# Empirical analysis of financial time series for piecewise homoscedasticity by means of wavelet MSML algorithm.

**ABSTRACT**

It is a well-known phenomenon that the time series of the [log]returns on equities are nonstationary, in particular, due to volatility clustering. On the other hand, the volatility is an important quantity for many risk measures and for position sizing. For practical purposes it is sufficient to have a piecewise stationarity (or at least homoscedasticity) on *some* time intervals, on which an investor wants to trade. During these stages an investor can use the historical volatility as a consistent estimate of the future volatility (until the next structural break comes).

To check whether the financial time series are homoscedastic at least on some time intervals I apply the "multiscale and multilevel technique for consistent segmentation of nonstationary time series", which is based on the locally stationary processes in wavelet domain (LSW-processes). I analyze the historical data of approximately 3000 stocks. It follows that approximately on 40% of segments the stock-logreturns are homoscedastic. Moreover, the number of homoscedastic segments does not significantly differ between crisis times and calm market times.

Vasily <<YetAnotherQuant>> Nekrasov

April, 15 2012

finanzmaster ȧt gmx DOT net
http://www.yetanotherquant.de

## 1. Introduction

Volatility is an important quantity for setting the take-profit and stop-loss orders and for position sizing: for example the glorified Turtle Traders used "volatility-based constant percentage risk position sizing algorithm" (s. "Turtles"). It is well-known that the volatility changes over time[1]. However, the Figure 1 shows that within the green area the time series of log-returns is likely to be stationary or at least homoscedastic.

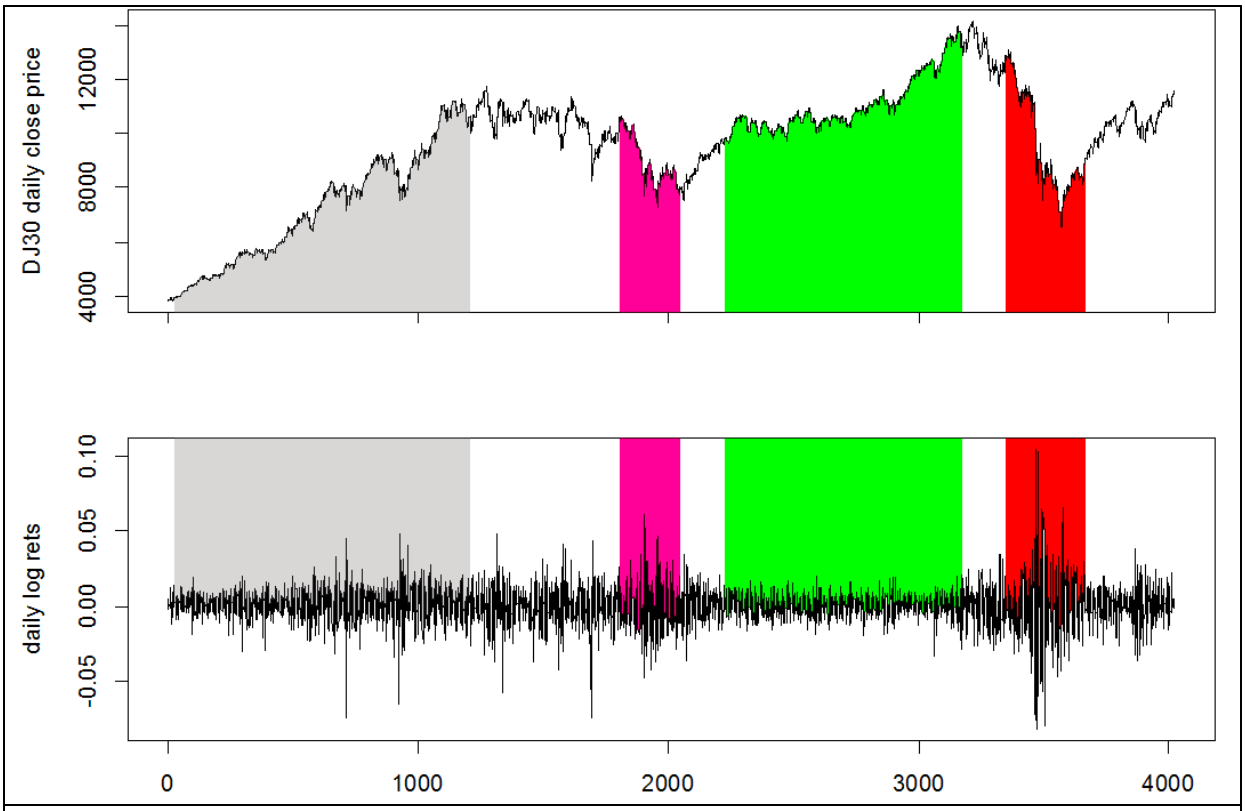



**Figure 1:** Dow Jones daily close and log-returns from Jan. 1995 to Dec. 2011

Moreover, we see that the green area is a period of the strong uptrend. So a trend following entry/exit strategy, combined with volatility based position sizing would be promising here. The red area on the right, in which the volatility is especially high, is the time of the Bank crisis'2008. Notably, there are two warning signals between the green and red areas: the trend reversal and the growing volatility.
On the other hand we can see a strong uptrend by heteroscedastic volatility within the grey area, after which the trend stagnates and the Dotcom crisis'2002 (pink area) occurs. However, the volatility does not grow before this crisis and even during the crisis its increase is moderate.

Our major goal is to analyze how often the phases of stationary (or at least homoscedastic) returns occurs and whether they last sufficiently long. The minor goal is to assess (so far qualitatively) how these phases are related to trends.

[1] For economic analysis of this phenomenon (which is not the key point of this paper) the readers are referred to Schwert(1989).

For this I assume that the daily log-returns follow an LSW-process and run the MSML algorithm (described in the next paragraph) on historical data of approximately 3000 stocks[2], for which there are at least 2049 entries[3] available. At first I log the detected breakpoints and generate the charts like Fugure 2 for each stock. Then I screen the generated charts to qualitatively assess the MSML algorithm and the trend/volatility dependence. Finally, I run stationarity and homoscedasticity tests for every segment for each time series.

