

Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery

ANNA LOUISE SCHRÖDER AND PIOTR FRYZLEWICZ*

Low-frequency financial returns can be modelled as centered around piecewise-constant trend functions which change at certain points in time. We propose a new stochastic time series framework which captures this feature. The main ingredient of our model is a hierarchically-ordered oscillatory basis of simple piecewise-constant functions. It differs from the Fourier-like bases traditionally used in time series analysis in that it is determined by change-points, and hence needs to be estimated from the data before it can be used. The resulting model enables easy simulation and provides interpretable decomposition of nonstationarity into short- and long-term components. The model permits consistent estimation of the multiscale change-point-induced basis via binary segmentation, which results in a variable-span moving-average estimator of the current trend, and allows for short-term forecasting of the average return.

KEYWORDS AND PHRASES: Financial time series, Adaptive trend estimation, Change-point detection, Binary segmentation, Unbalanced Haar wavelets, Frequency-domain modelling.

1. INTRODUCTION

In this work, we consider the problem of statistical modelling and forecasting of daily financial returns based on past observations, but the methodology we propose will also be of relevance to financial data at other frequencies. More generally, it leads to a new generic approach to statistical time series analysis, via adaptive oscillatory bases induced by change-points, which will be of interest in other fields of application beyond finance.

Given a time series P_t of daily speculative prices on risky financial instruments, such as equities, equity indices, commodities, or currency exchange rates, their daily returns X_t are defined by $X_t = \ln(P_t) - \ln(P_{t-1})$. Forecasting future values of X_t based on its own past is of major interest to quantitative finance practitioners, but presents an extremely challenging task due to the perceived low predictive content of past returns with respect to the future. The importance and difficulty of the problem have led to the

use of a large number of statistical and data analytic techniques to tackle it. In particular, we mention return forecasting based on traditional ARMA time series modelling with heteroscedastic innovations (Berkowitz and O'Brien, 2002; Garcia et al., 2005), methods stemming from the technical analysis of price series such as those based on moving average cross-overs, breakout and other technical signals (Katz and McCormick, 2000), as well as various machine-learning techniques such as those based on support vector machines (Kim, 2003) and neural networks for return forecasting (Catalão et al., 2007; Kim and Shin, 2007).

Our approach rests on the observation that the logarithmic price $\ln(P_t)$ can meaningfully and interpretably be modelled as fluctuating around a trend which started at a certain unknown time in the past, having a positive or negative linear slope. The points in time at which the slope changes will be referred to as change-points. The movements of $\ln(P_t)$ around the trend resemble random walk with heteroscedastic innovations. After differencing, this pattern translates to a piecewise-constant trend function in the return domain, plus serially uncorrelated deviations from it. Change-points in the slope of the linear trend in $\ln(P_t)$, or alternatively in the magnitude of the piecewise-constant trend in X_t , can be related to structural changes, coinciding for example with regulatory alterations, macroeconomic announcements, technological innovation or an economy's transition from recovery to recession or vice versa.

Trend detection in financial returns is a much-studied topic and a range of methods are widely applied in practice; we mention simple moving-average and other one-sided kernel smoothing of the returns, moving-average cross-overs at the level of logarithmic prices, L_2 and L_1 filtering, Kalman filtering, local polynomial smoothing, spline smoothing and nonlinear wavelet shrinkage as prime examples. These and other techniques are reviewed from a practitioner's perspective in Bruder et al. (2008); see also the references therein.

One contribution of this paper is to advocate a trend-detection methodology for financial returns that works by detecting change-points in the returns series and taking the current trend estimate to be the average return between the most recent estimated change-point and the current time. This amounts to averaging over the current estimated interval of stationarity in the conditional mean; related but different adaptive procedures for volatility (as opposed to trend) estimation appeared e.g. in Fryzlewicz, Sapatinas

*Corresponding author.

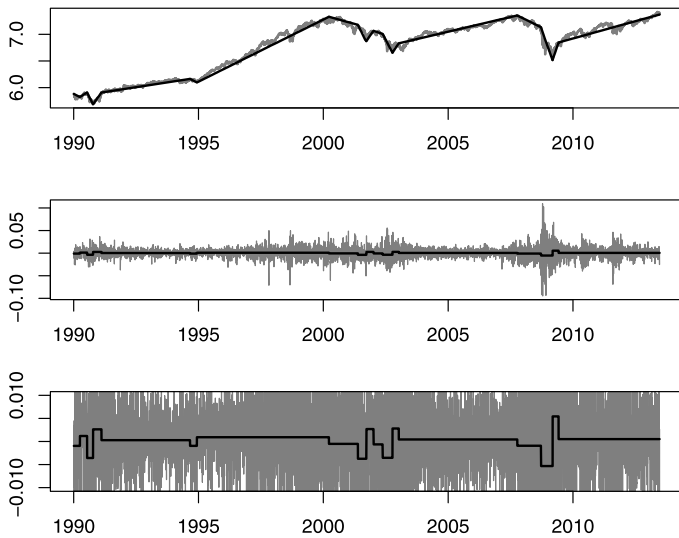


Figure 1. Daily closing values of the S&P 500 equity index between January 1990 and June 2013; top: log-price (grey) and the cumulatively integrated fit $\sum_{s=1}^t \hat{f}_s$ from our model (black), middle: log-return (grey) and our model fit \hat{f}_t (black), bottom: the same as middle but with a shorter range on the y-axis. Model fit shown for threshold parameter $C = 0.3$ and imposing a minimum change-point distance of 60 days.

and Rao (2006), Spokoiny (2009) and Čížek, Härdle and Spokoiny (2009). The first stage of our procedure is the segmentation of the returns series. Although many of the available techniques for time series segmentation could be used (Boysen et al., 2009; Lavielle and Moulines, 2000; Lebarbier, 2005; Lee, 1995; Pan and Chen, 2006; Yao and Au, 1989), we propose a modification of the Binary Segmentation procedure (Bai, 1997; Cho and Fryzlewicz, 2012; Fryzlewicz, 2012; Fryzlewicz and Subba Rao, 2013; Venkatraman, 1992; Vostrikova, 1981) and justify this choice below. An example of our model fit, using one particular value of the threshold parameter, involving daily closing values of the S&P 500 index, is in Figure 1.

However, the contribution of this work goes beyond merely advocating change-point detection as a useful approach to local trend estimation in financial returns. Our main objective is to propose a new approach to statistical time series analysis, whereby the time series, generically denoted here by X_t , is spanned by an orthonormal oscillatory basis induced by change-points in the conditional mean value of X_t . In this paper, the representation basis is assumed to be unknown to the analyst and needs to be estimated from X_t using a change-point detection procedure; hence we will occasionally refer to such a basis as ‘data-driven’ or ‘adaptive’. This in contrast to classical spectral approaches to time series analysis, which use a particular fixed basis that is known to the user, e.g. the Fourier basis in the classical spectral theory (Priestley, 1983), or a fixed

wavelet system in Nason, von Sachs and Kroisandt (2000). The SLEX approach of Ombao et al. (2002), although also using a data-driven basis principle, is fundamentally different from ours in that it models the second-, not the first-order structure of X_t and is limited to change-points occurring at dyadic locations. In financial applications, spectral analysis has occasionally been related to agents trading at different time horizons (Baron, Brogaard and Kirilenko, 2012; Gençay, Selçuk and Whitcher, 2001; Hasbrouck and Sofianos, 1993), although this point of view is of no primary relevance to us.

The building blocks used in our basis construction are the Unbalanced Haar (UH) wavelet vectors (Baek and Pipiras, 2009; Delouille, Franke and von Sachs, 2001; Fryzlewicz, 2007; Girardi and Sweldens, 1997; Timmermans, Delsol and von Sachs, 2012), which have the advantage of being particularly simple, well-suited to the task of change-point modelling, and hierarchically organised into a multiscale system, which is useful for the interpretability of the estimated change-point locations and basis vectors, and facilitates their arrangement according to their importance. We use the UH basis vectors to define the Unbalanced Haar time series model. Binary Segmentation is a natural tool for the estimation of UH basis vectors from the data due to the hierarchical structure of this procedure (Fryzlewicz, 2007), although we emphasise that other change-point detection methods could also be used for this purpose.

Our new adaptive basis approach to time series modelling opens up many interesting avenues. It leads to a formal stochastic time series model, used here to model local trends in financial returns but also applicable more widely, with interpretable nonstationarities in the autocovariance structure and in the conditional mean of X_t . It allows for a decomposition of the nonstationarity in the variance of X_t into longer-term trends and short-term outbursts of volatility. Finally, it yields a family of forecasting operators for X_t , parameterised by a single threshold parameter, which can be adjusted flexibly depending on the forecast horizon or the error criterion.

The paper is structured as follows. Section 2 motivates and defines the model, and studies its probabilistic properties. Section 3 describes the methodology and theory of change-point detection and basis recovery, as well as the implied methodology for current trend estimation and forecasting. Section 4 illustrates basis recovery in a numerical study. Section 5 shows the estimated bases for data examples from various asset classes and performs a forecasting competition between our method and the benchmark moving window approach. Proofs are in the appendix.

