim is to draw attention to the relevance of these topics and provide some simple yet useful strategies that can be easily adapted to the selection of onset scaling. For further references, a summary of regression diagnostics can be found in Belsley et al. (1980) and more development on change point analysis is given in Csörgő and Horváth (1997), Chen and Gupta (2000) and Wu (2005).

In view of asymptotics, we fix $j_2 = J$, say, the largest possible value and focus on the selection of j_1. This is not a serious restriction as the proposed method below can easily be extended to search both ends.

Consider a regression model given by

$$\tilde{Y}_j = \beta_0 + \beta_1 \tilde{x}_j + \tilde{\varepsilon}_j\,, \qquad j = 1, \cdots, J$$

where $\mathrm{Var}(\tilde{\varepsilon}_j) = \sigma^2/n_j$. This can be written as

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}\,, \tag{3.1}$$

where $\tilde{X}$ is an $J \times p$ matrix. Let $W = \mathrm{diag}\{w_1, \cdots, w_J\}$, where $w_j = n_j$ and define

$$\mathbf{y} = W^{1/2}\tilde{\mathbf{y}}\,, \quad X = W^{1/2}\tilde{X}\,, \quad \boldsymbol{\varepsilon} = W^{1/2}\tilde{\boldsymbol{\varepsilon}}\,.$$

Then

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.2}$$

where $\boldsymbol{\varepsilon} \sim (0, \sigma^2 I)$. Hence, the weighted least squares estimates from model (3.1) is equivalent to the ordinary least squares estimates from model (3.2). From now on, our formulation will be given based on model (3.2).

Write $\mathbf{x}_j$ to be the jth row vector of X. Let $\mathbf{b}$ be the estimate of $\boldsymbol{\beta}$ from the full model (3.2) and $\mathbf{b}(j)$ be the estimate from a reduced model with the jth row removed. Then

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y} \equiv H\mathbf{y}\,,$$

where H is the hat matrix with h_{ij} being the (i,j)th element of H. Denote the jth residual by e_j, which is given by

$$e_j = Y_j - \hat{Y}_j = Y_j - \mathbf{x}_j^T \mathbf{b}\,.$$

The relationship between estimates $\mathbf{b}$ and $\mathbf{b}(j)$ is summarized in Lemma 1.

Lemma 1

$$\mathbf{b} - \mathbf{b}(j) = \frac{(X^T X)^{-1}\mathbf{x}_j e_j}{1 - h_{jj}}$$

This quantity, along with many others, are used as regression diagnostic tools to check whether the jth observation is influential in the estimation. A similar idea can be applied to the selection of onset scaling. When the selection region is controlled by j_1, we are looking for a stable region where the estimates do not vary much. When j_1 moves one step ahead, the estimates on which the decision will be based change from $\mathbf{b}$ to $\mathbf{b}(1)$, as scales are fixed and ordered. If the difference is dramatic, we move forward. Lemma 1 shows how those sequential estimates are related and suggests an alternative goodness-of-fit measure.

A usual strategy in testing nested models is to compare relative improvement in the fit. For linear models, this is often measured by sum of squared residuals (SSR), defined by

$$SSR = \sum_i (Y_i - \hat{Y}_i)^2 .$$

Let s^2 be the usual estimate of σ^2, which is obtained by dividing SSR by an appropriate degrees of freedom;

$$s^2 = \frac{SSR}{n-p} = \frac{1}{n-p}\sum_{i=1}^{J}(Y_i - \mathbf{x}_i^T\mathbf{b})^2 .$$

Similarly $s^2(j)$ is defined by

$$s^2(j) = \frac{1}{n-p-1}\sum_{i\neq j}\{Y_i - \mathbf{x}_i^T\mathbf{b}(j)\}^2 .$$

The relationship between s^2 and $s^2(j)$ is given in Lemma 2.

Lemma 2

$$(n-p-1)s^2(j) = (n-p)s^2 - \frac{e_j^2}{1-h_{jj}}$$

Combined with the idea of Lemma 1, this forms the basis to construct a test statistic in Section 3.3.

For derivation and proofs we refer to Belsley et al. (1980).

3.1 Alternative formulation of the problem

The procedure can be viewed as sequential hypotheses testing problem. Suppose that

$$Y_j = \begin{cases} f(\mathbf{x}_j) + \varepsilon_j\,, & j < j_1^* \\ \mathbf{x}_j^T \boldsymbol{\beta}_0 + \varepsilon_j\,, & j \geq j_1^* \end{cases} \tag{3.3}$$

where $f(\mathbf{x}_j) \neq \mathbf{x}_j^T \boldsymbol{\beta}_0$ is unspecified.

$$H_0 : j_1^* = 1 \qquad H_1(j) : j_1^* = j, \quad j \geq 2\,. \tag{3.4}$$

Because of generality of the framework in (3.3), it is possible to come up with many different types of test statistic that would be considered appropriate. When there are several competing test statistics, it might be of interest to compare powers. Though related, our main interest is not so much of constructing a *best* test statistic that tells us that there occurs a change as estimating the change point directly through behavior of test statistics. A desirable property of test statistics is thus to reflect the change in noticeable way in the sequential comparison.

3.2 Selection of onset scaling by Veitch et al. (2003)

We review the main features of the approach presented in Veitch et al. (2003). For each j, fit the regression model with $\{(i, Y_i) : i = j, \ldots, J\}$ only. Define

$$T_0(j) = \sum_{i=j}^{J} (Y_i - \hat{Y}_i)^2 / \sigma^2\,.$$

Here $\hat{Y}_i = \hat{Y}_i(j)$, are estimates under the restricted model. Under the assumption that Y_js are Gaussian, the test statistic follows a chi-square distribution with degrees of freedom $N(j)-2$, where $N(j) = J-j+1$ is the number of observations included.

Veitch et al. (2003) proposed to search among candidate models by comparing the test statistic $\{T_0(j),\, j = 1, \ldots, J-2\}$. Let $p_0(j)$ be the p-value calculated at the observed value at $T_0(j)$. This has been used as an indication of change and a *best* model is defined as one that has the largest change in p-value. For a given j, under the null hypothesis, the test statistic $T_0(j)$ follows a chi-square

distribution with degrees of freedom $N(j) - 2$. They argue that the procedure can be viewed as a bias-variance tradeoff. The algorithm implemented in the paper searches for a point at which a dramatic increase in the p-value occurs as an indication of stabilization. In other words, the change point j_0^* is estimated by

$$\widehat{j_0^*} = arg\max_{j\geq 2} \frac{p_0(j)}{p_0(j-1)}\,.$$

The chi-square statistic was aimed at utilizing estimation of $\log_2(E[D(j,k)])$, which is possible to obtain for some well-known processes such as FARIMA and FGN processes. To extend the idea to unknown processes, we propose a general strategy of model selection using linear models.

