UC Santa Cruz
UC Santa Cruz Previously Published Works

Title

Long memory and changepoint models: a spectral classification procedure.

Permalink

https://escholarship.org/uc/item/2q63442q

Journal

Statistics and Computing, 28(2)

ISSN

0960-3174

Authors

Norwood, Ben

Killick, Rebecca

Publication Date

2018

DOI

10.1007/s11222-017-9731-0

Peer reviewed

CrossMark

Long memory and changepoint models: a spectral classification procedure

Ben Norwood[1] · Rebecca Killick[1]

Abstract Time series within fields such as finance and economics are often modelled using long memory processes. Alternative studies on the same data can suggest that series may actually contain a 'changepoint' (a point within the time series where the data generating process has changed). These models have been shown to have elements of similarity, such as within their spectrum. Without prior knowledge this leads to an ambiguity between these two models, meaning it is difficult to assess which model is most appropriate. We demonstrate that considering this problem in a time-varying environment using the time-varying spectrum removes this ambiguity. Using the wavelet spectrum, we then use a classification approach to determine the most appropriate model (long memory or changepoint). Simulation results are presented across a number of models followed by an application to stock cross-correlations and US inflation. The results indicate that the proposed classification outperforms an existing hypothesis testing approach on a number of models and performs comparatively across others.

Keywords Classification · Long memory · Changepoint · Wavelet spectrum · Non-stationarity

1 Introduction

It is not often the case that a given data set has a known explicit model from which it is generated. Analysts will look

✉ Rebecca Killick
    r.killick@lancs.ac.uk

    Ben Norwood
    b.norwood@lancs.ac.uk

1   Department of Mathematics and Statistics, Lancaster
    University, LA1 4YF Lancaster, UK

to fit an appropriate model to such a series in the hopes of understanding the underlying mechanisms or to make predictions into the future. The models proposed are expected to be distinct in their properties such that there is a clear prevalence of a suitable model for the data. However, models with certain structural features have been known to have similar properties to other models (Granger and Hyung 2004). This overlap will be here referred to as an 'ambiguity' between the models. This is such that either model may appear similar to one another in some metrics, but provide very different interpretations on the data generating process, and lead to different predictions into the future.

In this paper, we consider the ambiguity between long memory and changepoint models. This ambiguity has been documented in fields such as finance and economics which are modelled using long memory models (Granger and Ding 1996; Pivetta and Reis 2007) and changepoint models (Levin and Piger 2004; Starica and Granger 2005). Thus, it is reasonable to assert that there is an element of ambiguity between these two models. Following the discussion and in-depth analysis within Diebold and Inoue (2001), it has been shown that both models share some similar properties, especially within the spectrum. Often a decision on a model cannot be made with the 'luxury' of prior knowledge, and as such assuming the data derives from either of these models comes at a risk of mis-specification.

Existing work in Yau and Davis (2012) conducts a hypothesis test to determine between the changepoint and long memory model. The authors choose to use the changepoint model as a null model with the justification that this is the more plausible model. However, in some circumstances this may not be the case, so it leads to the question as to which model should be the null model. It would be entirely feasible to choose the changepoint model as the null model, not reject H_0 and then flip to have the long memory model as the

null model and also not reject H_0. This does not give a clear answer to the question of an appropriate model.

As an alternative this paper introduces a classifier, which places no such assumptions on which model is preferred. Instead, the purpose of a classifier is only to give a measure of which category provides the best fit. In the context here, it can measure which model best describes a time series, without assuming that this model is where the data were originally generated from. Classification of time series has been previously used in Grabocka et al. (2012) and Krzemieniewska et al. (2014). It was shown in Yau and Davis (2012) that the autocorrelation function and periodogram of data generated from a changepoint model and a long memory model exhibit similar structures (i.e. slow decay in the autocorrelation and spectral pole at zero). However, if we consider a time-varying periodogram, then the stationarity of a long memory model can be seen (constant structure over time), whilst a changepoint model exhibits the piecewise stationarity expected [see for example Killick et al. (2013)]. As the time-varying spectrum shows evidence of a difference between these models, we use it as the basis for our classification procedure.

The structure of this article is as follows. The background and methods to our approach are given in detail in Sect. 2. A simulation study of the proposed classification method, with a comparison to the likelihood ratio test from Yau and Davis (2012), can be found in Sect. 3. Applications of the classifier are then given using US price inflation and stock cross-correlations in Sect. 4. Finally, concluding remarks and a discussion is given in Sect. 5.

2 Methods

2.1 Changepoint and long memory models

The aim of our method is to distinguish between data which arise from either a changepoint or a long memory model. To define these, we first define the general autoregressive integrated moving average (ARIMA) model, characterised by its autoregressive (AR) parameters $\boldsymbol{\phi} \in \mathbb{R}^p$, moving average (MA) parameters $\boldsymbol{\theta} \in \mathbb{R}^q$ and the integration (I) parameter $d \in \mathbb{N}$. For random variables $X_1, X_2, \ldots, X_n$ this is formally defined as,

$$\left(1 - \sum_{k=1}^{p} \phi_k B^k\right)(1 - B)^d X_t = \left(1 + \sum_{k=1}^{q} \theta_k B^k\right)\epsilon_t$$

where $\epsilon_t \sim N(0, \sigma^2)$ and B is the backward shift operator such that $BX_t = X_{t-1}$ and $B\epsilon_t = \epsilon_{t-1}$. A variation of this, autoregressive fractional integrated moving average (ARFIMA), is such that $d \in \mathbb{R}$, allowing it to be fractional. This modification allows long memory behaviour to be cap-

tured through dependence over a large number of previous observations.

For the purpose of this paper, we define the changepoint and long memory models as:

$$X_t \sim \begin{cases} \mu_1 + \text{ARMA}(\boldsymbol{\phi_1}, \boldsymbol{\theta_1}) & \text{if } t = 1, 2, \ldots \tau \\ \mu_2 + \text{ARMA}(\boldsymbol{\phi_2}, \boldsymbol{\theta_2}) & \text{if } t = \tau+1, \tau+2, \ldots n. \end{cases}$$

(1)

$$X_t \sim \mu + \text{ARFIMA}(\boldsymbol{\phi}, d, \boldsymbol{\theta}) \quad t = 1, 2, \ldots, n$$

(2)

Note that we depict a single changepoint $\tau = \lfloor n\lambda \rfloor$ for notational ease, but the software we provide (see Sect. 5) contains the generalisation to multiple changes through use of the PELT algorithm (Killick et al. 2012) and extending Eq. (1) to include multiple τ. Other models such as ARCH models and fractional Gaussian noise (Molz et al. 1997) could also be used, but we restrict our consideration to ARFIMA here. In the general case, we allow $p, q \in \mathbb{N}$, but in the simulations and applications given in Sects. 3 and 4 we restrict their range for computational reasons.

2.2 Wavelet spectrum

The ambiguity present between diagnostics of the competing models given in Eqs. (1) and (2) can cause issues in identifying the correct model. Figure 1 shows the average empirical periodograms from realisations of long memory [ARFIMA(0, 0.4, 0)] and changepoint (AR(1), $\lambda = 0.5$, $\phi_1 = 0.1$, $\phi_2 = 0.4$, $\mu_1 = 0$, $\mu_2 = 1$) models. It can be seen that the periodogram for the changepoint model has a pole at zero and shows similar behaviour to that of long memory.