2. THE MODEL

2.1 Motivation and basic ingredients

As illustrated in Figure 1, piecewise-linear modelling of trends in $\ln(P_t)$ results in the average of the returns series

X_t oscillating around zero in a piecewise-constant fashion. We wish to embed this feature into a rigorous stochastic framework by formulating a time series model for X_t that captures this oscillatory behaviour.

Time series modelling using oscillatory building blocks is a well-established technique in time series analysis. Turning first to the frequency domain, every covariance-stationary process Y_t admits the Fourier representation

$$(1) \quad Y_t = \int_{(-\pi, \pi]} A(\omega) \exp(i\omega t) dZ(\omega), \quad t \in \mathbb{Z},$$

where $\exp(i\omega t)$ is a complex exponential oscillating at frequency ω , and $A(\omega)$ and $dZ(\omega)$ are, respectively, the associated amplitude and the corresponding orthonormal infinitesimal increment of a stochastic process $Z(\cdot)$. The support of the complex exponential $\exp(i\omega t)$ is the set of all integers t and its oscillations exhibit a homogeneous behaviour over this domain, for all ω . To facilitate similar representations for covariance-nonstationary processes Y_t , efforts have been made to localise the oscillatory building blocks with respect to time. For example, Nason, von Sachs and Kroisandt (2000) propose the Locally Stationary Wavelet model

$$(2) \quad Y_t = \sum_{j=1}^{\infty} \sum_{k \in \mathbb{Z}} w_{j,k} \psi_{t-k}^j \xi_{j,k}, \quad t \in [1, T],$$

where j is the scale parameter, analogous to frequency in (1), and ψ^j are compactly-supported wavelet vectors, oscillatory in the sense that $\sum_u \psi_u^j = 0$, and such that the length of their support increases, but the speed of their oscillation decreases, with j . The parameters $w_{j,k}$ are amplitudes, localised over time-location k , and $\xi_{j,k}$ are mutually uncorrelated increments. We refer the reader to Vidakovic (2009) and Nason (2008) for overviews of the use of wavelets in statistical modelling.

The above two approaches could be used to model returns processes X_t such as that illustrated in Figure 1, but are not ideal. Our signal of interest, the piecewise-constant average return, is blocky, and oscillates around zero in an inhomogeneous fashion in the sense that there is significant variation in the lengths of the constant intervals. However, with the exception of Haar wavelets, which we discuss below, both complex exponentials and wavelet vectors arise from continuous functions, and so are ill-suited for our purpose of modelling piecewise constancy. Also, crucially, both of the above approaches use bases which are not data-adaptive in the sense that they are fixed before the analysis rather than being tailored to, or estimated from, the data. In particular, one could possibly entertain the thought of using the piecewise-constant Haar wavelets for our purpose, but they would only permit change-points at dyadic locations $kT2^{-j}$, $j = 1, 2, \dots$, $k = 1, \dots, 2^j - 1$, where T is the sample size.

In our data, change-points occur at arbitrary locations and we hope to be able to capture this feature by the use

of a suitably flexible oscillatory basis that permits adaptive choice of change-points in the basis vectors, allowing for a sparse representation of the piecewise-constant trend. Arguably the simplest such construction is furnished by the Unbalanced Haar (UH) wavelets. With $\mathbb{I}(\cdot)$ denoting the indicator function, the generic UH vector $\psi^{s,b,e}$ is defined as

$$\psi_t^{s,b,e} = \begin{cases} \frac{1}{b-s+1} - \frac{1}{e-s+1} & \mathbb{I}(s \leq t \leq b) \\ -\left\{ \frac{1}{e-b} - \frac{1}{e-s+1} \right\} & \mathbb{I}(b+1 \leq t \leq e), \end{cases}$$

where s and e are, respectively, the start- and end-point of its support, and b is the location of a change-point. $\psi^{s,b,e}$ is constant and positive before the change-point, constant and negative after the change-point, and such that $\sum_t \psi_t^{s,b,e} = 0$ and $\sum_t (\psi_t^{s,b,e})^2 = 1$. A set of $T-1$ UH vectors plus one constant vector constitutes an orthonormal basis of \mathbb{R}^T if it is constructed as follows: define the first UH basis vector $\psi^{b_{0,1}} = \psi^{1,b_{0,1},T}$; the change-point $b_{0,1}$ needs to be chosen and we later say how. Then, recursively repeat this construction on the two parts of the domain determined by $b_{0,1}$: that is, provided that $b_{0,1} \geq 2$, define $\psi^{b_{1,1}} = \psi^{1,b_{1,1},b_{0,1}}$, and provided that $T - b_{0,1} \geq 2$, define $\psi^{b_{1,2}} = \psi^{b_{0,1}+1,b_{1,2},T}$. The recursion then continues in the same manner for as long as feasible, with each vector $\psi^{b_{j,k}}$ having at most two children vectors $\psi^{b_{j+1,2k-1}}$ and $\psi^{b_{j+1,2k}}$. Additionally, we define a vector $\psi^{b_{-1,0}}$ with elements $\psi_l^{b_{-1,0}} = T^{-1/2} \mathbb{I}(1 \leq l \leq T)$. The indices j, k are scale and location parameters, respectively. Small and large values of j can be thought of as corresponding to coarse and fine scales, respectively, as in the classical wavelet theory.

Example We consider an example of a set of UH vectors for $T = 6$. The rows of the matrix \mathbf{W} defined below contain (from top to bottom) vectors $\psi^{b_{-1,0}}$, $\psi^{b_{0,1}}$, $\psi^{b_{1,2}}$, $\psi^{b_{2,3}}$, $\psi^{b_{2,4}}$ and $\psi^{b_{3,7}}$ determined by the following set of change-points: $(b_{0,1}, b_{1,2}, b_{2,3}, b_{2,4}, b_{3,7}) = (1, 3, 2, 5, 4)$.

$$\mathbf{W} = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ \sqrt{5}/6 & -1/\sqrt{30} & -1/\sqrt{30} & -1/\sqrt{30} & -1/\sqrt{30} & -1/\sqrt{30} \\ 0 & \sqrt{3}/10 & \sqrt{3}/10 & -\sqrt{2}/15 & -\sqrt{2}/15 & -\sqrt{2}/15 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{2}/3 \\ 0 & 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}$$

In the above example it is not possible to create further vectors $\psi^{b_{j,k}}$.

2.2 Definition and examples

Motivated by the above discussion, our model for the returns series X_t is defined as follows.

Definition 2.1. A stochastic process X_t , $t = 1, 2, \dots, T$ is called the *Unbalanced Haar* process if it has a representation

$$(3) \quad X_t = T^{1/2} \sum_{(j,k) \in \mathcal{I}} A_{j,k} \psi_t^{b_{j,k}} + \sigma_t \varepsilon_t, \quad t \in [1, T],$$

where \mathcal{I} is a set of indices, of finite dimensionality $|\mathcal{I}| = N+1 < \infty$, such that $(-1, 0) \in \mathcal{I}$, and connected in the sense that if a child index $(j+1, 2k-1)$ or $(j+1, 2k)$ is in \mathcal{I} , then so is their parent (j, k) . The random variables $\{A_{j,k}\}_{(j,k) \in \mathcal{I}}$ are mutually independent, drawn from continuous distributions and satisfy $E(A_{j,k}) = 0$ and $E(A_{j,k}^2) < \infty$. The vectors $\psi^{b_{j,k}}$ are UH vectors defined in Section 2.1. The constants σ_t are such that $0 < \underline{\sigma} < \sigma_t < \bar{\sigma} < \infty$, and $\{\varepsilon_t\}_t$ is a sequence of independent standard normal variables, also independent of $A_{j,k}$.

Denote the elements of the sequence $\{b_{j,k}\}_{(j,k)}$, sorted in increasing order, by $\{\eta_i\}_{i=1}^N$. We assume that the parameters η_i are fixed in rescaled time in the sense that for each i , we have $\eta_i = \lfloor T v_i \rfloor$ and $\{v_i\}_{i=1}^N$ is an increasing sequence of constants in $(0, 1)$. For completeness, denote $\eta_0 = 1$, $v_0 = 0$, $\eta_{N+1} = T$, $v_{N+1} = 1$.

The model in (3) contains two additive parts, interpretable, respectively, as signal (or trend) and noise. The signal, $T^{1/2} \sum_{(j,k) \in \mathcal{I}} A_{j,k} \psi_t^{b_{j,k}}$, is designed to model the piecewise-constant average return and provides a multiscale representation of this quantity in terms of the basis functions $\psi^{b_{j,k}}$. Heuristically speaking, it is composed of a constant vector $\psi^{b_{-1,1}}$ plus N UH vectors $\psi^{b_{j,k}}$, each multiplied by its own independent amplitude $A_{j,k}$. This mimics the construction used in (1) and (2), in which the process in question is also composed of oscillatory vectors at different frequencies or scales, with random amplitudes. The fact that \mathcal{I} is connected leads to $T^{1/2} \sum_{(j,k) \in \mathcal{I}} A_{j,k} \psi_t^{b_{j,k}}$ being a random, piecewise-constant signal with N change-points located at $\{\eta_i\}_{i=1}^N$. As in (1), the basis vectors $\psi^{b_{j,k}}$ also change with the sample size, but our notation does not reflect this for simplicity. The factor $T^{1/2}$ is required to keep the scale of the amplitudes $A_{j,k}$ constant with respect to T .

