IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 4/2010

Using Wavelets for Time Series Forecasting – Does it Pay Off?

Stephan Schlüter
University of Erlangen-Nuremberg
Carola Deuschle
University of Erlangen-Nuremberg

ISSN 1867-6707

Friedrich-Alexander-Universität
IWQW
Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Using Wavelets for Time Series Forecasting – Does it Pay Off?

Stephan Schlüter, Carola Deuschle

Department of Statistics and Econometrics, University of Erlangen-Nuremberg, Lange Gasse 20, Nuremberg, D-90403
Email: stephan.schlueter@wiso.uni-erlangen.de

Abstract

By means of wavelet transform a time series can be decomposed into a time dependent sum of frequency components. As a result we are able to capture seasonalities with time-varying period and intensity, which nourishes the belief that incorporating the wavelet transform in existing forecasting methods can improve their quality. The article aims to verify this by comparing the power of classical and wavelet based techniques on the basis of four time series, each of them having individual characteristics.

We find that wavelets do improve the forecasting quality. Depending on the data's characteristics and on the forecasting horizon we either favour a denoising step plus an ARIMA forecast or an multiscale wavelet decomposition plus an ARIMA forecast for each of the frequency components.

JEL-Classification: C22, C53.

Keywords: Forecasting; Wavelets; ARIMA; Denoising; Multiscale Analysis.

1 Introduction

Forecasting prices of stocks, commodities or derivatives on liquid markets is in a large part guesswork. However, we can try to use the information contained in historical data to estimate future developements, which is what parametric statistical models do: It is assumed that the given time series is the realisation of an underlying stochastic process with a certain specification. The forecast is generated by extrapolation while eliminating the random element by taking the expectation.

A widely used approach is the Autoregressive Moving Average (ARIMA) model, which captures intertemporal linear dependence in the data itself as well as in the error term. Existing trends are treated by modeling not the data but the differences which is then called Autoregressive Integrated Moving Average (ARIMA) model (see McNeil et al., 2006). In contrary to these technical approaches, Majani (1987) assumes that the time series can be split into a sum of some deterministic components, e.g. seasonal oscillation and linear trend, and a stochastic error term. There are different ways to estimate the individual components. The seasonal component, for example, can be identified using the Kalman filter or via Fourier transform. However, the quality of both methods suffers, if the season has a changing period and/or intensity.

This is why Wong et al. (2003) use the wavelet transform, which is able to capture dynamics in period and intensity, to model both the trend and the seasonality. By means of the wavelet transform we can decompose a time series into a linear combination of different frequencies. With some restrictions we are able to quantify the influence of a pattern with a certain frequency at a certain time on the price. Having such a feature the wavelet transform is likely to improve the quality of forecasting. Besides estimating the components of Majani's model there are various concepts: Donoho & Johnstone (1994) apply them to filter the error term's influence (denoising). Conejo et al. (2005) decompose the time series into a sum of processes with different frequencies, and forecast the individual time series before adding up the results. It is assumed that the motions on different frequencies follow different underlying processes and that treating them separately should increase the forecasting quality. A rather technical concept is the locally stationary wavelet process (see Fryzlewicz, 2005), where the price process is written as a linear combination using wavelets as basis functions.

There are, however, some shortcomings: Using wavelets increases the model complexity, more parameters have to be estimated and a wavelet function has to be chosen. The number of approximation steps increases, i.e. there are more error sources. This paper intends to verify empirically wether it does really pay off to use the wavelet transform for forecasting. We choose four data sets with different characteristics: From the oil prices, where the long-term structure dominates, via the Euro-Dollar exchange rate and the Deutsche Bank stock price to the UK day ahead power price which shows heavy daily, i.e. short-term oscillations (see Figure 2). We perform a one day ahead and one week ahead forecast using the three classical methods as well as each of these techniques in combination with a wavelet denoising step and a wavelet decomposition scheme. In addition to that we try the approach of Wong et al. (2003) and compute forecasts based on locally stationary wavleet processes.

What we see is: It pays off to use wavelets to reduce forecasting errors, however, there is no method performing best across all scenarios. The optimal choice depends on both the time series characteristics like volatility or existence of long-term trends and the forecasting horizon. The methods to choose from are wavelet based denosining plus ARIMA forecast and wavelet based time series decomposition plus an ARIMA forecast of the individual frequency components. Using locally stationary wavelet processes fails completely and so does estimating Majani's structural component model via wavelets.

The paper is structured as follows: We present the classical forecasting techniques in Section 2. In Section 3 we give an introduction into wavelet transform and describe the corresponding forcasting methods. In Section 4 we present the data sets and discuss the results of our empirical study. Section 5 concludes.

2 Classical Forecasting Models

We present three well-known classical forecasting methods: the structural time series model (STSM), the autoregressive moving average (ARMA) model and its extension, the autoregressive integrated moving average (ARIMA) model.

2.1 Structural Time Series Model

The structural time series model describes a process $(X_t)_{t\in\mathbb{Z}}$ at time t as a sum of a long-term trend T_t, a seasonal component S_t and a random (noise) term U_t (see Majani, 1987):

$$X_t = T_t + S_t + U_t \tag{2.1}$$

An analogous version ,in which X_t is the product of the above factors, can be obtained by applying the logarithm. Trend and season are expected to be deterministic, but we can also design them to be stochastic (see Harvey, 1989).

The exact shape of T_t and S_t depends on how both components are estimated. Common methods are the moving average method, Fourier transform, Kalman filter or exponential smoothing. A more sophisticated version is to explain T_t as a function $f(t;\beta_1,\ldots,\beta_n)$ with parameters $\beta_1,\ldots,\beta_n \in B$, where B is the parameter domain. Examples for f are $f(t) = \beta_1 f_1(t) + \ldots + \beta_n f_n(t) + \epsilon_t$ or $f(t) = f_1(t)^{\beta_1} + \ldots + f_n(t)^{\beta_n} + \epsilon_t$, where ϵ_t is a noise term and $f_1,\ldots,f_n$ some functions. The parameters are estimated via least squares method or in more complex scenarios via numerical methods like the Gauss-Newton algorithm. The seasonal component S_t is commonly estimated via Fourier transform or dummy variables (see Harvey, 1989). Both methods, however, require a true seasonal pattern with fixed period and intensity to provide sound estimation results. Forecasting with (2.1) is done by extrapolating both T_t and S_t, and expecting $E(U_t) = c \in \mathbb{R}$.

2.2 Autoregressive Moving Average

The autoregressive moving average model (ARMA) of order $(p, q) \in \mathbb{N}^2$ is a linear model which comprises of an autoregressive and a moving average term; it describes a process $(X_t)_{t\in\mathbb{Z}}$ of the form

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \quad \phi_i, \theta_k \in \mathbb{R} \ \ \forall i \in 1, \dots, p \ j = 1, \dots, q. \tag{2.2}$$

The ϵ_t is a random variable with a given distribution F and $\mu \in \mathbb{R}$ is the drift (see McNeil et al., 2006). The moving average part is weakly stationary by definition and the autoregressive part is weakly stationary if for all $z \in \mathbb{C}$ that fulfill

$$1 - \phi_1 z - \dots - \phi_p z^p = 0$$

holds: $|z| > 1$. Having weakly stationarity and a Gaussian random term causes that the whole process is mean and covariance ergodic. We can then compute consistent estimatees for the mean and covariance of X_t based on the historical data.