3.3 F statistic for linearity

Another disadvantage of using the chi-square statistic in the sequential linear model is that it does not directly account for *linearity* in the comparison. We are mainly interested in the linear model with a significant slope. Therefore, for the selection of an onset scale, we can view this problem as selection of a submodel that shows the strongest linearity. For each submodel indexed by j, we compare

$$H_0(j) : EY = constant\,, \qquad H_1(j) : EY = linear\,. \tag{3.5}$$

By including all the linear models in the alternative, we don't presume that there is *any* linearity in the model. If any, it is more likely that the null hypothesis will be rejected, which would result in a smaller p-value. Contrary to the previous case, the decision of rejecting the null hypothesis as strong as possible will be desirable. Denote the sum of squared residuals under the null model at step j by $SSR_j(old)$ and that under the alternative model by $SSR_j(new)$. To test the significance of the nested models, we adopt a commonly used F test statistic, defined as

$$T_1(j) = \frac{(SSR_j(\text{old}) - SSR_j(\text{new}))/1}{SSR_j(\text{new})/(N(j)-2)} = \frac{\sum_{i=j}^{J}(\bar{Y} - \hat{Y}_i)^2}{\sum_{i=j}^{J}(Y_i - \hat{Y}_i)^2/(N(j)-2)}\,,$$

which follows a F distribution with degrees of freedom $(1, N(j)-2)$. We select one that has the smallest p-value. Let $p_1(j)$ be the p-value evaluated at the observed $T_1(j)$. Define

$$\widehat{j_1^*} = arg\min_{j\geq 1}\{p_1(j)\}\,.$$

Now p-values can be interpreted as a measure of how strong the linearity holds. This utilization of p-values, which conforms to common sense of interpretation, also allows to employ a direct search method for maximum in implementation. In addition, the magnitude of a p-value is closely related to goodness-of-fit and thus can be used as an indication of violation of linearity assumption. When all the p-values are relatively large, we suspect that there is no significant linear relationship. This is a noticeable feature because in practice presence of linearity itself may be in doubt, in which case the selected model would clearly indicate a large p-value.

Where the estimates are stabilized, p-values tend to be close to zero and the minimum is not more meaningful than the second minimum. Thus, the proposed principle can be relaxed to allow small fluctuation within the range by setting for a fixed $\alpha > 0$

$$\widehat{j_1^{**}} = \min\{j \geq 1 : p_1(j) < \alpha\}\,.$$

3.4 F statistic as diagnostics

We may view the selection of scales as regression diagnostics, where detecting outliers and influential points are of main interests. If j_1 has to move up one by one, that means the first observation may be considered as an outlier in the original regression and thus has to be removed. For submodels indexed by j, consider

$$H_0(j) : j_1^* = j \qquad H_1(j) : j_1^* = j + 1\,.$$

When this test is applied sequentially, we may expect that improvements made by deleting one row will be most dramatic when j crosses the true change point from $j_1^* - 1$ to j_1^*. Indeed, we show that a F test statistic can be constructed based on this idea and p-values can be used to detect the change point. With slightly different motivation, the statistic appears as part of the regression diagnostic methods developed in Belsley et al. (1980). We borrow their arguments to present here as Lemma and for derivation and proofs we refer to Belsley et al. (1980).

Lemma 3

$$T = \frac{[SSR(old) - SSR(new)]/1}{SSR(new)/(n-p-1)}$$
$$= \frac{(n-p)s^2 - (n-p-1)s^2(j)}{s^2(j)} = \frac{e_j^2}{s^2(j)(1-h_{jj})}$$

where new model is one without the jth row. If $\mathbf{y}$ *follows Gaussian distribution, then*

$$T \sim F(1, n-p-1).$$

Denote $SSR(j) = SSR_j(new)$, sum of squared residuals calculated with $(1, \ldots, j)$ rows removed. Define

$$T_2(j) = \frac{SSR(j-1) - SSR(j)}{SSR(j)/(N(j+1)-2)} \qquad j = 1, \ldots, J-2$$

and let $p_2(j)$ be the p-value evaluated at the observed value of $T_2(j)$. Define

$$\widehat{j_2^*} = arg \max_{j \geq 1} \left\{ \frac{p_2(j-1)}{p_2(j)} \right\}.$$

One can also choose the scaling set based on this criterion but we do not implement this approach in our data analyses in Sections 4 and 5.

3.5 Comparison of regression models and selection criteria

We observe, within our limited experiences as shown in Sections 4 and 5, dramatic improvements in performance of estimators with the new regression model and wonder where the *robustness* really comes from. However, when evaluating estimators, it is not easy to single out the source between regression models and model selection criteria. Here we separate the issues as an attempt to make some comparisons to existing methods.

3.5.1 Comparison of regression models

For regression models, one way of measuring robustness would be to consider the *influence function* of the estimator, which measures how sensitive the regression coefficients are to the outlier (Belsley et al., 1980 or McKean, 2004). For the standard regression model with $Var(\varepsilon_i) = \sigma^2$, replacing $Var(\varepsilon_i) = \sigma^2/w_i$ for

the specific ith observation only, differentiating with respect to w_i evaluated at $w_i = 1$ gives

$$\left.\frac{\partial b(w_i)}{\partial w_i}\right|_{w_i=1} = (X^T X)^{-1}\mathbf{x}_i^T e_i\,,$$

assuming the unified notation introduced earlier. Since the design matrices for both regression models are identical, one might suspect that the effect of outlier should be similar unless the variance of e_i or ε_i is dramatically different.