The noise, $\sigma_t \varepsilon_t$, models the random movements of X_t around the trend and, for technical simplicity, is assumed to be Gaussian. In this work, we do not dwell on the issue of estimating the volatility parameters σ_t , but treat them as constant in our theoretical considerations. Naturally, in practice, they need to be estimated from the data; we later specify what estimators we use.

Model (3) enables easy simulation of sample paths of X_t . For example, the simulated sample paths of X_t from model (3), displayed in Figure 2, use the canonical basis (see Section 3.2 for details) and σ_t estimated from the real data example from Figure 1, and the coefficients $A_{j,k}$ drawn from the normal distribution with mean zero and variances matching the corresponding empirical variances from the data.

2.3 Unconditional properties of the model

We start by exhibiting and discussing some simple unconditional probabilistic properties of X_t defined by (3). We use the term unconditional to mean that we do not condition on particular values of the random coefficients $A_{j,k}$. Were we

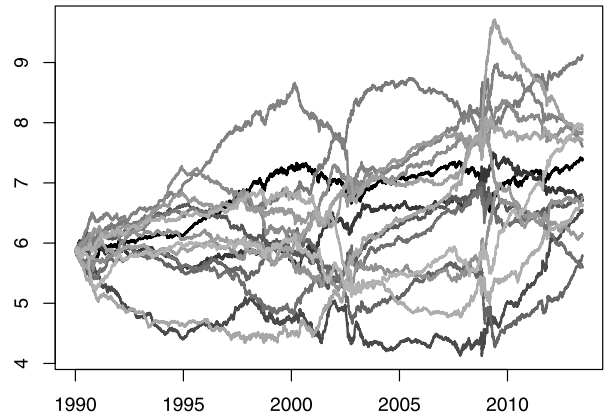


Figure 2. Daily closing values of the S&P 500 equity index (log-price, black line) between January 1990 and June 2013 and simulated sample paths (grey lines) from the estimated canonical basis $\{\hat{b}_{j,k}\}_{(j,k)}$; coefficients $A_{j,k}$ are drawn from the normal distribution with mean zero and variances matching the corresponding empirical variances, σ_t is estimated from the data. Model estimated with a threshold parameter $C = 0.3$ and a minimum change-point distance of 60 days.

to do this, our analysis would amount to considering a particular, rather than random, piecewise-constant signal plus noise, a set-up which we investigate in Section 2.4 below.

$E(A_{j,k}) = 0$ and $E(\varepsilon_t) = 0$ imply $E(X_t) = 0$. The variance and autocovariance of X_t admit the following decompositions:

$$(4) \quad \text{Var}(X_t) = T \sum_{(j,k) \in \mathcal{I}} E(A_{j,k}^2) (\psi_t^{b_{j,k}})^2 + \sigma_t^2$$

$$(5) \quad \text{Cov}(X_t, X_{t+\tau}) = T \sum_{(j,k) \in \mathcal{I}} E(A_{j,k}^2) \psi_t^{b_{j,k}} \psi_{t+\tau}^{b_{j,k}}, \tau \neq 0.$$

A few remarks are in order.

Stationarity Clearly, X_t is stationary in the mean, but is variance- and covariance-nonstationary. The nonstationarity in the variance arises not just because of the σ_t^2 term, but also because of the term $T \sum_{(j,k) \in \mathcal{I}} E(A_{j,k}^2) (\psi_t^{b_{j,k}})^2$, which captures the variability of longer-term trends, as opposed to the daily variability captured by σ_t^2 . The daily variability σ_t could also, in principle, be modelled as a GARCH-type process with a conditional, rather than unconditional, heteroscedasticity, but we do not pursue this option in this work because of the technical requirement of the boundedness of σ_t from above.

Variance Firstly, we observe that $\psi_t^{b_{j,k}} = 0$ if t is outside the support of $\psi^{b_{j,k}}$, and the values of $\psi_t^{b_{j,k}}$ differ depending on whether t falls within the positive or the negative part of the support of $\psi^{b_{j,k}}$; the shorter the relevant part of the support, the higher the value of $(\psi_t^{b_{j,k}})^2$, and hence formula (4)

indicates that shorter trends contribute more to the variability of X_t . In simple heuristic terms, this can be interpreted as shorter-term trends being more variable. Secondly, the average variance admits the following decomposition:

$$T^{-1} \sum_{t=1}^T \text{Var}(X_t) = \sum_{(j,k) \in \mathcal{I}} E(A_{j,k}^2) + T^{-1} \sum_{t=1}^T \sigma_t^2,$$

due to the fact that $\sum_t (\psi_t^{b_{j,k}})^2 = 1$. Therefore, the average variance of the signal has a simple representation in terms of the variances of the amplitudes $A_{j,k}$.

Autocovariance If both t and $t + \tau$ are within the same, positive or negative, part of the support of $\psi^{b_{j,k}}$, then $\psi_t^{b_{j,k}} \psi_{t+\tau}^{b_{j,k}}$ reduces to $(\psi_t^{b_{j,k}})^2$, and by formulae (4) and (5), the contribution of the term indexed (j, k) to $\text{Var}(X_t)$ and $\text{Cov}(X_t, X_{t+\tau})$ is the same and equal to $E(A_{j,k}^2)(\psi_t^{b_{j,k}})^2$. On the other hand, if t and $t + \tau$ are within the support of $\psi^{b_{j,k}}$ but on two different sides of the change-point $b_{j,k}$, then the term $E(A_{j,k}^2) \psi_t^{b_{j,k}} \psi_{t+\tau}^{b_{j,k}}$ contributes negatively to $\text{Cov}(X_t, X_{t+\tau})$, which reflects the fact that t and $t + \tau$ belong to two opposing trends at scale j .

Unbalanced Haar spectrum By formulae (4) and (5), the term $T \sum_{(j,k) \in \mathcal{I}} E(A_{j,k}^2) (\psi_t^{b_{j,k}})^2$ in the variance of X_t as well as the autocovariance of X_t have a representation in terms of $E(A_{j,k}^2)_{(j,k) \in \mathcal{I}}$. Therefore, it is natural to introduce a separate definition for this quantity.

Definition 2.2. Let X_t follow model (3). The sequence $E(A_{j,k}^2)_{(j,k) \in \mathcal{I}}$ is referred to as the *Unbalanced Haar spectrum* of X_t with respect to the basis $\psi^{b_{j,k}}$.

We emphasise that the Unbalanced Haar spectrum is defined with respect to a fixed basis $\psi^{b_{j,k}}$; a different basis for the same process X_t would result in a different Unbalanced Haar spectrum. In particular, if the Unbalanced Haar spectral approach is used to compare two or more time series, a common basis should be used in order for the comparison to be meaningful, unless it is of interest to study feature misalignment between the different series. We do not pursue multivariate Unbalanced Haar processes in this work. For univariate Unbalanced Haar processes, we discuss the important issue of basis selection later on.

The Unbalanced Haar spectrum is an analogue of the spectral density in the Fourier representation (1) and the wavelet spectrum in the LSW model (2) in the sense that it also arises as the sequence of variances of the random amplitudes associated with the oscillatory building blocks used in the construction of the process. However, the role of the UH building blocks in model (3) is subtly different from those of the Fourier exponentials in (1) and wavelets in (2): the latter two directly model the contribution of the corresponding oscillations to the autocovariance structures of the respective processes, while our primary aim in using the UH building blocks is to model the piecewise-constant

trend in X_t rather than represent the autocovariance of X_t . The fact that the autocovariance structure of X_t also has a representation in terms of $\psi^{b_{j,k}}$ is nothing but a useful and interpretable by-product of this modelling approach.

If the basis $\psi^{b_{j,k}}$ is known to the analyst, then the Unbalanced Haar periodogram, defined below, provides an asymptotically unbiased but inconsistent estimate of the Unbalanced Haar spectrum, in the same way as the classical periodogram is asymptotically unbiased but inconsistent for the spectral density in the classical Fourier theory. The case of an unknown basis is more delicate and will be discussed in more detail in Sections 2.4 and 3.2, as will be the issue of restoring, in a certain sense, the consistency of spectral estimation in model (3). In the remainder of the paper, $\langle \cdot, \cdot \rangle$ denotes inner product between two vectors.

Definition 2.3. Let X_t follow model (3) and denote $X = (X_1, \dots, X_T)$. The sequence of statistics defined by

$$(6) \quad I_{j,k} = T^{-1} \langle X, \psi^{b_{j,k}} \rangle^2 = T^{-1} \left(\sum_{t=1}^T X_t \psi_t^{b_{j,k}} \right)^2$$

is called the *Unbalanced Haar periodogram* of X_t with respect to the basis $\psi^{b_{j,k}}$.

The following result quantifies the asymptotic unbiasedness of $I_{j,k}$ for $E(A_{j,k}^2)$.