Before discussing the wavelet spectrum, we provide a brief background to wavelets and the specific spectrum we propose to use.

Wavelets capture properties of the data through a location–scale decomposition using compactly supported oscillating functions. Through dilation and translation, a wavelet is applied across a number of a scales and locations to capture behaviour occurring over different parts of a series. Further information on them and their application can be found in Daubechies (1992) and Nason (2010). In this work, we use the model framework of the locally stationary wavelet process which provides a stochastic model for second-order structure using wavelets as building blocks.

We follow the definition in Fryzlewicz and Nason (2006) for a locally stationary wavelet (LSW) process.

Definition 1 Define the triangular stochastic array $\{X_{t,N}\}_{t=0}^{N-1}$ which is in the class of LSW processes given it has the mean-square representation
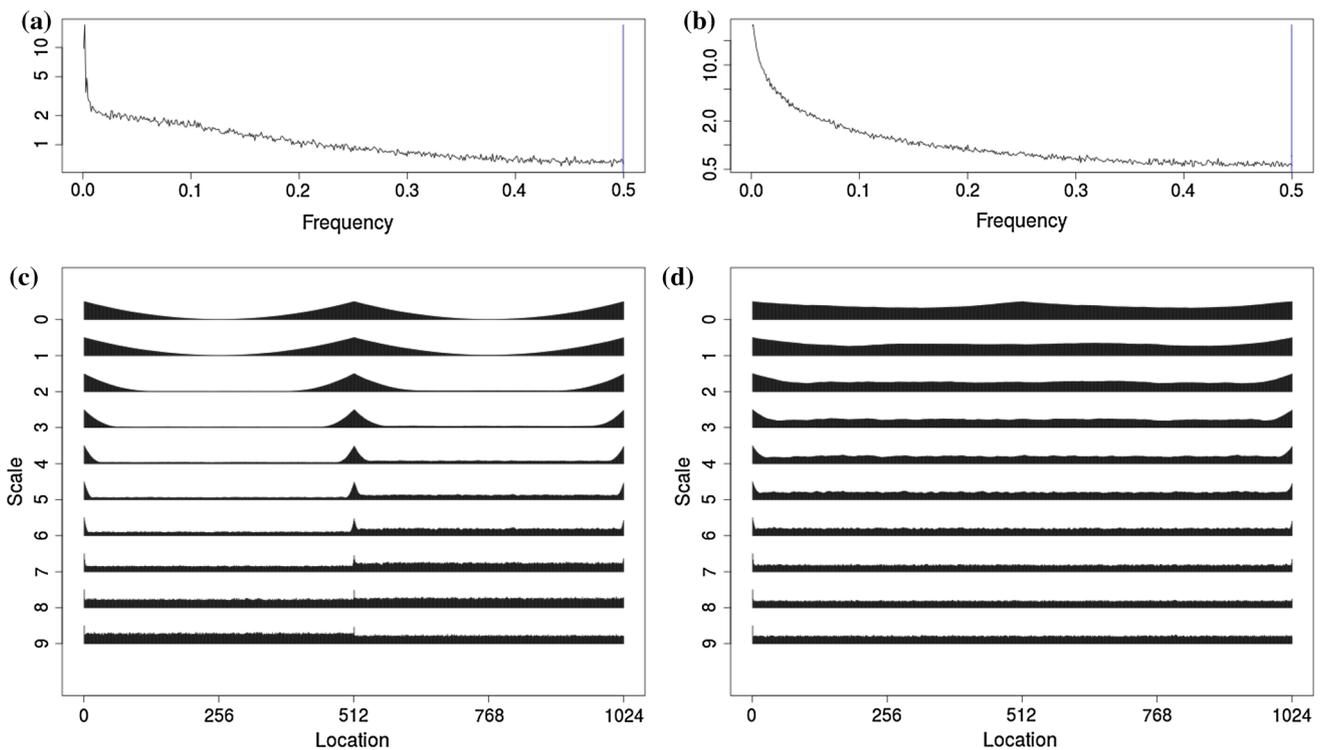
Fig. 1 Empirical periodogram and wavelet spectrum averaged over 500 realizations. **a** Changepoint periodogram. **b** Long memory periodogram. **c** Changepoint wavelet spectrum. **d** Long memory wavelet spectrum

$$X_{t,N} = \sum_{j=1}^{\infty} \sum_{k} W_j\left(\frac{k}{n}\right) \psi_{j,k-t}\xi_{j,k},$$

where $j \in 1, 2, \ldots$ and $k \in \mathbb{Z}$ are scale and location parameters, respectively, $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,L_j-1})$ are discrete, compactly supported, real-valued non-decimated wavelet vectors of support length L_j. If the ψ_j are Daubechies wavelets (Daubechies 1992) then $L_j = (2^j - 1)(N_h - 1) + 1$ where N_h is the length of the Daubechies wavelet filter, finally the $\xi_{j,k}$ are orthonormal, zero-mean, identically distributed random variables. The amplitudes $W_j(z) : [0, 1] \rightarrow \mathbb{R}$ at each $j \geq 1$ are time-varying, real-valued, piecewise constant functions which have an unknown (but finite) amount of jumps. The constraints on $W_j(z)$ are such that if $\mathcal{P}_j$ are Lipschitz constants representing the total magnitude of jumps in $W_j^2(z)$, then the variability of $W_j(z)$ is controlled by

- $\sum_{j=1}^{\infty} 2^j \mathcal{P}_j < \infty$,
- $\sum_{j=1}^{\infty} W_j^2(z) < \infty$ uniformly in z.

As in the traditional Fourier setting, the spectrum is the square of the amplitudes and as such the evolutionary wavelet spectrum can be defined as

$$S_j\left(\frac{k}{N}\right) = \left| W_j\left(\frac{k}{N}\right) \right|^2$$

which changes over both scale (frequency band) j and location (time) k.

Considering both scale and location, the two dimensions allow the differences between the proposed models to be seen. Examples of the differences in these spectra are given in Fig. 1 for both the changepoint and long memory models. To interpret the wavelet spectrum: scale corresponds to frequency bands with high frequency at the bottom to low frequency at the top. Further details on the spectrum and its applicability can be found in Fryzlewicz and Nason (2006), Nason (2010) and Killick et al. (2013). Note that there is a clear difference between the wavelet spectra of the two models with the changepoint model being piecewise stationary (pre- and post-change), with the change occurring in the spectrum where the change occurs in the data. In contrast the long memory model remains flat across each scale and time reflecting the stationarity of the original series.

Due to the fact that the wavelet spectrum gives a distinction between the two models, we propose to use this as the basis for our inference regarding the most appropriate model. Whilst the Fourier spectrum could be used here as in Janacek et al. (2005), we choose to use the evolutionary wavelet spectrum. As shown in Fig. 1, this is advantageous for characterising the non-stationarity changepoint data due to the scale–location transformation used. This is since the $W_j(z)$ are constant for stationary models, but for non-stationary

models the break in the second-order structure of the original data causes breaks in the wavelet spectra, as described in Cho and Fryzlewicz (2012).

