Managerial Economics
2014, vol. 15, no. 1, pp. 107–131
http://dx.doi.org/10.7494/manage.2014.15.1.107

Stephan Schlüter*, Carola Deuschle*

Wavelet-based forecasting of ARIMA time series – an empirical comparison of different methods

1. Introduction

Forecasting prices of stocks or commodities on liquid markets is mainly guesswork. To reduce this insecurity about future price developments, we can try to use the information contained in historical data. This can be done, for example, by using parametric statistical models: it is assumed that the given data is the realization of an underlying stochastic process with a certain specification, and historical data is used to calibrate the process parameters. The forecast is then the result of an extrapolation step while eliminating the random element by taking the expectation.

Simple but powerful parametric models are autoregressive formulas where the current value is determined partly by the value of the previous time step and partly by a random term. If the model includes autoregression for the random term as well, we speak of an autoregressive moving average (ARMA) model. A more advanced concept is the autoregressive integrated moving average (ARIMA) model, which captures intertemporal dependence in the data itself as well as in the error term (cf. McNeil *et al.*, 2006). Neither model, however, can capture seasonal effects – these have to be filtered or modeled by an extra component.

Two well-known filter mechanisms are the Kalman filter and the Fourier transform (cf. Hamilton 1995). However, the quality of both methods suffers if the season has a variable period and/or intensity. Contrary to the previous methods, wavelet transform is able to capture these dynamics; this is why wavelet

* Both authors from University of Erlangen-Nuremberg. University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg.

transform is interesting for time series analysis. By means of this function, we can decompose the process into a linear combination of different frequencies. We can – with some restrictions – quantify the influence of a pattern with a certain frequency at a certain time on the price. Having such a feature, it is very promising that wavelet transform can help to improve the quality of forecasting.

There is already a broad range of work on this topic: A. Wong *et al*. (2003) use wavelets to fit a structural time series model to exchange rates. D. Donoho and I. Johnstone (1994), L. Breiman (1996), J. Bruzda (2013), H.Y. Gao and A.G. Bruce (1997), R.M. Alrumaih and M.A. Al-Fawzan (2002) and G.P. Nason (2008) use wavelets to eliminate random noise. A.J. Conejo *et al*. (2005), C.M. Lee and C.N. Ko (2011), Y. Chen et al. (2013) as well as K. Kriechbauer *et al*. (2014) decompose the time series into a sum of processes with different frequencies and forecast the individual time series before adding up the results. M. Shafie-Khah *et al*. (2011) proceed in a similar way, but add a neural network component to their toolbox. Given such a variety of approaches, it remains to be seen which model with which specification performs best in which scenario.

In order to conduct such an analysis we choose four time series, each with its own individual characteristics: oil prices, where the long-term structure dominates, Euro-Dollar exchange rates and Deutsche Bank stock prices, where we see both long- and short-term patterns, and UK day-ahead power prices, which show a distinct daily oscillation. We perform day and week-ahead out-of-sample forecasts using models from the literature listed above. The results are compressed by computing standard error measures like the root mean squared error. To validate each model's performance, we generate two benchmarks: one using a simple ARIMA model (which does not model seasonality) and one using the Census X-12 ARIMA method. Census X-12 ARIMA was developed and is used by the U.S. Census Bureau to identify and model seasonal patterns and trends.

We come to the conclusion that the utilization of wavelets improves the accuracy of forecasting, especially for forecasting horizons larger than one-day-ahead. However, there is no single method that is best in all scenarios. The performance of each wavelet-based method varies with the data set and the forecasting horizon. Depending on the scenario, we recommend applying wavelets either for denoising purposes or using the method of A.J. Conejo *et al*. (2005). The concept of H. Wong *et al*. (2003) is outperformed in all scenarios.

We structure this paper as follows: first relevant definitions of time series analysis are given and the basic models are presented. In Section 3, we introduce wavelet transform and explain how to use wavelet in time series forecasting. In Section 4, we introduce the data sets and perform an empirical comparison of the presented wavelet-based forecasting methods. Section 5 summarizes this paper.

2. Some basics of time series analysis

In our analysis, we assume that the observed data is the realization of an unknown stochastic process. A stochastic process is a family of random variables $(X_t)_{t \in I}$, where $I \subset \mathbb{R}$ is interpreted as time index. As we analyze discrete-time data sets, $I = \mathbb{Z}$. There are various parametric stochastic process models. Two widely used concepts, which are presented in the sequel, are the ARMA model and its extension, the ARIMA model. Eventually, we introduce another concept, the structural time series model (STSM) as well as widely used implementation, the Census X-12 ARIMA method. The STSM distinguishes between three components of a time series: a deterministic trend, a deterministic seasonality, and a stochastic noise term.

2.1. Autoregressive moving average models

The autoregressive moving average model of order $(p,q) \in \mathbb{N}^2$ is a linear time series model which describes a process $(X_t)_{t \in \mathbb{Z}}$ of the form

$$X_t = \mu_t + \sum_{i=1}^{p} \phi_i \left(X_{t-i} - \mu\right) + \sum_{j=1}^{q} \theta_i \epsilon_{t-j} + \epsilon_t \tag{1}$$

where $\phi_i, \theta_k \in \mathbb{R} \forall i \in 1, \ldots p, j \in 1, \ldots, q$. The $\mu_t \in \mathbb{R}$ is the long-term drift and by default $\epsilon_t \sim N\left(0, \sigma^2\right), \sigma > 0$ (cf. McNeil et al., 2006). Other distribution assumptions for the innovations ϵ_t are possible as well. The first sum represents the autoregressive (AR) part; i.e., the current value of X_t is partly determined by its own past. The second sum is the moving average (MA) part, which introduces autoregression for ϵ_t .

Important functions to characterize X_t are its mean $\mu_t = E(X_t)$, its variance function $\sigma_t^2 = E(X_t - \mu_t)^2$ and its autocorrelation function $\rho(s,t) = Cov(X_s, X_t) / \sqrt{\sigma_s^2 \cdot \sigma_t^2}$, where $s \neq t$ and $s,t \in \mathbb{Z}$. Having obtained a set of observations $(X_1, \ldots, X_T)$ we can calculate these functions for every time step (which is quite cumbersome for large data sets, though). One way to reduce this effort is to demand the process to be stationary. Stationarity describes a certain invariance of the shape of a process. We speak of strict stationarity, if the conjoint distribution of a subset $(X_t)_{t \in W}, W \subset I$ is invariant under time shifts. A process is called weakly (or wide-sense) stationary, if its mean and variance function are constant over time, and if the covariance is only a function of the distance $(s - t)$. We focus on this kind of stationarity, and will omit the adjective ”weak” in the following. If a stochastic process shows this feature, the number of parameters that have to be computed is significantly reduced. An even more important consequence of stationarity is the following: if an ARMA process with Gaussian innovations is stationary, then it is ergodic regarding mean and variance (cf. Green 2008). Ergodicity means that we can estimate the process parameters consistently using time-series data. If ergodicity is not given (e.g. due to trends or seasonality), the process parameter estimates are biased. Time series forecasts based on the estimated parameters would then be biased as well.