The parameters of (2.2), including those of F, are estimated using the regression method of Durbin (1960), 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 are necessary. For determining the lag-order (p, q) of the process, methods like overfitting or measures like the Bayesian information criterion, that punishes a high number of variables, can be used (see McNeil et al., 2005). A comparison of different criteria is given by Kereisha & Pukkila (1995).

The h-step forecast $(h \in \mathbb{N})$ $\hat{X}_{t+h}$ is obtained by computing the expected value of (2.2) conditional on the filtration up till time t denoted by $\mathcal{F}_t$:

$$\hat{X}_{t+h} = E\left[\mu + \sum_{i=1}^{p} \theta_i X_{T+h-p} + \epsilon_{T+h} + \sum_{j=1}^{q} \theta_j \epsilon_{T+h-j} \middle| \mathcal{F}_t\right]. \tag{2.3}$$

Because the conditional expectation is a linear function we additionally know

$$E\left[X_{T+j}\right] = \begin{cases} X_{T+j} & \text{if } j \leq 0, \\ \hat{X}_{T+j} & \text{else,} \end{cases} \qquad E\left[U_{T+j}\right] = \begin{cases} X_{T+j} - \hat{X}_{T+j} & \text{if } j \leq 0, \\ \hat{X}_{T+j} & \text{else.} \end{cases} \tag{2.4}$$

Thereby we see that the prediction converges to the long run mean for $h \to \infty$.

2.3 Autoregressive Integrated Moving Average

If the ARMA(p,q) process is not weakly stationary (e.g. because of a trend), one can try to achieve stationarity by computing differences and then modeling the new time series $\Delta X_t = X_t - X_{t-1}$. This procedure can be repeated and we speak of an autoregressive integrated moving average (ARIMA) process with integration order $d \in \mathbb{N}$, if $\Delta^d X_t$ is weakly stationary. For more sophisticated versions like the seasonal ARIMA model (SARMA) or the fractional integrated ARIMA (ARFIMA) model, where the model has a long-term memory, we refer to Granger & Joyeux (1980) and Hosking (1981).

The optimal h-step forecast ($h \in \mathbb{N}$) for an ARIMA(p,d,q) model is computed in two steps. First, we forecast $Y_t = \Delta^d X_t$ applying (2.3) and (2.4) to obtain an estimate for the differences $\hat{Y}_{t+h}$, then we use the relation $Y_{t+h} = (1-B)^d X_{t+h}$, with $BX_{t+h} = X_{t+h-1}$, to obtain a forecast for X_{t+h} (see McNeil et al., 2005).

3 Wavelet Based Forecasting

As suggested in the introduction we can try to improve the forecasting accuracy of the above methods by using wavelets. Before presenting the individual concepts we give a few basic definitions of wavelet theory.

3.1 A Brief Introduction Into Wavelet Theory

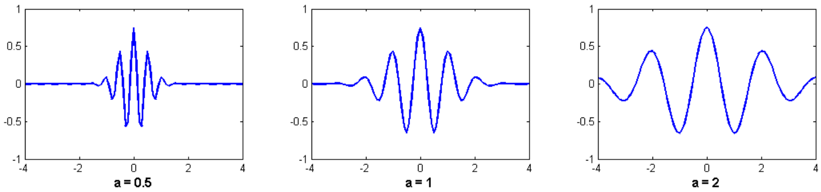
The continuous wavelet transform (CWT) generalizes the Fourier transform and is, unlike the latter, able to detect seasonal oscillations with time-varying intensity and frequency. Stationarity of the process is not required but square-integrability is (see Mallat, 2003). In the sequel we focus on the CWT, for an introduction into the discrete wavelet transform see Kaiser (1994) or Jensen & Cour-Harbo (2001).

The CWT is a complex-valued function $\Psi(t) \in \mathcal{L}^1(\mathbb{R}) \cap \mathcal{L}^2(\mathbb{R})$ that fulfills the admissibility condition $C_\Psi = \int_{-\infty}^{\infty} |\hat{\Psi}(\omega)|^2 / |\omega| \, d\omega < \infty$, where the hat denotes the Fourier transform. Each Ψ has a fixed mean and frequency. To make it more flexible, set $\Psi_{a,b} = \Psi\left(t/a - b\right)/\sqrt{a}$, which translates Ψ by $b \in \mathbb{R}$ and a scaling factor $a > 0$ that is inverse proportional to the frequency. The CWT is the orthogonal projection of a process $(S_t)_{t\in\mathbb{R}}$ on $\Psi_{a,b}$, i.e.

$$WT_S(a,b) = \langle S, \Psi_{a,b} \rangle = \int_{\mathbb{R}} S_t \overline{\Psi_{a,b}(t)} dt, \tag{3.1}$$

where the overline denotes the conjugate complex (see Mallat, 2003). The $WT_S(a,b)$ indicates how much of S is explained by the local oscillation Ψ in time b with scale a.

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



The inverse transform is therefore a linear combination of Ψ and in the continuous case a double integral (see Mallat, 2003) of the form

$$S(t) \quad = \quad \frac{1}{C_\Psi}\int_0^{a_0}\int_{-\infty}^{\infty} WT(a,b)\frac{1}{a^2\sqrt{a}}\Psi\left(\frac{t-b}{a}\right)dbda. \tag{3.2}$$

We can simplify (3.2) significantly when having a discrete data set, e.g. daily commodity prices. Then 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 (see Shannon, 1949).

When identifying the influence of patterns with a certain scale/frequency (seasonalities, for example), we have to respect the uncertainty principle of time-frequency analysis, which says that not both scale and location of a signal can be exactly specified simultaneously (see Lau & Weng, 1995). We are limited to an analysis of time-frequency windows and the only lever we can pull is to choose an adequate wavelet function. For various selection criterions see e.g Ahuja et al. (2005). The best one regarding window size is the Morlet wavelet $\Psi_M(t)$ with

$$\Psi_M(t) = c_{\omega_0}\pi^{-1/4}e^{-t^2/2\sigma^2}\left(e^{i\omega_0 t} - e^{-\frac{1}{2}\omega_0^2}\right), \;\; c_{\omega_0} = \left(1 - e^{-\omega_0^2} - 2e^{-\frac{3}{4}\omega_0^2}\right)^{-\frac{1}{2}},$$

where $\omega_0 > 0$ denotes the basis frequency and $\sigma > 0$ (see Daubechies, 1992). It is plotted in Figure 1 at three different scales for $b = 0$, and its time-frequency 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 the a set called cone of influence (CoI) that reads as $[b - \lceil (s_u - s_l)a \rceil, b + \lfloor (s_u - s_l)a \rfloor]$, where $[s_l, s_u] \subseteq \overline{\mathbb{R}}$ is the support of Ψ (see Lau & Weng, 1995). If data within the CoI are missing for a time t and a scale $a > 0$, $WT_S(a,t)$ from (3.1) is skewed, which especially holds for the edge regions of a finite data set. For methods to reduce this effect see Meyers et al. (1993), Jensen & Cour-Harbo (2001) or Torrence & Compo (1998).