In the next section, we detail how to use the wavelet spectrum of the two models in a classification procedure.

2.3 Classification

Testing whether a long memory or changepoint model is more appropriate whilst under model uncertainty comes with the hazard of mis-specification. A formal hypothesis test places assumptions on the underlying model in both the null and alternative, but the allocation of the null is hazardous— should the changepoint model be the null or alternative? It would be entirely feasible to choose the changepoint model as the null model, not reject H_0 and then flip to have the long memory model as the null model and also not reject H_0. Given the absence of a clear null model, which result to proceed with is unclear. Instead, it may be preferable to quantify the evidence for each model separately. A classification method such as the one proposed here gives a candidate series a measure of distance from a number of groups, which can then be used to select the most appropriate group.

In the previous subsection, it was demonstrated that the wavelet spectrum can be used to distinguish the changepoint model from the long memory model, and the classifier proposed here builds on this. However, to begin a classification method must first 'teach' itself on the structure of the classes through sets of training data. These are data sets already determined to be in each category and are the basis for calculating the distances from each group. This previous knowledge allows for determination of patterns and features of each category (that are unique from other categories) for comparison to the candidate data set. A common example is the spam filter on mailboxes, which is trained on previous spam emails so that it can classify if a new email that arrives is spam or not. The decision is made by comparing it to a number of patterns already determined to be features in spam email for example, short messages or hidden sender identities. Further information on classification methods and training them can be found within Michie et al. (1994).

In our example, we only have a single data set of length n, the classifier has no previous information to train on. To remedy this we create training data through simulation. Given a candidate series we first fit the competing models in Eqs. (1) and (2) choosing the best fit for each model. For the changepoint model the best fit uses the ARMA likelihood within the PELT multiple changepoint framework to identify multiple changes in ARMA structure (Hyndman and Khandakar 2008; Killick et al. 2012). When considering fitted long memory models, a number of ARFIMA models are fitted (Veenstra

2012) and selection occurs according to Bayesian information criterion [following Beran et al. (1998)].

Following the identification of the best changepoint and long memory models, the training data are then simulated as (Monte Carlo) realisations from these, denoted by

$$X_m^g = \left\{ X_{i,m}^g \right\}_{i=1,2,\ldots,n} \quad \begin{aligned} m &= 1, 2, \ldots, M. \\ g &= 1, 2. \end{aligned}$$

where the group, $g = 1$ for changepoint simulations and $g = 2$ for long memory simulations, M is the number of simulated series and n is the length of the original series. Note that we are not sampling from the original series, we are generating realizations from the fitted models.

Now we have the training data and the observed data, denoted X^o, a measure of distance of the observed data from each group is calculated. As discussed previously, we will use a comparison of their evolutionary wavelet spectra as the distance metric. Before detailing the metric, we first define the wavelet spectrum of the original series as

$$S^o = \left\{ S_k^o \right\}_{k=1,2,\ldots n*J}$$

where we remove the index over scale j by concatenating scales, hence $k = 1, 2, \ldots n * J$, where $J = \lfloor \log_2(n) \rfloor$. Similarly we define the spectra for each simulated series:

$$S_m^g = \left\{ S_{k,m}^g \right\}_{k=1,2,\ldots n*J}.$$

To obtain a group spectra, an average is then taken over the M simulated series at each position of each scale for each group,

$$\bar{S}^g = \left\{ \frac{1}{M} \sum_{m=1}^{M} S_{k,m}^g \right\}_{k=1,2,\ldots n*J}.$$

Based on these spectra, the distance metric proposed is a variance-corrected squared distance, across all spectral coefficients as proposed in Krzemieniewska et al. (2014),

$$D^g = \frac{M}{(M+1)} \sum_{k=1}^{n*J} \frac{(S_k^o - \bar{S}_k^g)^2}{\sum_{m=1}^{M} (S_{k,m}^g - \bar{S}_k^g)^2} \quad (3)$$

Note that the variance correction occurs within the denominator to account for potentially different variability seen across simulations for each group. This is modified from Krzemieniewska et al. (2014) to allow different variances within each group. The theoretical consistency of the classification was shown in Theorem 3.1 from Fryzlewicz and Ombao (2009) where the error for misclassifying two spectra $\left\{S_k^{(1)}\right\}_k$ and $\left\{S_k^{(2)}\right\}_k$ (whose difference summed over k is larger than CN) is bounded by $\mathcal{O}\left(N^{-1}\log_2^3 N + N^{1/\{2\log_2(a)-1\}-1}\log_2^2 N\right)$. However, this result requires a short memory assumption that is clearly not satisfied for our long memory processes. Thus, we prove a similar bound under the assumption that the spectra are created from ARFIMA processes. We first replicate the required assumptions from Fryzlewicz and Ombao (2009) for completeness:

Assumption 2.1 (*Assumption 2.1 from* Fryzlewicz and Ombao (2009)) The set of those locations z where (possibly infinitely many) functions $S_j(z)$ contain a jump is finite. In other words, let $\mathcal{B} := \left\{z : \exists j \lim_{u \to z^-} S_j(u) \neq \exists j \lim_{u \to z^+}\right\}$. We assume $B := \#\mathcal{B} < \infty$.

Assumption 2.2 (*Assumption 2.2 from* Fryzlewicz and Ombao (2009)) There exists a positive constant C_1 such that for all j, $S_j(z) \leq C_1 2^j$.

Theorem 1 *Suppose that assumptions 2.1 and 2.2 hold, and that the constants $\mathcal{P}_j$ from definition 1 decay as $\mathcal{O}(a^j)$ for $a > 2$. Let $S_j^{(1)}(z)$ and $S_j^{(2)}(z)$ be two non-identical wavelet spectra from ARFIMA processes. Let $I_{k,N}^{(J)}$ be the wavelet periodogram constructed from a process with spectrum $S^{(1)}(z)$, and let $L_{k,N}^{(j)}$ be the corresponding bias-corrected periodogram, with $J^* = \log_2 N$. Let*

$$\sum_{j,k}\left\{S_j^{(1)}(k/N) - S_j^{(2)}(k/N)\right\}^2 = \mathcal{O}(N).$$

The probability of misclassifying $L_{k,N}^{(j)}$ as coming from a process with spectrum $S_j^{(2)}(z)$ can be bounded as follows:

$$P(D_1 > D_2) = \mathcal{O}\left(\log_2^2 N\left[N^{-1} + N^{\frac{1}{(2\log_2 a - 1)} - 1}\right]\right)$$

Proof The proof is given in Appendix 1.

A summary of the proposed procedure is given in Algorithm 1.

Initialization:

$\mathbf{X} : \{X_i\}_{i=1}^n$ observed series.
n : Length of series
M : Number of bootstrap simulations
$\bar{\mathbf{S}}^1, \bar{\mathbf{S}}^2$: Empty Spectra 1, 2.