Consider the ideal case of Gaussian assumptions discussed in Section 2. Let $U_k, k = 1, \ldots, n_j$ be i.i.d χ^2 random variables with 1 degree of freedom. From (2.2) and (2.5), we may write

$$\tilde{\varepsilon}_j \overset{d}{=} \log_2\left(\frac{1}{n_j}\sum_{k=1}^{n_j} U_k\right), \quad \text{and} \quad \varepsilon_j \overset{d}{=} \frac{1}{n_j}\sum_{k=1}^{n_j}\log_2 U_k\,.$$

At first glance, taking the logarithm first seems to greatly reduce variability. This would be the case if variables take values mostly greater than 1 but for the χ^2 random variables U_k, with mean 1 and variance 2, the log-transformation can amplify variability for values between 0 and 1. Also, observe that

$$\tilde{\varepsilon}_j \overset{d}{=} \log_2\left(\Gamma(\frac{n_j}{2}, \frac{n_j}{2})\right),$$

with $E[\Gamma(\frac{n_j}{2}, \frac{n_j}{2})] = 1$ and $Var[\Gamma(\frac{n_j}{2}, \frac{n_j}{2})] = \frac{2}{n_j}$. Here $\Gamma(r, a)$ represents a Gamma random variable with a density $f_{r,a}(x) = \frac{a^r}{\Gamma(r)}x^{r-1}e^{-ax}$. This is also reflected in the variance. Recall that

$$Var[\tilde{\varepsilon}_j] = \frac{\zeta(2, n_j/2)}{(\ln 2)^2}\,, \quad \text{and } Var[\varepsilon_j] = \frac{\zeta(2, 1/2)}{n_j(\ln 2)^2}\,.$$

Veitch and Abry (1999) derived an asymptotic form as $\zeta(2, n_j/2) \sim 2/n_j$ for large n_j, which shows asymptotic equivalence in order of magnitude. Moreover, assuming $n_j = 2k, k \geq 1$, it can be shown that

$$\zeta\left(2, \frac{2k}{2}\right) = \zeta(2) - \left\{\frac{1}{1^2} + \ldots + \frac{1}{(k-1)^2}\right\} = \frac{\pi^2}{6} - \left\{\frac{1}{1^2} + \ldots + \frac{1}{(k-1)^2}\right\}$$

and

$$\frac{\zeta(2, 1/2)}{n_j} = \frac{3\zeta(2)}{n_j} = \frac{\pi^2}{4k}\,.$$

Thus, both variances converge to zero, as n_j grows, with no strict inequality on either direction and thus the impact of taking logarithm first is not as dramatic as it appears.

What makes the new model more appealing is that by taking the logarithm first it removes the need of correcting bias by subtracting g_j in (2.4), which is highly dependent on distributional assumptions. Therefore, for processes close to the ideal Gaussian processes, performance of both estimators should be similar. When the processes show departure from the assumption, the new estimator is expected to perform more effectively and thus *robustly*.

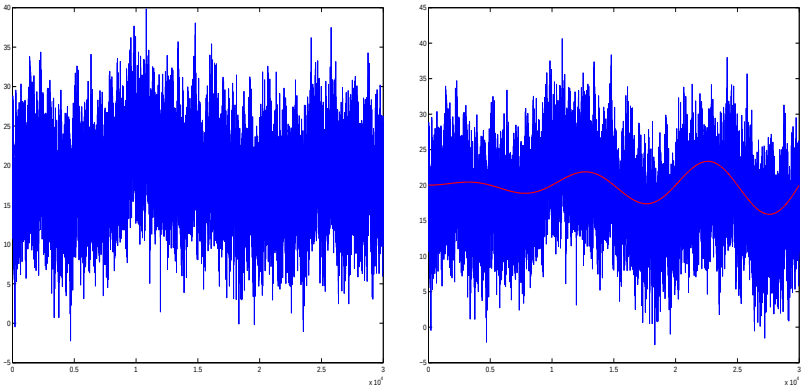
3.5.2 Comparison of model selection criteria

In general when the model selection criteria is concerned, the chi-square statistic appears as an estimate of the error variance, often termed as $\hat{\sigma}^2$. Although it is a best unbiased estimate of σ^2 for linear models, the statistic alone may not be adequate to detect a *true* model when the number of parameters or observations vary. Here we have a fixed number of parameters with varying sample size. Most model selection criteria such as AIC or BIC were introduced to take into account the varying size by an adding additional penalty term, controlling the number of parameters estimated against the number of observations used. In a slightly different context, BIC was used to select a *best* model in Shen, Zhu, and Lee (2007). Although it would be possible to consider AIC or BIC type model selection criteria for the situation considered here, we consider the F type statistic because it arises as a natural choice for linear models.

It is worth mentioning the difference in the formulation of hypotheses testing. When these test statistics are computed sequentially, in the first case with (3.4), emphasis lies in how consistent the estimate of the slope would be when reducing the region of interest. In contrast, the second formulation in (3.5) allows the possibility of having no clear linear relationship and thus the bias and variance trade off comes into play only after linearity becomes effective.

4. Simulation study

In this section, we test the robustness of the proposed wavelet spectrum by using four simulated examples analyzed in Stoev et al. (2005). The examples are displayed in Figure 3. Each example has 100 realizations of $N = 30000$ time series points. Using the examples we compare Abry and Veitch's method (**AV**) and the proposed method (**New**). As explained Section 3, there are two important dif-



(a) Example 1 (b) Example 2

(c) Example 3 (d) Example 4

Figure 3: Simulated examples. True signals are overlaid onto FGN.

ferences between the existing and the proposed methods. The proposed method takes the logarithm of wavelet coefficients first and averages them later, and uses the F statistic, instead of χ^2, for model selection. It would be interesting to see an effect of each difference. Thus, we add another version of wavelet estimator (**Ad-hoc**) to the comparison, which takes the logarithm first but utilizes the χ^2 test statistic for model selection. We use the Daubechies wavelet with $M = 3$ for constructing wavelet spectra as Veitch and Abry (1999) suggested.

Example 1: Fractional Gaussian Noise (FGN)

Consider first an ideal situation when the data is a sample of FGN with $H = 0.9$. Figure 4 compares the three wavelet estimators. The top panels show the 100 H estimates of each method (solid lines) along with pointwise 95% confidence intervals (dotted lines). While all the three methods contain the true $H = 0.9$ in most of their confidence intervals, the AV tends to underestimate the true value compared to the other two. The Ad-hoc estimation has the highest variation since its confidence intervals are the widest.

The middle panels show the selected j_1 of each method. The proposed method impressively chooses $j_1 = 1$ for every repetition, which can be regarded as the true value since long-range dependence should appear at all scales for a FGN process. The AV has a small variation between $j_1 = 1$ and 2, and the Ad-hoc has the highest variation. The bottom panels show the goodness-of-fit measure of each method. For the New and Ad-hoc it is easier to understand what this values mean (p-values are close to 0), but it is not meaningful to interpret the goodness-of-fit measure value itself for the AV and it varies much from one simulation to another. The overall performance of the proposed method is satisfactory over the other two in this simulation.