Applying this theory to a data set $S_t, t = 1, \ldots, T$, with $dt = 1$ we set $b \in \mathbb{Z}$. For discretizing the scale grid, most authors (e.g. Torrence & Compo, 1998) use a dyadic approach to form a set of scales $A = \{a_1, \ldots, a_J\}$, which reads as

$$a_j = a_0 2^{j\delta j} \;\; j = 0, 1, \ldots, J, \;\text{ and }\; J = \left\lfloor \delta j^{-1} \log_2 \left(T/a_0\right)\right\rfloor + 1, \tag{3.3}$$

where $a_0, \delta j \in \mathbb{R}^+$. The grid is finer for lower scales, which means that for higher frequencies we look closer at the process. This makes sense as we expect to have more detail information there than in lower frequencies (i.e. higher scales). This is also why we want to avoid computing (3.2) for higher scales as this implies unnecessary high numerical effort. The goal is to aggregate the influence of all scales larger than $a^* > 0$ on S_t and we can do this by introducing a scaling function ϕ that behaves like a low-pass filter. There is a huge variety of scaling functions (see e.g. Ahuja et al., 2005) but when operating together with a wavelet Ψ it has to fulfill

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

Now allowing for a rescale and a shift, i.e. set $\phi_{a,b}(t) = \phi((t-b)/a)/\sqrt{a}, a > 0, b \in \mathbb{Z}$, we are able write for a time series S_t and a scale $a^* \in A$

$$S_t = \frac{1}{C_\Psi} \sum_{b \in \mathbb{Z}} \langle S, \phi_{a^*,b} \rangle \, \phi_{a^*,b}(t) + \frac{1}{C_\Psi} \sum_{b \in \mathbb{Z}} \sum_{a \in A \wedge a > a^*} \langle S, \Psi_{a,b} \rangle \, \Psi_{a,b}(t) \frac{1}{a^2} \quad \forall\, t \tag{3.4}$$

(see Mallat, 2003). In (3.4) we see the reduced effort as for scales larger than a^* the double sum is substituted by a simple sum.

However, the CWT is still computationally very intensive. Depending on the application scenario a more efficient technique, the à trous algorithm of Holschneider et al. (1989), may be applied. Below we describe how it works: Let $(S_t)_{t \in \mathbb{Z}}$ be a discrete time process and $(g_k, h_k)_{k \in \mathbb{Z}}$ filter banks with $d_n^m = \langle S, \Psi_{m,n} \rangle, c_n^m = \langle S, \phi_{m,n} \rangle$ and

$$d_n^m = \sum_{k \in \mathbb{Z}} g_k c_k^{m-1}, \quad c_n^m = \sum_{k \in \mathbb{Z}} h_k c_k^{m-1}$$

for a set of scaling functions $\left\{\phi_{m,n}(\bullet) = \phi\left(\bullet/2^m - n\right)/\sqrt{2^m} : m, n \in \mathbb{Z}\right\}$ and a corresponding set of wavelet functions $\left\{\Psi_{m,n}(\bullet) = \Psi\left(\bullet/2^m - n\right)/\sqrt{2^m} : m, n \in \mathbb{Z}\right\}$. Set $d^m = \{d_n^m : n \in \mathbb{Z}\} \in \mathcal{L}^2(\mathbb{Z})$ and $c^m = \{c_n^m : n \in \mathbb{Z}\} \in \mathcal{L}^2(\mathbb{Z})$. Let now h^r, g^r be recursive filters with $g^0 = h, h^0 = g$. The 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

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

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

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

For the reconstruction of the time series we start with $M, c^M, d^m, m = 1, \dots, M$ and gradually compute $\forall \ m = M, M-1, \dots, 1 : c^{m-1} = H^{m*} c^m + G^{m*} d^m$. The result is c^0 and from that we yield the time series via inversion of the respective convolution.

3.2 Wavelet Based Forecasting Methods

Basically there are four different methods: One is to use wavelets for eliminating noise in the data or to estimate the compoents in a STSM. Another one is to do the forecasting directly on the wavelet generated time series decomposition, or we can use locally stationary wavelet processes.

3.2.1 Wavelet Denoising

Wavelet denoising is based on the assumption that a data set $(S_t)_{t=1,\dots,T}$ is the sum of a deterministic function Y_t and a white noise component $\epsilon_t \sim \mathcal{N}(0, \sigma^2), \sigma > 0$, i.e. $S_t = Y_t + \epsilon_t$. Reducing the noice yields a modified S_t on which the forecasting methods from Section 2 can be applied (see Alrumaih & Al-Fawzan, 2002).

The denoising is accomplished as follows: Initially, the wavelet transform is applied to S_t with a scale discretization $A = \{a_1, \dots, a_n\}$ and $b = 1, \dots, T$ with $n \in \mathbb{N}$. The result is a matrix of wavelet coefficients $WT \in \mathbb{R}^{n \times T}$. These indicate how much of S_t is described by a scaled and translated wavelet $\Psi_{a,b}$. Nason (2008) shows that per definition the noise term has impact on each coefficient, whereas the information of Y_t is concentrated only in a few coefficients. So if $WT(a, b)$ from (3.1) 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. Therefore we set all coefficients below a certain threshold $\lambda \geq 0$ to zero, and eventually invert the modified coefficients $WT'(a,b)$ using (3.2) to obtain the modified time series S_t'. Donoho & Johnstone (1994) propose two different thresholds:

$$(a) \quad WT'(a,b) = WT(a,b)\mathbf{1}_{\{|WT(a,b)|>\lambda\}} \qquad \text{(hard threshold)},$$

$$(b) \quad WT'(a,b) = sgn(WT(a,b))(|WT(a,b)| - \lambda)\mathbf{1}_{\{|WT(a,b)|>\lambda\}} \qquad \text{(soft threshold)},$$

where *sgn* denotes the signum function. The larger λ, the more noise but also more of Y_t is cut out, and vice versa. Donoho & Johnstone (1994) propose $\lambda_{universal} = \hat{\sigma}\sqrt{2\log T}$, where $\hat{\sigma}$ is an estimator for σ. 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.

Donoho & Johnstone (1994) derive another threshold based on the SURE[1] estimation method developed by Stein (1981). They derive for a scale a the optimal threshold λ_{SURE} by solving

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

$$\text{with} \qquad SURE(WT,\lambda) = T - \#\{t : |WT(a,t)| \leq \lambda\} + \sum_{t=1}^{T} \min(|WT(a,t)|,\lambda)^2.$$

As the SURE-method works not very good for sparsely ocuppied matrices, Donoho & Johnstone (1994) unite both concepts in the SureShrink method, which uses $\lambda_{universal}$ as threshold if $\sum_t (WT(a,t)^2 - 1) \leq \log_2 T^{3/2}$ for $a \in A$ and λ_{SURE} otherwise. Gao & Bruce (1997) or Breiman (1996) propose further threshold rules.