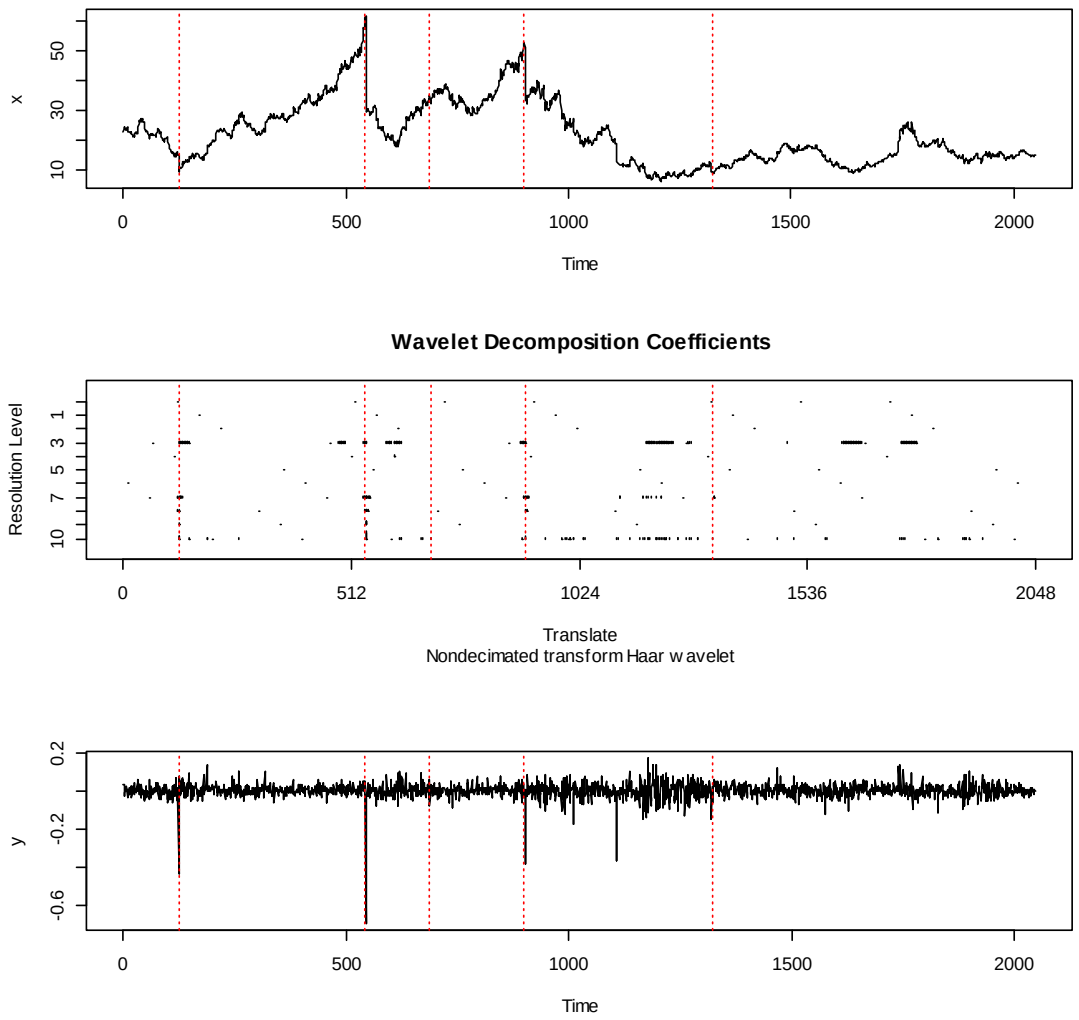



**Figure 2:** An instance of generated charts for NVIDIA stock.
Up: daily close prices, bottom: daily log-returns,
mid: wavelet decomposition of daily log-returns.
Red vertical dotted lines are detected breakpoints.

[2] Exactly speaking 2947 stocks. The key properties of the historical data are described in my previos report " A review of literature for private traders and a backtest of the recommended strategies "

[3] 2049-1 = 2048 = $2^{11}$, the length of input time series for the MSML algorithm must be a power of two, additionally I need one more entry since I analyze diff(log(X)), where X is the input time series.

## 2. Haar wavelets, LSW processes and MSML algorithm

LSW (locally stationary wavelet processes) were introduced by Nason et al(2000), whereas they are based on idea of the local stationarity by Dahlhaus(1997). The idea of local stationarity is a trade-off between departure from stationarity and an ability to estimate the model parameters from a single realization of a stochastic process. One assumes that the model parameters (e.g. AR-coefficients, frequency spectrum, etc) change over time, however, sufficiently slowly.
For reader convenience I give an informal introduction to LSW processes and MSML algorithm. For the [pretty hard-going] formal proofs the readers are referred to Nason et al(2000), and Cho and Fryzlewicz(2012).

It is well-known that a stationary time series can be represented in frequency domain as a sum of sines and cosines with random uncorrelated amplitudes. One can readily replace "big waves" (sines and cosines) with "small waves" (wavelets). The main advantage of such approach that the "small" waves [at finer scales] are well-localized both in time and frequency, which w.r.t. time series analysis, allows to capture the nonstationarity. The simplest wavelet is the Haar wavelet, defined as

$$\varphi(t) = \begin{cases} 1 & t \in [0, \frac{1}{2}) \\ -1 & t \in [\frac{1}{2}, 1) \\ 0 & \text{else} \end{cases} \qquad \textbf{(2.1)}$$

$\varphi(t)$ can be translated (by $-k$) , dilated (by $2^j$) and rescaled($\sqrt{2^j}$) to obtain a doubly-indexed sequence of functions $\{\varphi_{jk}(t)\}_{j\in\mathbb{N}_0,k=0,\dots,2^j-1} = \sqrt{2^j}\varphi(2^j(t-k))$ or explicitly

$$\{\varphi_{jk}(t)\}_{j\in\mathbb{N}_0,k=0,\dots,2^j-1} = \sqrt{2^j} \begin{cases} 1 & t \in [2^{-j}k, 2^{-j}(k+0.5)) \\ -1 & t \in [2^{-j}(k+0.5), 2^{-j}(k+1)) \\ 0 & \text{else} \end{cases} \qquad \textbf{(2.2)}$$

One can easily check that $\{\varphi_{jk}(t)\} \cup f(t) = 1$ is an orthonormal basis in Hilbert space $L^2[0,1]$. In particular, the elements of this basis do not overlap (Figure 3)
Analogously one can define an ON-basis in a finite dimensional space[4], e.g. for N=4 dimensions we have a Haar wavelet ON-basis as follows

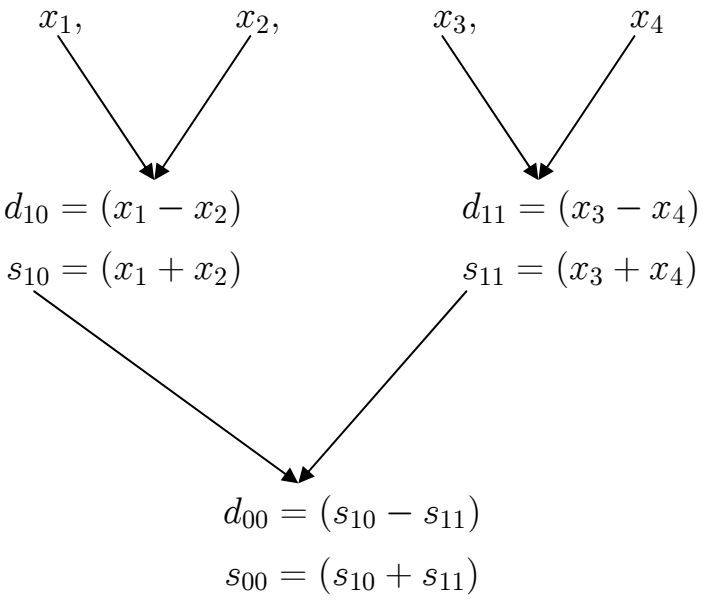
$$\mathbb{W} = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \sim \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ \varphi_{0,0}(t) \\ \varphi_{1,0}(t) \\ \varphi_{1,1}(t) \end{bmatrix} \qquad \textbf{(2.3)}$$

where "$\sim$" in (2.3) means how to relate rows of $\mathbb{W}$ with elements of $\{\varphi_{jk}(t)\}$, whereas for a better understanding one should look the left part of Fig. 3, where black dots at OX-axis depict the elements of $x$.

The [discrete] wavelet transform (DWT) of a vector $x := [x_1, x_2, x_3, x_4]^T$ is the vector $[w_1, w_2, w_3, w_4]^T = \mathbb{W}x$.

[4] Number of dimensions must be a power of two.

One can readily see that $w_4 = 0.5\sqrt{2}(x_3 - x_4)$, $w_3 = 0.5\sqrt{2}(x_1 - x_2)$, $w_2 = 0.5(x_1 + x_2 - x_3 - x_4)$ and $w_1 = 0.5(x_1 + x_2 + x_3 + x_4)$.
This fact not just shows that the Haar wavelet coefficients are "smoothes and differences" of the original vector but also enables a fast computation algorithm (pyramid algorithm) with complexity $O(n)$. "Pyramid" means the following:




$d_{10}$ stands for "detail at level $j = 1$, location $k = 0$", $s_{10}$ is respectively "smooth at level $j = 1$, location $k = 0$". One sees that $Wx = [s_{00}, d_{00}, d_{10}, d_{11}]$. The pyramid algorithm allows to introduce the partial wavelet transform $[s_{10}, s_{11}, d_{10}, d_{11}]$, which is useful for *multiresolution analysis* of the time series (roughly speaking, it separates irregular details from a smooth part at coarser and coarser levels).