Proposition 2.1. *We have*

$$|E(I_{j,k}) - E(A_{j,k}^2)| \leq \bar{\sigma}^2 T^{-1}.$$

2.4 Properties of the model conditional on $A_{j,k}$

Having observed a sample path of X_t , it is of interest to establish the number and locations of change-points in the observed signal $T^{1/2} \sum_{(j,k) \in \mathcal{I}} A_{j,k} \psi_t^{b_{j,k}}$. From the point of view of this task, it is helpful to treat $A_{j,k}$ as already observed, and therefore carry out the change-point analysis conditioning on their values $A_{j,k} = a_{j,k}$, for all $(j, k) \in \mathcal{I}$. Define $f_t = T^{1/2} \sum_{(j,k) \in \mathcal{I}} a_{j,k} \psi_t^{b_{j,k}}$, suppressing the dependence of f_t on $a_{j,k}$ and $b_{j,k}$, for simplicity of notation. In fact, neither $a_{j,k}$ nor $b_{j,k}$ are observed directly, but instead we observe a noisy version of f_t , that is

$$(7) \quad X_t = f_t + \sigma_t \varepsilon_t,$$

as in model (3) conditional on $A_{j,k} = a_{j,k}$. We now briefly discuss some properties of f_t . Let P^A be the probability measure induced by the random variables $A_{j,k}, (j, k) \in \mathcal{I}$. The signal f_t satisfies the following properties.

- Property 2.1.** (i) f_t is piecewise-constant with at most N change-points, and bounded.
(ii) The magnitude of the i th change-point in f_t is of the form $|\sum_{j,k} \alpha_{j,k,i} a_{j,k}|$, where $\alpha_{j,k,i}$ are scalars.

- (iii) Let A_0 be the set of $N + 1$ -tuples of those values of $a_{j,k}$, $(j, k) \in \mathcal{I}$, for which f_t has fewer than N change-points. We have $P^A(A_0) = 0$.
- (iv) With probability one with respect to P^A , change-points in f_t are located at $b_{j,k}$; sorted in increasing order, their locations are $\{\eta_i\}_{i=1}^N$.

Naturally, conditioning on $A_{j,k}$ changes the first- and second-order properties of X_t . From (7), we have $E(X_t|A_{j,k} = a_{j,k}) = f_t$, $\text{Var}(X_t|A_{j,k} = a_{j,k}) = \sigma_t^2$ and $\text{Cov}(X_t, X_{t+\tau}|A_{j,k} = a_{j,k}) = 0$ for $\tau \neq 0$.

The following section discusses the estimation of various aspects of the conditional signal f_t .

3. ESTIMATION AND FORECASTING

3.1 Change-point detection

Given observations from the conditional model (7), it is of interest to estimate the number N and locations $\{\eta_i\}_{i=1}^N$ of change-points in f_t . Multiple change-point detection has been widely studied in literature, and some of the existing techniques reviewed in Section 1 could be used for this purpose, possibly with some modifications. We propose the use of the Binary Segmentation method, mainly because it allows for simultaneous estimation of the canonical basis, which is discussed in Section 3.2 below. Another benefit of Binary Segmentation is rate optimality for the estimators of change-point locations in the particular setting of the spacing between change-points being of order $O(T)$, as is the case in our model. The Binary Segmentation algorithm is best defined recursively and hence described by pseudocode. We first define

$$\tilde{X}_{s,e}^b = \langle X, \psi^{s,b,e} \rangle$$

$$\tilde{\sigma}_s^e = \left(\frac{1}{e-s+1} \sum_{t=s}^e \sigma_t^2 \right)^{1/2}.$$

The main function is defined as follows.

```

function BINSEG( $s, e, \zeta_T$ )
  if  $e - s < 1$  then
    STOP
  else
     $b_0 := \arg \max_b |\tilde{X}_{s,e}^b|$ 
    if  $|\tilde{X}_{s,e}^{b_0}| / \tilde{\sigma}_s^e > \zeta_T$  then
      add  $b_0$  to the set of estimated change-points
      BINSEG( $s, b_0, \zeta_T$ )
      BINSEG( $b_0 + 1, e, \zeta_T$ )
    else
      STOP
    end if
  end if
end function

```

The standard Binary Segmentation procedure is launched by the call $\text{BINSEG}(1, T, \zeta_T)$, where ζ_T is a threshold parameter. We have the following result regarding the consistency

of the Binary Segmentation procedure for the number and locations of change-points.

Theorem 3.1. *Let f_t in (7) be constructed with parameters $a_{j,k}$ lying outside set A_0 from Property 2.1. Let N and η_1, \dots, η_N denote, respectively, the number and locations of change-points in f_t . Let \hat{N} denote the number, and $\hat{\eta}_1, \dots, \hat{\eta}_N$ the locations, sorted in increasing order, of the change-point estimates obtained by the Binary Segmentation algorithm. Let the threshold parameter satisfy $\zeta_T \geq c_1 \log^p T$ ($p > 1/2$) and $\zeta_T \leq c_2 T^\theta$ ($\theta < 1/2$), for any positive constants c_1, c_2 . Then there exists a positive constant C_1 such that $P(\mathcal{A}_T) \rightarrow 1$, where*

$$\mathcal{A}_T = \{ \hat{N} = N; \max_{i=1, \dots, N} |\hat{\eta}_i - \eta_i| \leq C_1 \epsilon_T \}$$

with $\epsilon_T = O(\log T)$, where $P(\cdot)$ is the probability measure induced by $\{\epsilon_t\}_t$.

3.2 Basis recovery

Given any piecewise-constant signal f_t with $N > 1$ change-points at $\{\eta_i\}_{i=1}^N$, there is no unique way to represent it in a UH basis $\psi^{b_{j,k}}$, as any of the N change-points can be assigned to scale $j = 0$; that is, there are N mutually exclusive possibilities: $b_{0,1} = \eta_1, b_{0,1} = \eta_2, \dots, b_{0,1} = \eta_N$. If $N > 2$, then there is similar choice at the following scales $j = 1, 2, \dots$.

Some of the possible UH bases are more interpretable and useful than others. Here, we define and focus on one particular type of UH basis, termed the *canonical UH basis*, which enables the partial ordering of the change-points $\{\eta_i\}_i$ according to what can be interpreted as their importance.

Recall the construction of any UH basis, described in Section 2.1. Given a current interval of interest $[s, e]$, a basis selection procedure will be completely specified if one specifies how to choose $b_{j,k}$ on that interval. The canonical basis is defined as follows.

Definition 3.1. Given a piecewise-constant signal f_t , $t = 1, \dots, T$ with $N > 0$ change-points at η_1, \dots, η_N , a *canonical UH basis* is one for which $b_{j,k}$ on the current interval $[s, e]$ (on which f_t is non-constant) is assigned as

$$b_{j,k} = \arg \max_{\eta_i \in [s,e], i=1, \dots, N} |\langle f, \psi^{s, \eta_i, e} \rangle|,$$

where $f = (f_1, \dots, f_T)$. If there is more than one such bases, any of them is referred to as canonical.

In other words, applying Definition 3.1 sequentially, a canonical basis for f_t is chosen as follows. At scale 0, $b_{0,1}$ is chosen as the η_i that maximises $|\langle f, \psi^{1, \eta_i, T} \rangle|$; that is, the η_i which defines the step function with one change-point that fits f_t best in the least-squares sense. Such η_i is not necessarily unique; however, ties are not an issue for us as we clarify further below. Having identified $b_{0,1}$, canonical basis selection then proceeds sequentially as described in

Section 2.1, at each scale fitting the best approximation in the L_2 sense to f_t on the relevant interval $[s, e]$ by means of a step function with one change-point, until all change-points in f_t have been accounted for.

We now give a result that specifies the uniqueness of a canonical basis for f_t from model (7).

Proposition 3.1. *There is a set B_0 of $N + 1$ -tuples $a_{j,k}, (j, k) \in \mathcal{I}$ with $P^A(B_0) = 0$ such that for $\{a_{j,k}\}_{j,k}$ from outside B_0 , f_t from model (7) has a unique canonical basis.*

Summarising the above discussion, for a signal f_t defined by model (7), there exists a unique canonical UH basis for it with probability 1 with respect to P^A . The unique canonical basis partially orders the change-points in f_t according to their importance in terms of explaining f_t in the L_2 sense. The change-point $b_{0,1}$ can be interpreted as the most important, with change-points as subsequent finer scales being interpretable as gradually less important.

In this setting, we now state that it is possible to use Binary Segmentation to reconstruct the canonical basis $b_{j,k}$ of f_t from the noisy observations defined by model (7), as well as the corresponding canonical basis coefficients $\bar{a}_{j,k}$. The latter fact can be interpreted as the conditional consistency of spectral estimation in model (3) with respect to each unique canonical basis.