In the case of the ARMA model, stationarity is relatively easy to verify (cf. McNeil *et al.* 2005): the moving average part of Eq. (1) is weakly stationary by definition, and the autoregressive part is weakly stationary if $|z| > 1$ for all $z \in \mathbb{C}$ that fulfill

$$1 - \phi_1 z - \ldots \;-\; \phi_p z^p = 0. \tag{2}$$

The optimal forecast for the model in Eq. (1) is obtained by minimizing the forecasting error regarding a chosen goodness of fit measure. If we opt for the mean square error, it can be shown that the optimal h-step forecast ($h \in \mathbb{N}$) $\hat{X}_{t+h}$ is the expected value of Eq. (1) given the filtration until time t , which is denoted by $\mathcal{F}_t$ (cf. Hamilton 1995):

$$\hat{X}_{t+h} = E\left[\mu + \sum_{i=1}^{p} \phi_i \left(X_{T+h-p} - \mu\right) + \sum_{j=0}^{q} \theta_j \epsilon_{T+h-j} \mid \mathcal{F}_t \right]. \tag{3}$$

As the conditional expectation is a linear function, Eq. (3) can be simplified. Because we assume that the innovations have zero mean, we obtain

$$\begin{aligned} E\left[X_{t+j} \mid \mathcal{F}_t\right] &= \begin{cases} X_{T+j} & \text{if } j \leq 0, \\ \hat{X}_{T+j} & \text{otherwise.} \end{cases} \\ E\left[\epsilon_{t+j} \mid \mathcal{F}_t\right] &= \begin{cases} X_{T+j} - \hat{X}_{T+j} & \text{if } j \leq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

The h-step forecast for an ARMA(1,1) model, for example, reads as follows:

$$\hat{X}_{t+h} = E\left[X_{t+h} \mid \mathcal{F}_t\right] = \mu + \phi^h \left(X_t - \mu\right) + \phi^{h-1} \theta \epsilon_t. \tag{5}$$

Fitting a process $X_t \in \mathbb{Z}$ of the form of Eq. (1) to a data set means estimating the lag order, the coefficients, and the parameters of $\mathcal{F}$. For determining the lag order (p, q) we test various lag order combinations and choose the best one using information criteria that punish a higher number of variables. An example for such a criterion is the Bayesian information criterion, but there are others as well. For an overview, refer to S.G. Koreisha and T.A. Pukkila (1995).

The further parameters of Eq. (1), including those of F, can be estimated using Durbin's (1960) regression method, the conditional or unconditional least squares method, or by maximizing the likelihood function. As this is a nonlinear optimization problem, numerical methods like the Berndt-Hall-Hall-Hausmann algorithm or the Newton-Raphson algorithm come to be applied (cf. McNeil *et al*. 2005 or Hamilton 1995).

2.2. Autoregressive integrated moving average models

we mainly distinguish between instationarity in the mean and instationarity in the variance. For forecasting it is crucial to avoid the first one. Instationarity in the mean is caused, for example, by linear trends which can be eliminated by modeling $\Delta X_t = X_t - X_{t-1}$ instead of X_t. This procedure can be repeated to treat trends of higher polynomial order, and we speak of an autoregressive integrated moving average process with integration order $d \in \mathbb{N}$, if $\Delta^d X_t$ is stationary. Thereby $\Delta^d = \Delta_t^d - \Delta_{t-1}^{d-1}, d \in \mathbb{N} \setminus \{1\}$.

The optimal h-step forecast ($h \in \mathbb{N}$) for an ARIMA(p,d,q) model is computed in two steps: first, we compute expectations according to Eq. (3) and (4) for $Y_t = \Delta^d X_t$ and obtain an estimate for $\hat{Y}_{t+h}$. Second, we use the relation $Y_{t+h} = (1-B)^d X_{t+h}$ with $B^d X_{t+h} = X_{t+h-d}$, $d \in \mathbb{N}$, to obtain a forecast for X_{t+h} (cf. McNeil *et al.* 2005).

To estimate the integration order d, we use tests on instationarity, e.g. the augmented Dickey Fuller (ADF) test (cf. Dickey, Fuller 1979) or the Phillips-Peron (PP) test (cf. Phillips, Peron 1988). If we find instationarity in X_t, we proceed as follows: we compute the first differences and perform the unit root test. If the test still indicates instationarity, we compute the second differences and apply the test again. We continue with this procedure until we find a difference $\Delta^d X_t$ which is stationary.

The ARIMA model is able to capture trends, and there are also extensions to include seasonality and long-term dependence. For these versions please refer to (Granger, Joyeux 1980), (Hosking 1981).

2.3. The structural time series model

The structural time series model consists of three major components. A process $X_t \in \mathbb{Z}$ at time t is described as a sum of a long-term trend T_t, a seasonal component S_t, and a random (noise) term ϵ_t (cf. Majani 1987):

$$X_t = T_t + S_t + \epsilon_t. \tag{6}$$

By means of the exponential function, we can transform the additive model from Eq. (6) into a multiplicative one. Trend and season are expected to be deterministic, but we can also design them to be stochastic (cf. Harvey, 1989).

The exact shape of T_t and S_t depends on how both components are estimated. Common methods of identifying T_t are the moving average method, the Fourier transform, the Kalman filter or exponential smoothing. A more sophisticated version would be to see T_t as a function $f(t;\beta_1,\ldots,\beta_n)$ with parameters $\beta_1,\ldots,\beta_n \in B$, where $B \subseteq \mathbb{R}$ denotes their domain. Examples for f are $f(t) = \beta_1 f_1(t) + \ldots + \beta_n f_n(t) + \epsilon_t$ or $f(t) = f_1(t)^{\beta_1} + \ldots + f_n(t)^{\beta_n} + U_t$, where U_t is a noise term and $f_t,\ldots,f_n$ are functions of t. The parameters can be estimated applying the least squares method, i.e. by solving

$$\min_{\beta_1,\ldots,\beta_n} \sum_{t \in T} \left(X_t - f(t) \right)^2, \tag{7}$$

where $t = 1,\ldots,T \in \mathbb{N}$ is the index of our observations. In more complex scenarios we can use numerical methods like the Gauss–Newton algorithm. The seasonal component S_t is commonly estimated using the Fourier transform or dummy variables (cf. Harvey, 1989). However, both methods require a true seasonal pattern with fixed period and intensity to provide sound estimation results. For ϵ_t various stochastic processes (e.g. an ARIMA model) can be assumed. For producing forecasts, both T_t and S_t are extrapolated and the forecast of ϵ_t is evaluated.

An implementation of the STSM is the Census X-12 ARIMA method developed by the U.S. Census Bureau. It defines season as constantly repeating intra-year variation and patterns with a longer period as trend. A further component for daily features can be added. Seasonal and trend adjustment is done by applying different moving averages iteratively. What is left is then modeled by an ARIMA process (cf. Findley *et al*. 1998).

3. Wavelet-based forecasting

As suggested in the introduction, wavelets may be used to extend the methods from Section 2 in order to improve forecasting accuracy. Before presenting three possible extensions, we give a few basic definitions of wavelet theory.

3.1. A brief introduction to wavelet theory

A wavelet is a complex-valued function $\Psi(t) \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$ that fulfills the admissibility condition

$$C_\Psi = \int_{-\infty}^{+\infty} \frac{\left| \widehat{\Psi}(\omega) \right|^2}{|\omega|} < \infty, \tag{8}$$

where the hat denotes the Fourier transform. Each Ψ has a fixed mean and frequency. To make it more flexible, set $\Psi_{a,b} = \Psi\left((t-b)/a\right)$, which translates Ψ by $b \in \mathbb{R}$ and scales Ψ by a scaling factor $a > 0$ that is inverse proportional to the frequency (cf. Mallat 2003).

The continuous wavelet transform (CWT) generalizes the Fourier transform and is, unlike the latter, able to detect seasonal oscillations with time-varying intensity and frequency. While stationarity of the process is not required, square-integrability is needed (see Mallat 2003). In the following, we focus on the CWT. For an introduction to the discrete wavelet transform please refer to G.A. Kaiser (1994) or A. Jensen and A. Cour-Harbo (2001). The CWT is the orthogonal projection of a process $(X_t)_{t \in \mathbb{R}}$ on $\Psi_{a,b}$, i.e.

$$WT_X(a,b) = X, \Psi_{a,b} = \int_{\mathbb{R}} X_t \frac{1}{\sqrt{a}} \overline{\Psi}_{a,b}(t)\,dt, \tag{9}$$

where the overline denotes the conjugate complex (cf. Mallat, 2003). The $WT_X(a,b)$ indicates how much of X_t is explained by a local oscillation Ψ at scale a in time b . The inverse transform is therefore a linear combination of Ψ and in the continuous case a double integral of the form (cf. Mallat 2003)

$$X(t) = \frac{1}{C_\Psi} \int_0^\infty \int_{-\infty}^\infty WT(a,b) \frac{1}{a^2\sqrt{a}} \psi\left(\frac{t-b}{a}\right) db\, da. \tag{10}$$

We can simplify Eq. (10) significantly for a discrete data set, e.g. for daily commodity prices. In this case Shannon‘s sampling theorem states that the signal can be exactly reconstructed using only a discrete set of scales; i.e., the above integration is reduced to a sum (cf. Shannon 1949).