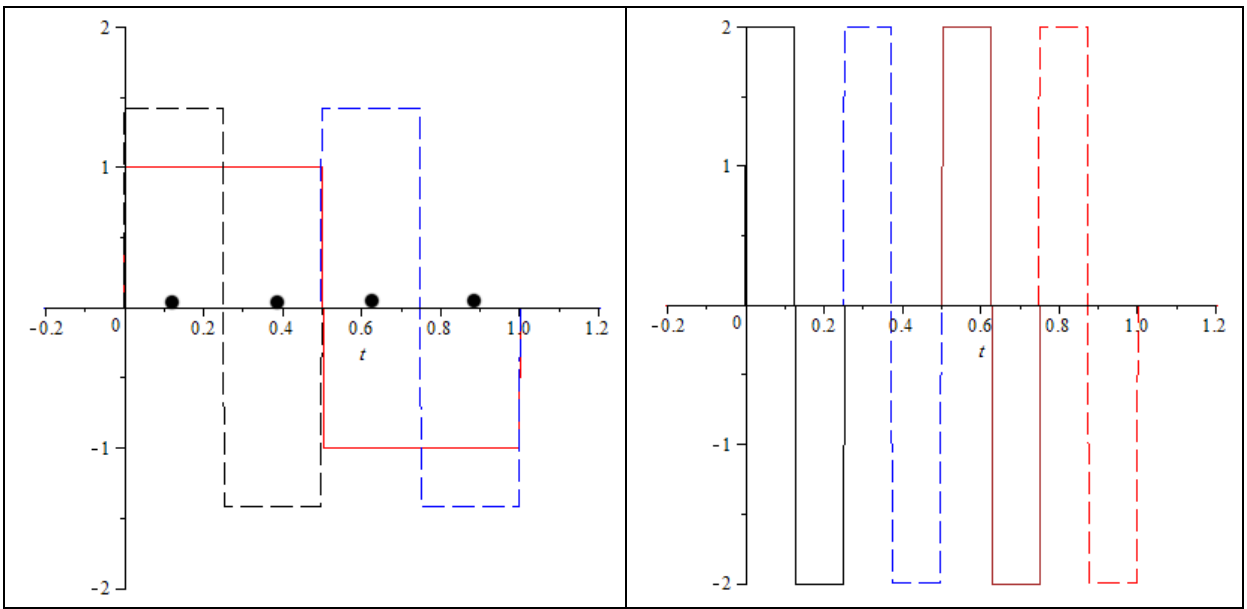



**Figure 3**: Some elements of $\{\varphi_{jk}(t)\}$
On the left $\varphi_{0,0}(t)$ solid red, $\varphi_{1,0}(t)$ dashed black and $\varphi_{1,1}(t)$ dashed blue.
On the right: $\varphi_{2,0}(t)$ black , $\varphi_{2,1}(t)$ dashed blue, $\varphi_{2,2}(t)$ brown and $\varphi_{2,3}(t)$ dashed red

Unfortunately, there is a problem with wavelet transform: it is not invariant to a cyclic shift of the input vector[5]. Or other way around: in our example we consider $x_1 - x_2$ and $x_3 - x_4$ but not $x_2 - x_3$ and not $x_4 - x_1$. Probably we loose some interesting details?!

To avoid this problem several authors (see e.g. Persival and Walden(2000) and Nason(2008)) have introduced the the non-decimated wavelet transform (NDWT), also known as maximum overlap wavelet transform (MODWT) or stationary wavelet transform (SWT). Following Nason, I will further write "NDWT". By NDWT each row of the $\mathbb{W}$ is shifted to all possible locations, i.e.

$$\widetilde{\mathbb{W}} = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ (1 & 1 & -1 & -1) \\ (-1 & 1 & 1 & -1) \\ (-1 & -1 & 1 & 1) \\ (1 & -1 & -1 & 1) \\ \{\sqrt{2} & -\sqrt{2} & 0 & 0\} \\ \{0 & \sqrt{2} & -\sqrt{2} & 0\} \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 0 & 0 & \sqrt{2} \end{bmatrix} \qquad \textbf{(2.4)}$$

Now there are N=4 coefficients at each level, so the NDWT is redundant and thus is not uniquely invertible. However, LSW processes (which are based on NDWT) do have the unique characterizations in terms of a function, which controls the amplitude of the NDWT-coefficients.

Consider stochastic process, represented in the mean-square sense as

$$X_t = \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{+\infty} \eta_{jk} \varphi_{jk}(t) \qquad t = 0, ..., T-1 \qquad \textbf{(2.5)}$$

where $\eta_{jk}$ are $\mathcal{N}(0,1)$ i.i.d. random variables. ($X_t$ is so far not an LSW process in full generality, it is "plain-vanilla" stationary and is the sum of MA-processes, as we will see).

Note that in (2.5) the wavelets run [at each scale $j$] over the whole real line: contrary to (2.2) $k$ is now independent of $j$, so there are the same number of shifted locations $k$ at each scale $j$, thus it is a NDWT (and not a DWT) representation.

Also, following the original notation of Nason et al(2000), I let $j \in (-\infty, -1]$ (not $\mathbb{N}_0$ as in (2.2)). The finest = the shortest = the most localized wavelets live now at scale $j = -1$.
On other words $\{\varphi_{jk}(t)\}_{j \in \mathbb{Z}^-, k=-\infty,...,+\infty} = \sqrt{2^j}\varphi(2^j(t-k))$.
The advantage of such notation is that as number of observations $T$ grows, we involve the wavelets at coarser and coarser scales, which characterize the time series "more and more globally".

[5] Actually the Fourier basis is the only ON basis with the shift-invariance property.

The Haar wavelets are now defined as

$$\varphi_{-1,0}(t) = \begin{cases} \frac{1}{\sqrt{2}} & t = 0 \\ -\frac{1}{\sqrt{2}} & t = 1 \\ 0 & else \end{cases} \qquad \textbf{(2.6)}$$

$$\varphi_{-1,-1}(t) = \begin{cases} \frac{1}{\sqrt{2}} & t-1 = 0 \\ -\frac{1}{\sqrt{2}} & t-1 = 1 \\ 0 & else \end{cases} = \begin{cases} \frac{1}{\sqrt{2}} & t = 1 \\ -\frac{1}{\sqrt{2}} & t = 2 \\ 0 & else \end{cases} \qquad \textbf{(2.7)}$$

$$\varphi_{-2,0}(t) = \begin{cases} \frac{1}{2} & t = 0, 1 \\ -\frac{1}{2} & t = 2, 3 \\ 0 & else \end{cases} \qquad \textbf{(2.8)}$$

and so on.

Let us try to depict graphically a sample realization of $X_t$, starting observation (according to the model) from $t = 0$

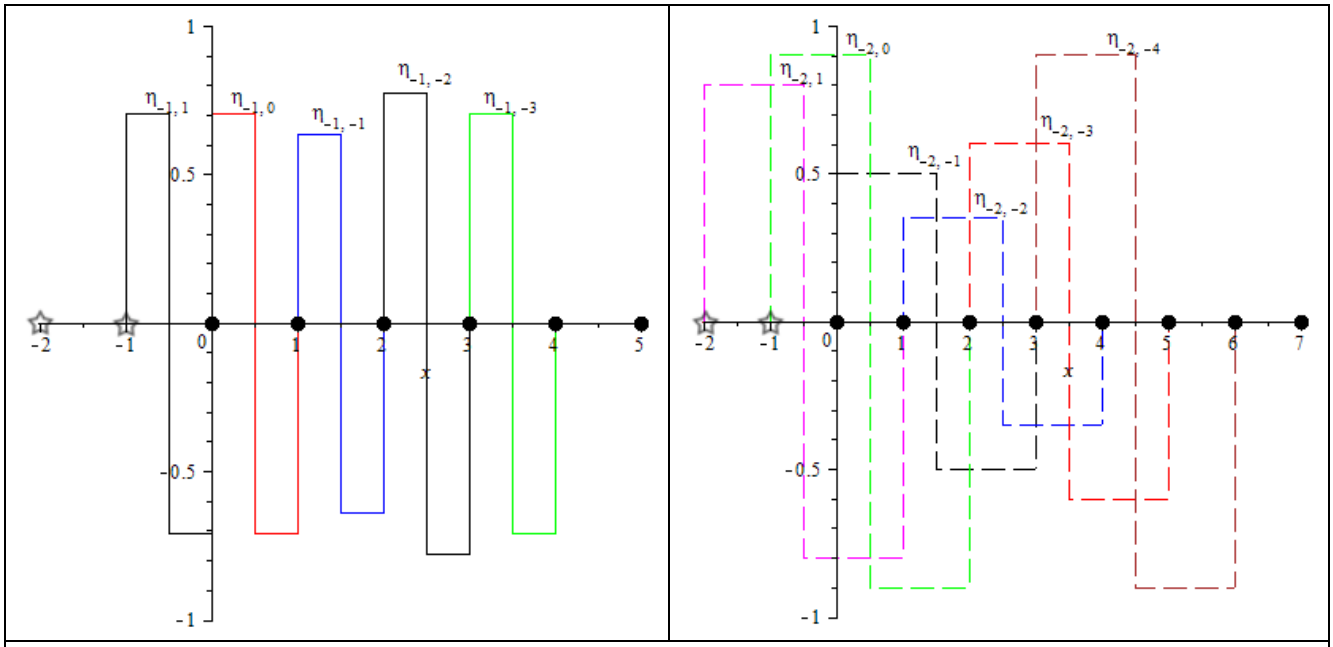



**Figure 4**: Ingredients of $X_t$ at levels $j = -1$ (left) and $j = -2$ (right)

Wavelet functions are drawn with different heights (amplitudes), which depend on the realizations of $\eta_{jk}$. As usual the black dots mean where our discretely observed data live w.r.t. the location of the wavelet functions. The pentagrams mean, in turn the *unobserved* values of $X_t$ from the past. However, these values are irrelevant for us, what is relevant are

the realizations of $\eta_{jk}$. They are not observed directly but rather we have the *state space representation* - , i.e. though we do not observe the realizations of $\eta_{jk}$ directly[6], they influence the realizations of $X_t$ which we do observe.