Algorithm:

1. Fit: $\mathcal{M}_1$ - best changepoint model (Equation (1)) to $\mathbf{X}$.
2. Fit: $\mathcal{M}_2$ - best long memory model (Equation (2)) to $\mathbf{X}$.
3. Calculate training spectra
   for $m = 1, 2, \ldots, M$ **do**
   Simulate n observations from $\mathcal{M}_1$, denote as $\mathbf{Y_1}$
   Calculate Evolutionary Wavelet Spectra $\mathbf{S}_m^1$ of $\mathbf{Y_1}$
   Let $\bar{\mathbf{S}}^1 = \bar{\mathbf{S}}^1 + \mathbf{S}_m^1$

   Simulate n observations from $\mathcal{M}_2$, $\mathbf{Y_2}$
   Calculate Evolutionary Wavelet Spectra $\mathbf{S}_m^2$ of $\mathbf{Y_2}$
   Let $\bar{\mathbf{S}}^2 = \bar{\mathbf{S}}^2 + \mathbf{S}_m^2$
   end
4. Calculate the average Evolutionary Wavelet Spectra for each group $\bar{\mathbf{S}}^1 = \frac{\bar{\mathbf{S}}^1}{M}$, $\bar{\mathbf{S}}^2 = \frac{\bar{\mathbf{S}}^2}{M}$.
5. Calculate Evolutionary Wavelet Spectrum of $\mathbf{X}$, $\mathbf{S}^o$.
6. Compute the distance D_1, D_2, between $\mathbf{S}^o$ and $\bar{\mathbf{S}}^1$, $\bar{\mathbf{S}}^2$ respectively (Equation (3)).

Output: Distances D_1, D_2.

Algorithm 1: Wavelet Classifier Algorithm

3 Simulation study

To test the empirical accuracy of our proposed approach, simulations were conducted over a number of models. Here, these models are chosen over a number of parameter magnitudes and combinations to show the effectiveness of the approach outlined in Sect. 2. A number of these models also appear in Yau and Davis (2012) which uses a likelihood ratio method to test the null hypothesis of a changepoint model. Their results for these models are correspondingly given as a comparison.

For each model given in the tables below, 500 realisations of each model were generated and classified, using $M = 1000$ training simulations for each fit. For computational efficiency, the maximum order of the fitted models is constrained to $p, q \leq 1$. Three different time series lengths were computed for each model; 512, 1024 and 2048. It is expected that as a series grows larger, more evidence of long memory features will become prevalent, and as such the effect of length of series on accuracy is investigated.

Table 1 Changepoint observations results with likelihood ratio comparison (Yau and Davis 2012)

Ref	Model parameters						Classification rate			Y and D Likelihood ratio	
	λ	μ	ϕ_1	θ_1	ϕ_2	θ_2	$n = 512$	$n = 1024$	$n = 2048$	$n = 500$	$n = 1000$
1	0.5	1	0.1	0.3	0.4	0.2	1.00	1.00	1.00	0.99	0.97
2	0.5	2	0.1	0.3	0.4	0.2	1.00	1.00	1.00	0.95	0.93
3	0.5	1	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.97	0.99
4	0.5	2	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.94	0.95
5	0.7	1	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.94	0.94
6	0.7	2	0.1	0.3	0.8	0.2	1.00	1.00	1.00	0.91	0.93

We have used $n = 2^J$ as the length of the series as the wavelet decomposition software (Nason 2016b) requires that the series transformed is of dyadic length. This is not a desirable trait as data sets come in many different sizes. Thus, we overcome this using a standard padding technique (Nason 2010) that adds 0's to the left of each series until the data are of length 2^J. The extended wavelet coefficients are then removed before calculating the distance metric.

3.1 Changepoint observations

For the changepoint models, we used the simulations given in Yau and Davis (2012). Table 1 gives the parameters used in Eq. (1) along with the correct classification rate. The results show that if the data follow a changepoint model then we have a 100% classification rate. A movement of the changepoint to a later part of the series, as in models 5 and 6, does not appear to have an effect upon classification rates unlike for the Yau and Davis method. It is not really a surprise that we are receiving 100% classification rates as if a changepoint occurs then it is a clear feature within the spectrum.

It should be noted that as the Yau and Davis method is a hypothesis test we would expect results around 0.95 for a 5% type I error.

3.2 Long memory observations

In contrast to the changepoint models, the classification of a long memory model is expected to be less clear. This is due to the variation within the wavelet spectrum of long memory series that could be interpreted as different levels and hence a changepoint model would be more appropriate. To demonstrate the effect of the classifier on long memory observations, a larger number of models were considered. We simulated long memory models with differing levels of long memory as measured by the d parameter, values close to 0 are closer to short memory models and values close to 0.5 are stronger long memory models (values >0.5 are not stationary and thus not considered).

The results in Table 2 give an indication of the accuracy of the classifier in a number of different situations. Overall,

as the length of the time series increases we see an increase in classification accuracy. This is to be expected as evidence of long memory will be more prevalent in longer series. Similarly as we increase the long memory parameter d from 0.1 to 0.4 we improve the classification rate.

Some interesting things to note include, when there are strong AR parameters (ϕ) such as models 7–10 and 19–22 we require longer time series to achieve good classification rates. However, in contrast if there are strong MA components as in the remaining models the classifier performs better. A larger effect is found when the MA parameter is negative, seen through models 11–14 where the classifier performs strongly even at $n = 512$. This effect is further exemplified by models 23–26 which include a further MA parameter and achieve near 100% classification at $n = 512$. Here the maximum used p, q was 2.

Comparing our results to that of Yau and Davis we note that the opposite performance is seen. For the likelihood ratio method there is high power for models with strong AR components and poor performance for strong MA components. Notably the strong MA performance is much worse than our method on the strong AR components.

4 Application

To further demonstrate the usage of our approach, two applications to real data are given in this section. The first is an economics example based on US price inflation and this is followed by financial data on stock cross-correlations. A sensitivity analysis was conducted over the possible maximum values of p, q. It was found that no additional parameters were required beyond maximum $p, q = 4$, thus these results are presented here.

4.1 Price inflation

US price inflation can be determined using the GDP index. The data set used here is available from the Bureau of Economic Analysis, based on quarterly GDP indexes, denoted P_t, from the first quarter of 1947 to the third quarter of