When identifying the influence of patterns with a certain scale or frequency (e.g. annual seasonality), we have to consider the uncertainty principle of time-frequency analysis. It says that both scale and location of a signal cannot be exactly specified simultaneously (cf. Lau, Weng 1995). Thus, we are limited to an analysis of time-frequency windows and the only lever we can pull is the choice of an appropriate wavelet. For various selection criteria, please refer to N. Ahuja *et al.* (2005). The best wavelet regarding window size is the Morlet wavelet, which is a function $\Psi_M(t|\sigma,\omega_0): \mathbb{R} \to \mathbb{C}$ with

$$\Psi_M\left(t|\sigma,\omega_0\right) = c_{\omega_0}\pi^{-1/4}e^{-t^2/2\sigma^2}\left(e^{i\,\omega_0 t} - e^{-1/2\,\omega_0^2}\right), \tag{11}$$

where

$$c_{\omega_0} = \left(1 - e^{\omega_0^2} - 2e^{-\frac{3}{4}\omega_0^2}\right)^{-1/2} \tag{12}$$

and $\omega_0 > 0$ denotes the basis frequency and $\sigma > 0$ (cf. Daubechies 1992). It is plotted in Figure 1 at three different scales for $b = 0$ and its time-frequency window can be found in Appendix A. In Figure 1, we can clearly see the influence of the scale parameter and the character of a local oscillation. It is diminishing outside a set called cone of influence (CoI) that reads as $\left[b-\left(s_u - s_l\right)a, b+\left(s_u - s_l\right)a\right]$, where $[s_l, s_u] \subseteq \overline{\mathbb{R}}$ is the support of Ψ (cf. Lau, Weng 1995). If data within the CoI is missing for time t and scale $a > 0$, the coefficient $WT_{X(a,t)}$ from Eq. (9) is biased, which especially holds for the edge regions of a finite data set. Methods to reduce this effect are given by S.D. Meyers *et al.* (1993), A. Jensen and A. Cour-Harbo (2001), or C. Torrence and G.P. Compo (1998).

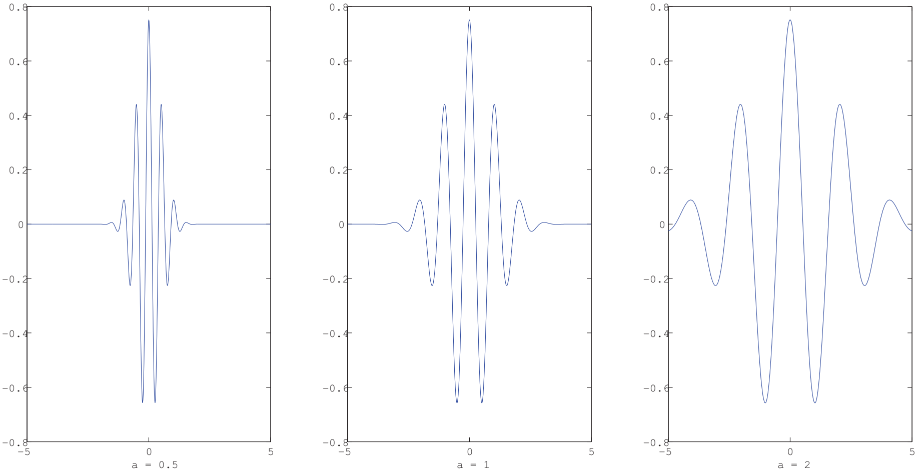

Figure 1. The Real Part of the Morlet Wavelet at Different Scales

In this paper, we analyze daily data $X_t, t = 1,\dots,T$. Hence, we set $dt = 1$ and $b \in \mathbb{Z}$. The scale grid has to be discretized as well. Most authors (e.g. Torrence, Compo 1998) use a dyadic approach to form a set of scales $A = \left\{a_0, a_1, \dots, a_J\right\}$. We construct A likewise:

$$a_j = 2^{1+j\delta j}, j = 0,1,\ldots,J, and \quad J = \delta j^{-1} log_2\left(\frac{T}{a_0}\right)+1, \tag{13}$$

where $\delta j \in \mathbb{R}^+$ determines the resolution of wavelet transform. The grid is finer for lower scales. This is reasonable as information is more concentrated in the lower scales than in the higher (i.e. lower frequencies). It is likely, for example, that a process has a weekly and a monthly oscillation, but less likely to find an annual oscillation together with an oscillation having a period of a year and 20 days. As a consequence, we can aggregate the influence of larger scales without losing relevant information. For this purpose we introduce the wavelet scaling function ϕ that behaves like a low-pass filter and aggregates the influence of all scales larger than $a^* > 0$ on X_t (cf. Mallat 2003). There is a huge variety of scaling functions we can use (cf. Ahuja *et al*. 2005) but when operating together with a wavelet Ψ it has to fulfill at least

$$\left|\hat{\phi}(\omega)\right|^2 = \int_{\omega}^{\infty} \frac{\left|\widehat{\Psi}(\xi)\right|^2}{\xi} d\xi. \tag{14}$$

Just as Ψ, each scaling function has a certain frequency and is centered around a certain $x \in \mathbb{R}$. Thus, we define a rescaled and shifted version of ϕ by

$$\phi_{a,b}(t) = \frac{1}{\sqrt{a}}\phi\left(\frac{t-b}{a}\right), a > 0, b \in \mathbb{Z}. \tag{15}$$

Eventually, we are able to split up a process $X_t \in \mathbb{Z}$ for a scale $a^* \in A$ as follows (cf. Mallat, 2003):

$$X_t = \frac{1}{C_\Psi}\sum_{b\in\mathbb{Z}} X,\phi_{a^*,b} + \frac{1}{C_\Psi}\sum_{b\in\mathbb{Z}} \sum_{a\in A \wedge a > a^*} \left\langle X,\Psi_{a,b}\right\rangle \Psi_{a,b}(t)\frac{1}{a^2} \quad \forall t. \tag{16}$$

The first addend represents the long-term trend and the second addend contains short-term information of X_t. In Eq. (16) we can see that the effort is reduced because for scales larger than a^*, the double sum is substituted by a simple sum. However, the CWT is still computationally very intensive. One way to reduce the effort is to use the á trous algorithm of M. Holschneider *et al.* (1989) for decomposition purposes. The main idea of this algorithm is that the wavelet of a certain scale $a_j \in A$ is not computed exactly, but interpolated using the wavelets of scale a_{j-1}. The result is a cascade of filter banks. In Appendix B, we present the algorithm in detail.

3.2. Wavelet-Based Forecasting Methods

Essentially, there are three different wavelet-based forecasting methods. One is to use wavelets for eliminating noise in the data, and one uses wavelets to estimate the components in a STSM. Another method performs the forecasting based on the wavelet generated time series decomposition. In the following we briefly describe each of these methods.

3.2.1. Wavelet Denoising

Wavelet denoising is based on the assumption that a data set $\left(X_1,\ldots,X_T\right)$ is the sum of a deterministic function Y_t and a white noise component $\epsilon_t \sim N\left(0,\sigma^2\right)$, i.e., $X_t = Y_t + \epsilon_t$. By means of wavelets, the noise is reduced and the standard forecasting methods from Section 2.1 can be applied to the modified data set (cf. Alrumaih, Al-Fawzan 2002).

The denoising is accomplished as follows: initially, the CWT is applied to X_T with a scale discretization of $A = \left\{a_0,\ldots,a_n\right\}$ and $b = 1,..,T$ with $n \in \mathbb{N}$. The result is a matrix of wavelet coefficients $\mathbb{R}^{n\times T}$. The CWT for a pair of parameters $\left(a,b\right)$ is an orthogonal projection of X_t on the wavelet $\Psi_{a,b}$. Thus, each $WT\left(a,b\right)$ indicates how much of X_t is described by $\Psi_{a,b}$. G.P. Nason (2008) shows that the noise term has an impact on each coefficient, while the information of Y_t is concentrated only in a few. So, if $WT\left(a,b\right)$ is relatively large, it contains information about both Y_t and ϵ_t, whereas small coefficients indicate a motion solely caused by the noise term. If we now set all coefficients below an appropriate threshold $\lambda \geq 0$ to zero and invert the modified coefficients $WT'\left(a,b\right)$, we obtain a noise adjusted time series X'_T.