3.2.2 Wavelet-based Estimation of the Three Components Model

Wong et al. (2003) use wavelets to estimate the components in the STSM from Section 2.1, i.e. they model a process $(S_t)_{t\in\mathbb{Z}}$ as the sum of a trend T_t, a season X_t and a noise ϵ_t, i.e.

$$S_t = T_t + X_t + \epsilon_t \quad t \in \mathbb{Z}.$$

They give estimators $\hat{T}_t, \hat{X}_t$ for trend and seasonality, and do the forecasting by extrapolating from polynomial functions fitted to $\hat{T}_t$ and $\hat{X}_t$. To $\hat{\epsilon}_t = S_t - \hat{T}_t - \hat{X}_t$ they fit an ARMA(1,0) to compute a forecast.

[1] Stein's Unbiased Risk Estimator

The $\hat{T}_t$ is computed by aggregating the high-scale patterns using a scaling function ϕ as described in Section 3.1, i.e.

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

which is for discret-time data a linear combination of the price's observations, as the convolution integral is approximated by a sum. It remains to choose a scaling function and the optimal scale a^*, which should be small enough to capture the whole trend, but large enough not to cut through some short-term oscillations.

For estimating X_t, Wong et al. (2003) use the hidden periodicity analysis (see Appendix B).

3.2.3 Forecasting Based on a Wavelet Decomposition

The time series $(S_t)_{t=1,\ldots,T}$ is transformed via (3.1) to obtain the wavelet coefficients $WT(a,b), a \in A, b = 1,\ldots,T$, where A denotes a scale discretization. For each a the corresponding vector $WT(a) = (WT(a,1),\ldots,WT(a,T))$ is treated as a time series, and standard techniques like ARMA-based forecasting are applied to obtain wavelet coefficient forecasts, which are subsequently added to the matrix WT (see Conejo et al., 2005, or Yousefi et al., 2005). Renauld et al. (2005) use only specific coefficients for this forecast which is very efficient but increases the forecasting error. The extended matrix WT' is then inverted according to (3.4) and we yield a forecast $\hat{S}_{t+1}$ for the S_t in the time space.

There are a few arguments that speak in favour of this way of forecasting. Among others, Soltani et al. (2000) show that when decomposing time series with long-term memory, the processes of wavelet coefficients at each scale lack this feature. There is also no long-term dependence between different scales. Abry et al. (1995) come to a similar result for fractional brownian motions.

3.2.4 Locally Stationary Wavelet Processes

The locally stationary wavelet process is a model developed to handle second-order dependence structure, i.e. time-varying variance, and is based on a general class of instationary processes developed by Dahlhaus (1997). Let $(S_t)_{t=1,\ldots,T}$ be a set of realizations of a process $(S_t)_{t\in\mathbb{Z}}$. For a fixed $T \in \mathbb{N}$ we define a locally stationary wavelet process for a wavelet Ψ with compact support of length $s_m \in \mathbb{R}$ as

$$S_{t,T} = \sum_{m=1}^{M(t)} \sum_{n \in \mathbb{Z}} \omega_{m,n;T} \Psi_{m,n}(t) \xi_{m,n}, \quad \text{with } S_{t,T} = a\left(\frac{t}{T}\right) S_{t-1,T}. \tag{3.5}$$

Thereby $a(\bullet) \in (0,1], \omega_{m,n} \in \mathbb{R}$, $M(T) = \max\{m \in \mathbb{N} : s_m \leq T\}$ and $E[S_{t,T}] = 0 \; \forall \; t, T$. The random variable $\xi_{m,n}$ and the whole process have to fulfill some regularity conditions (see Appendix C and Nason et al., 2000). The problem now is that (3.5) is not unique as Ψ doesn't have to be orthogonal. We therefore can't simply use (3.5) to compute the coeffients $(b_{1,T}, \ldots, b_{N,T})$ of our linear forecasting function

$$\hat{S}_{N+1,T} = \sum_{n=1}^{N} b_{N+1-n,T} S_{n,T}, \quad N < T. \tag{3.6}$$

Instead of Ψ we use the local autocovariance fuction

$$c(z,\tau) = \sum_{m=1}^{\infty} WS_m(z)\Psi_m(\tau), \quad \Psi_m(\tau) = \sum_{n\in\mathbb{N}} \Psi_{m,n}(0)\Psi_{m,n}(\tau), \;\; z \in (0,1], \tau \in \mathbb{Z},$$

which is based on the unique wavelet spectrum $WS_m(z), m \in \mathbb{N}, z \in (0,1]$ and converges against the autocovariance function of the process itself (see Nason et al., 2000). This autocovariance is estimated by Nason et al. (2000) via

$$\hat{c}\left(\frac{n}{T}, \tau\right) = \sum_{m=1}^{M(T)} \left(\sum j = 1^{M(T)} A^{-1}_{m,j} \Psi_j(\tau)\right) WT^2_{m,n;T}, \;\; n = 1, \ldots, N,$$

with $A_{m,j} = \sum_\tau \Psi_m(\tau)\Psi_j(\tau)$ and $\Psi_m(\tau) = \sum_{n\in\mathbb{N}} \Psi_{m,n}(0)\Psi_{m,n}(\tau)$. Fryzlewicz et al. (2003) now show that the parameter vector which minimizes the mean squared error is the vector which minimizes

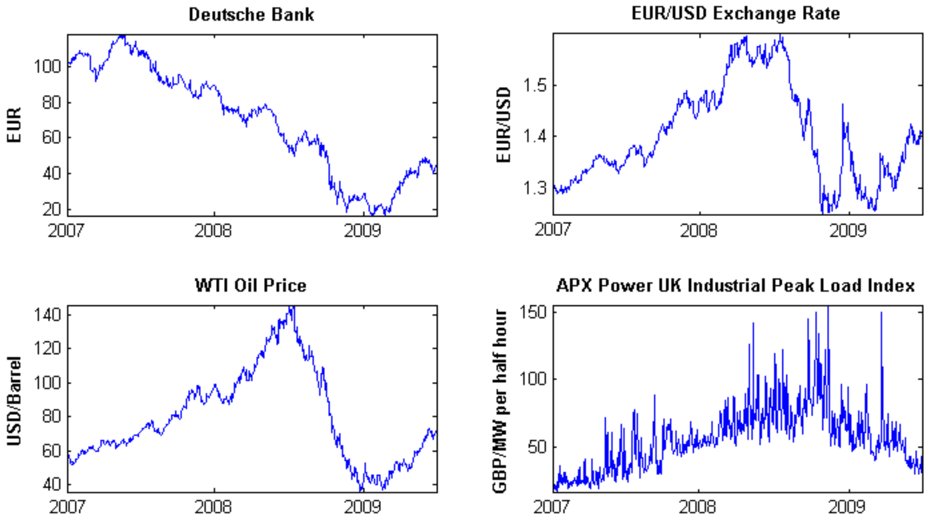
$$\sum_{k=1}^{N} b_{N+1-k;T} c\left(\frac{k+p}{2T}, k-p\right) = c\left(\frac{N+1+p}{2T}, N+1-p\right), \;\; p = 1, \ldots, N. \tag{3.7}$$

The forecasting procedure is as follows: Initially, the wavelet transform is applied to the time series, the local autocovariance is estimated and used to solve (3.7). The minimum argument of this equation is an estimate for $(b_{1,T}, \ldots, b_{N,T})$ in (3.6).