Table 2 Long memory observations result with likelihood ratio comparison (Yau and Davis 2012)

| Ref	Model parameters				Classification rate			Y and D LR power
	ϕ	d	θ_1	θ_2	$n = 512$	$n = 1024$	$n = 2048$	$n = 500$
7	−0.8	0.1	0.6		0.42	0.61	0.79	0.63
8	−0.8	0.2	0.6		0.56	0.83	0.94	0.97
9	−0.8	0.3	0.6		0.66	0.90	0.96	0.98
10	−0.8	0.4	0.6		0.75	0.88	0.96	0.96
11	0.1	0.1	−0.8		0.74	0.87	0.95	0.08
12	0.1	0.2	−0.8		0.84	0.96	0.99	0.09
13	0.1	0.3	−0.8		0.89	0.98	1.00	0.15
14	0.1	0.4	−0.8		0.88	0.99	1.00	0.32
15	0.1	0.1	0.8		0.54	0.78	0.90	
16	0.1	0.2	0.8		0.61	0.85	0.91	
17	0.1	0.3	0.8		0.62	0.87	0.95	
18	0.1	0.4	0.8		0.63	0.87	0.98	
19	0.6	0.1	−0.8		0.33	0.45	0.65	
20	0.6	0.2	−0.8		0.38	0.62	0.83	
21	0.6	0.3	−0.8		0.44	0.63	0.87	
22	0.6	0.4	−0.8		0.39	0.59	0.86	
23	0.0	0.1	0.7	−0.7	0.94	0.97	0.99	
24	0.0	0.2	0.7	−0.7	1.00	0.99	1.00	
25	0.0	0.3	0.7	−0.7	1.00	1.00	1.00	
26	0.0	0.4	0.7	−0.7	1.00	0.99	1.00	
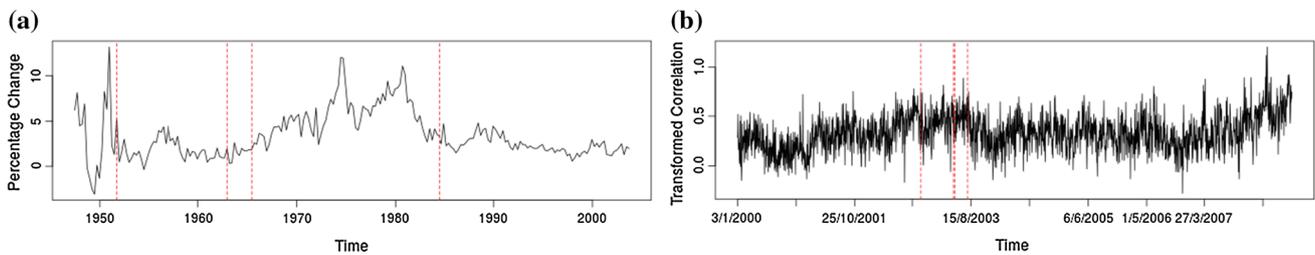
Fig. 2 Real data examples. **a** Time series of US price inflation. **b** Time series of the cross-correlations of American Express and Home Depot.

2006 (227 data points). Price inflation is calculated as $\pi_t = 400 \ln(P_t/P_{t-1})$ (thus $n = 226$). A plot of the inflation is given below in Fig. 2a. Studies of the persistence of this data have been conducted to determine the level of dependence within the series. A high amount of persistence, indicating long memory, was found in Pivetta and Reis (2007). However Levin and Piger (2004) found a structural break, which when accounted for showed the series to have low persistence, indicating the presence of changepoints with short memory segments. Applying our classification approach to this series will give an additional indication as to which model is statistically more appropriate.

The parameters of the fitted changepoint and long memory models are given in Table 3. Diagnostic autocorrelation and partial autocorrelation function plots are given in Fig. 3. The level shifts are given in respect to their position in the series,

but correspond to 1951 Q3, 1962 Q4, 1965 Q2, 1984 Q2. The classifier returns a changepoint classification for this series.

4.2 Stock cross-correlations

Stock cross-correlation data have been obtained from the supplementary material of Chiriac and Voev (2011). The data consist of open to close stock returns for 6 companies from January 1st 2001 to 30th July 2008 ($n = 2156$). The data are first transformed using a Fisher transformation, then correlations are calculated between each stock. Here analysis will look at the correlation between American Express and Home Depot.

These data have been analysed previously by Bertram et al. (2013) to determine between fractional integration (long memory behaviour) and level shifts and are given in Fig. 2b. Parameters for the models fitted by the algorithm are also

Table 3 Model fits and scores for US inflation (inflation) and stock cross-correlations (stock)

	Long Memory Model	Score	Changepoint Model	Score
Data				
Inflation	ARFIMA$(4, 0, 0)$		AR(1)	Segment 1
	$\phi = (0.30, 0.20, 0.20, -0.16)$		$\phi = 0.68, \quad \tau = 18, \quad \mu = 0$	
	$d = 0.31$		AR(1)	Segment 2
	$\mu = 3.29$		$\phi = 0.63, \quad \tau = 63, \quad \mu = 1.70$	
			ARMA$(2, 1)$	Segment3
			$\phi = (-0.43, -0.33)$	
			$\theta = 0.47, \quad \tau = 73, \quad \mu = 1.69$	
			AR(1)	Segment 4
			$\phi = 0.82, \quad \tau = 149, \quad \mu = 5.43$	
			AR(3)	Segment 5
			$\phi = (0.60, 0.19, 0.18)$	
		40147	$\tau = 226, \quad \mu = 0$	**38750**
Stock	ARFIMA$(1, 0, 1)$		ARMA$(4, 4)$	Segment 1
	$\phi = 0.30$		$\phi = (-0.05, -0.002, 0.07, 0.94)$	
	$\theta = 0.58$		$\theta = (0.17, 0.14, 0.04, -0.83)$	
	$d = 0.47, \quad \mu = 0.35$		$\tau = 715, \quad \mu = 0.31$	
			AR(1)	Segment 2
			$\phi = 0.29, \quad \tau = 841, \quad \mu = 0.42$	
			ARMA$(2, 1)$	Segemnt 3
			$\phi = (0.09, 0.09)$	
			$\theta = -0.58, \quad \tau = 847, \quad \mu = 0.30$	
			MA(1)	Segment 4
			$\theta = -0.72, \quad \tau = 896, \quad \mu = 0.50$	
			ARMA$(3, 1)$	Segment 5
			$\phi = (1.12, -0.06, -0.07)$	
		12128	$\theta = -0.90, \quad \tau = 2156, \quad \mu = 0.35$	5348047

Bold scores are the minimum. Each segment ending at τ is separated by a dotted line
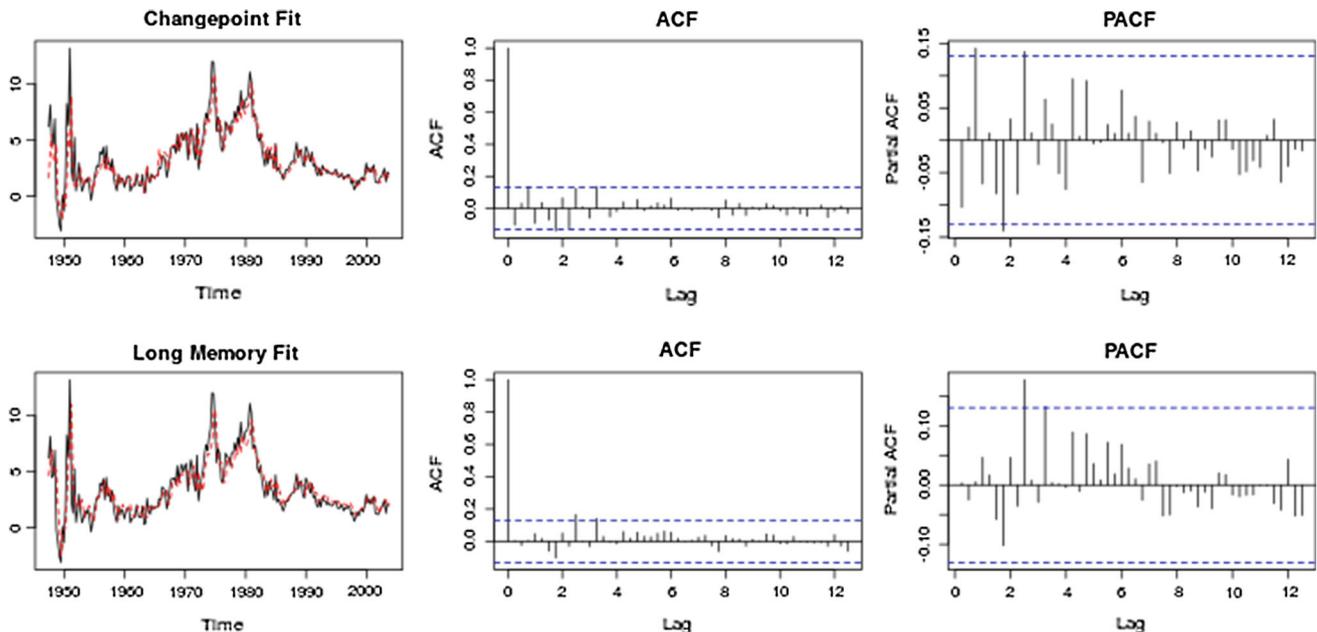