The question of how to choose λ remains. D. Donoho and I. Johnstone (1994) propose two different thresholds:

$$(a) \qquad WT'\left(a,b\right) = WT\left(a,b\right)1_{\{|WT(a,b)|>\lambda\}} \qquad (\textit{hard threshold})$$

$$(b) \qquad WT'\left(a,b\right) = sgn\left(WT\left(a,b\right)\right)\left(\left|WT\left(a,b\right)\right| - \lambda\right)1_{\{|WT(a,b)|>\lambda\}} \qquad (\textit{soft threshold})$$

where sgn denotes the signum function. The larger the λ, the more noise; but at the same time, more of Y_t is cut out and vice versa. D. Donoho and I. Johnstone (1994) suggest $\lambda_{universal} = \hat{\sigma}\sqrt{2logT}$ for λ, where $\hat{\sigma}$ is an estimator for the standard deviation σ of the wavelet coefficients at resolution level a_0. Using $\lambda_{universal}$ in the hard threshold function is called *VisuShrink*. This procedure is quite smoothing, as it cuts off a relatively large number of coefficients.

D. Donoho and I. Johnstone (1995) propose a further threshold based on the SURE[1] estimation method developed by C.M. Stein (1981). For a scale a, they derive the optimal threshold λ_{SURE} by solving

$$\lambda_{SURE} = \arg \min_{0 \leq \lambda \leq \lambda_{universal}} SURE\left(WT, \lambda\right) \tag{17}$$

$$SURE\left(WT, \lambda\right) = T - \#\left\{t : \left|WT\left(a,t\right)\right| \leq \lambda\right\} + \sum_{t=1}^{T} \min\left(\left|WT\left(a,t\right)\right|, \lambda\right)^2 . \tag{18}$$

This method does not work very well for sparsely occupied matrices. Therefore D. Donoho and I. Johnstone (1994) unite both concepts in the SureShrink method, which uses $\lambda_{universal}$ as threshold if

$$\sum_t \left(WT\left(a,t\right)^2 - 1\right) \leq \log_2 T^{3/2} \tag{19}$$

for $a \in A$ and λ_{SURE} otherwise. H.Y. Gao and A.G. Bruce (1997) or L. Breiman (1996) propose further threshold rules.

3.2.2. Wavelet-based estimation of a structural time series model

In Eq. (16), we break up a process $\left(X_t\right)_{t\in\mathbb{Z}}$ into a long-term component and a short-term part by means of a scaling function and a wavelet. H. Wong *et al.* (2003) make use of this fact to estimate the components of the STSM from Section 2.3, which models X_t as the sum of the trend T_t , the season S_t and the noise ϵ_t, i.e.

$$X_t = T_t + S_t + \epsilon_t, \qquad t \in \mathbb{Z} . \tag{20}$$

First, they estimate trend and seasonality $\hat{T}_t, \hat{S}_t$ from the data. Second, they produce forecasts for T_t and S_t by extrapolation polynomials fitted to $\hat{T}_t$ and $\hat{S}_t$. To $\hat{\epsilon}_t = X_t - \hat{T}_t - \hat{S}_t$ they fit an ARMA(1,0) model and generate a forecast as well.

The $\hat{T}_t$ is computed by aggregating the high-scale patterns using a scaling function ϕ as described in Section 3.1, which is for discrete-time data a linear combination of the observations, as the convolution integral is approximated by a sum:

$$\hat{T}_t = \frac{1}{C_\Psi} \sum_{b\in\mathbb{Z}} \left\langle S, \phi_{a*,b} \right\rangle \phi_{a*,b}\left(t\right). \tag{21}$$

[1] Stein's Unbiased Risk Estimate (SURE) is an unbiased estimator of the mean squared error.

It remains to choose a scaling function and the optimal scale a^*, which depends on the analyzed data set. This scale should be small enough to capture the whole trend, but large enough not to cut through some short-term oscillations.

For estimating X_T, H. Wong et al. (2003) use the hidden periodicity analysis, which is described in Appendix C.

3.2.3. Forecasting based on a wavelet decomposition

We can motivate this procedure using the commodity market as an example: prices are determined by different traders, each with their individual intentions and investment horizons. People might trade because they need the commodity for production purposes, while trading is pure speculation for others. Using wavelets we intend to “unbundle” the influence of traders with different investment horizons, i.e., split the price process into a sum of processes with different frequencies. The underlying assumption is that we can model and forecast these individual patterns more precisely.

Further, there are technical arguments in favor of this method. Among others, S. Soltani *et al.* (2000) show that we can avoid (existing) long-term memory by modeling the multivariate process of wavelet coefficients instead of the process itself. They also show that there is no long-term dependence between different scales. P. Abry *et al.* (1995) come to a similar result for fractional Brownian motions.

The procedure is as follows: The time series $\left(X_t\right)_{t=1,\ \dots,\ T}$ is transformed according to Eq. (9) to obtain a matrix of wavelet coefficients $WT\left(a,b\right), a \in A, b=1,\ \dots,\ T$, where A denotes a scale discretization. For each a, the corresponding vector $WT\left(a\right)=WT\left(a,1\right),\ \dots,\ WT\left(a,T\right)$ is treated as a time series. Standard forecasting techniques like those from Section \ref{btsa} are applied to obtain forecasted wavelet coefficients, which are then added to WT in order to obtain an extended matrix $WT\,'$ (cf. Conejo *et al.* 2005, or Yousefi *et al*. 2005). O. Renaud *et al.* (2005) use only specific coefficients for this forecast, which is very efficient but increases the forecasting errors. The extended matrix $WT\,'$ is then inverted according to Eq. (16), and we finally obtain a forecast $\hat{X}_{t+1}$ for X_t in the time space.

4. An empirical comparison of different forecasting methods

The wavelet-based forecasting techniques from Section 3 are applied to four data sets in order to evaluate their performance. To check whether the additional effort is worthwhile, we do also compute forecasts using the classic methods from Section 2. Below, we present the chosen time series, and then describe the test design and comment on the estimation results.

4.1. The data sets

We analyze four different time series which are displayed in Figure 2: the Deutsche Bank (DB) stock price, the Euro-Dollar exchange rate, the West Texas Intermediate (WTI) oil price, and the APX Power UK Peak Load Index (provided by the APX Group), i.e., the average UK day-ahead power price. Each of these time series has its own individual characteristics. The WTI, which represents commodities in our study, has a comparatively strong long-term pattern which dominates the short-term oscillation. The DB stock prices show a long-term trend as well, but also some medium-term oscillations and a few price jumps. The EUR/USD exchange rate, which represents the foreign exchange market, has a visible long-term component, a less important short-term structure and shows some distinct price jumps. The UK power prices represent the recently evolving electricity markets. They show only a minor upward trend, but a strong daily oscillation.

For the first three time series, we have weekday closing prices whereas the UK power prices include weekends. Initially, we apply both the ADF and the PP test to our data sets as well as to the first differences to identify the integration order. The alternative hypothesis for both tests is stationarity. The corresponding p-values for the time series and their first differences, which are displayed in Table 1, are constructed from the tables in Banerjee *et al.* 1993. The ADF test indicates an integration order (d) of one for all time series. The PP test shows similar results, except for the UK where the test is indifferent between $d = 1$ and $d = 0$. This coincides with Figure 2 as the power prices' long-term pattern (i.e. its trend) is comparably weak. In Table 1 we also give the empirical standard deviation $\hat{\sigma}$, which is computed from the empirical error of an ARIMA(1,1,1) model.

This parameter has a positive influence on forecast volatility. The larger the $\hat{\sigma}$, the larger the probability that the real value will deviate from the forecasted one. In Table 1 we observe that the power prices' standard deviation is substantially higher than the standard deviation of the other time series. The EUR/USD exchange rate has the lowest standard deviation; i.e., the weakest oscillation. Therefore we expect that the forecasts of the exchange rate are better than those of the power prices.