Theorem 3.2. *Let f_t in (7) be constructed with parameters $a_{j,k}$ lying outside set $A_0 \cup B_0$. Let N and η_1, \dots, η_N denote, respectively, the number and locations of change-points in f_t , and let $b_{j,k}$ define the canonical basis of f_t , ordered according to increasing j , and with the $b_{j,k}$'s within each scale j sorted in increasing order. Let $\bar{a}_{j,k}$ be the UH coefficients with respect to $b_{j,k}$; that is, $\bar{a}_{j,k} = T^{-1/2} \langle f, \psi^{b_{j,k}} \rangle$. Let \hat{N} denote the number, and $\hat{\eta}_1, \dots, \hat{\eta}_N$ the locations, sorted in increasing order, of the change-point estimates obtained by the Binary Segmentation algorithm, and let $\hat{b}_{j,k}$ be the estimated change-points in the order returned by the Binary Segmentation algorithm. Let $\hat{a}_{j,k} = T^{-1/2} \langle X, \psi^{\hat{b}_{j,k}} \rangle$. Let the threshold parameter satisfy $\zeta_T \geq c_1 \log^p T$ ($p > 1/2$) and $\zeta_T \leq c_2 T^\theta$ ($\theta < 1/2$), for any positive constants c_1, c_2 . Then there exist positive constants C_1, C_2 such that $P(\mathcal{A}_T) \rightarrow 1$, where*

$$\mathcal{A}_T = \{ \hat{N} = N; \quad \max_{(j,k) \in \mathcal{I}} |\hat{b}_{j,k} - b_{j,k}| \leq C_1 \epsilon_T; \\ \max_{(j,k) \in \mathcal{I}} |\hat{a}_{j,k} - \bar{a}_{j,k}| \leq C_2 T^{-1/2} \log^{1/2} T \}$$

with $\epsilon_T = O(\log T)$.

The result of Theorem 3.2 goes one step further than the change-point detection result of Theorem 3.1. It states that it is not only possible to detect the number and location of the change-points, but also their importance as defined by the partial ordering specified by the canonical basis, as well as the coefficient values with respect to that basis. Another way of viewing this result is that the canonical basis $\psi^{b_{j,k}}$

can be estimated from the noisy observations X_t . The implication of this result is that Binary Segmentation can be used to estimate not only the change-points in f_t , but also their relative importance with respect to each other. This is attractive from the point of view of both the interpretability of the detected change-points, and their use in forecasting, as we demonstrate further below.

We end this section by remarking that although it is tempting to attempt to define a ‘‘canonical UH spectrum’’ in model (3), this does not appear to be a straightforward task. The reason for this is that the canonical basis is defined for each f_t separately, and therefore two different sets of $\{a_{j,k}\}$, leading to two different realisations of f_t , can result in two different canonical bases. Therefore, there is no such thing as a single specific canonical basis for the random generator $T^{1/2} \sum_{(j,k) \in \mathcal{I}} A_{j,k} \psi_t^{b_{j,k}}$ of signal f_t . Hence, defining a canonical UH spectrum, being the set of variances of $A_{j,k}$ with respect to a canonical basis, is not particularly obvious here.

3.3 Forecasting

We define our forecasting task in model (3) as follows. Having observed X_1, \dots, X_T , we are interested in predicting the value of f_{T+h} , the conditional mean of X_{T+h} , where $h \geq 1$ is the forecasting horizon. However, we note that if h is small with respect to T , which we formally quantify as $h = o(T)$, then our model guarantees that $f_{T+h} = f_T$. This is because the change-points locations η_i in our model satisfy $\eta_i = \lfloor T v_i \rfloor$ where the v_i 's are constants, so the next change-point after η_N is not expected until $T(1 + \delta) > T + h$, with δ being a positive constant. Therefore, the task of forecasting f_{T+h} in model (3) is equivalent to the task of estimating f_T .

The condition $f_T = f_{T+h}$ for $h = o(T)$ carries an implicit assumption that the asset modelled by X_t is of a *trend following* type: we assume that the current average return f_T will not change in the near future. However, our framework may also be a useful starting point for the modelling and forecasting of *mean-reverting* assets, in which f_{T+h} is likely to be different from f_T ; perhaps negative if $f_T > 0$, or vice versa. In the latter case, one possibility is to assume that $f_{T+h} = g(f_T)$ where g is possibly different from identity and needs to be estimated from the data. Although rigorous treatment of this case is beyond the scope of this work as it is not technically covered by our modelling framework, we discuss it from the practical point of view in Section 5.

Representing f_t in its canonical basis, we obtain

$$f_t = T^{1/2} \sum_{(j,k) \in \mathcal{I}} \bar{a}_{j,k} \psi_t^{b_{j,k}} = \sum_{(j,k) \in \mathcal{I}} \langle f, \psi^{b_{j,k}} \rangle \psi_t^{b_{j,k}}.$$

A natural estimator for f_t is

$$\hat{f}_t = T^{1/2} \sum_{(j,k) \in \hat{\mathcal{I}}} \hat{a}_{j,k} \psi_t^{\hat{b}_{j,k}} = \sum_{(j,k) \in \hat{\mathcal{I}}} \langle X, \psi^{\hat{b}_{j,k}} \rangle \psi_t^{\hat{b}_{j,k}},$$

where $\hat{\mathcal{I}}$ is the set of estimated indices of the estimated change-points; note that by Theorem 3.2, we have $\hat{\mathcal{I}} = \mathcal{I}$ with high probability. Therefore, \hat{f}_t is the orthogonal projection of the data X onto the space spanned by the estimated canonical UH vectors $\psi^{\hat{b}_{j,k}}$. Hence, \hat{f}_T reduces to

$$\hat{f}_T = \frac{1}{T - \hat{\eta}_N} \sum_{t=\hat{\eta}_N+1}^T X_t,$$

where $\hat{\eta}_N$ is the most recent estimated change-point. Thus, \hat{f}_T can be interpreted as an adaptive, as opposed to fixed-span, moving average of the recent values of X_t , where the adaptation is with respect to the estimated change-point structure in the data.

Our estimation procedure is parameterised by the threshold parameter ζ_T . The permitted theoretical range of ζ_T is specified in Theorem 3.2, and the practical choice of the constants in ζ_T is discussed in Section 5. The lower the value of ζ_T , the later the Binary Segmentation procedure is likely to stop, and therefore the closer to T the final detected change-point $\hat{\eta}_N$ is likely to lie. Therefore, lower (higher) values of ζ_T are likely to lead to shorter (longer) average spans $T - \hat{\eta}_N$. In other words, recalling the importance interpretation of the detected change-points discussed earlier, higher values of ζ_T lead to forecasts based on more important detected change-points, whereas the lower the value of ζ_T , the higher the chance of basing the forecasts on less important detected change-points.

4. SIMULATION STUDY

In this section, we briefly exhibit the change-point detection and canonical basis recovery capabilities of our Binary Segmentation algorithm. We generate trends $\{f_t\}_{t=1}^T$, with $T = 1,000$, as follows. For each trend, the number N of change-points is drawn from the Poisson(5) distribution and their locations η_i are drawn uniformly on $[1, T]$. The jump sizes are simulated independently as $N(0, V_{\eta_i})$. We repeat the trend generation 100 times for $V_{\eta_i} = 1$ and 100 times for $V_{\eta_i} = 2$. We denote the canonical basis of f_t by $\{b_{j,k}\}_{(j,k)}$. For each trend f_t , we simulate 1,000 sample paths of $X_t = f_t + \sigma_t \varepsilon_t$ with $\sigma_t = 1$. The resulting sample paths tend to have low signal-to-noise ratios and are challenging from the point of view of change-point detection.

In the Binary Segmentation algorithm, we use the threshold $\zeta_T = \sqrt{C} \log^p T$ with $p = 1/2$, which is the lower end of the permitted theoretical range from Theorem 3.2. The study is repeated for $C = \{0.5, 0.75, 1, 1.25, 1.5, 2\}$. For robustness, we estimate the local volatility $\tilde{\sigma}_s^e$ using the Median Absolute Deviation estimator for the Gaussian distribution, both here and in the remainder of the paper. The estimated number of change-points is denoted by \hat{N} .

To judge the quality of change-point detection and basis recovery, we use three statistics, one of which is $N - \hat{N}$. To

gauge the distance between the estimated and true canonical bases, we take all those sample paths for which $\hat{N} = N$. For those sample paths, we define the ‘Not Assigned’ (NA) and ‘Scale Difference’ (SD) statistics, which measure, respectively, the number of estimated change-points that cannot be assigned to any true one in terms of their location, and the sum of differences in scales between the estimated and true change-points if assignment is possible. Define $\hat{\mathcal{B}}_{j,k} = \{\hat{b}_{i,l} : \hat{b}_{i,l} \in (b_{j,k} - \Delta_T, b_{j,k} + \Delta_T)\}$, where $\Delta_T = 5$, and let Υ denote the set of all estimated change-points $\hat{b}_{j,k}$. Set $\text{SD} = 0$. For all (j, k) , from coarser to finer scales and from left to right, if $|\hat{\mathcal{B}}_{j,k}| = 1$, then the matching for that (j, k) is completed. If $|\hat{\mathcal{B}}_{j,k}| > 1$, choose as the closest match the $\hat{b}_{i,l} \in \hat{\mathcal{B}}_{j,k}$ that minimises $|i - j|$. If there are multiple such $\hat{b}_{i,l}$ ’s, choose the one closest to $b_{j,k}$. Delete the matched estimated change-point from Υ . Add $|i - j|$ to SD . After considering all (j, k) ’s, set $\text{NA} = |\Upsilon|$.