So at zero, i.e. at time $t = 0$ there are only two wavelets with non-zero support at scale $j = -1$ : they are $\varphi_{-1,1}$ and $\varphi_{-1,0}$. The result, which we obtain at scale $j = -1$ is thus

$$X_0^{j=-1} = \eta_{-1,1}\varphi_{-1,1} + \eta_{-1,0}\varphi_{-1,0} = \frac{1}{\sqrt{2}}(\eta_{-1,0} - \eta_{-1,1}) \qquad \textbf{(2.8)}$$

At $t = 1$ we have $X_1^{j=-1} = \frac{1}{\sqrt{2}}(\eta_{-1,-1} - \eta_{-1,0})$ and so on - so at scale $j = -1$ lives nothing else but the MA(1) [moving average of the order 1] process.

Analogously at scale $j = -2$ we have an MA(2) process:

$$X_0^{j=-2} = \frac{1}{2}\left(-\eta_{-2,2} - \eta_{-2,1} + \eta_{-2,0} + \eta_{-2,-1}\right) \qquad \textbf{(2.9)}$$

Now it should be clear why it is reasonable to use the non-decimated wavelets instead of an ON-wavelet basis. The elements from an ON-wavelet basis do not overlap, so the class of our process would not include the MA-processes.
Moreover, since every stationary process can be represented as an infinite sum of MA-processes (Wald representation?!) - our class includes all stationary processes.

Now we start departing from the stationarity and assume that the amplitudes of $\varphi_{jk}(t)$ are influenced not only by the realizations of the random variables $\eta_{jk}$ but also buy deterministic coefficients $w_{jk}$. So where we had the MA-processes before, we will have the MA-processes *with time-varying coefficients*, which are locally stationary in the sense of Dahlhaus(1997).

But if we do not impose any restrictions on $w_{jt}$ there is no hope to infer them from a *single* realization of $X_t$. That why we assume that they vary sufficiently slowly, i.e. for each scale there are sufficiently bounded functions $W_j(z) \quad z \in (0,1)$ and the constants $C_j$ such that

$$sup_k \left| w_{jk;T} - W_j\left(\frac{k}{T}\right)\right| < \frac{C_j}{T} \qquad \textbf{(2.10)}$$

Note that $W_j$ are defined on $(0,1)$ and in (2.10) there are $k/T \in (0,1)$ and $C_j/T$. The idea behind is as follows: we *rescale* our observation time to $(0,1)$ and assume that the more data we observe, the closer are our coefficients $w_{jk;T}$ to $W_j(k/T)$.

Finally, we can define a class of locally stationary wavelet (LSW) processes as a *triangular sequence of doubly-indexed* stochastic processes $\{X_{t,T}\}_{t=0,\dots,T-1,T=2^J\geq 1}$ with mean-square representation as

$$X_{t,T} = \sum_{j=-J}^{-1}\sum_{k=0}^{T-1} w_{jk;T}\varphi_{jk}(t)\eta_{jk} \qquad \textbf{(2.11)}$$

[6] Do not forget, that in reality we observe $\sum_j X_t^j$ and not the separate components $X_t^j$

However, it is important to understand that for $S \neq T$ the $X_{t,T}$ and the $X_{t,S}$ are pretty different processes. In particular if, say, $S > T$ then the sigma algebra generated by $X_{t,T}$ is *not* a subset of the sigma algebra generated by $X_{t,S}$. In simple words, knowing the values of $X_{t,S}$ up to time $S$ does not give us exhaustive information about $X_{t,T}$ up to time $T$.
On the other hand if $min(S,T) \to \infty$ then (for every $j,k$) $w_{jk;T} \to w_{jk;S}$.
This fact allows to set up the rigorous theory of asymptotic parameter estimation - that's why such complicated settings we introduced[7].

In practice we observe not a triangular two-dimensional array but just a single realization. Still, there is a way to infer the *wavelet spectrum* (squared $w_{jk;T}$) from a single realization, which we now discuss.

Let us calculate the autocovariance function of $X_{t,T}$.

$$c_T(z,\tau) = cov(X_{[zT],T}, X_{[zT]+\tau,T}) \qquad \textbf{(2.12)}$$

Since (due to varying $w_{jk;T}$) $X_{t,T}$ is no more stationary, $c_T$ is a function of two parameters. Because $\eta_{jk}$ are i.i.d. only those of $\varphi_{jk}(t)$ and $\varphi_{jk}(t+\tau)$ that overlap[8] do contribute to $c_T(z,\tau)$. Thus we can analyze $c_T(z,\tau)$ *separately* for each scale. Starting from the finest scale $j=-1$ we yield according to (2.6), (2.7) and Figure 4 the following:

| $\tau$ | $cov\left(X^{j=-1}_{[zT],T}, X^{j=-1}_{[zT]+\tau,T}\right)$ |
|---|---|
| 0 | $\varphi^2_{-1,k}(0)\mathbb{E}\eta^2_{jk} = \left((\frac{1}{\sqrt{2}})^2 + (-\frac{1}{\sqrt{2}})^2\right)\mathbb{E}\eta^2_{jk} = w^2_{jk;T}$ |
| 1 | $\varphi_{-1,0}(0)\varphi_{-1,0}(1)\mathbb{E}\eta^2_{jk} = \left((\frac{1}{\sqrt{2}})*0 + (-\frac{1}{\sqrt{2}})(\frac{1}{\sqrt{2}}) + (-\frac{1}{\sqrt{2}})*0\right)\mathbb{E}\eta^2_{jk} = -\frac{1}{2}w^2_{jk;T}$ |
| -1 | As in case of $\tau = 1$ due to symmetry of the Haar wavelets |
| > 1 | 0 |
| < -1 | 0 |

Note that the contribution of the wavelets does *not* depend on $k$, since the wavelet is just shifted but remains the same. Only the relative shift $\tau$ matters for the wavelet contribution (but for $w^2_{jk;T}$ certainly both matter).

That's why it is reasonable to introduced the so-called *autocorrelation wavelets*:

$$\Psi_j(\tau) := \sum_k \varphi_{jk}(0)\varphi_{jk}(\tau) \qquad \textbf{(2.13)}$$

We have already calculated $\Psi_1(\tau)$. Nason(2008) and Nason et al(2000) show how to calculate $\Psi_j(\tau)$ for arbitrarily $j$.

To make the idea of the autocorrelation wavelets even more clear, let us calculate $\Psi_{-3}(2)$, Figure 4a.

[7] Probably the best way to understand these ideas is to choose some functions $W_j(z)$ and simulate the LSW processes for different $T$'s. Nason(2008) explains how to do it with R.
[8] Every wavelet $\varphi_{jk}(t)$ has a compact support and is associated with a random variable $\eta_{jk}$. If $\tau$ is big, $\varphi_{jk}(t)$ and $\varphi_{jk}(t+\tau)$ do not overlap (compact support).