4.2. Test design and goodness of fit measures

We compute day-ahead and week-ahead forecasts, which is a step of seven days for the power prices and a step of five days for the other three data sets (as these exclude weekends). Out-of-sample forecasts for the last n data points of each time series are calculated, where n is 14 for the power prices and 10 for the rest. The results are evaluated using three different error measures, namely the mean absolute deviation (MAD), the root mean squared error (RMSE), and the mean average percentage error (MAPE). These measures are defined as follows:

$$\begin{aligned} MAD\left(X,\hat{X}\right) &= \sum_{i=T-n+1}^{T} \left|X_i - \hat{X}_i\right| / n, \\ RMSE\left(X,\hat{X}\right) &= \sum_{i=T-n+1}^{T} \left(X_i - \hat{X}_i\right)^2 / n, \\ MAPE\left(X,\hat{X}\right) &= \sum_{i=T-n+1}^{T} \left|X_i - \hat{X}_i\right| / (X_i n), \end{aligned} \tag{22}$$

for data $X_t, t = 1,\ldots,T$ and estimates $\hat{X}_t, t = 1,\ldots,T$. The MSE penalizes large deviations more than the MAD. The MAPE focuses on the relative deviation, i.e. it allows larger deviations if X_t itself is large at time t. Using these different measures allows us to evaluate forecasting method performance from different points of view.

The first forecasts we compute are based on an ARIMA(p,1,q)-model. We use ARIMA instead of ARMA, showing in Table 1 (Appendix) that each time series is integrated of order 1. The pair $(p,q) \in \mathbb{N}^2$ is identified as described in Section 2.1. Moreover, we apply the Census X-12 method of the U.S. Census Bureau (briefly X-12) as an implementation of the STSM from Section 2.3.

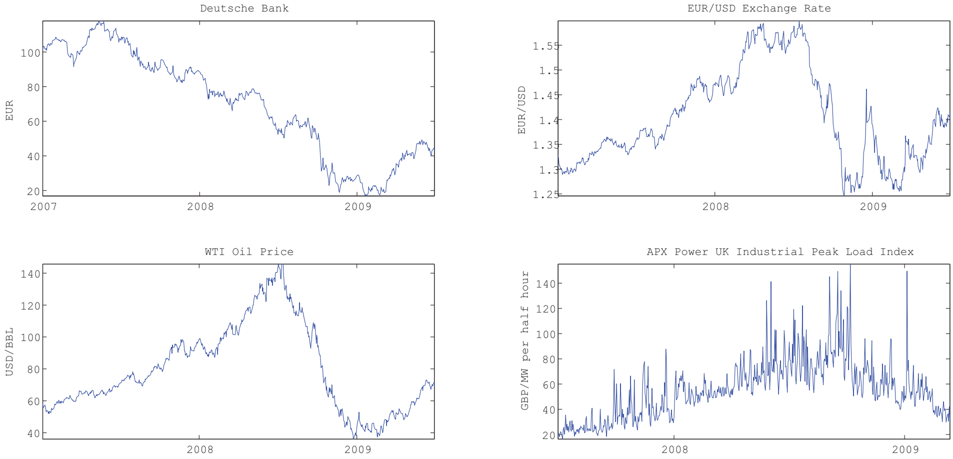

Figure 2. The Analyzed Data Sets

To implement the wavelet-based methods, we choose three widely used functions: the Haar wavelet (see Appendix D), which is the simplest wavelet and orthogonal to a scale-dependent moving average (cf. Stollnitz *et al*. 2005), the Morlet wavelet, which has the best time-frequency resolution, and the (orthogonal) Daubechies D4 wavelet (see Appendix D), which is easy to implement and works well with efficient techniques like the à trous algorithm (cf. Daubechies 1992). We follow C. Torrence & G.P. Compo (1998) when constructing a scale grid according to Eq. (13) and set $a_0 = 2, \delta j = 0.6$ in case of the Morlet wavelet and $a_0 = 2, \delta j = 1$ in case of Haar's function. We apply the Haar wavelet with all wavelet-based methods, and further evaluate whether it pays off to use more complex wavelet functions. Morlet's wavelet, for example, is chosen for performing a multiscale forecast as described in Section 3.2.3. For the denoising procedure, we follow D. Donoho and I. Johnstone (1994) and apply the Daubechies D4 wavelet.

Eventually, we apply the concept of Wong et al. (2003) and generate forecasts with a wavelet-based STSM (see Section 3.2.2). For this purpose, only the Haar wavelet is chosen.

4.3. Presentation and Evaluation of the Estimation Results

The forecasting results differ from time series to time series. We find that the performance of each wavelet-based method varies with the data set and the forecasting horizon, and there is no single forecasting method which would be applicable to all time series (see Table 2, Appendix). In the following we briefly summarize the results of our study for each time series before drawing an overall conclusion. Tables containing the exact results of all error measures can be found in Appendix E.

The Deutsche Bank stock price. Looking at the day-ahead forecast, the Haar-based multiscale method combined with the ARIMA model performs best regarding all three error measures, although its APE is as low as the APE of the classic ARIMA model. Moreover, both MAD and RMSE of the classic ARIMA model are less than 1% worse than those of the Haar-based multiscale method. The classic Census X-12 method proves to be inadequate in this scenario, as all wavelet-based methods show lower values for MAD, RMSE and APE.

In case of the week-ahead forecast, the difference between classic and wavelet-based methods is stronger. The Morlet wavelet-based multiscale decomposition with Census X-12 forecasting turns out to be the best method. Its MAD is 4.5% lower, its RMSE is 10% lower, and its APE is 9% lower than the best classic forecasting method, which is the ARIMA model.

The Euro/Dollar exchange rate. In the day-ahead forecast the Haar wavelet-based multiscale method and the ARIMA model performs best. They make it possible to reduce the forecasting error by about 4–7% (depending on the error measure) compared to the classic ARIMA model, which still produces better forecasts than the Census X-12 method. In the week-ahead scenario the results are different. Now Haar or Daubechiet D4 wavelet-based denoising in combination with the ARIMA model allows us to reduce the errors significantly (by 12–22%) compared to the ARIMA model, which is again the best among the classic methods.

The WTI oil price. The Haar/Daubechies D4 wavelet-based denoising method in combination with the ARIMA model performs best regarding all three error measures in the day-ahead forecast. MAD and RMSE of the best classic forecasting method (ARIMA) can be lowered by 8% and the APE even by 25%. In the week-ahead forecast the Morlet multiscale decomposition combined with the Census X-12 method generates the best forecasts regarding MAD and APE. The RMSE favors the same decomposition method except for the ARIMA model. If we use the Census X-12 method on the decomposed time series instead of an ARIMA model, we are able to reduce MAD, RMSE, and APE by 13% (MAD) –25% (APE).

The UK power prices. The best day-ahead forecast regarding MAD and RMSE is generated by a Morlet wavelet-based multiscale decomposition combined with the Census X-12 method. The APE favors the simple ARIMA model. However, the difference between both methods regarding MAD and RMSE is less than 3%. So, for the one-day-ahead forecast, the classic ARIMA model provides sound results. This also holds true for the week-ahead forecasts, where only the MAD can be lowered by less than 1% when using the Haar/Daubechies D4 wavelet-based denoising plus the Census X-12 method instead.

The estimation results above indicate that there is not one single “outstanding” wavelet-based method. Sometimes denoising is preferred, and sometimes the multiscale forecasting method provides sound results. Even the optimal wavelet varies with the data set and the forecasting horizon. Denoising, where switching from Haar’s to Morlet’s wavelet has a minimal effect on forecasting errors, stands out as an exception.

What we find is that it generally pays off to use wavelet-based forecasting methods. The UK power prices are an exception, though. There, the classic ARIMA model is sufficient, which is reasonable as this time series consists mainly of a dominating short-term oscillation. Wavelet transform is applied because we want to make use of certain structures within the data set. If there is no significant structure, the payoff is small. We can see the opposite in the results of the WTI oil prices and the exchange rate, both having a significant medium- and long-term structure. Using wavelet-based methods leads to a considerable reduction of computed errors.

Wavelet-based methods are more powerful for longer forecasting horizons, as our results indicate (excluding the power price scenario). The errors of the week-ahead forecasts are reduced to a higher extent than in the day-ahead scenario.

To explain this fact, we use the same argument as above. Wavelets are applied to make use of certain structures in the time series. Identifying these structures is more important for longer time horizons than for short ones. In the day-ahead scenario, autoregression is able to capture a large part of these structures; in the week-ahead forecast it is not sufficient.