Example 2: FGN plus a smooth trend

One major advantage of wavelet methods for estimating the Hurst parameter is that it can ignore smooth polynomial trends in the data owing to the vanishing moments in (2.1). This example has the setting of

$$\tilde{Y}_2(t_i) = Y(t_i) + P_l(t_i), \quad i = 1, 2, \ldots, N,$$

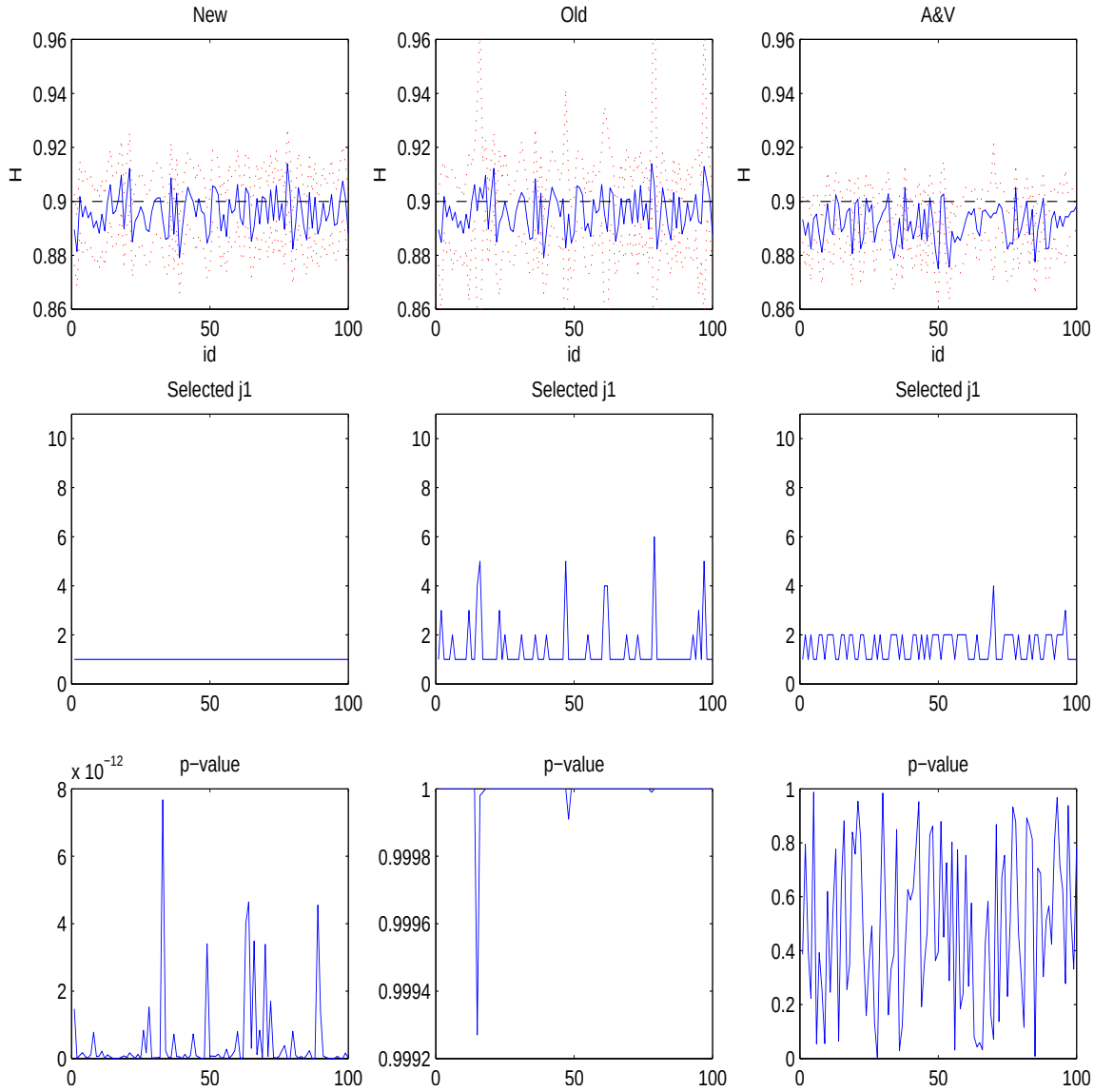


Figure 4: FGN ($H = 0.9$)