Table 1 summarizes the results over various parameter specifications. As expected, \hat{N} is closer to N if jump sizes have a larger variance V_{η_i} . If ζ_T is too small, many spurious change-points are detected, while if it is too large, too few are identified. The average NA and SD measures are small in value, which shows the closeness of the true and estimated canonical bases, provided \hat{N} is estimated correctly.

5. DATA ANALYSIS

5.1 Data

We analyse the performance of our model in terms of basis recovery and forecasting in an application to 16 financial time series, 4 from each of the 4 asset classes: equity indices, single-name stocks, foreign exchange rates and commodity futures. We consider their daily closing prices between 1 January 1990 and 21 June 2013, obtained from Bloomberg. Details of the assets are provided in Table 4. We ignore days on which no price data are available. Correspondingly, we obtain $T_a^* \in [4, 774, \dots, 6, 124]$ data points per each asset a .

5.2 Interpretation of change-point importance

Figure 3 shows log-price series and our integrated model fit $\sum_{s=1}^t \hat{f}_s$ for one asset from each asset class (General Electric, GBP-USD exchange rate, WTI crude oil) apart from equity indices (S&P 500), for which we refer back to Figure 1. While the patterns exhibited in the three series from Figure 3 differ, two of the more pronounced trend changes take place for all three series at the start of the global financial crisis in early 2008 and following the dot-com bubble burst in 2000/2001. However, the estimated canonical bases reveal that both of these events were more important in the evolution of the General Electric share price than in those of the GBP-USD exchange rate or the crude oil: the earliest change-point in the General Electric price series that can be attributed to the recent global financial crises is estimated

Table 1. Simulation results for various V_{η_i} and C ; the numbers are averages over all signals and sample paths. Columns 3–5 show the proportion of the number of estimated change-points \hat{N} that equal or fall within a small range of N . Columns 6 and 7 show $N - \hat{N}$ and $|N - \hat{N}|$, respectively. Columns 8 and 9 show the NA and SD measures, respectively

V_{η_i}	C	Prop $\hat{N} = N$	Prop \hat{N} in $[N - 1, N + 1]$	Prop \hat{N} in $[N - 2, N + 2]$	Avg $(N - \hat{N})$	Avg $(N - \hat{N})$	Avg NA	Avg SD
1	0.50	0.02	0.05	0.08	-15.47	15.52	2.29	0.29
	0.75	0.13	0.35	0.51	-3.15	3.75	2.21	0.40
	1.00	0.22	0.56	0.76	-0.14	1.76	1.85	0.42
	1.25	0.22	0.57	0.78	0.96	1.57	1.49	0.38
	1.50	0.20	0.51	0.74	1.47	1.71	1.18	0.33
	2.00	0.16	0.42	0.66	1.97	2.02	0.83	0.25
2	0.50	0.01	0.03	0.05	-17.15	17.17	1.20	0.15
	0.75	0.12	0.29	0.43	-4.25	4.48	1.25	0.25
	1.00	0.26	0.60	0.77	-1.09	1.79	1.11	0.31
	1.25	0.32	0.71	0.88	0.05	1.18	0.91	0.32
	1.50	0.33	0.73	0.90	0.58	1.08	0.74	0.33
	2.00	0.30	0.67	0.87	1.05	1.20	0.51	0.31

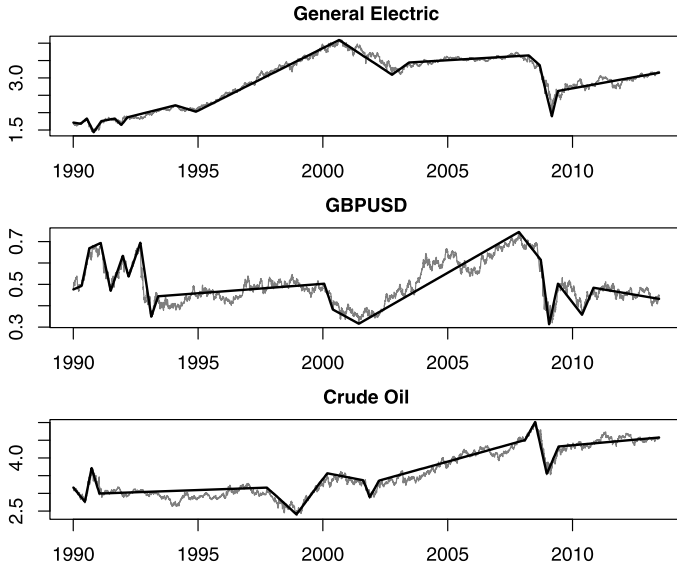


Figure 3. Daily closing values of the General Electric share price, the GBP-USD exchange rate and WTI crude oil between January 1990 and June 2013; log-price (grey) and the cumulatively integrated fit $\sum_{s=1}^t \hat{f}_s$ from our model (black), shown for threshold parameter $C = 0.3$ and imposing a minimum distance between change-points of 60 trading days.

at scale $j = 2$ of the canonical basis, while corresponding change-points for GBP-USD and crude oil only appear at scales $j = 7$ and $j = 8$, respectively. The dot-com burst corresponds to the change-point at scale $j = 1$ for the GE data and hence this event can be interpreted as the most important for this series. For the GBP-USD exchange rate, other events are more important: the change-point at scale $j = 1$

is detected in 1990, corresponding to the British currency’s joining of the European Exchange Rate Mechanism. For the oil price, the most important change-point is detected early in time and can be related to the price shock due to the Iraqi invasion of Kuwait in the summer of 1990. Outside of the period 1990/1991, the Asian financial crisis in 1997/1998 triggered some important changes in the oil price trend.

5.3 Forecast evaluation

Our model’s ability to predict returns is evaluated in a rolling window forecast analysis conducted for each of the 16 assets. Formally, for observed returns on each asset a ending at time $\bar{T} + s$, we evaluate the h -day cumulative return forecast $\hat{f}_{s,\bar{T}}^{a,h} = \sum_{t=1}^h \hat{f}_{s+\bar{T}+t}^a = h\hat{f}_{s+\bar{T}}^a$, and compare it to the observed return $X_{s,\bar{T}}^{a,h} = \sum_{t=1}^h X_{s+\bar{T}+t}^a$. To avoid overlapping forecast windows, the forecasts are evaluated for $s = \{1, 1 + h, 1 + 2h, \dots, 1 + (H_{a,h} - 1)h\}$, where $H_{a,h} = \lfloor (T_a^* - \bar{T})/h \rfloor$ is a function of the forecast horizon $h = \{1, 2, 5, 10, 15, 20, 40\}$. The length of the estimation period is set to $\bar{T} = 1,750$ days, or around 7 trading years, which roughly corresponds to the typical length of a US business cycle since 1980 (NBER, 2013). The parameters of the Binary Segmentation procedure are as in Section 4, with the threshold constant C taking values in the set $\{0.1, 0.2, \dots, 3\}$.

In the following, the index a is suppressed and it is understood that the forecast performance is evaluated for each asset a . The forecast $\hat{f}_{s,\bar{T}}^h$ equals h -times the moving average of the most recent observations X_t , with a span $\bar{T} + s - \hat{\eta}_{\hat{N}}$ chosen data-adaptively from the data as described in Section 3.3. Hence, a natural benchmark is a non-adaptive moving-average estimator that forecasts $\hat{f}_{s,\bar{T},w}^h = h/w \sum_{t=s+\bar{T}-w+1}^{s+\bar{T}} X_t$. As the choice of w will clearly affect the performance, we compare our model’s forecast $\hat{f}_{s,\bar{T}}^h$ with an optimized threshold constant C against the

Table 2. Relative success ratio RSR in percent for our model with optimized threshold value and the benchmark model with optimized moving average window length. If the relative success ratio has the same sign for both models, the larger absolute value is in bold. Forecast horizons are grouped with $h = 1, 2$, $h = 5, 10, 15$ and $h = 20, 40$ representing short-, medium- and long-term forecasts, respectively

	$h = 1, 2, 5, 10, 15, 20, 40$		$h = 1, 2$		$h = 5, 10, 15$		$h = 20, 40$	
	Model	Benchmark	Model	Benchmark	Model	Benchmark	Model	Benchmark
DAX	6.8	6.8	3.0	2.2	7.1	6.1	10.0	12.6
FTSE	5.6	2.6	1.7	-1.9	5.0	-0.4	10.4	11.6
HSI	2.5	4.6	1.6	3.0	4.2	4.2	0.7	6.9
S&P	7.2	6.2	2.4	0.2	7.7	6.6	11.3	11.5
LHA	-0.1	4.4	-2.0	0.6	0.1	5.9	1.4	6.0
GE	-1.1	-3.4	-2.3	-2.0	-1.2	-0.8	0.2	-8.9
JNJ	-1.2	0.9	-1.7	-1.7	-1.4	0.4	-0.2	4.1
MSFT	-1.8	-1.7	-2.1	-0.3	-2.2	0.1	-1.0	-5.8
GBPUSD	-4.6	-3.2	-1.3	-0.1	-3.9	-3.3	-8.9	-6.0
USDJPY	0.2	1.7	-1.5	-0.1	1.8	1.7	-0.5	3.4
AUDUSD	3.9	4.8	2.1	-0.3	3.3	4.5	6.6	10.3
USDMXN	-4.1	-0.4	-2.3	1.7	-5.4	1.0	-4.2	-4.6
Oil	6.6	1.8	2.9	0.1	6.2	1.7	10.8	3.5
Gold	2.5	1.6	3.3	-2.2	4.1	0.1	-0.6	7.4
Live Cattle	2.0	-6.1	-1.7	-2.9	0.8	-5.8	7.5	-9.7
Sugar	-2.8	3.4	-2.0	0.3	-1.5	4.1	-5.5	5.5

best forecast from a range of moving-average models with $w = \{h, 2h, \dots, 10h\}$ for each forecast horizon h .