Another observation is that wavelet-based methods are superior to the classic Census X-12 model, which is shown for all data sets. Nevertheless, the X-12 method is still useful. However, looking at the results of this study indicate that it makes sense to integrate the X-12 method into a wavelet-based procedure. The wavelet-based STSM proposed by H. Wong *et al*. (2003) is outperformed in all tested data sets. Thus, from our results we cannot recommend using it.

5. Conclusion

The purpose of this paper is to evaluate the power of wavelet-based forecasting methods. Wavelets are used mainly in the context of data preprocessing. The actual forecast is done using one of the existing forecasting techniques, of which we presented the ARMA/ARIMA model and the Census X-12 method. We also gave a brief introduction to wavelet theory and then described how wavelets are used for forecasting purposes. For our empirical study, we chose four time series with different characteristics. Two different forecasting horizons (one day, one week) are tested, and the results are compared using three standard error measures.

Evaluating the results we come to the conclusion that using wavelet-based forecasting methods pays off, as long as there is some structure in the data. If a time series consists to a large part of short-term oscillation, the gain of using wavelets is small or even negative. However, for data with existing medium- and long-term structure we were able to reduce the errors of the day-ahead forecasts substantially and further reduce the errors of the week-ahead forecast. One has to note, though, that there is nothing like a general method applicable to all scenarios, as performances vary with the data and time horizon.

References

[1] Abry P., Goncalves P., Flandrin P. 1995, *Wavelets, spectrum analysis and 1/f processes*, in: *Wavelets and statistics*, A. Antoniadis (ed.), Springer, New York, pp. 15–30.

[2] Ahuja N., Lertrattanapanich S., Bose N.K. 2005, *Properties determining choice of mother wavelet*, IEEE Proceedings – Vision, Image & Signal Processing, 152(5), pp. 659–664.

[3] Alrumaih R.M., Al-Fawzan M.A. 2002, *Time series forecasting using wavelet denoising: an application to Saudi stock index*, "Journal of King Saud University, Engineering Sciences", 2(14), pp. 221–234.

[4] Banerjee A., Dolado J.J., Galbraith J.W., Hendry D.F. 1993, *Cointegration, error correction, and the econometric analysis of non-stationary data*, Oxford University Press, Oxford.

[5] Breiman L. 1996, *Heuristics of instability and stabilization in model selection*, "The Annals of Statistics", 24(6), pp. 2350–2383.

[6] Bruzda J. 2013, *Forecasting Via Wavelet Denoising – The Random Signal Case*, Working Paper, Nicolaus Copernicus University.

[7] Chen Y., Shi R., Shu S., Gao W. 2013, *Ensemble and enhanced PM10 concentration forecast model based on stepwise regression and wavelet analysis*, "Atmospheric Environment", 74, pp. 346–359.

[8] Conejo A.J., Plazas M.A., Espinola R., Molina A.B. 2005, *Day-ahead electricity price forecasting using the wavelet transform and ARIMA models*, IEEE Transactions on Power Systems, 20(2), pp. 1035–1042.

[9] Daubechies I. 1992, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

[10] Dickey D.A., Fuller W.A. 1979, *Distribution of the estimators for autoregressive time series with a unit root*, "Journal of the American Statistical Association", 74, pp. 427–431.

[11] Donoho D., Johnstone I. 1994, *Ideal spatial adaptation via wavelet shrinkage*, "Biometrika", 81, pp. 425–455.

[12] Donoho D., Johnstone I. 1995, *Adapting to unknown smoothness via wavelet shrinkage*, "Journal of the American Statistical Association", 90, pp. 1200–1224.

[13] Durbin J. 1960, *The fitting of time series models*, "The International Statistical Review", 28, pp. 233–244.

[14] Fabert O. 2004, *Effiziente Wavelet Filterung mit hoher Zeit-Frequenz-Auflösung*, Verlag der Bayerischen Akademie der Wissenschaften, Munich.

[15] Findley D.F., Monsell B.C., Bell W.R., Otto M.C., Chen B.-C. 1998, *New capabilities and methods of the X-12-ARIMA seasonal adjustment program*, "Journal of Business and Economic Statistics", 16, pp. 127–176.

[16] Gao H.Y., Bruce A.G. 1997, *WaveShrink with firm shrinkage*, "Statistica Sinica", 7, pp. 855–874.

[17] Granger C.W.J., Joyeux R. 1980, *An introduction to long-memory time series models and fractional differencing*, "Journal of Time Series Analysis", 1(1), pp. 15–19.

[18] Green W.H. 2008, *Econometric analysis*, 6th edition, Prentice Hall International, Upper Saddle River, NJ.

[19] Hamilton J. 1995, *Time series analysis*, Princeton University Press, Princeton.

[20] Harvey A.C. 1989, *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

[21] Holschneider M., Kronland-Martinet R., Morlet J., Tchamitchian P., *Wavelets, time-frequency methods and phase space*, Springer, Berlin.

[22] Hosking J.R.M. 1981, *Fractional Differencing*, "Biometrika", 68(1), pp. 165–176.

[23] Jensen A., Cour-Harbo A. 2001, *Ripples in mathematics, the discrete wavelet transform*, Springer, Berlin.

[24] Kaiser G.A. 1994, *Friendly guide to wavelets*, Birkhäuser, Boston.

[25] Koreisha S.G., Pukkila T.A. 1995, *Comparison between different order-determination criteria for identification of ARIMA models*, "Journal of Business & Economic Statistics", 13(1), pp. 127–131.

[26] Kriechbauer T., Angus A., Parsons D., Casado M.R. 2014, *An improved wavelet-ARIMA approach for forecasting metal prices*, "Resources Policy" 39, pp. 32–41.

[27] Lau K.-M., Weng H. 1995, *Climate signal detection using wavelet transform: how to make a time series sing*, "Bulletin of the American Meteorological Society", 76(12), pp. 2391–2402.

[28] Lee C.-M., Ko C.-N. 2011, *Short-term load forecasting using lifting scheme and ARIMA models*, "Expert Systems with Applications", 38, pp. 5902–5911.

[29] Li Y., Xie Z. 1997, *The wavelet detection of hidden perodicities in time series*, "Statistics and Probability Letters", 35(1), pp. 9–23.

[30] Majani B.E. 1987, *Decomposition methods for medium-term planning and budgeting*, in: *The handbook of forecasting: A manager's guide*, S. Makridakis, S. Wheelwright (ed.), Whiley, New York, pp. 219–237.

[31] Mallat S.A. 2003, *Wavelet tour of signal processing*, 2nd edition, Academic Press, Manchester.

[32] McNeil A.J., Frey R., Embrechts P. 2005*, Quantitative risk management: concepts, techniques, and tools*, Princeton University Press, Princeton.

[33] Meyers S.D., Kelly B.G., O'Brien J.J. 1993, *An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of yanai waves*, "Monthly Weather Review", 121(10), pp. 2858–2866.

[34] Nason G.P. 2008, *Wavelet methods in statistics with R*, Springer, New York.

[35] Phillips P.C.B., Perron P. 1988, *Testing for a unit root in time series regression*, „Biometrika", 75, pp. 335–346.

[36] Renaud O., Starck J.L., Murtagh F. 2005, *Wavelet-based combined signal filtering and prediction*, IEEE Transactions on Systems, Man, and Cybernetics, B – Cybernetics, 35(6), pp. 1241–1251.

[37] Shannon C.E. 1949, *Communication in the presence of noise*, Proceedings of the Institute of Radio Engineers, 37(1), pp. 10–11.

[38] Shafie-Khah M., Moghaddam M.P., Sheikh-El-Eslami M.K. 2011, *Price forecasting of day-ahead electricity markets using a hybrid forecast method*, "Energy Conversion and Management", 52, pp. 2165–2169.

[39] Soltani S., Boichu D., Simard P., Canu S. 2000, *The long-term memory prediction by multiscale decomposition*, "Signal Processing", 80(10), pp. 2195–2205.

[40] Stein C.M. 1981, *Estimation of the mean of a multivariate Normal distribution*, "The Annals of Statistics", 9(6), pp. 1135–1151.

[41] Stollnitz E.J., DeRose T.D., Salesin D.H. 1995, *Wavelets for computer graphics: a primer*, part 1. IEEE Computer Graphics and Applications, 15(3), pp. 76–84.