Fig. 3 Inflation diagnostics. (*Top*) *Left* Original data with fitted changepoint model; *Middle* Autocorrelation function of changepoint model residuals; *Right* Partial autocorrelations of changepoint model residuals. (*Bottom*) *Left* Original data with fitted long memory model; *Middle* Autocorrelation function of long memory model residuals; *Right* Partial autocorrelations of long memory model residuals
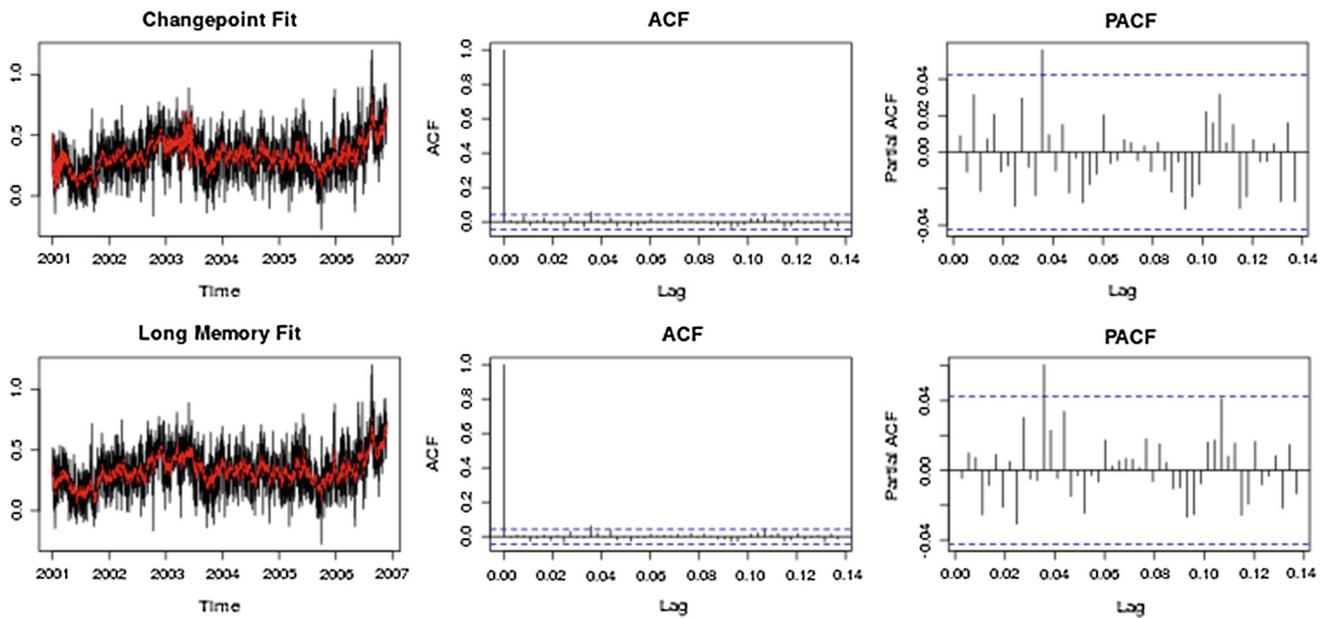
Fig. 4 Stock diagnostics. (*Top*) *Left* Original data with fitted change-point model; *Middle* Autocorrelation function of changepoint model residuals; *Right* Partial autocorrelations of changepoint model residu-als. (*Bottom*) *Left* Original data with fitted long memory model; *Middle* Autocorrelation function of long memory model residuals; *Right* Partial autocorrelations of long memory model residuals

in Table 3. It can be seen that one of the AR coefficients is close to 1 indicating an element of non-stationarity; however, we conducted a test of stationarity on this segment using the `locits` R package (Nason 2016a) which implements the test of stationarity from Nason (2013) (no rejections) and also the `fractal` R package (Constantine and Percival 2016) which implements the Priestley–Subba Rao (PSR) test (Priestley and Rao 1969) (time- varying p value 0.061). This coupled with autocorrelation and partial autocorrelation function plots given in Fig. 4 means we conclude that the segment is stationary. Here the estimated changepoints at times 715, 841, 847 and 896 correspond 15/12/2002, 20/04/2003, 26/04/2003 and 14/06/2003. The distance scores given by the classifier indicate a strong preference for long memory over changepoints. This result stands against that found in Bertram et al. (2013) which indicated a preference for a model with similarly 4 changepoints. The difference is likely due to the fact that in Bertram et al. (2013) the changepoint model does not contain any short memory dependence and we have shown here that if that short memory structure is correctly taken into account within the sub-series then the series shows greater evidence of long memory properties.

5 Conclusion

The wavelet classification process presented within this paper provides the user a distinct choice over a number of proposed

models, and when explicitly applied to an ambiguity such as long memory or a changepoint as in Sect. 3, it provides an additional piece of information to aid decision-making. The accuracy of the classifier over a number of simulated models has been presented within Sect. 3 and applied to data from the financial and economic fields in Sect. 4.

The evolutionary wavelet spectrum provides a representation of non-stationarity which is lacking in the commonly used (averaged over time) spectrum. This gives an advantage when drawing comparisons between non-stationary and stationary series, since the wavelet spectrum may appear substantially different. Quantifying this visual difference allows for a direct comparison between the series and each proposed model.

The variance-corrected squared distance metric used in the proposed classifier has been demonstrated to be quite accurate under the ambiguity of long memory and changepoint models. It is particularly effective at identifying changepoint models correctly, as the results in Table 1 demonstrate. It was noted that there is relatively lower variation between the simulations generated for the changepoint than the long memory model, which reduces the distance metric significantly even though it is variance corrected.