4 Comparing Forecasting Methods Empirically

The presented methods are applied to real data in order to evaluate how wavelet based techniques perform compared to the classical ones. We first present the chosen time series, describe the test design and then comment on the estimation results.

Figure 2: The Analyzed Data Sets



4.1 The Data Sets

The forecasting is done based on four time series displayed in Figure 2, each having its own characteristics. One data set is the West Texas Intermediate (WTI) oil price which serves as an example for commodity prices. It shows a comparatively strong long-term pattern which dominates the short-term oscillation. The Deutsche Bank stock prices clearly show a long-term trend, only a few minor price jumps and some medium-term oscillations. From the foreign exchange market we take the Euro/Dollar exchange rate which has a visible long-term component, a minor medium-term structure but some distinct price jumps. The UK day ahead power prices (provided by the APX Group) 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 from January 1st 2007 until June 30th 2009. The UK power prices include weekends and range from July 7th 2007 until March 13th 2009.

4.2 Test Design and Goodness of Fit Measures

We compute one day ahead and one week ahead forecasts, which is a step of seven days for the power prices and of five days for the other three data sets as these exclude weekends. We use the Census X-12 method of Findley et al. (1988) to implement the STSM, and test the ARMA and the ARIMA model. To separate ARIMA from ARMA we set the difference order to one.

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-dependend moving average (see Stollnitz et al., 2005), the Morlet wavelet, which has the best time-frequency resolution, and the Daubechies D4 wavelet (see Appendix D) which works well with efficient techniques like the à trous algorithm (see Daubechies, 1992). For the Morlet wavelet we set $a_0 = 2, \delta j = 0.6$ in (3.3) and for Haar's function we choose $a_0 = 1, \delta j = 1$.

We try the denoising and use the SureShrink method once together with the Haar and once with Morlet's wavelet. To the modified time series we apply the three classical forecasting methods. In addition we try the multiscale forecasting, which is done based on the Haar and the Daubechies D4 wavelet as both work with the à trous algorithm. To forecast the decomposed time series we use Census X-12, ARMA and ARIMA. Eventually we try a wavelet based STSM and locally stationary wavelet processes. For both techniques we use the Haar wavelet.

We do out of sample forecasts for the last n data points of each time series, where $n = 14$ for the power prices and 10 for the rest. The results are evaluated using three error measures, namely the root mean squared error (RMSE), the mean absolute deviation (MAD) and the mean average percentage error (MAPE). The latter two are defined as

$$MAD(S, \hat{S}) = \sqrt{\sum_{i=T-n+1}^{T} \left| S_i - \hat{S}_i \right| / n}, \quad MAPE(S, \hat{S}) = \sqrt{\sum_{i=T-n+1}^{T} \left| S_i - \hat{S}_i \right| / S_i n},$$

for $S_t, t = 1, \ldots, T$ being a data set and $\hat{S}_t, t = 1, \ldots, T$ the corresponding estimates.

4.3 Presenting and Evaluating the Estimation Results

The result of our empirical study is given in Appendix E in tabular form. We see that for each data set the methods perform differently:

Deutsche Bank stock price: The Haar multiscale ARIMA model is the best in the one step ahead forecast regarding all three error measures, although the APE is equally low for the simple ARIMA model and the Haar multiscale ARMA forecast. Actually, the classical ARIMA forecast gives comparably well forecasts, which are better than all wavelet based ones except those mentioned above. The results change a bit when looking at the one week ahead forecast. The best method is now the Daubechies D4 multiscale Census X-12 forecast, although the same procedure but an ARIMA forecast produces only slightly higher error values.

Euro/Dollar exchange rate: In the one day ahead forecast all Haar multiscale decompositions perform better than the classical ones. The lowest errors produce the multiscale ARIMA forecast. The Daubechies D4 multiscale forecasts perform about equal to the classical methods, whereas all further techniques show higher errors. Things change for the one week ahead forecast. Only wavelet based denoising procedures achieve lower errors than the classical forecating methods. Among those, a Morlet or Haar based denoising plus ARMA forecast performs best.

The WTI oil price: In the one day ahead forecast, denoising proves to be the method of choice whereby the ARIMA forecast gives the lowest errors. The Haar multiscale ARMA forecast does also reasonably well, whereas the other wavelet methods perform not as good as the classical ones. In the one week ahead forecast, denoising still helps to lower the error, but now the Daubechies D4 multiscale forecast shows the smallest errors whereby ARIMA and Census X-12 forecasting produce almost equally low errors.

UK power prices: In case of the UK day ahead power prices, the classical ARIMA model does fairly well both in the one day ahead and in the one week ahead forecast. In the one step ahead forecast it achieves the lowest APE and is slightly worse than a Daubechies D4 multiscale Census X-12 forecast regarding MAD and RMSE. The further wavelet based methods perform worse. In the one week ahead forecast it again shows the smallest RMSE and APE, and only the MAD of a Haar based denoising plus Census X-12 forecast is lower.

The above estimation results indicate a few things about the tested forecasting techniques. We see that the classical STSM and the Census X-12 technique as its numerical implementation produces less exact forecasts than an ARIMA model. This shows that trends and seasonality are better treated by computing differences than modeling them as individual components. Wavelet based denoising doesn't help to improve the power of the STSM. Combined with an ARMA or ARIMA forecast, however, denoising substantially helps to improve the models quality as we can see for the oil prices and the exchange rate time series.

Multiscale forecasts also improve the forecasting quality in almost all scenarios. However, it varies from time series to time series which wavelet and which forecasting method gives the best estimation results. ARIMA is doing reasonably well in contrary to the Census X-12, whose performance is unstable: in some cases it helps to lower the forecasting errors but in most scenarios it is even not as good as the classical ARIMA model.

The forecasting power of locally stationary wavelet processes and wavelet based STSMs is not convincing. The error measures are always and in most cases significantly higher than those of the classical methods and also of the other wavelet based techniques.

5 Summary and Conclusion

The purpose of this paper was to compare the power of the main classical forecasting methods and wavelet based extensions of them. For this purpose we first presented the classical structural time series model and the ARMA/ARIMA approach. We also gave a brief introduction into wavelet theory and described how wavelet functions can be used in time series forecasting. For our empirical empirical study we chose four time series with different characteristics. We tested two different forecasting horizons (one day, one week) and compared the results using three standard error measures.