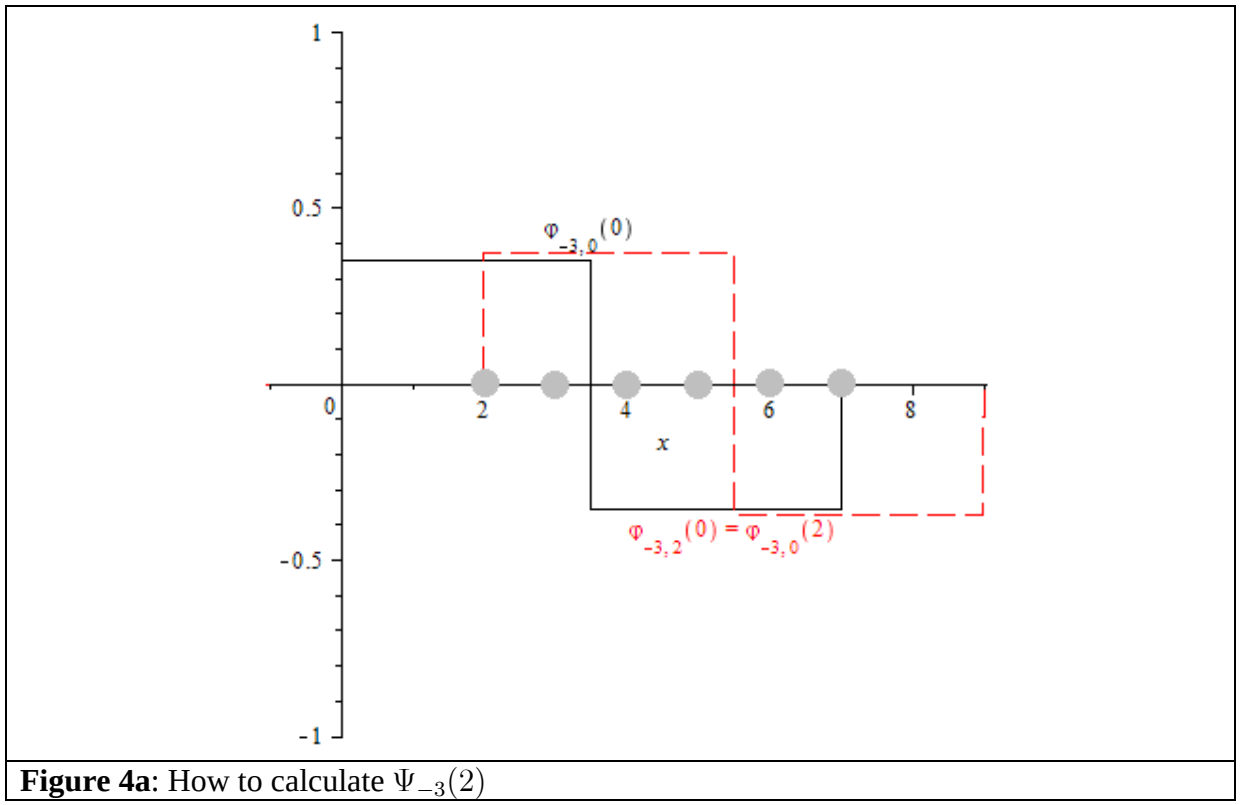



**Figure 4a**: How to calculate $\Psi_{-3}(2)$

Index $k$ is responsible for the location, i.e. according to (2.13) we must sum over all locations, at which both wavelets are not zeros (gray dots). So

$$\Psi_{-3}(2) = \left(\frac{2}{2\sqrt{(2)}}\right)^2 (\underbrace{1\cdot 1}_{k=2} + \underbrace{1\cdot 1}_{k=3\ldots} + 1\cdot(-1) + 1\cdot(-1) + (-1)\cdot(-1) + \underbrace{(-1)\cdot(-1)}_{\ldots k=7})$$

Once again - it would change nothing if we shift *together both* wavelets.

Defining the "true" evolutionary wavelet spectrum(EWS) as $S_j(z) := W_j^2(z)$ we can define the local autocovariance (LACV) $c(z,\tau)$ of an LSW process by

$$c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau) = \vec{S}(z)\vec{\Psi}(\tau) \qquad \tau \in \mathbb{Z}, z \in (0,1) \qquad \textbf{(2.14)}$$

Due to (2.10) $|c_T(z,\tau) - c(z,T)| = O(T^{-1})$ (see Nason et al(2000) for proof details).

Now our goal to try to invert (2.14) in order to obtain $\vec{S}(z)$, i.e. $S_j(z)$ for all $j$. For that Nason et al(2000) define the operator $A = (A_{jl})_{j,l<0}$ by

$$A_{jl} := < \Psi_j, \Psi_l > = \sum_{\tau} \Psi_j(\tau)\Psi_l(\tau) \qquad \textbf{(2.15)}$$

Since in practice we deal with finite number of scales, we consider the $J$-dimensional matrix $A_J = (A_{jl})_{j,l=-1,\ldots,-J}$ .

One should consider (2.15) as follows: at first we calculate $\Psi_j(\tau)$ and $\Psi_l(\tau)$ for all $\tau$'s (there is only finite number of them at which they are not zero and we have seen that for the scale $j=-2$ there are more such $\tau$'s as for $j=-1$, i.e. the coarser the scale the more $\tau$'s there are). Then we multiply $\Psi_j(\tau)$ and $\Psi_l(\tau)$ , and sum this product over all $\tau$'s at which *both* factors are not zero, i.e. the *finer* scale determines how many such $\tau$'s there are.

Further define a vector $\vec{\kappa}(\tau)=A^{-1}\vec{\Psi}(\tau)$ where $A^{-1}$ is an inverse of $A$. Nason et al (2008) prove that $A^{-1}$ is well-defined (since $A$ is an operator in *infinite*-dimensional space the existence of the inverse is not a trivial question but we will not delve into details).

And finally there is an inversion formula for $S_j(z)$ !

$$S_j(z)=\sum_{\tau} c(z,\tau)\kappa_j(\tau) \qquad \textbf{(2.16)}$$

Nason et al(2008) give the rigourous proof of (2.16) but for the "engineering understanding" it is better not to reproduce this proof here but rather to consider the simplest case when we have only two scales in our process.
In this case we rewrite the 2*2 symmetrical(!) matrix $A$ is the following compact notation

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} =: \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

and obtain

$$A^{-1}=\frac{1}{ac-b^2}\begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$

Further

$$\kappa_1(\tau)=\frac{1}{ac-b^2}\begin{bmatrix} c & -b \end{bmatrix}\begin{bmatrix} \Psi_1(\tau) \\ \Psi_2(\tau) \end{bmatrix}=\frac{1}{ac-b^2}\left(c\Psi_1(\tau)-b\Psi_2(\tau)\right)$$

And finally

$$\sum_{\tau}\left(\begin{bmatrix} S_1(z) & S_2(z) \end{bmatrix}\begin{bmatrix} \Psi_1(\tau) \\ \Psi_2(\tau) \end{bmatrix}\frac{1}{ac-b^2}\left(c\Psi_1(\tau)-b\Psi_2(\tau)\right)\right)$$

$$=\frac{1}{ac-b^2}\sum_{\tau}\left(\begin{bmatrix} S_1(z) & S_2(z) \end{bmatrix}\begin{bmatrix} c\Psi_1(\tau)\Psi_1(\tau)-b\Psi_1(\tau)\Psi_2(\tau) \\ c\Psi_2(\tau)\Psi_1(\tau)-b\Psi_2(\tau)\Psi_2(\tau) \end{bmatrix}\right)$$

$$=\frac{1}{ac-b^2}\begin{bmatrix} S_1(z) & S_2(z) \end{bmatrix}\begin{bmatrix} ca-bb \\ cb-bc \end{bmatrix}=\begin{bmatrix} S_1(z) & S_2(z) \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}=S_1(z)$$

Analogously we proceed with $S_2(z)$. Notably that we did not need the explicit values of $A^{-1}$ coefficients, we just relied on the property of symmetrical matrices. However, to infer the $\vec{S}(z)$ from a single realization of an LSW process we do need them. They are calculated in Nason et al(2000), see also Nason(2008). The calculation is just a little bit cumbersome algebra exercise.

Now we come to the main question: how to infer the process parameters from a *single* realization?