As mentioned in Sect. 1 there are many series that can be found in fields such as economics and finance which show evidence of the ambiguity investigated here. This classification is not intended to propose a final model for these series, but instead give additional information, treated perhaps as a

diagnostic. This could be to begin investigation of a series, or to confirm a previously found model fit. As this is not a formal test, the lack of assumptions allows for more flexibility in how the classification can be used. This work, however, is not restricted only to the ambiguity mentioned here, further work could extend it to determine between other features, such as local trends and seasonal behaviour or combining the behaviour of both models, i.e. a long memory model with a changepoint.

An aspect not covered in this paper is the precise form of ARMA and long memory models in the LSW paradigm, i.e. how the model coefficients relate to the $W_{j,k}$'s. This is an interesting area for future research which would cement the LSW model as an encompassing model but is beyond the scope of this paper.

An R package (LSWclassify) is available from the authors that implements the method from the paper.

Appendix: Proof of Theorem 1

Proof We replicate the steps of the proof within the Appendix of Fryzlewicz and Ombao (2009) up until (A.6), where following this step the short memory condition is used. To briefly summarise previous steps,

$$P(D_1 - D_2 > 0) = P(X - t > 0) \leq E(\tilde{X}^2)/t^2,$$
$$\text{(by Chebyshev's Inequality)}$$
$$E(\tilde{X}^2) =: I + II,$$
$$I \leq C J_0 J^* \sum_{j=-1}^{-J0} \sum_{i=-1}^{-J^*} 2^{i+j} E\left\{b_{i,j}^2\right\},$$
$$E\left\{b_{i,j}^2\right\} =: 2A + 2B.$$

Definitions for these components can be found in the original proof. Component A is where we alter the proof.

Recall $I_{k,N}^{(i)}$ is the wavelet periodogram at a fixed scale i, at position k with total length N, with $d_{k,N}^{(i)}$ the wavelet coefficient corresponding to it through the relationship $I_{k,N}^{(i)} = (d_{k,N}^{(i)})^2$. We continue the proof from (A.6) using the ARFIMA assumption instead. Following from above (A.6):

$$A = E\left\{\sum_{k=1}^{N}\left\{I_{k,N}^{(i)} - E\left(I_{k,N}^{(i)}\right)\right\}c_{j,k}\right\}^2$$
$$\leq 2^{2j}\sum_{k,k'=1}^{N}\left|\text{cov}\left(I_{k,N}^{(i)}, I_{k',N}^{(i)}\right)\right|$$
$$= 2^{2j}\sum_{k,k'=1}^{N} 2\text{cov}^2\left(d_{k,N}^{(i)}, d_{k',N}^{(i)}\right)$$
$$\text{(by Isserli's Theorem)} \tag{4}$$

Jensen (2000) gives bounds for the covariance of wavelet coefficients;

$$\text{cov}\left(d_{k,N}^{(m)}, d_{n,N}^{(j)}\right) = C_1|\alpha|^{2d-1-2M} + R_{2M+1}$$
$$\alpha = 2^{m-j}k - n, \quad m \geq j$$
$$|R_{2M+1}| \leq C_2|\alpha|^{2d-2-2M},$$

where $M \geq 1$ is the number of vanishing moments in the wavelet used. Using $|\alpha| = |2^{i-i}k - k'| = |k - k'| \geq 1$ and substituting into Eq. (4):

$$A = 2^{2j+1}\sum_{k,k'=1}^{N}(C_3|\alpha|^{2d-1-2M} + R_{2M+1})^2$$
$$= 2^{2j+1}\sum_{k,k'=1}^{N}|C_3|\alpha|^{2d-1-2M} + R_{2M+1}|^2$$
$$\leq 2^{2j+1}\sum_{k,k'=1}^{N}(\left|C_3|\alpha|^{2d-1-2M}\right| + |R_{2M+1}|)^2$$
$$\leq 2^{2j+1}\sum_{k,k'=1}^{N}(C_4|\alpha|^{2d-1-2M} + C_5|\alpha|^{2d-2-2M})^2$$

As $|\alpha|^{2d-2-2M} \leq |\alpha|^{2d-1-2M}$ we have:

$$A \leq 2^{2j+1}\sum_{k,k'=1}^{N}\left(C_6|k - k'|^{2d-1-2M}\right)^2$$
$$= 2^{2j+1}\sum_{k,k'=1}^{N}C_7\frac{1}{|k - k'|^{-2(2d-1-2M)}}$$
$$= 2^{2j+1}\sum_{s=1}^{N-1}(N - s)C_7\frac{1}{s^{-2(2d-1-2M)}}$$
$$= 2^{2j+1}C_7\left[N\sum_{s=1}^{N-1}\frac{1}{s^{-2(2d-1-2M)}} - \sum_{s=1}^{N-1}\frac{1}{s^{-4(d-M)+1}}\right]$$

Given that $|d| < 0.5$ and $M \geq 1$ then $4 < -2(2d - 1 - 2M) = \delta_1$ and $3 < -4(d - M) + 1 = \delta_2$. We can then

replace the sums using the definition of generalised harmonic numbers and their convergence:

$$H_{n,m} = \sum_{k=1}^{n} \frac{1}{k^m}$$
$$H_{n,m} = \mathcal{O}(1) \quad \text{as } n \to \infty \qquad (m > 1)$$

Thus

$$A \le 2^{2j+1} C_7 \left(N H_{N-1,\delta_1} - H_{N-1,\delta_2} \right) \quad = 2^{2j+1} C_7 \mathcal{H}_N,$$

where $\mathcal{H}_N = N H_{N-1,\delta_1} - H_{N-1,\delta_2}$. Returning to consider (A.4) from Fryzlewicz and Ombao (2009), we find a bound for component I, where $J_0, J^* = \log_2 N$ and $\Delta = \frac{1}{(2 \log_2 a - 1)}$:

$$
\begin{aligned}
I &= C_8 J_0 J^* \sum_{j=-1}^{-J_0} \sum_{i=-1}^{-J^*} 2^{i+j} \left[2^{2j+1} C_7 \left(N H_{N-1,\delta_1} \right. \right. \\
&\quad \left. \left. - H_{N-1,\delta_2} \right) + N^{1+\Delta} 2^j \right] \\
&= C_8 \log_2^2 N \sum_{j=-1}^{-J_0} 2^j \left[2^{2j+1} C_7 \mathcal{H}_N \right. \\
&\quad \left. + N^{1+\Delta} 2^j \right] \left(1 - 2^{J^*} \right) \\
&= C_8 \log_2^2 N \left[\sum_{j=-1}^{-J_0} C_7 2^{3j+1} \left(1 - 2^{J^*} \right) \mathcal{H}_N \right. \\
&\quad \left. + \sum_{j=-1}^{-J_0} 2^{3j} \left(1 - 2^{J^*} \right) N^{1+\Delta} \right] \\
&= C_9 \log_2^2 N \left(1 - 2^{-J^*} \right) \mathcal{H}_N \sum_{j=-1}^{-J_0} 2^{3j+1} \\
&\quad + C_8 \log_2^2 N \left(1 - 2^{-J^*} \right) N^{1+\Delta} \sum_{j=-1}^{-J_0} 2^{3j} \\
&= C_9 \log_2^2 N \left(1 - 2^{-J^*} \right) \mathcal{H}_N \frac{2}{7} \left(1 - 2^{-3J_0} \right) \\
&\quad + C_8 \log_2^2 N \left(1 - 2^{-J^*} \right) N^{1+\Delta} \frac{1}{7} \left(1 - 2^{-3J_0} \right) \\
&= \log_2^2 N \left(1 - N^{-1} \right) \left(1 - N^{-3} \right) \\
&\quad \times \left[C_{10} \mathcal{H}_N + C_{11} N^{1+\Delta} \right] \\
&= \log_2^2 N \left(1 - N^{-3} - N^{-1} + N^{-4} \right) \\
&\quad \times \left[C_{10} \mathcal{H}_N + C_{11} N^{1+\Delta} \right] \\
&\le C_{12} \log_2^2 N \left(\mathcal{H}_N + N^{1+\Delta} \right)
\end{aligned}
$$