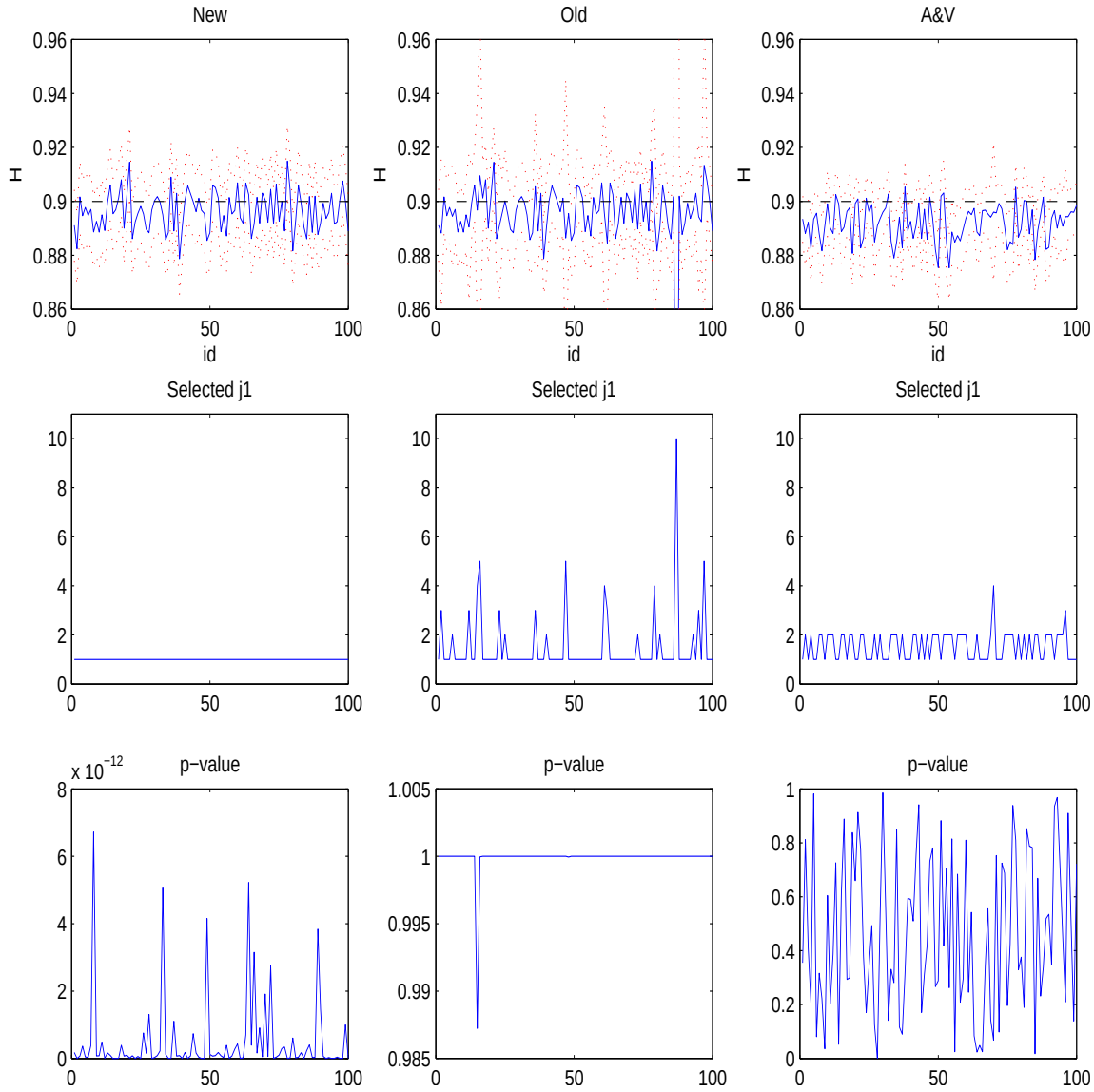


Figure 5: FGN ($H = 0.9$) plus a smooth trend

where $Y(t)$ is a FGN with $H = 0.9$ and $P_l(t) = a_0 t^l + \cdots + a_{l-1}t + a_l$, $t \in R$, is a polynomial of degree l. Hence, theoretically, the estimators of H, based on the wavelet coefficients of the perturbed process $\tilde{Y}$, will be identical to those based on the process Y as long as the vanishing moment M is sufficiently large. This is true in the sense that Figure 5 is not much different from Figure 4. Therefore, the lessons learned from Example 1 remain the same.

Example 3: FGN plus a high-frequency oscillating trend

Even though wavelet estimators are robust to a large class of smooth low-frequency trends, they can be quite sensitive to high-frequency deterministic oscillations. This example has the setting of

$$\tilde{Y}_3(t_i) = Y(t_i) + h_\nu(t_i), \quad i = 1, 2, \ldots, N,$$

where $h_\nu(t) = \sin(2\pi\nu t/N)$, $\nu > 0$. Here ν corresponds to the number of oscillations of h_ν in the interval $[0, N]$. If $\nu << M$, where M is the number of zero moments of ψ, then the function $h_\nu(t)$ can be essentially interpolated by a polynomial of degree $l < M$, and hence the wavelet estimator of H remains unaffected as seen in Example 2. However, a large M is not practically recommended (we use $M = 3$ in our analysis) and the high-frequency behavior then has a big impact on the estimation.

The top panels of Figure 6 show that the New and Ad-hoc overestimate the true H. Although they produce biased results, the estimates are stable through the repetitions. However, the H estimates by the AV shoot up and down and show big variations. This happens because the selected j_1 in the AV is always 10 (middle panel), which results in only a couple of points for estimating H in a regression setting. On the contrary, the New and Ad-hoc always choose $j_1 = 1$ despite the appearance of high-frequency oscillation trends. This implies that the robustness of the proposed method mainly comes from by taking the logarithm first.

Park et al. (2004) shows that the Sat1300 time series shown in Figure 1 has a high-frequency behavior inside the big spike in the middle. This simulated example clearly explains why the AV method does not work properly as shown in Figure 2.