The results cannot confirm the statements of Wong et al. (2003) or Fryzlewicz et al. (2003), who say that wavelet based STSMs or locally stationary wavelet processes improve the forecasting quality. Besides that, in all scenarios, we were able to find a wavelet based method that performs better than the classical techniques, however, the method of choice depends on the time series characteristics. We see that for time series with a strong random component like the UK power prices wavelets generate only little improvements. If the long-term structure is more important than the short-term oscillation, as we can see in the oil prices, then denoising plus ARIMA forecasting is the method of choice. If again, the prices consist of a mixture-term structure and an important oscillation – see the exchange rates and the Deutsche Bank stock prices– then the Multiscale ARIMA forecasting delivers quite good results.

References

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

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

[3] Alrumaih RM, Al-Fawzan MA. 2002. Time series forecasting using wavelet denoising: an application to Saudi Stock Index. Journal of King Saud University, Engineering Sciences **2**(14): 221-234.

[4] Breiman L. 1996. Heuristics of instability and stabilization in model selection. The Annals of Statistics **24**(6): 2350-2383.

[5] Conejo AJ, Plazas MA, Espinola R, Molina AB. 2005. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Transactions on Power Systems **20**(2): 1035-1042.

[6] Dahlhaus R. 1997. Fitting time series models to nonstationary processes. The Annals of Statistics **25**(1): 1-37.

[7] Daubechies I. 1992. Ten lectures on wavelets. Society for Industrial and Applied Mathematics: Philadelphia.

[8] Donoho D, Johnstone I. 1994. Ideal spatial adaptation via wavelet shrinkage. Biometrika **81**: 425-455.

[9] Durbin J. 1960. The fitting of time series models. The International Statistical Review **28**: 233-244.

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

[11] Findley DF, Monsell BC, Otto MC, Pugh MG. 1988. Towards X-12-ARIMA. Proceedings of the Fourth Annual Research Conference, Washington, D.C.; U.S. Bureau of the Census: 591–624.

[12] Fryzlewicz P. 2005. Modelling and forecasting financial log-returns as locally stationary wavelet processes. Journal of Applied Statistics **32**: 503-528.

[13] Fryzlewicz P, van Bellegem S, von Sachs R. 2003. Forecasting non-stationary time series by wavelet process modelling. Annals of the Institute of Statistical Mathematics **55**: 737-764.

[14] Gao HY, Bruce AG. 1997. WaveShrink with firm shrinkage. Statistica Sinica **7**: 855-874.

[15] Granger CWJ, Joyeux R. 1980. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis **1**(1): 15-29.

[16] Harvey AC. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press: Cambridge.

[17] Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P. 1989. Wavelets, time-frequency methods and phase space. Springer: Berlin; 289–297.

[18] Hosking, JRM. 1981. Fractional Differencing. Biometrika **68**(1): 165-176.

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

[20] Kaiser G. A friendly guide to wavelets. 1994. Birkhäuser: Boston.

[21] Koreisha, SG. Pukkila T. 1995. A Comparison between different order- determination criteria for identification of ARIMA models. Journal of Business & Economic Statistics **13**(1): 127-131.

[22] 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): 2391–2402.

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

[24] Majani, BE. 1987. Decomposition methods for medium-term planning and budgeting. Makridakis S, Wheelwright S (edt.): The handbook of forecasting: A manager's guide; Whiley: New York; 219–237.

[25] Mallat S. 2003. A wavelet tour of signal processing, 2nd edition. Academic Press: Manchester.

[26] McNeil AJ, Frey R, Embrechts P. 2005. Quantitative risk management: concepts, techniques, and tools. Princeton University Press, Princeton.

[27] Meyers SD, Kelly BG, O'Brien JJ. 1993. An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of yanai waves. Monthly Weather Review **121**(10): 2858-2866.

[28] Nason GP, von Sachs R, Kroisandt G. 2000. Wavelet processes and adaptive estimation of evolutionary wavelet spectra. Journal of the Royal Statistical Society. Series B **62**: 271-292.

[29] Nason GP. 2008. Wavelet methods in statistics with R. New York: Springer; 83–132.

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

[31] Shannon CE. 1949. Communication in the presence of noise. Proceedings of the Institute of Radio Engineers **37**(1): 10-21.

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

[33] Stollnitz EJ, DeRose TD, Salesin DH. 1995. Wavelets for computer graphics: a primer, part 1. IEEE Computer Graphics and Applications **15**(3): 76–84.

[34] Torrence C, Compo GP. 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society **79**(1): 61–78.

[35] Wong H, Ip WC, Xie Z, Lui X. 2003. Modelling and forecasting by wavelets, and the application to exchange rates. Journal of Applied Statistics **30**(5): 537-553.

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

A The Time-Scale Window of Morlet's Wavelet

For $a > 0$ and $b \in \mathbb{R}$ the time-scale window of $\Psi_M(t)$ is (see Fabert, 2004)

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

B Hidden Periodicity Analysis

Let $(S_t)_{t=1,\dots,T}$ be a time series with an estimated trend $\hat{T}$. Let $\hat{Y}_t = S_t - \hat{T}_t$ and assume

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

with a complex random variable $\alpha_n, n = 1, \dots, N$, which has finite variance, no autocorrelation and it holds $0 < \alpha < \|\alpha_n\|^2, \alpha \in \mathbb{C}$. The random variable ξ_t is a linear combination of ergodic processes η_t: $\xi_t = \sum_{i=1}^{\infty} \beta_i \eta_{t-i}$ with $\sum_{j=1}^{\infty} \sqrt{j}|\beta_j| < \infty$. Having T observations for $\hat{\nu}_t$ Wong et al. (2003) identify hidden periodicities 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 $I_T(\lambda) = \left|\sum \hat{Y}_t e^{-it\lambda}\right|^2 /2\pi T$ for $\lambda \in [-\pi, \pi]$. Large coefficients for a specific scale indicate a hidden perodicity. Wong et al. (2003) use a dyadic wavelet decomposition scheme similar to (3.3), i.e. their set of scales is $A = \{2^m, m \in \mathbb{Z}\}$. The algorithm to identify hidden periodicities is as follows:

(1) Let $M = \left\{0, 1, \ldots, 2^{|m|} - 1\right\}$. Compute $\{WT_{I_N}(m, b_m) : m = m_0, m_0 - 1, \ldots, -\infty, b_m \in M\}$ for a $m_0 \in \mathbb{Z}$. Set $n = 1$.

(2) Let $b(m) = arg \max_{b\in M} \left(WT_{I_N}(m, b)\right), \quad MW(m) = \max_{b\in M} \left(WT_{I_N}(m, b)\right).$

  (i) If $MW(m) \sim c$ with $m = m_0, m_0 - 1, \ldots, -\infty$ and a constant $c \in (R)$, then $\hat{\lambda}_n = 2^{m'+1}\pi b(m) - 0.5$ where $m' \in \mathbb{Z}$ is sufficiently small. Go to Step (3).

  (ii) If $MW(m) \to 0$ for $m = m_0, m_0 - 1, \ldots, -\infty$, then there are no further perodicities. 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 $\hat{Y}_t' = \hat{Y}_t - \hat{\alpha}_t e^{i\hat{\lambda}_n t}$ $\forall\, t = 1, \ldots, T$. Go to Step (1).