Following this, using results in Fryzlewicz and Ombao (2009) the probability of misclassification is:

$$P(X > t) = \mathcal{O} \left(\log_2^2 N \left[N^{-1} + N^{\Delta-1} \right] \right).$$

References

Beran, J., Bhansali, R.J., Ocker, D.: On unified model selection for stationary and nonstationary short and long-memory autoregressive processes. Biometrika **85**(4), 921–934 (1998). doi:10.1093/biomet/85.4.921. http://biomet.oxfordjournals.org/content/85/4/921.abstract

Bertram, P., Kruse, R., Sibbertsen, P.: Fractional integration versus level shifts: the case of realized asset correlations. Stat. Pap. **54**(4), 977–991 (2013). https://ideas.repec.org/a/spr/stpapr/v54y2013i4p977-991.html

Chiriac, R., Voev, V.: Modelling and forecasting multivariate realized volatility. J. Appl. Econom. **26**(6), 922–947 (2011). doi:10.1002/jae.1152

Cho, H., Fryzlewicz, P.: Multiscale and multilevel technique for consistent segmentation of nonstationary time series. Statistica Sinica **22**(1), 207–229 (2012). http://www.jstor.org/stable/24310145

Constantine, W., Percival, D.: fractal: Fractal Time Series Modeling and Analysis. R package version 2.0-1 (2016). https://CRAN.R-project.org/package=fractal

Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)

Diebold, F.X., Inoue, A.: Long memory and regime switching. J. Econom. **105**(1), 131 – 159 (2001). doi:10.1016/S0304-4076(01)00073-2. http://www.sciencedirect.com/science/article/pii/S0304407601000732. Forecasting and empirical methods in finance and macroeconomics

Fryzlewicz, P., Nason, G.P.: Haar–Fisz estimation of evolutionary wavelet spectra. JRSSB **68**(4), 611–634 (2006). doi:10.1111/j.1467-9868.2006.00558.x

Fryzlewicz, P., Ombao, H.: Consistent classification of nonstationary time series using stochastic wavelet representations. J. Am. Stat. Assoc. **104**(485), 299–312 (2009). doi:10.1198/jasa.2009.0110

Grabocka, J., Nanopoulos, A., Schmidt-Thieme, L.: Invariant time-series classification. In: P.A. Flach, T. De Bie, N. Cristianini (eds.) Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol. 7524, pp. 725–740. Springer Berlin Heidelberg (2012). doi:10.1007/978-3-642-33486-3_46

Granger, C.W., Ding, Z.: Varieties of long memory models. J. Econom. **73**(1), 61–77 (1996). doi:10.1016/0304-4076(95)01733-X. http://www.sciencedirect.com/science/article/pii/030440769501733X

Granger, C.W., Hyung, N.: Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. J. Empir. Finance **11**(3), 399–421 (2004). doi:10.1016/j.jempfin.2003.03.001

Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. J. Stat. Softw. **26**(3), 1–22 (2008). http://ideas.repec.org/a/jss/jstsof/27i03.html

Janacek, G., Bagnall, A., Powell, M.: A likelihood ratio distance measure for the similarity between the fourier transform of time series. In: T. Ho, D. Cheung, H. Liu (eds.) Advances in Knowledge Discovery and Data Mining, *Lecture Notes in Computer Science*, vol. 3518, pp. 737–743. Springer Berlin Heidelberg (2005). doi:10.1007/11430919_85

Jensen, M.J.: An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. J. Econ. Dyn. Control **24**(3), 361–387 (2000). doi:10.1016/S0165-

1889(99)00010-X. http://www.sciencedirect.com/science/article/pii/S016518899900010X

Killick, R., Eckley, I.A., Jonathan, P.: A wavelet-based approach for detecting changes in second order structure within nonstationary time series. Electron. J. Stat. **7**, 1167–1183 (2013). doi:10.1214/13-EJS799

Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. **107**(500), 1590–1598 (2012). doi:10.1080/01621459.2012.737745

Krzemieniewska, K., Eckley, I., Fearnhead, P.: Classification of nonstationary time series. Stat **3**(1), 144–157 (2014). doi:10.1002/sta4.51

Levin, A.T., Piger, J.M.: Is inflation persistence intrinsic in industrial economies? Working Paper Series 0334, European Central Bank (2004). https://ideas.repec.org/p/ecb/ecbwps/20040334.html

Michie, D., Spiegelhalter, D.J., Taylor, C.C., Campbell, J.: Machine Learning, Neural and Statistical Classification. Ellis Horwood, Upper Saddle River, NJ, USA (1994)

Molz, F.J., Liu, H.H., Szulga, J.: Fractional brownian motion and fractional gaussian noise in subsurface hydrology: a review, presentation of fundamental properties, and extensions. Water Resour. Res. **33**(10), 2273–2286 (1997). doi:10.1029/97WR01982

Nason, G.: Wavelet Methods in Statistics with R. Springer Science & Business Media, Berlin (2010)

Nason, G.: A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. J. R. Stat. Soc: Ser. B (Stat. Methodol.) **75**(5), 879–904 (2013). doi:10.1111/rssb.12015

Nason, G.: locits: Tests of stationarity and localized autocovariance (2016). R package version 1.7.1

Nason, G.: wavethresh: Wavelets Statistics and Transforms (2016). https://CRAN.R-project.org/package=wavethresh. R package version 4.6.8

Pivetta, F., Reis, R.: The persistence of inflation in the united states. J. Econ. Dyn. Control **31**(4), 1326–1358 (2007)

Priestley, M.B., Rao, T.S.: A test for non-stationarity of time-series. J. R. Stat. Soc. Ser B **31**(1), 140–149 (1969). http://www.jstor.org/stable/2984336

Starica, C., Granger, C.: Nonstationarities in stock returns. Rev. Econ. Stat. **87**(3), 503–522 (2005). https://ideas.repec.org/a/tpr/restat/v87y2005i3p503-522.html

Veenstra, J.Q.: Persistence and anti-persistence: Theory and software. Ph.D. thesis, Western University (2012)

Yau, C.Y., Davis, R.A.: Likelihood inference for discriminating between long-memory and change-point models. J. Time Ser. Anal. **33**(4), 649–664 (2012). doi:10.1111/j.1467-9892.2012.00797.x