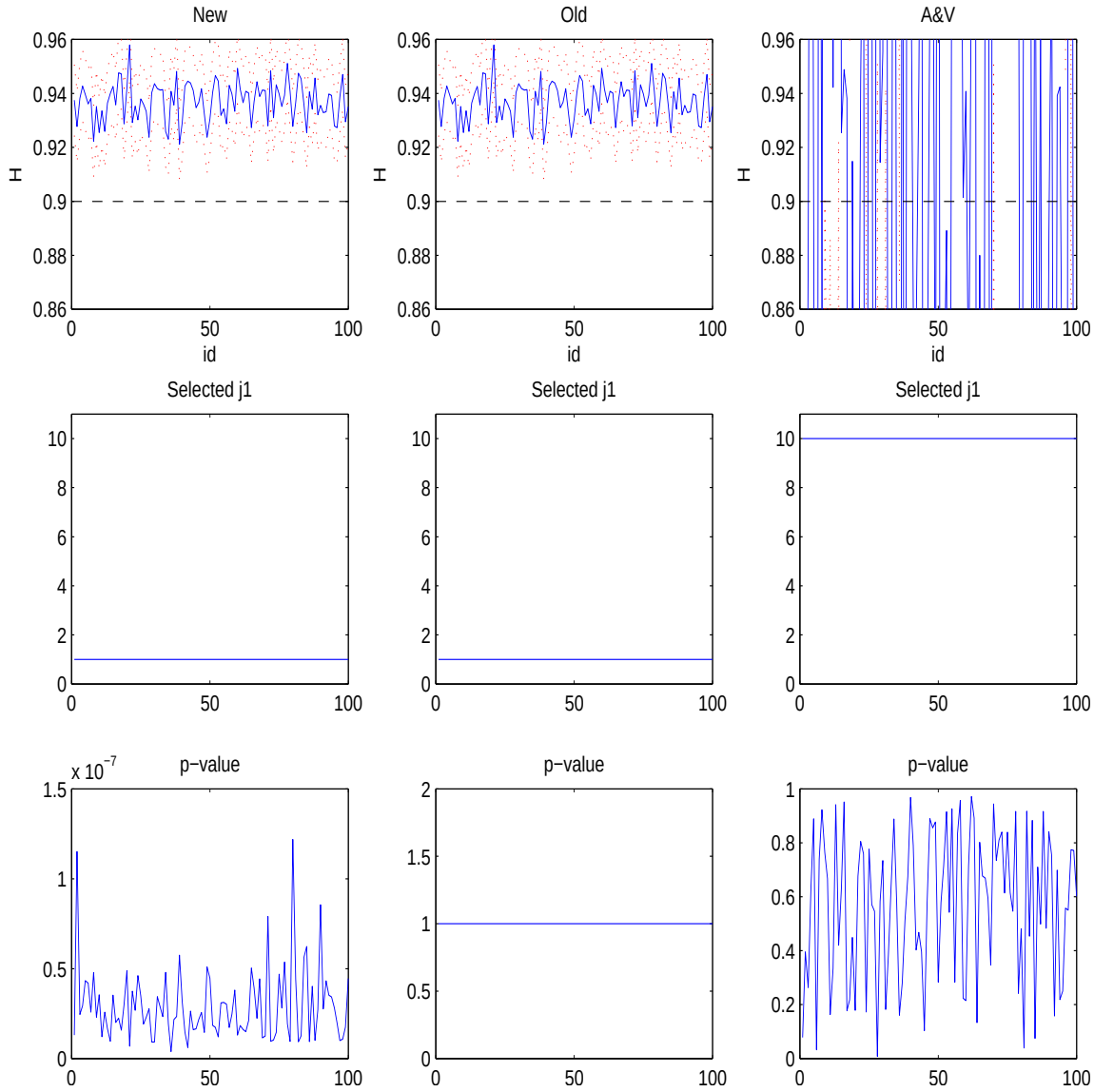


Figure 6: FGN ($H = 0.9$) plus a high-frequency oscillating trend

Example 4: FGN plus breaks

The wavelet spectrum of a time series can be influenced by breaks or shifts in the mean. The last example has the form of

$$\tilde{Y}_4(t_i) = Y(t_i) + h(t_i), \quad i = 1, 2, \ldots, N,$$

where the function $h(t)$ is a linear combination of indicator functions. Since the perturbation has a low degree of polynomial functions, all the three methods perform well in this case. Again the AV tends to underestimate the true value and the Ad-hoc has the highest variations. The proposed method has the least bias and the variation in the estimation and also produces the consistent $j_1 = 1$.

From the four simulations we can see that the robustness is achieved by taking the logarithm of wavelet coefficients first. The F statistic provides more stable estimation over the χ^2 statistic.

5. Real data example

In this section, we analyze two more Internet traffic packet counts data sets collected from the UNC link in 2002.

Figure 8 (a) displays a time series measured at the link of UNC on April 13 Saturday, from 7:30 p.m. to 9:30 p.m., 2002 (Sat1930). Figure 8 (b) displays a time series measured at the link of UNC on April 11 Thursday, from 1 p.m. to 3 p.m., 2002 (Thu1300). Again, they were originally measured every 1 millisecond but aggregated by a factor of 1000 (at 1 second) for better displays of trends. The Sat1930 time series shows one peak in the middle but the time series looks stationary in general. The Thu1300 time series shows a few spikes shooting up and down. Especially, the first downward spike hits the zero, which means no signal. This dropout lasts 8 seconds as shown in Park et al. (2007a).

Figures 9 (a) and (b) compare the Abry and Veitch's and the proposed methods using the Sat1930 time series. They produce similar estimates, $\hat{H} = 0.89$ (with 95% confidence interval $[0.86, 0.92]$) and $\hat{H} = 0.88$ (with 95% confidence interval $[0.86, 0.90]$), respectively. Also, they choose similar ranges of the scale, $j_1 = 12$ and $j_1 = 10$, respectively. We can see that the two methods produce similar estimates in the case of a stationary process as seen in Example 1 of Section 4.

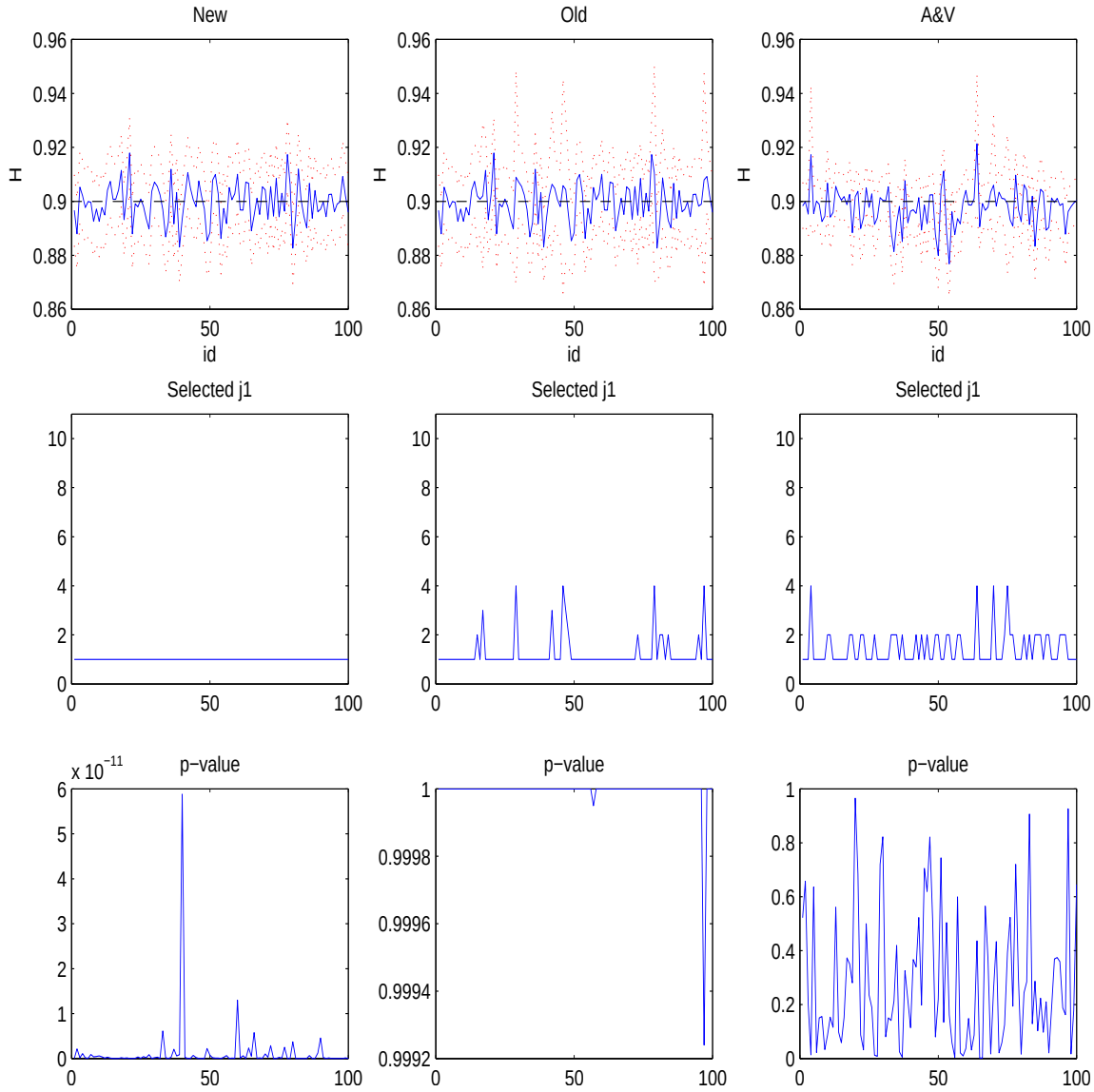
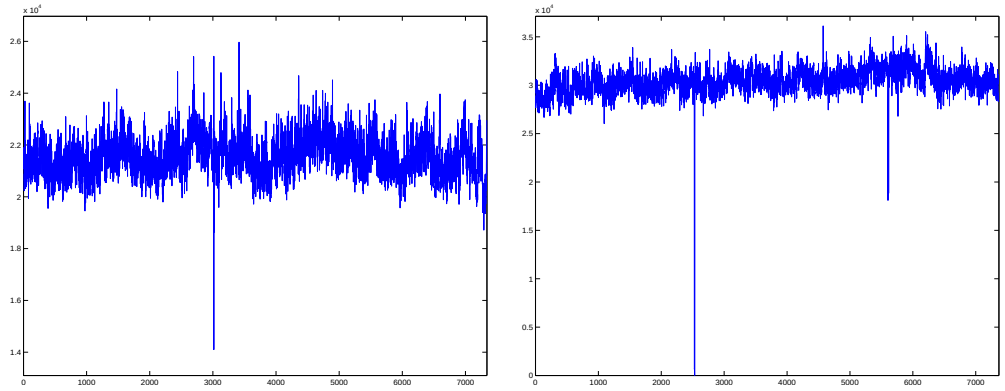