Our success criterion is sign predictability, defined as the proportion of correctly predicted signs. Some authors, e.g. Leitch and Tanner (1991), argue that this approach provides more robust results than statistics based on the level of predicted returns and provide evidence that in financial applications under the objective of profit maximization, the proportion of correctly predicted signs beats, amongst others, the mean square prediction error as a criterion for choosing forecasts. We define the ‘Relative Success Ratio’ for a model with forecast \hat{f}^h at forecast horizon h as $RSR_h = 1/H_h \sum_{i=0}^{H_h-1} \mathbb{I}(X_{1+ih, \bar{T}}^h \hat{f}_{1+ih, \bar{T}}^h > 0) - 0.5$.

A large positive RSR_h provides evidence for predictability and can be expected if an asset behaves in a trend following way. However, as mentioned in Section 3.3, a large negative RSR_h can be related to mean reverting behaviour and also provides a meaningful trading signal. We only compare the two models in terms of their RSR_h ’s if both provide the same directional signal, that is if both suggest trend following or mean reversion. Table 2 summarizes the general results as well as results for forecast horizons grouped into short, medium and long term. The results are individually optimized in terms of the threshold constant C (for our model) and the moving average span w (for the benchmark) for each forecast horizon to yield the largest absolute RSR_h , i.e. the model fit with the proportion of predicted signs deviating furthest from 0.5.

Over all forecast horizons combined, our model and the benchmark are performing comparably well, with our model offering better performance for 7 assets, ahead of the benchmark, which performs better for 5 assets. The difference in

performance appears to be due to commodities, where the sign predictability of our model is clearly more pronounced for gold and oil, which show trend-following behaviour. Considering forecast horizons in more detail, our model tends to outperform the benchmark in short- and medium-term prediction, but performs worse in long-run predictions. Table 3 summarizes the RSR’s during times of strong movements. We classify time periods as containing strong movements if their h -day cumulative return lies in the top or bottom 10% of their historic distribution. Formally, we consider only those forecasts for which $X_{s, \bar{T}-h+1}^h < Q_{s, \bar{T}, h}^{10\%}$ or $X_{s, \bar{T}-h+1}^h > Q_{s, \bar{T}, h}^{90\%}$ with $Q_{s, \bar{T}, h}^{c\%}$ denoting the c -percentile of non-overlapping h -day returns in the period $[s, s + \bar{T}]$. In this subset our model does well particularly in currencies, where the apparent mean-reverting behaviour can be captured better by our adaptive predictor. Otherwise the pattern is similar to that in Table 2, but even more favourable to our adaptive predictor. In particular, for the long-term forecasts, our model now performs comparably well to the benchmark.

APPENDIX A. PROOFS

Proof of Proposition 2.1. Using the orthonormality of the basis $\psi^{b_j, k}$, we have

$$E(I_{j,k}) = T^{-1} E \left(T^{1/2} \sum_{(i,l) \in \mathcal{I}} A_{i,l} \sum_{t=1}^T \psi_t^{b_{i,l}} \psi_t^{b_j, k} + \sum_{t=1}^T \sigma_t \varepsilon_t \psi_t^{b_j, k} \right)^2$$

Table 3. Relative success ratio RSR in percent for our model with optimized threshold value and the benchmark model with optimized moving average window length. If the relative success ratio has the same sign for both models, the larger absolute value is in bold. Results for forecasts taking place when the most recent h -day cumulative return is in the top or bottom decile of its historical distribution. Forecast horizons are grouped with $h = 1, 2$, $h = 5, 10, 15$ and $h = 20, 40$ representing short-, medium- and long-term forecasts, respectively

	$h = 1, 2, 5, 10, 15, 20, 40$		$h = 1, 2$		$h = 5, 10, 15$		$h = 20, 40$	
	Model	Benchmark	Model	Benchmark	Model	Benchmark	Model	Benchmark
DAX	7.0	3.0	-0.1	2.6	7.7	4.7	13.1	0.9
FTSE	2.4	-3.4	-2.7	-2.4	3.0	-7.6	6.8	1.9
HSI	3.0	9.5	4.0	3.2	2.9	9.1	2.1	16.4
S&P	0.9	4.3	-4.1	-4.4	2.2	4.3	3.8	12.9
LHA	-4.2	0.1	-3.5	3.4	-10.1	-6.0	3.8	5.8
GE	-3.1	-2.3	-0.7	-0.3	-5.4	-5.4	-2.1	0.5
JNJ	-0.3	2.1	-4.4	-6.2	1.4	-1.0	1.2	15.0
MSFT	-3.4	2.2	-3.4	4.4	-2.8	2.1	-4.4	0.3
GBPUSD	-7.4	-2.0	-3.4	0.2	-4.2	2.4	-16.4	-11.0
USDJPY	-2.1	-0.4	-0.4	0.0	-1.9	-0.6	-3.9	-0.7
AUDUSD	1.3	-2.3	-0.6	-4.3	3.4	-1.3	0.0	-1.8
USDMXN	-3.1	-2.1	-0.5	3.8	-3.0	0.3	-5.7	-11.4
Oil	4.5	4.4	0.7	-1.4	5.3	10.4	6.9	1.3
Gold	-5.6	-5.3	3.3	-0.7	-6.9	-0.6	-12.8	-16.8
Live Cattle	-5.4	-6.1	-2.1	-2.5	-9.1	-4.1	-3.0	-12.8
Sugar	-1.9	5.8	-3.8	-3.2	3.4	8.8	-8.0	10.4

$$\begin{aligned}
&= T^{-1} E \left(T^{1/2} A_{j,k} + \sum_{t=1}^T \sigma_t \varepsilon_t \psi_t^{b_{j,k}} \right)^2 \\
&= E(A_{j,k}^2) + T^{-1} E \left(\sum_{t=1}^T \sigma_t \varepsilon_t \psi_t^{b_{j,k}} \right)^2,
\end{aligned}$$

which gives

$$|E(I_{j,k}) - E(A_{j,k}^2)| \leq T^{-1} \bar{\sigma}^2 \sum_{t=1}^T (\psi_t^{b_{j,k}})^2 = \bar{\sigma}^2 T^{-1}. \quad \square$$

Proof of Property 2.1. (i) results from the fact that f_t is a realisation of a piecewise-constant random process with at most N change-points and that by (ii), their magnitudes are $|\sum_{j,k} \alpha_{j,k,i} a_{j,k}|$, which is a finite quantity. For (ii), note that $f_t - f_{t-1} = T^{1/2} \sum_{(j,k) \in \mathcal{I}} a_{j,k} (\psi_t^{b_{j,k}} - \psi_{t-1}^{b_{j,k}})$ and that $T^{1/2} |\psi_t^{b_{j,k}} - \psi_{t-1}^{b_{j,k}}|$ is bounded in T as $\psi_t^{b_{j,k}} = O(T^{-1/2})$ due to the fact that the spacings between change-points satisfy $\min_{i=1, \dots, N+1} |\eta_i - \eta_{i-1}| = O(T)$. (iii) is implied by the fact that $P^A(|\sum_{j,k} \alpha_{j,k,i} A_{j,k}| = 0) = 0$ since the distributions of $A_{j,k}$ are continuous and mutually independent. Outside of the set A_0 , (iv) holds because f_t can only have change-points at $b_{j,k}$ as this is where the change-points in the basis vectors $\psi^{b_{j,k}}$ are located and the ends of their supports coincide with their parents' change-points as the set \mathcal{I} is connected. \square

Proof of Theorem 3.1. The proof is based on the proof of Theorem 3.1 from Fryzlewicz (2012). Firstly, we observe that our conditional signal f_t satisfies the assumptions of that Theorem, since $\{a_{j,k}\}_{j,k} \notin A_0$. Further, Lemmas A.1 and A.2 from Fryzlewicz (2012) hold if λ_1 in those Lemmas is replaced by $\bar{\sigma} \lambda_1$. Lemma A.3 holds with $\lambda_2 = O(\log T)$ and

$\epsilon_T = O(\log T)$ since in our case, $\delta_T = O(T)$. Lemmas A.4 and A.5 hold with the respective changes to λ_1 , λ_2 and ϵ_T , as above. Hence, the proof of Theorem 3.1 proceeds in the same way as the proof of Theorem 3.1 in Fryzlewicz (2012). \square

Proof of Proposition 3.1. The proof proceeds similarly to that of Property 2.1 (iii) by noting that $\langle f, \psi^{e, \eta_i, e} \rangle$ is a linear combination of $\{a_{j,k}\}_{j,k}$, different for each i such that $\eta_i \in [s, e]$. \square