C Regularity Conditions for Locally Stationary Wavelet Processes

- For $\xi_{m,n}$ holds $E[\xi_{m,n}] = 0 \ \ \forall\, m \in \mathbb{N}, n \in \mathbb{Z}$ and $Cov[\xi_{m,n}, \xi_{o,p}] = \delta_{m,o}\delta_{n,p}$ $\forall\, m, o \in \mathbb{N}, n, p \in \mathbb{Z}$, where δ is the Kroenecker delta,

- For all $m = 1, \ldots, M(T)$ exits a function $W_m(z)$ which is Lipschitz-continuous on $(0, 1]$ with constant $L_m \in \mathbb{R}_+$ and

  - $\sum_{m=1}^{\infty} \|W_m(z)\|^2 < \infty$ for all $z \in (0, 1]$;
  - $\exists (C_m)_{m\in\mathbb{N}} \in \mathbb{R}^+$ s.t. each T fulfills $\forall\, m \in \mathbb{N}, n \in \mathbb{Z}$:

$$\sup_{k=1,\ldots,T} |\omega_{m,n;T} - W_m(k/T)| \leq C_m/T,$$

  and for C_m, L_m holds $\sum_{m=1}^{\infty} s_m(C_m + s_m L_m) < \infty$.

D The Haar and Daubechies D4 Wavlet

The Haar scaling function ϕ_H and the corresponding Haar wavelet Ψ_H are defined as (see Stollnitz et al., 1995)

$$\phi_H(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{else} \end{cases} \quad , \quad \phi_H(x) = \begin{cases} 1 & 0 \leq x < 1/2 \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{else.} \end{cases}$$

The Ψ_H is in fact part of the wavelet family introduced by Daubechies (1992), the Daubechies D2 wavelet. Another one is the Daubechies D4 wavelet Ψ_D and its corresponding scaling function ϕ_D, for which no closed form is be given. It is defined iteratively using the relations

$$\begin{aligned} h(n) &= \frac{1}{\sqrt{2}} \left\langle \phi_D \left(\frac{t}{2} \right), \phi_D(t-n) \right\rangle, \quad n = 0, \ldots, 1, \\ \frac{1}{\sqrt{2}} \phi_D \left(\frac{t}{2} \right) &= \sum_{n=0}^{3} h(n) \phi_D(t-n), \quad \frac{1}{\sqrt{2}} \Psi_D \left(\frac{t}{2} \right) = \sum_{n=3}^{3} (-1)^{1-n} h(1-n) \phi_D(t-n), \end{aligned}$$

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

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

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

E Estimation Results

Table 1: Errors of the One Day Ahead Forecast of Deutsche Bank Stock Prices

h=1			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	1,6077	2,8988	0,0606
ARMA	1,1712	1,7396	0,0323
ARIMA	1,1620	1,5672	**0,0316**
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	1,2713	2,2147	0,0383
Denoising (ARMA)	1,1791	1,8287	0,0331
Denoising (ARIMA)	1,1633	1,7893	0,0322
Multiscale Forecasting (X-12)	1,3185	1,9284	0,0402
Multiscale Forecasting (ARMA)	1,1595	1,5682	**0,0316**
Multiscale Forecasting (ARIMA)	**1,1587**	**1,5619**	**0,0316**
Wavelet based STSM	3,4215	13,4808	0,2275
Locally stationary wavelet process	4,4468	19,8275	0,4618
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	1,2713	2,2147	0,0383
Denoising (ARMA)	1,1791	1,8287	0,0331
Denoising (ARIMA)	1,1633	1,7893	0,0332
Multiscale Forecasting (X-12)	1,2175	1,8147	0,0344
Multiscale Forecasting (ARMA)	1,2165	1,8130	0,0344
Multiscale Forecasting (ARIMA)	1,2166	1,8132	0,0344

Table 2: Errors of the One Week Ahead Forecast of Deutsche Bank Stock Prices

h=5			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	2,6679	12,9264	0,1643
ARMA	2,7552	9,7621	0,1753
ARIMA	1,6705	3,2222	0,0655
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	2,6659	8,8941	0,1663
Denoising (ARMA)	2,6799	8,9580	0,1702
Denoising (ARIMA)	1,7231	3,4029	0,0702
Multiscale forecasting (X-12)	2,3495	6,8411	0,1288
Multiscale forecasting (ARMA)	6,4417	55,0096	0,9713
Multiscale forecasting (ARIMA)	3,5912	14,9590	0,3019
Wavelet based STSM	2,4375	7,3828	0,1380
Locally stationary wavelet process	9,2196	160,9949	1,9856
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	2,6659	8,8941	0,1663
Denoising (ARMA)	2,6799	8,9580	0,1702
Denoising (ARIMA)	1,7231	3,4029	0,0702
Multiscale forecasting (X-12)	**1,5947**	**2,8977**	**0,0594**
Multiscale forecasting (ARMA)	6,7320	46,3410	1,0574
Multiscale forecasting (ARIMA)	1,6473	3,0939	0,0637

Table 3: Errors of the One Day Ahead Forecast of the Euro/Dollar Exchange Rate

h=1			
Classical Methods:	**MAD**	**RMSE**	**APE**
X-12	0,1020	0,0126	0,0075
ARMA	0,0863	0,0084	0,0053
ARIMA	0,0873	0,0085	0,0054
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	0,0970	0,0111	0,0067
Denoising (ARMA)	0,1008	0,0124	0,0072
Denoising (ARIMA)	0,1008	0,0124	0,0072
Multiscale forecasting (X-12)	0,0846	0,0092	0,0051
Multiscale forecasting (ARMA)	0,0843	0,0080	0,0051
Multiscale forecasting (ARIMA)	**0,0840**	**0,0080**	**0,0050**
Wavelet based STSM	0,1477	0,0234	0,0156
Locally stationary wavelet process	0,0990	0,0118	0,0070
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	0,0970	0,0111	0,0067
Denoising (ARMA)	0,1008	0,0124	0,0072
Denoising (ARIMA)	0,1008	0,0124	0,0072
Multiscale forecasting (X-12)	0,0871	0,0086	0,0054
Multiscale forecasting (ARMA)	0,0866	0,0085	0,0054
Multiscale forecasting (ARIMA)	0,0866	0,0085	0,0054

Table 4: Errors of the One Week Ahead Forecast of the Euro/Dollar Exchange Rate

h=5			
Classical Methods:	**MAD**	**RMSE**	**APE**
X-12	0,1797	0,0419	0,0232
ARMA	0,1092	0,0144	0,0085
ARIMA	0,1129	0,0153	0,0091
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	0,1293	0,0240	0,0119
Denoising (ARMA)	**0,0949**	**0,0114**	**0,0064**
Denoising (ARIMA)	0,0996	0,0123	0,0071
Multiscale forecasting (X-12)	0,1587	0,0312	0,0181
Multiscale forecasting (ARMA)	0,4390	0,2466	0,1382
Multiscale forecasting (ARIMA)	0,2765	0,1037	0,0548
Wavelet based STSM	0,1403	0,0266	0,0141
Locally stationary wavelet process	0,1130	0,0156	0,0091
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	0,1293	0,0240	0,0119
Denoising (ARMA)	**0,0949**	**0,0114**	**0,0064**
Denoising (ARIMA)	0.0996	0.0123	0,0071
Multiscale forecasting (X-12)	0,1103	0,0147	0,0087
Multiscale forecasting (ARMA)	0,6091	0,3754	0,2655
Multiscale forecasting (ARIMA)	0,1207	0,0158	0,0104