Figure 7: FGN ($H = 0.9$) plus breaks



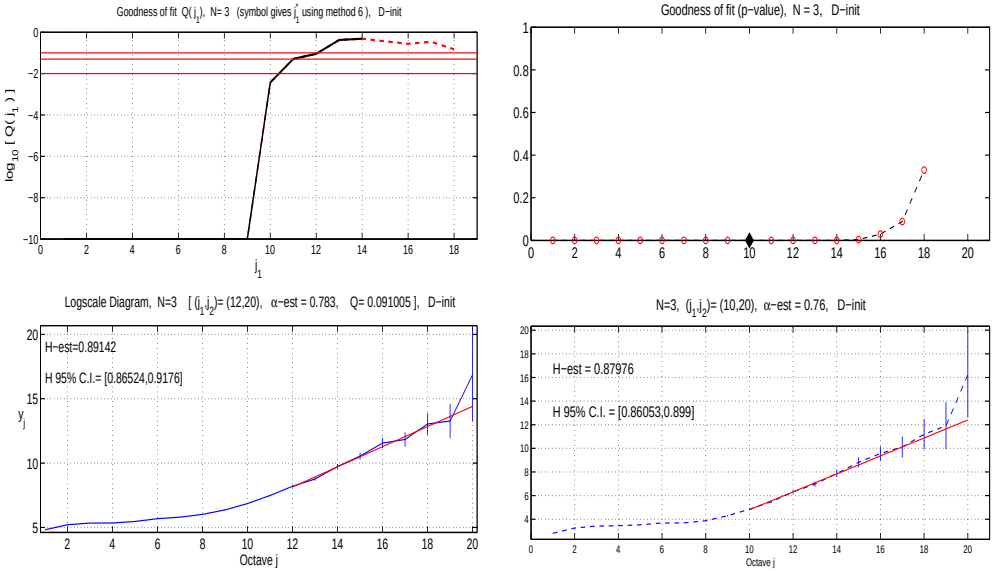
(a) Sat1930 (b) Thu1300

Figure 8: Packet count time series of aggregated traffic at 1 second: (a) Sat1930 and (b) Thu1300.

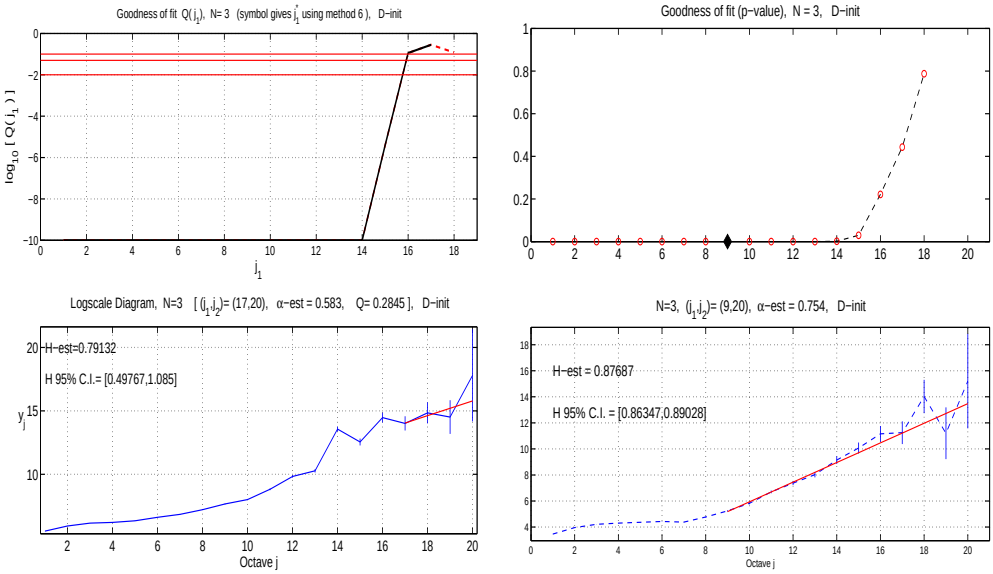
Figures 9 (c) and (d) compare the Abry and Veitch's and the proposed methods using the Thu1300 time series. They produce very different estimates, $\hat{H} = 0.79$ (with 95% confidence interval $[0.50, 1.09]$) and $\hat{H} = 0.88$ (with 95% confidence interval $[0.86, 0.89]$), respectively. The wide confidence of the Abry and Veitch's method is caused by the selection of the scale range, $j_1 = 17$. Note that the proposed method has $j_1 = 9$ and thus a narrower confidence band. The wavelet spectrum in Figure 9 (c) shows two bumps which force the method to choose the large j_1. Park et al. (2007a) showed that these bumps were created by the dropout. If this 8-second segment of the time series where the dropout occurs is excluded and the remaining parts are concatenated, then $\hat{H}$ is around 0.9, which is close to our estimate. This example clearly shows the robustness of the proposed method.