Proof of Theorem 3.2. The proof is a straightforward modification of the proof of Theorem 3.1, which itself uses the proof of Theorem 3.1 from Fryzlewicz (2012). It is sufficient to observe that in our context, a stronger version of Lemma A.2 in the latter work holds, whereby η_{p_0+r} achieves the unique maximum of $|\langle f, \psi^{s, \eta, e} \rangle|$ over $\eta \in [s, e]$. This is because if the unique maximum were achieved by $\eta_{p_0+q} \neq \eta_{p_0+r}$, then $|\langle f, \psi^{s, \eta_{p_0+q}, e} \rangle| > |\langle f, \psi^{s, \eta_{p_0+r}, e} \rangle| = O(T^{1/2})$ by Lemma 1 of Cho and Fryzlewicz (2012), and therefore the b maximising $|\langle X, \psi^{s, b, e} \rangle|$ would have to fall near η_{p_0+q} , rather than η_{p_0+r} by Lemma A.1 from Fryzlewicz (2012). \square

For the estimation of $\bar{a}_{j,k}$, we have

$$\begin{aligned}
|\hat{a}_{j,k} - \bar{a}_{j,k}| &= T^{-1/2} |\langle f, \psi^{b_{j,k}} \rangle - \langle X, \psi^{\hat{b}_{j,k}} \rangle| \\
&\leq T^{-1/2} \left\{ |\langle f, \psi^{b_{j,k}} \rangle - \langle f, \psi^{\hat{b}_{j,k}} \rangle| \right. \\
&\quad \left. + |\langle X, \psi^{\hat{b}_{j,k}} \rangle - \langle f, \psi^{b_{j,k}} \rangle| \right\} \\
&\leq T^{-1/2} \left\{ O(\epsilon_T T^{-1/2}) + O(\log^{1/2} T) \right\} \\
&= O(T^{-1/2} \log^{1/2} T),
\end{aligned}$$

Table 4. Data series used in the empirical evaluation; data provider: Bloomberg; number of observations corresponds to number of days for which a quote is available between 1 January 1990 and 21 June 2013

Asset Class	Asset	Bloomberg Ticker	Primary Quote/Trading Venue	Observations
Equity Index	Dax Index	DAX Index	Germany - DB	5,936
Equity Index	FTSE 100 Index	UKX Index	United Kingdom - LSE	5,930
Equity Index	Hang Seng Index	HSI Index	Hong Kong - HKSE	5,802
Equity Index	S&P 500 Index	SPX Index	United States - NYSE	5,915
Equity	Deutsche Lufthansa	LHA GR Equity	Germany - Xetra	5,891
Equity	General Electric	GE US Equity	United States - NYSE	5,915
Equity	Johnson & Johnson	JNJ US Equity	United States - NYSE	5,915
Equity	Microsoft	MSFT US Equity	United States - Nasdaq	5,915
Currency	AUD/USD Spot rate	AUDUSD Curncy	London Composite	6,124
Currency	GBP/USD Spot rate	GBPUSD Curncy	London Composite	6,124
Currency	USD/JPY Spot rate	USDJPY Curncy	London Composite	6,124
Currency	USD/MXN Spot rate	USDMXN Curncy	London Composite	6,124
Commodity Future	Crude Oil	CL1 Comdty	New York Mercantile Exchange	5,895
Commodity Future	Gold	GC1 Comdty	CMX-Commodity Exchange	5,895
Commodity Future	Live Cattle	LC1 Comdty	Chicago Mercantile Exchange	5,923
Commodity Future	Sugar	SB1 Comdty	NYB-ICE Futures US Softs	5,880

where we use, respectively, the triangle inequality, a technique as in Lemma 2 of Cho and Fryzlewicz (2012) and Lemma A.1 of Fryzlewicz (2012).

APPENDIX B. DATA

Details of the assets used are in Table 4.

ACKNOWLEDGEMENTS

This work was supported by the Economic and Social Research Council [grant number ES/J00070/1].

Received 15 July 2013

REFERENCES

- BAEK, C. and PIPIRAS, V. (2009). Long range dependence, unbalanced Haar wavelet transformation and changes in local mean level. *International Journal of Wavelets, Multiresolution and Information Processing* **7** 23–58. [MR2582507](#)
- BAL, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. [MR1455175](#)
- BARON, M., BROGAARD, J. and KIRILENKO, A. (2012). The trading profits of high frequency traders. Technical Report, Foster School of Business, University of Washington.
- BERKOWITZ, J. and O'BRIEN, J. (2002). How accurate are value-at-risk models at commercial banks? *Journal of Finance* **57** 1093–1111.
- BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTECH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Annals of Statistics* **37** 157–183. [MR2488348](#)
- BRUDER, B., DAO, T. L., RICHARD, J. C. and RONCALLI, T. (2008). Trend filtering methods for momentum strategies. Technical Report, Lyxor Asset Management.
- CATALÃO, J. P. S., MARIANO, S. J. P. S., MENDES, V. M. F. and FERREIRA, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research* **77** 1297–1304.
- CHO, H. and FRYZLEWICZ, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica* **22** 207–229. [MR2933173](#)
- DELOUILLE, V., FRANKE, J. and VON SACHS, R. (2001). Nonparametric stochastic regression with design-adapted wavelets. *Sankhyā: The Indian Journal of Statistics Series A* **63** 328–366. [MR1897046](#)
- FRYZLEWICZ, P. (2007). Unbalanced Haar technique for nonparametric function estimation. *Journal of the American Statistical Association* **102** 1318–1327. [MR2412552](#)
- FRYZLEWICZ, P. (2012). Wild Binary Segmentation for multiple change-point detection. Preprint.
- FRYZLEWICZ, P., SAPATINAS, T. and RAO, S. S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika* **93** 687–704. [MR2261451](#)
- FRYZLEWICZ, P. and SUBBA RAO, S. (2013). BaSTA: Consistent multiscale multiple change-point detection. Preprint.
- GARCIA, R. C., CONTRERAS, J., VAN AKKEREN, M. and GARCIA, J. B. C. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems* **20** 867–874.
- GENÇAY, R., SELÇUK, F. and WHITCHER, B. J. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.
- GIRARDI, M. and SWELDENS, W. (1997). A new class of unbalanced Haar wavelets that form an unconditional basis for L_p on general measure space. *Journal of Fourier Analysis and Applications* **7** 457–474. [MR1468375](#)
- HASBROUCK, J. and SOFIANOS, G. (1993). The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance* **48** 1565–1593.
- KATZ, J. O. and MCCORMICK, D. (2000). *The Encyclopedia of Trading Strategies*. McGraw-Hill, New York.
- KIM, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing* **55** 307–319.
- KIM, H. J. and SHIN, K. S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing* **7** 569–576.
- LAVIELLE, M. and MOULINES, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* **21** 33–59. [MR1766173](#)
- LEBARBIER, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing* **85** 717–736.

- LEE, C. B. (1995). Estimating the number of change points in a sequence of independent normal random variables. *Statistics & Probability Letters* **25** 241–248. [MR1369518](#)
- LEITCH, G. and TANNER, J. E. (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review* **81** 580–590.
- NASON, G. P. (2008). *Wavelet Methods in Statistics with R*. Springer. [MR2445580](#)
- NASON, G. P., VON SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B* **62** 271–292. [MR1749539](#)
- NBER (2013). US Business Cycle Expansions and Contractions. National Bureau of Economic Research.
- OMBAO, H., RAZ, J., VON SACHS, R. and GUO, W. (2002). The SLEX model of a non-stationary random process. *Annals of the Institute of Statistical Mathematics* **54** 171–200. [MR1893549](#)
- PAN, J. and CHEN, J. (2006). Application of modified information criterion to multiple change point problems. *Journal of Multivariate Analysis* **97** 2221–2241. [MR2301636](#)
- PRIESTLEY, M. B. (1983). *Spectral Analysis and Time Series*. Academic Press.
- SPOKOINY, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Annals of Statistics* **37** 1405–1436. [MR2509078](#)
- TIMMERMANS, C., DELSOL, L. and VON SACHS, R. (2012). Using Bagdis in nonparametric functional data analysis: Predicting from curves with sharp local features. *Journal of Multivariate Analysis* **115** 421–444. [MR3004568](#)
- ČÍŽEK, P., HÄRDLE, W. and SPOKOINY, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *Econometrics Journal* **12** 248–271. [MR2562386](#)
- VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. PhD thesis, Department of Statistics, Stanford University. [MR2687536](#)
- VIDAKOVIC, B. (2009). *Statistical Modeling by Wavelets*. John Wiley & Sons. [MR1681904](#)
- VOSTRIKOVA, L. (1981). Detecting ‘disorder’ in multidimensional random processes. *Soviet Mathematics Doklady* **24** 55–59.
- YAO, Y. C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics Series A* **51** 370–381. [MR1175613](#)
- Anna Louise Schröder
 Department of Statistics
 London School of Economics
 London WC2A 2AE
 United Kingdom
 E-mail address: a.m.schroeder@lse.ac.uk
- Piotr Fryzlewicz
 Department of Statistics
 London School of Economics
 London WC2A 2AE
 United Kingdom
 E-mail address: p.fryzlewicz@lse.ac.uk