[42] Torrence C., Compo G.P. 1998, *A practical guide to wavelet analysis*, "Bulletin of the American Meteorological Society", 79(1), pp. 61–78.

[43] Wong H., Ip W.C., Xie Z., Lui X. 2003, *Modelling and forecasting by wavelets, and the application to exchange rates*, "Journal of Applied Statistics", 30(5), pp. 537–553.

[44] Yousefi S., Weinreich I., Reinarz D. 2005, *Wavelet-based prediction of oil prices*, Chaos, Solitons & Fractals, 25, pp. 265–275.

Appendix

A. The time-scale window of morlet's wavelet

For $a,\sigma,\omega_0 > 0$ and $\mathbb{R}$ the time-scale window of $\Psi_M(t\,|\,\sigma,\omega_0)$ is (cf. Fabert 2004)

$$\tau(a,b) = \left[b - \frac{a\sigma}{\sqrt{2}}, b + \frac{a\sigma}{\sqrt{2}}\right] \times \left[a\frac{2\sqrt{2}\pi\sigma}{\omega_0\sqrt{2}\sigma + 1}, a\frac{2\sqrt{2}\pi\sigma}{\omega_0\sqrt{2}\sigma - 1}\right].$$

B. The à trous algorithm

Let

$$\left\{\phi_{m,n}(\cdot) = \phi\left(\cdot/2^m - n\right)/\sqrt{2^m} \;: m,n,\in \mathbb{Z}\right\}$$

denote a set of scaling functions to a dyadic scale discretization. The corresponding set of wavelet functions reads as

$$\left\{\Psi_{m,n}(\cdot) = \Psi\left(\cdot/2^m - n\right)/\sqrt{2^m} \;: m,n,\in \mathbb{Z}\right\}.$$

If ϕ is chosen such that it generates for each scale an orthonormal basis, then according to (Mallat 2003) there is a vector $(h_n)_{n\in\mathbb{Z}}$ with

$$\phi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_n \phi(2t-n).$$

The $(h_n)_{n\in\mathbb{Z}}$ is called scaling filter. S.A. Mallat (2003) also shows that for the corresponding wavelet there is a vector $(g)_{n\in\mathbb{Z}}$ with $g_n = (-1)^n h_{1-n}$ such that

$$\Psi(t) = \sqrt{2}\sum_{n\in\mathbb{Z}} g_n \Psi(2t-n).$$

Let now $(X_t)_{t\in\mathbb{Z}}$ be a discrete-time process and define $d_n^m = \langle X, \Psi_{m.n}\rangle$, $c_n^m = \langle X, \phi_{m,n}\rangle$ Define

$$d^m = \left\{d_n^m : n \in \mathbb{Z}\right\} \in \mathcal{L}^2(\mathbb{Z}) \text{ and } c^m = \left\{c_n^m : n \in \mathbb{Z}\right\} \in \mathcal{L}^2(\mathbb{Z}).$$

We introduce recursive (filter) functions h^r, g^r, whereby $r \in \mathbb{N}$ indicates the approximation level (the higher r the coarser the approximation). Set $g^0 = h, h^0 = g$. In every filter step we want to obtain a coarser approximation of the time series. Therefore, the filters g^r, h^r are computed by introducing zeros between each component of g^{r-1}, h^{r-1}. Two operators G^r, H^r are defined as follows

$$G^r : \mathcal{L}^2(\mathbb{Z}) \to \mathcal{L}^2(\mathbb{Z}) \text{with}\, c \mapsto \left\{\left(G^r c\right)_n = \sum_{k\in\mathbb{Z}} g_{k-n}^r c_k\right\},$$

$$H^r : \mathcal{L}^2(\mathbb{Z}) \to \mathcal{L}^2(\mathbb{Z}) \text{with}\, c \mapsto \left\{\left(H^r c\right)_n = \sum_{k\in\mathbb{Z}} h_{k-n}^r c_k\right\}.$$

The adjoint functions G^{r^*}, H^{r^*} are defined analogously to invert this mapping. Given these definitions the à trous decomposition algorithm is performed as follows: As input we require $c^0 = \left\{c_n^0 : n \in \mathbb{Z}\right\}$ and a $M \in \mathbb{N}$ to determine the maximal scale 2^M. We then gradually compute for $m = 1,\ldots,M : d^m = G^{m-1}c^{m-1}, c^m = H^{m-1}c^{m-1}$ and yield $c^M, d^m, m = 1,\ldots,M$, i.e. a multiscale decomposition of the time series with c^M containing the information about the highest scale (that is the long-term component). For the reconstruction of the time series we start with $M, c^M, d^m, m = 1,\ldots,M$ and gradually compute

$$\forall m = M, M-1,\ldots,1 : c^{m-1} = H^{m*}c^m + G^{m*}d^m.$$

The result is c^0 from which we obtain the time series by inverting the corresponding convolution.

C. Hidden periodicity analysis

Here, just the algorithm is given. For a more detailed overview, see (Li, Xie 1997) or (Wong *et al*. 2003). Let $\left(X_t\right)_{t=1,\ldots,T}$ be a time series with an estimated trend $\hat{T}$. Let $\hat{Y}_t = X_t - \hat{T}$ and assume

$$\hat{Y}_t = \sum_{n=1}^{T} \alpha_n e^{n\lambda_n t} + \xi_t, \qquad -\pi < \lambda_1 < \ldots < \lambda_n < \pi, n \in \mathbb{N},$$

with a complex random variable $\alpha_n, n = 1,\ldots,N$, which has finite variance, no autocorrelation and for which holds $0 < \alpha <\| \alpha_n \|^2, \alpha \in \mathbb{C}$. The random variable ξ_t is a linear combination of ergodic processes $\eta_t : \xi_t = \sum_{i=1}^{\infty} \beta_i \eta_{t-i}$ with $\sum_{j=1}^{\infty} \sqrt{j}\left|\beta_j\right| < \infty$. Having T observations for $\hat{v}_t$, H. Wong *et al*. (2003) identify "hidden periodicities", i.e. regular patterns contained in the time series, using a wavelet function whose Fourier transform has finite support and integrates to a nonnegative but finite constant. The idea is to compute the wavelet coefficients of the periodogram $I_T\left(\lambda\right) = \left|\sum \hat{Y}_t e^{-it\lambda}\right|^2 / 2\pi T$ for $\lambda \in \left[-\pi, \pi\right]$. Then, large coefficients for a specific scale indicate a hidden periodicity. H. Wong *et al.* (2003) use a dyadic wavelet decomposition scheme similar to Eq. (13), i.e. the set of scales is $A = \left\{2^m, m \in \mathbb{Z}\right\}$. Their algorithm to identify hidden periodicities is as follows. Set $n = 1$:

1) Let $M = \left\{0,1,\ldots,2^{|m|} - 1\right\}$. Compute $\left\{WT_{I_T}\left(m, b_m\right) : m = m_0, m_0 - 1, \ldots, -\infty, b_m \in M\right\}$ for a $\mathbb{Z}$.
2) Let $b\left(m\right) = \arg\max_{b \in M}\left(WT_{I_T}\left(m,b\right)\right), MW\left(m\right) = \max_{b \in M}\left(WT_{I_T}\left(m,b\right)\right)$:
   a) If $MW\left(m\right) \sim c$ with $m = m_0, m_0 - 1, \ldots, -\infty$ and a constant $c \in \mathbb{R}$, then $\hat{\lambda}_n = 2^{m'+1}\pi b\left(m\right) - 0.5$ where $m' \in \mathbb{Z}$ is sufficiently small. Go to Step (3).
   b) If $MW\left(m\right) \to 0$ for $m = m_0, m_0 - 1, \ldots, -\infty$, then there are no further periodicities. Stop the algorithm.
3) Is $\hat{\lambda}_n$ an estimate for a hidden periodicity, then set $\hat{\alpha}_n = \sum_{t=1}^{T} \hat{Y}_t e^{-i\hat{\lambda}_n t}$ and $\hat{Y}_t' = \hat{Y}_t' - \hat{\alpha}_t e^{i\hat{\lambda}_n t}\ \forall t = 1,\ldots,T$. Set $n = n + 1$. Go to Step (1).