6. Concluding Remarks

We have shown that some issues with wavelet estimation of the Hurst parameter for long range dependent processes can be resolved by taking an alternative regression model, on which the estimator is based. The proposed wavelet estimator shows significant improvements in performance in various non-standard scenarios that standard estimators fail to reconcile. In addition, we have pro-



(a) Sat1930: Abry and Veitch (b) Sat1930: Proposed method



(c) Thu1300: Abry and Veitch (d) Thu1300: Proposed method

Figure 9: Wavelet spectra and the Hurst parameter estimates.

posed a new method of selecting an onset scaling, by making the link to the idea of regression diagnostics for linear models. These techniques are easy to implement and provide informative goodness of fit measures. There is accumulating evidence that the traffic exhibits much more versatile and dynamic behaviour than that can be described by a single parameter model. Thus, it is likely that there arises a situation where additional nonstationary phenomena need to be taken into account before the robust estimator or any other estimator can be employed. In the current framework, there may require different levels of preprocessing step to be able to justify the use of the Hurst parameter. Alternatively, we may adopt a view of modelling nonstationarity or local stationairy. Therefore it would be useful to develop a general framework where various nonstationary features can be incorporated so that the Hurst parameter itself can be a function of covariates such as time or other factors. We leave this consideration as future work.

Acknowledgment

The Internet traffic data we use here have been processed from logs of IP packets by members of the DIstributed and Real-Time (DIRT) systems lab at UNC Chapel Hill. This research was initiated when the authors were involved in the program of Network modeling for the Internet at Statistical and Applied Mathematical Sciences Institute in Research Triangle Park, North Carolina. The authors would like to thank them for providing an excellent environment and inspiring working atmosphere.

References

Abry, P., Flandrin, P., Taqqu, M. S., and Veitch, D. (2003). Self-similarity and long-range dependence through the wavelet lens. *in Theory and Applications of Long-range Dependence.* Birkhäuser, Boston, 527-556.

Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory* **44**, 2-15.

Bardet, J. M., Lang, G., Moulines, E., and Soulier, P. (2000). Wavelet estimator of long-range dependent processes. *Statistical Inference for Stochastic Processes* **3**, 85-99.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity.* John Wiley & Sons, New York-Chichester-Brisbane.

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

Chen, Jie and Gupta, A. K. (2000). *Parametric statistical change point analysis.* Birkhäuser, Boston.

Crovella, M. E. and Bestavros, A. (1996). Self-similarity in World Wide Web traffic evidence and possible causes. *Proceedings of the ACM SIGMETRICS 96* 160-169.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* John Wiley & Sons, New York.

Ferguson, T. S. (1996). *A course in large sample theory.* Chapman & Hall, London.

Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Tables of integrals, series and products.* Academic Press, 6th edition.

Hernández-Campos, F., Marron, J. S., Samorodnitsky, G. and Smith, F. D. (2004). Variable Heavy Tails in Internet Traffic. *Performance Evaluation* **58**, 261-284.

McKean, Joseph W. (2004). Robust Analysis of Linear Models. *Statistical Science* **19**, 562-570.

Park, C., Marron, J. S., and Rondonotti, V. (2004). Dependent SiZer: goodness-of-fit tests for time series models. *Journal of Applied Statistics* **31**, 999-1017.

Park, C., Godtliebsen, F., Taqqu, M., Stoev, S., and Marron, J. S. (2007a). Visualization and Inference Based on Wavelet Coefficients, SiZer and SiNos. *Computational Statistics and Data Analysis* **51**, 5994-6012.

Park, C., Hernández Campos, F., Le, L., Marron, J. S., Park, J., Pipiras, V., Smith F. D., Smith, R. L., Trovero, M., and Zhu, Z. (2007b). Long-range de-

pendence analysis of Internet traffic. *Under revision, Technometrics.* Web-available at http://www.stat.uga.edu/∼cpark/papers/LRDWebPage4.pdf

Paxson, V. and Floyd, S. (1995). Wide Area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking* **3**, 226-244.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* **23**, 1630-1661.

Shen, H., Zhu, Z., and Lee, T. (2007). Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Processing* **87**, 2111-2124.

Soltani, S., Simard, P., and Boichu, D. (2004). Estimation of the self-similarity parameter using the wavelet transform. *Signal Processing* **84**, 117-123.

Stoev, S. Pipiras, V., and Taqqu, M. S. (2002). Estimation of the self-similarity parameter in linear fractional stable motion. *Signal Processing* **82**, 1873-1901.

Stoev, S. and Taqqu, M. S. (2003). Wavelet estimation for the Hurst parameter in stable processes. In Rangarajan, G. and Ding, M. editors, *Processes with long-range correlations: theory and applications.* Springer-Verlag, Berlin, 61-87.

Stoev, S. and Taqqu, M. S. (2005). Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *Journal of Time Series Analysis* **26**, 211-249.

Stoev, S., Taqqu, M. S., Park, C., and Marron, J. S. (2005). On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of Internet traffic. *Computer Networks* **48**, 423-445.

Veitch, D. and Abry, P. (1999). A wavelet based joint estimator for the parameters of LRD. *IEEE Transactions on Information Theory* **45**, 878-897.

Veitch, D., Abry, P. and Taqqu, M. S. (2003). On the automatic selection of the onset of scaling. *Fractals* **11**, 377-390.

Wu, Y. (2005). *Inference for change-point and post-change means after a CUSUM test.* Springer, New York.

Department of Mathematics and Statistics, Lancaseter University, Lancaster, LA1 4YF, U.K.

E-mail: (juhyun.park@lancaster.ac.uk)

Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.

E-mail: (cpark@stat.uga.edu)