Table 5: Errors of the One Day Ahead Forecast of the WTI Oil Prices

h=1			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	1,5744	2,8091	0,0355
ARMA	1,1327	1,6256	0,0185
ARIMA	1,0768	1,4420	0,0167
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	1,2967	2,2076	0,0243
Denoising (ARMA)	0,9625	1,4243	0,0134
Denoising (ARIMA)	**0,9306**	**1,3208**	**0,0125**
Multiscale forecasting (X-12)	1,1203	1,5331	0,0179
Multiscale forecasting (ARMA)	1,0624	1,4420	0,0163
Multiscale forecasting (ARIMA)	1,0754	1,4880	0,0167
Wavelet based STSM	2,3485	6,6597	0,0793
Locally stationary wavelet process	2,7627	7,8032	0,1094
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	1,2967	2,2076	0,0243
Denoising (ARMA)	0,9625	1,4243	0,0134
Denoising (ARIMA)	**0,9306**	**1,3208**	**0,0125**
Multiscale forecasting (X-12)	1,1620	1,5536	0,0194
Multiscale forecasting (ARMA)	1,1595	1,5465	0,0193
Multiscale forecasting (ARIMA)	1,1598	1,5467	0,0193

Table 6: Errors of the One Week Ahead Forecast of the WTI Oil Prices

h=5			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	2,9017	14,4185	0,1193
ARMA	2,6720	7,3716	0,1025
ARIMA	1,6117	2,9530	0,0372
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	2,5311	7,6044	0,0914
Denoising (ARMA)	2,2159	5,5404	0,0704
Denoising (ARIMA)	1,5974	2,9820	0,0368
Multiscale forecasting (X-12)	1,6987	3,6149	0,0412
Multiscale forecasting (ARMA)	5,5166	36,7273	0,4318
Multiscale forecasting (ARIMA)	3,0311	11,1294	0,1317
Wavelet based STSM	6,9384	49,1278	0,6874
Locally stationary wavelet process	4,8573	42,8949	0,3399
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	2,5311	7,6044	0,0914
Denoising (ARMA)	2,2159	5,5404	0,0704
Denoising (ARIMA)	1,5974	2,9820	0,0368
Multiscale forecasting (X-12)	**1,3957**	2,3575	**0,0280**
Multiscale forecasting (ARMA)	8,6661	81,2465	1,0756
Multiscale forecasting (ARIMA)	1,4060	**2,2636**	0,0284

Table 7: Errors of the One Day Ahead Forecast of the UK Power Prices

h=1			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	3,0506	10,6172	0,3066
ARMA	2,1824	5,5884	0,1539
ARIMA	2,1462	5,4182	**0,1459**
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	3,0503	10,6041	0,3077
Denoising (ARMA)	2,2869	6,1913	0,1740
Denoising (ARIMA)	2,2438	5,6306	0,1605
Multiscale forecasting (X-12)	2,1715	6,6690	0,1575
Multiscale forecasting (ARMA)	2,2282	5,6552	0,1521
Multiscale forecasting (ARIMA)	2,2604	5,7186	0,1557
Wavelet based STSM	3,5868	15,6367	0,3823
Locally stationary wavelet process	4,1682	19,4502	0,5731
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	3,0573	10,6209	0,3092
Denoising (ARMA)	2,2869	6,1913	0,1740
Denoising (ARIMA)	2,2438	5,6306	0,1605
Multiscale forecasting (X-12)	**2,1237**	**5,2731**	0,1469
Multiscale forecasting (ARMA)	2,2921	6,1814	0,1566
Multiscale forecasting (ARIMA)	2,2921	6,1810	0,1566

Table 8: Errors of the One Week Ahead Forecast of the UK Power Prices

h=7			
Classical methods:	**MAD**	**RMSE**	**APE**
X-12	6,6911	74,2138	1,4268
ARMA	3,6601	14,6317	0,4461
ARIMA	2,5220	**7,4728**	**0,2092**
Haar wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	**2,5025**	88,4410	1,7801
Denoising (ARMA)	3,7977	15,4610	0,4752
Denoising (ARIMA)	2,6670	8,2682	0,2336
Multiscale forecasting (X-12)	5,1985	36,2447	0,8189
Multiscale forecasting (ARMA)	24,2213	806,9786	18,2407
Multiscale forecasting (ARIMA)	9,0070	111,1019	2,4665
Wavelet based STSM	9,2260	90,2059	2,6168
Locally stationary wavelet process	8,5246	148,7063	1,9484
Daubechies D4/ Morlet wavelet:	**MAD**	**RMSE**	**APE**
Denoising (X-12)	7,5025	88,4410	1,7801
Denoising (ARMA)	3,7977	15,4610	0,4752
Denoising (ARIMA)	2,6670	8,2682	0,2336
Multiscale forecasting (X-12)	2,6213	9,1848	0,2492
Multiscale forecasting (ARMA)	14,0224	217,6205	6,2754
Multiscale forecasting (ARIMA)	4,1623	34,8051	0,6011

Diskussionspapiere 2010
Discussion Papers 2010

01/2010 **Mosthaf, Alexander, Schnabel, Claus and Stephani, Jens:** Low-wage careers: Are there dead-end firms and dead-end jobs?

02/2010 **Schlüter, Stephan and Matt Davison:** Pricing an European Gas Storage Facility using a Continuous-Time Spot Price Model with GARCH Diffusion

03/2010 **Fischer, Matthias, Gao, Yang and Herrmann, Klaus:** Volatility Models with Innovations from New Maximum Entropy Densities at Work

Diskussionspapiere 2009
Discussion Papers 2009

01/2009 **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009 **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009 **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009 **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009 **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009 **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009 **Herrmann, Klaus:** Non-Extensitivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009 **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009 **Dewenter**, Ralf, **Haucap**, Justus and **Wenzel**, Tobias: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009 **Stuehmeier**, Torben and **Wenzel**, Tobias: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009 **Klein**, Ingo, **Köck**, Christian and **Tinkl**, Fabian: Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

12/2009 **Schlüter**, Stephan: Constructing a Quasilinear Moving Average Using the Scaling Function

13/2009 **Blien,** Uwe, **Dauth,** Wolfgang**, Schank,** Thorsten and **Schnabel,** Claus**:** The institutional context of an “empirical law”: The wage curve under different regimes of collective bargaining

14/2009 **Mosthaf**, Alexander, **Schank**, Thorsten and **Schnabel**, Claus: Low-wage employment versus unemployment: Which one provides better prospects for women?

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl**: Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl**: Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm**: Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner**: The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias**: On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus**: Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' “The Principal Averages and the Laws of Error which Lead to Them” - Refinement and Generalisation