D. The haar wavelet and daubechies D4 wavlet

The Haar scaling function ϕ_H and the corresponding wavelet Ψ_H are real-valued functions on $\mathbb{R}^+$ that are defined as follows (cf. Stollnitz *et al*. 1995)

$$\phi_H(x) = \begin{cases} 1 & 0 \le x < 1 \\ 0 & \mathit{otherwise} \end{cases}, \quad \Psi_H(x) = \begin{cases} 1 & 0 \le x < 1/2 \\ -1 & 1/2 \le x < 1 \\ 0 & \mathit{otherwise.} \end{cases}$$

The Ψ_H is in fact part of a wavelet family introduced by I. Daubechies (1992), and also called Daubechies D2 wavelet. Another representative of this family is the Daubechies D4 wavelet Ψ_D and its corresponding scaling function ϕ_D, for which no closed form is given. Both functions are defined iteratively using the relations

$$h(n) = \frac{1}{\sqrt{2}}\phi_D(t/2), \phi_D(t-n) \qquad ,n = 0,\ldots,1,$$

$$\frac{1}{\sqrt{2}}\phi_D\left(\frac{t}{2}\right) = \sum_{n=0}^{3} h(n)\phi_D(t-n), \; \frac{1}{\sqrt{2}}\psi_{\mathrm{D}}\left(\frac{t}{2}\right) = \sum_{n=3}^{3} (-1)^{1-n} h(1-n)\phi_D(t-n),$$

where for the coefficients $h(0),\ldots,h(3)$ holds

$$h(0) = \frac{1+\sqrt{3}}{4\sqrt{2}}, h(1) = \frac{3+\sqrt{3}}{4\sqrt{2}}, h(2) = \frac{1-\sqrt{3}}{4\sqrt{2}}, h(3) = \frac{3-\sqrt{3}}{4\sqrt{2}}.$$

For further properties or numerical issues refer to (Daubechies 1992) or (Mallat 2003).

Table 1

Characteristics of the Analyzed Time Series

Data Set	Start	End	#	ADF Test		PP Test		σ
				$X(t)$	$\Delta(t)$	$X(t)$	$\Delta(t)$	
DB	01-01-07	30-06-09	632	0.71	< 0.01	0.73	< 0.01	1.65
EUR/USD	01-01-07	30-06-09	636	0.62	< 0.01	0.85	< 0.01	0.01
WTI Oil	01-01-07	30-06-09	623	0.87	< 0.01	0.96	< 0.01	2.19
UK Power	07-07-07	13-03-09	623	0.29	< 0.01	< 0.01	< 0.01	12.78

Table 2

The Best Forecasting Method for Each Data Set

Data Set	Day-Ahead	Week-Ahead
DB	Multiscale (Haar + ARIMA)	Multiscale (Morlet + X-12)
EUR/USD	Multiscale (Haar + ARIMA)	Denoising + ARIMA
WTI	Denoising + ARIMA	Multiscale (Morlet + X-12)
UK Power Prices	Multiscale (Morlet + X-12)	ARIMA

Table 3

Forecasting Errors of the Deutsche Bank Stock Prices

	Day-Ahead			**Week-Ahead**		
Classic Methods:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
ARIMA	1.162	1.5672	0.0316	1.6705	3.2222	0.0655
X-12	1.6077	2.8988	0.0606	2.6679	12.9264	0.1643
Haar Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	1.1633	1.7893	0.0322	1.7231	3.4029	0.0702
Denoising + X-12 forecast	1.2713	2.2147	0.0383	2.6659	8.8941	0.1663
Multiscale forecast (ARIMA)	1.1587	1.5619	0.0316	3.5912	14.959	0.3019
Multiscale forecast (X-12)	1.3185	1.9284	0.0402	2.3495	6.8411	0.1288
Wavelet-based STSM	3.4215	13.4808	0.2275	2.4375	7.3828	0.1380
Daubechies/Morlet Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	1.1633	1.7893	0.0332	1.7231	3.4029	0.0702
Denoising + X-12 forecast	1.2713	2.2147	0.0383	2.6659	8.8941	0.1663
Multiscale forecast (ARIMA)	1.2166	1.8132	0.0344	1.6473	3.0939	0.0637
Multiscale forecast (X-12)	1.2175	1.8147	0.0344	1.5947	2.8977	0.0594

Table 4

Forecasting Errors of the Euro/Dollar Exchange Rate

	Day-Ahead			**Week-Ahead**		
Classic Methods:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
ARIMA	0.0873	0.0085	0.0054	0.1129	0.0153	0.0091
X-12	0.1020	0.0126	0.0075	0.1797	0.0419	0.0232
Haar Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	0.1008	0.0124	0.0072	0.0996	0.0123	0.0071
Denoising + X-12 forecast	0.0970	0.0111	0.0067	0.1293	0.0240	0.0119
Multiscale forecast (ARIMA)	0.0840	0.0080	0.0050	0.2765	0.1037	0.0548
Multiscale forecast (X-12)	0.0846	0.0092	0.0051	0.1587	0.0312	0.0181
Wavelet-based STSM	0.1477	0.0234	0.0156	0.1403	0.0266	0.0141
Daubechies/Morlet Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	0.1008	0.0124	0.0072	0.0996	0.0123	0.0071
Denoising + X-12 forecast	0.0970	0.0111	0.0067	0.1293	0.0240	0.0119
Multiscale forecast (ARIMA)	0.0866	0.0085	0.0054	0.1207	0.0158	0.0104
Multiscale forecast (X-12)	0.0871	0.0086	0.0054	0.1103	0.0147	0.0087

Table 5

Forecasting Errors of the WTI Oil Prices

	Day-Ahead			**Week-Ahead**		
Classic Methods:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
ARIMA	1.0768	1.4420	0.0167	1.6117	2.9530	0.0372
X-12	1.5744	2.8091	0.0355	2.9017	14.4185	0.1193
Haar Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	0.9306	1.3208	0.0125	1.5974	2.9820	0.0368
Denoising + X-12 forecast	1.2967	2.2076	0.0243	2.5311	7.6044	0.0914
Multiscale forecast (ARIMA)	1.0754	1.4880	0.0167	3.0311	11.1294	0.1317
Multiscale forecast (X-12)	1.1203	1.5331	0.0179	1.6987	3.6149	0.0412
Wavelet-based STSM	2.3485	6.6597	0.0793	6.9384	49.1278	0.6874
Daubechies/Morlet Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	0.9306	1.3208	0.0125	1.5974	2.9820	0.0368
Denoising + X-12 forecast	1.2967	2.2076	0.0243	2.5311	7.6044	0.0914
Multiscale forecast (ARIMA)	1.1598	1.5467	0.0193	1.4060	2.2636	0.0284
Multiscale forecast (X-12)	1.162	1.5536	0.0194	1.3957	2.3575	0.0280

Table 6

Forecasting Errors of the UK Power Prices

	Day-Ahead			**Week-Ahead**		
Classic Methods:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
ARIMA	2.1462	5.4182	0.1459	2.5220	7.4728	0.2092
X-12	3.0506	10.6172	0.3066	6.6911	74.2138	1.4268
Haar Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	2.2438	5.6306	0.1605	2.6670	8.2682	0.2336
Denoising + X-12 forecast	3.0503	10.6041	0.3077	2.5025	88.4410	1.7801
Multiscale forecast (ARIMA)	2.2604	5.7186	0.1557	9.0070	111.1020	2.4665
Multiscale forecast (X-12)	2.1715	6.6690	0.1575	5.1985	36.2447	0.8189
Wavelet-based STSM	3.5868	15.6367	0.3823	9.2260	90.2059	2.6168
Daubechies/Morlet Wavelet:	**MAD**	**RMSE**	**APE**	**MAD**	**RMSE**	**APE**
Denoising + ARIMA forecast	2.2438	5.6306	0.1605	2.6670	8.2682	0.2336
Denoising + X-12 forecast	3.0573	10.6209	0.3092	4.1623	34.8051	0.6011
Multiscale forecast (ARIMA)	2.2921	6.1810	0.1566	4.1623	34.8051	0.6011
Multiscale forecast (X-12)	2.1237	5.2731	0.1469	2.6213	9.1848	0.2492