Let $d_{jk,T} = \sum_{t=0}^{T-1} X_t \varphi_{jk}(t)$ be the empirical wavelet coefficient at scale $j$, location $k$ and define the *wavelet spectrum* (analogous to Fourier spectrum) as

$$I^j_{k,T} := |d_{jk,T}|^2 \qquad \textbf{(2.17)}$$

Nason et al(2000) claim[9] that

$$\mathbb{E}\left[I^j_{[zT],T}\right] = \sum_l A_{jl} S_l(z) + O(2^{-j}/T) \qquad \textbf{(2.18)}$$

If it is so, we can invert (2.18) just like we inverted (2.14) in (2.14). However, there might be some troubles with correctness of (2.18).

In the following we fix $T$ and drop it from notation. We consider the simplest [non trivial] case: let $S_j(z) = 0$ for $j \notin \{1, 2\}$, so there are only two scales that matter. Let us calculate $I^{j=-1}_{k=1}$.
Consider the Figure 4b. In the realization of $X_t$ the wavelet $\varphi_{-1,1}$ is the blue wavelet (with random factor $\eta_{-1,1}$ and spectrum $S_{-1}(1/T)$). It will fully contribute to $I^{j=-1}_{k=1}$, i.e. with factor $\Psi_{-1}(0)$. But (and this is a peculiarity of NDWT) the wavelets are not orthogonal, so the red wavelet on the left and the brown wavelet on the right will contribute too, however, with factors $\Psi_{-1}(-1) = \Psi_{-1}(1)$.

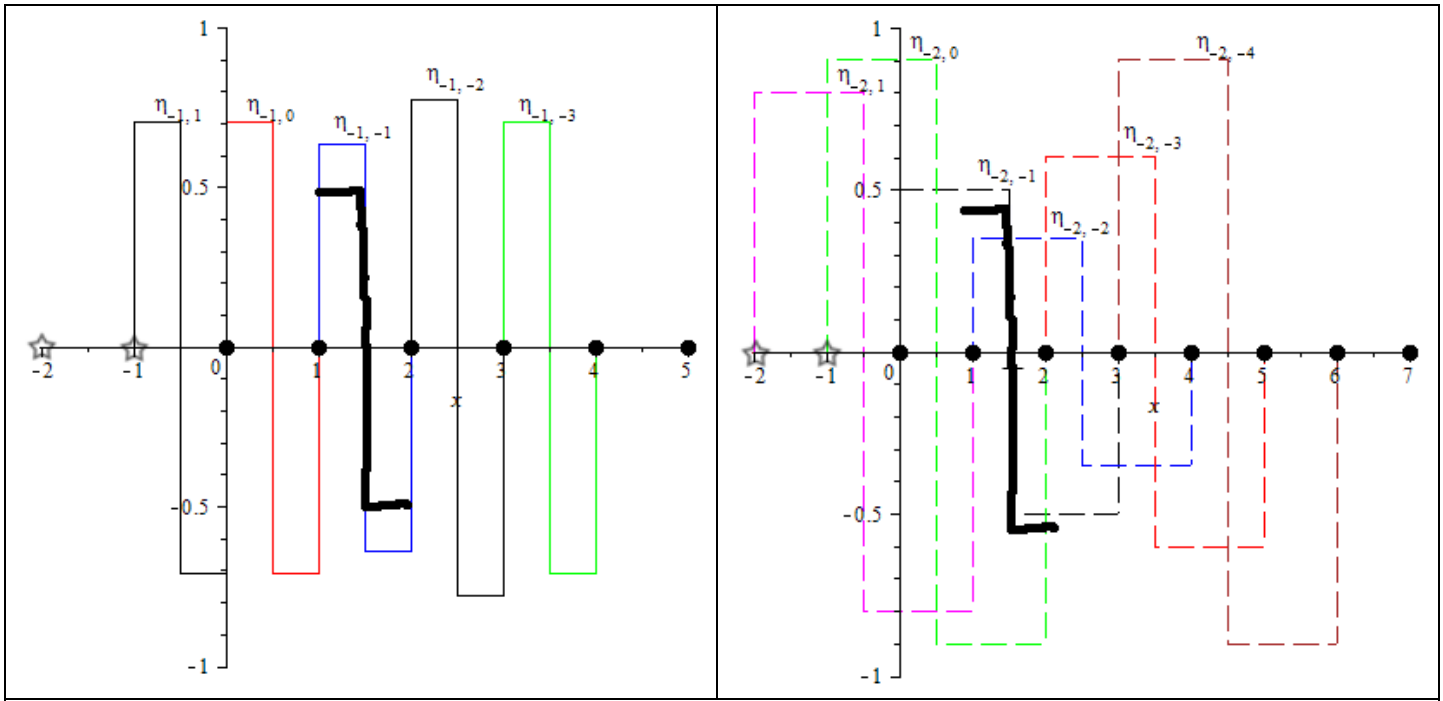



**Figure 4b**: Ingredients of $X_t$ at levels $j = -1$ (left) and $j = -2$ (right) and calculation of $I^{j=-1}_{k=1}$

These wavelet have also different spectrum $S_{-1}(0/T)$ and $S_{-1}(2/T)$ but since the wavelet coefficients are close to each other, their spectra are approximately the same.

So we have

[9] But do not give an explicit proof

$$\mathbb{E}\left[I^{j=-1}_{k=1,contrib:j=-1}\right] = \mathbb{E}\left[\left(W_{-1}(1/T)\left[\Psi_{-1}(-1)\eta_{-1,0} + \Psi_{-1}(0)\eta_{-1,-1} + \Psi_{-1}(1)\eta_{-1,-2}\right]\right)^2\right]$$

$$= W^2_{-1}(1/T)\left[\Psi^2_{-1}(-1) + \Psi^2_{-1}(0) + \Psi^2_{-1}(1)\right] \qquad \textbf{(2.19)}$$

since $\mathbb{E}[\eta_{jk}\eta_{lm}]$ is 1 for $j = l, k = m$ and 0 otherwise. But $S_{-1}(1/T) = W^2_{-1}(1/T)$ and $\left[\Psi^2_{-1}(-1) + \Psi^2_{-1}(0) + \Psi^2_{-1}(1)\right] =< \Psi_{-1}, \Psi_{-1} >= A_{11}$ .

So far, so gut. But now consider the contribution of the scale $j = -2$. If it were the DWT, it would not contribute at all (orthogonality of DWT) but for NDWT it does.
There are four wavelets of scale $j = -2$ at $k = 1$, they are magenta, blue, green and brown. All they are the shifted version of the same wavelet, but $\varphi_{-1,1}$ is not shifted here, so we cannot reduce the contribution of these wavelets to $< \Psi{-}2(\tau), \Psi_{-1}(\tau) >= A_{21}$ !
instead we need something like *cross*-correlation wavelets (Nason et al(2000) do *not* introduce anything like this) defined as

$$C_{jl}(\tau) := \sum_k \varphi_{jk}(0)\varphi_{lk}(\tau) \qquad \textbf{(2.20)}$$

In our case $j = -1$ and $l = -2$, so analogously to (2.19) we will get

$$\mathbb{E}\left[I^{j=-1}_{k=1,contrib:j=-2}\right] =$$
$$W^2_{-2}(1/T)\left[C^2_{-1,-2}(-2) + C^2_{-1,-2}(-1) + C^2_{-1,-2}(0) + C^2_{-1,-2}(1) + C^2_{-1,-2}(2)\right]$$

But $C^2_{jl}(\tau) = \sum_\tau \left(\sum_k \varphi_{jk}(0)\varphi_{lk}(\tau)\right)^2 \neq \sum_\tau \left(\sum_k \varphi_{jk}(0)\varphi_{jk}(\tau)\right)\left(\sum_k \varphi_{lk}(0)\varphi_{lk}(\tau)\right) = A_{jl}$
An this is somehow the problem! But not a big problem!

Actually, we got

$$\mathbb{E}\left[I^{j=-1}_{k=1}\right] = \mathbb{E}\left[I^{j=-1}_{k=1,contrib:j=-1} + I^{j=-1}_{k=1,contrib:j=-2}\right] \qquad \textbf{(2.21)}$$

Now let us introduce the matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{21} \end{bmatrix} \quad \text{where} \quad \Sigma_{jl} = \sum_\tau C^2_{jl}(\tau) \qquad \textbf{(2.22)}$$

Not that for the Haar wavelets $\Sigma$ is symmetric, since the Haar wavelet are symmetric and it does not matter whether we shift $\varphi_{jk}$ w.r.t. $\varphi_{lk}$ or vice versa. For some other wavelets the symmetry is probably lost but it does not matter as long as $\Sigma$ remains invertible.

So now we can rewrite (2.21) in terms of $\Sigma$ as

$$\mathbb{E}\left[I^{j=-1}_{k=1}\right] = S_1(k/T)\cdot\Sigma_{11} + S_2(k/T)\cdot\Sigma_{12} \qquad \textbf{(2.24)}$$

**Thus instead of (2.18) we should yield something like**

$$\mathbb{E}\left[I^{j}_{[zT],T}\right] = \sum_l \Sigma_{jl}S_l(z) + O(2^{-j}/T) \qquad \textbf{(2.25)}$$

Or in Matrix form

$$\mathbb{E}\left[\vec{I}_{[zT],T}\right] = \Sigma \cdot S(\vec{z}) + somethingVanishing \quad \textbf{(2.26)}$$

And as long as we know $\Sigma^{-1}$ we can invert (2.24) !

Final note is that the variance of the raw estimated wavelet periodogram does not asymptotically vanish (just as in the classical Fourier case). So it must be smoothed. Smoothing can be understood as a way to extract the common properties from the wavelets, which (properties) are less localized than wavelets. Since in case of LSW processes the wavelet spectrum varies slowly, some smoothing should be meaningful.

As to the MSML algorithm by Cho and Fryzlewicz(2012), the realization is complicated but the idea is very similar. We infer the spectra for each level from (2.18), find according to some statistic the structural breaks within each spectra and these are, roughly speaking, our breakpoints.
There is also a post-processing step, which removes "false break points". Roughly speaking, we define a window with length $\tau$, so that if several breakpoints at *different* scales occur within such window, they are treated as one. The breakpoint, which occurs at the finest scale, is preferred.

## *Empirical analysis and the results*

I use daily historical quotes, available free of charge from yahoo.finance.com
To proceed with the segmentation according to the MSML algorithm I use original R-code by Fryzlewicz and Cho(2012). The MSML algorithm requires the length of input data to be a power of two, so I decide for $2^{11}$=2048, which corresponds to approx. 8 years (since it is 2048 business days).

As an application example, Fryzlewicz and Cho(2012) analyze the historic time series of the Dow Jones closing values.
I, however, always analyze the log returns, not the prices: first of all because it is the usual practice and secondly because the LSW-processes, on which the MSML algorithm is based, are assumed by construction to be trendless, which is clearly not the case for stock prices but very plausible for the stock log returns.

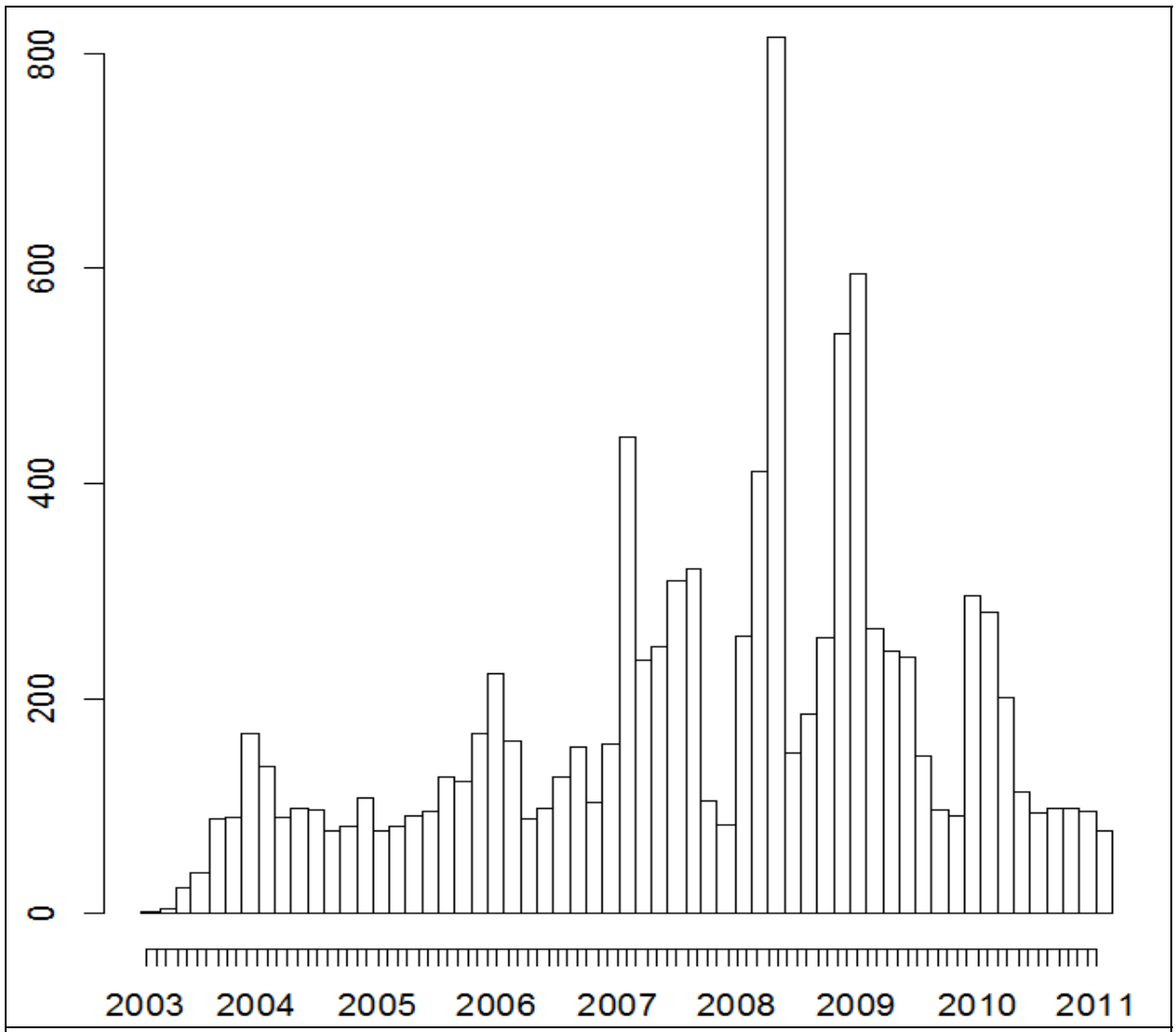


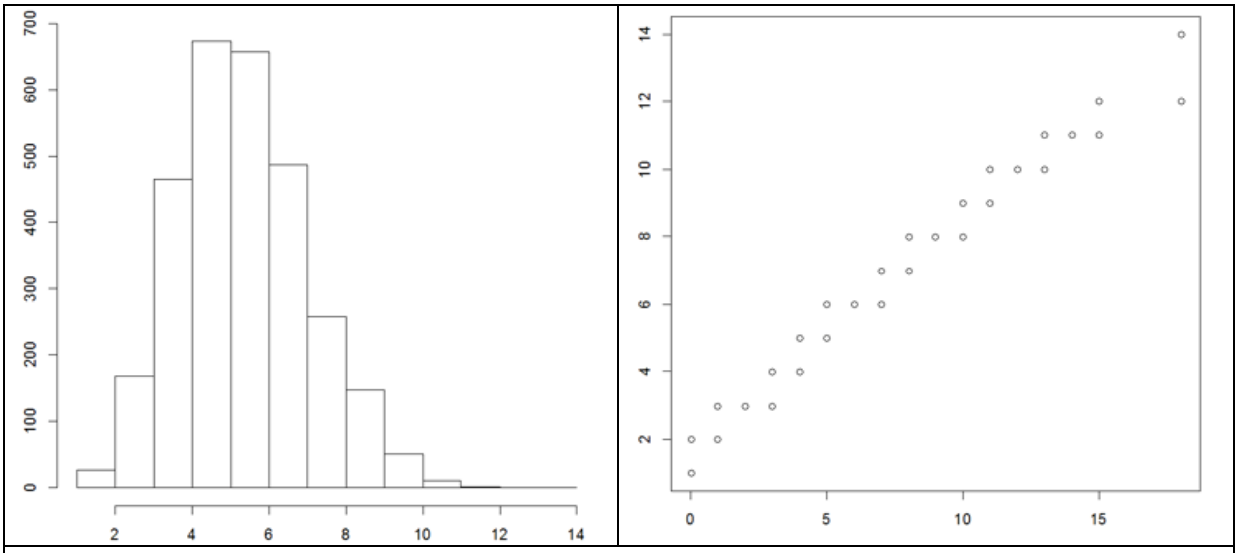

**Figure 5**: Temporal distribution of the breakpoints (X-axis is graduated monthly)

The MSML algorithm is computationally-intensive: to proceed with a stock takes about 5 minutes on a conventional PC. Surprisingly, the function compilation into R-bytecode does

not help[10]. However, the modern PC-processors have many kernels, so one can split the database into equal parts and process them simultaneously. In such a way it took two days to complete the segmentation.

The temporal distribution of the breakpoints[11] over all assets is according to the Figure 5. The peak in year 2008 is not surprising (the acutest phase of the financial crisis) but closer to the year 2011 we have fewer breakpoins, which might speak for a fading of turbulence[12].

In turn, the number of the breakpoints per asset (under condition that there is at least one breakpoint for an asset) is distributed according to Figure 5a.




| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 1.000 | 5.000 | 6.000 | 5.831 | 7.000 | 14.000 |

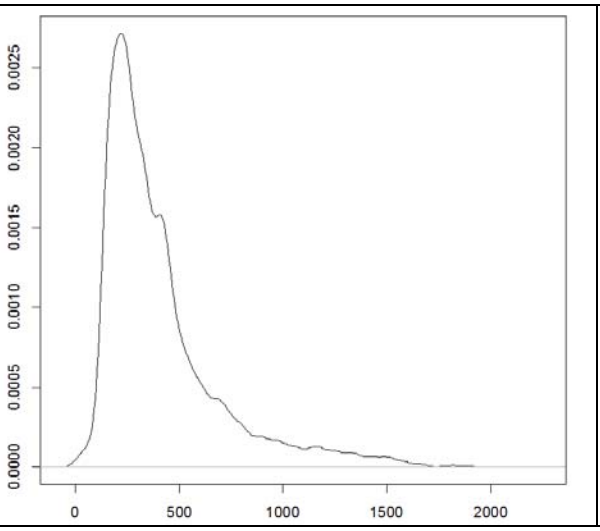
**Figure 5a**: Distribution of the number of the breakpoints per asset with summary statistics and QQ-plot against Poi(5.831)-distribution

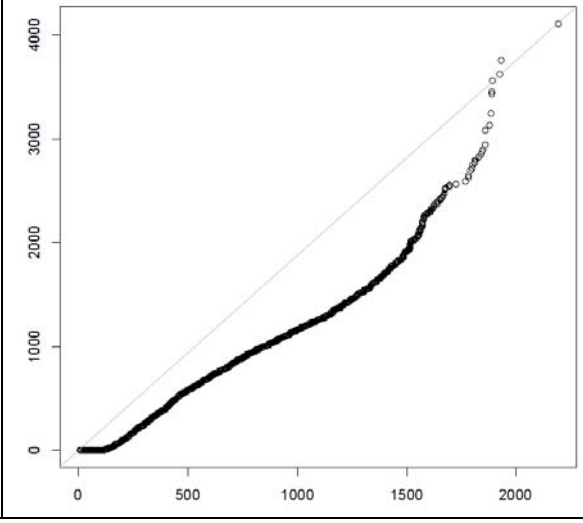
Six breakpoint on average within eight years seems to be really few.
However, the screening of the charts[13] confirms that the segmentation is plausible. As a matter of fact, a sharp change in stock prices does not necessarily mean a structural break and in such cases the market often calms down quickly.
From technical point of view though the MSML algorithm often detects a break point near price jumps(drops) it is not critically sensitive to them. For example, the famous jump in Volkswagen stock price(as Porsche tried to acquire it) is not a breakpoint (Figure 6).

The empirical distribution suites the Poisson distribution sufficiently well, which indirectly confirms that the breakpoint occurrence is a rare event.

However the time between the breakpoints is probably not $exp(\lambda)$ distributed (as Figure 5b shows), at least if we assume $\lambda = const$. But a time-dependent $\lambda(t)$ is plausible, since in crisis time the breakpoints occurs by more assets.

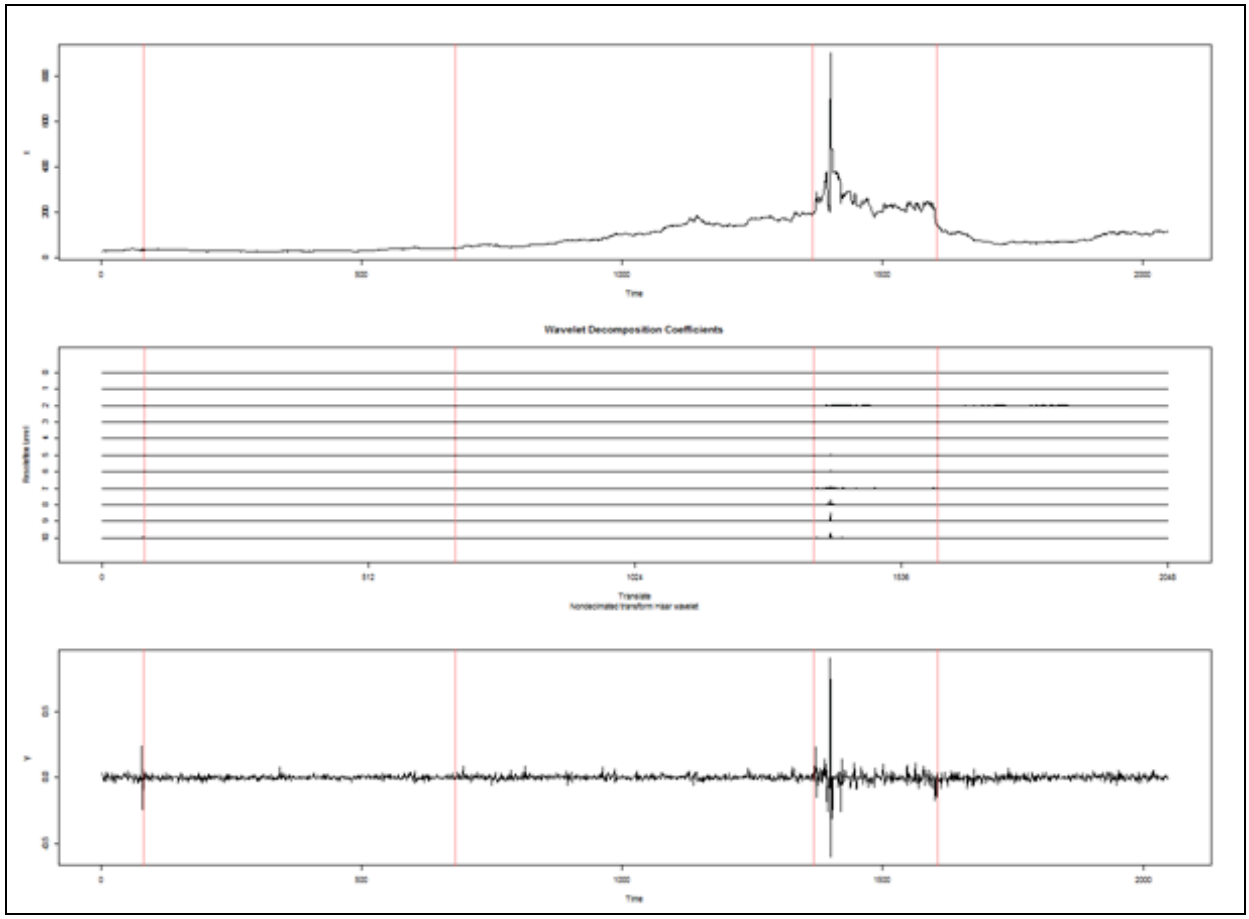
[10] checked with benchmark routine from R-package "rbenchmark"
[11] For the time span from July 2003 to August 2011
[12] Note, however, that the time period after Aug. 2011, as S&P decreased USA rating from 'AAA' to 'AA' was not considered in our analysis.
[13] An example of these charts is Figure 2




| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 4 | 218 | 322 | 410 | 488 | 2189 |

**Figure 5b**: Distribution of the number of the calendar days between two adjacent breakpoints (left). It can be considered exponential only very approximately , as the QQ-plot against the exp-distribution shows (right).



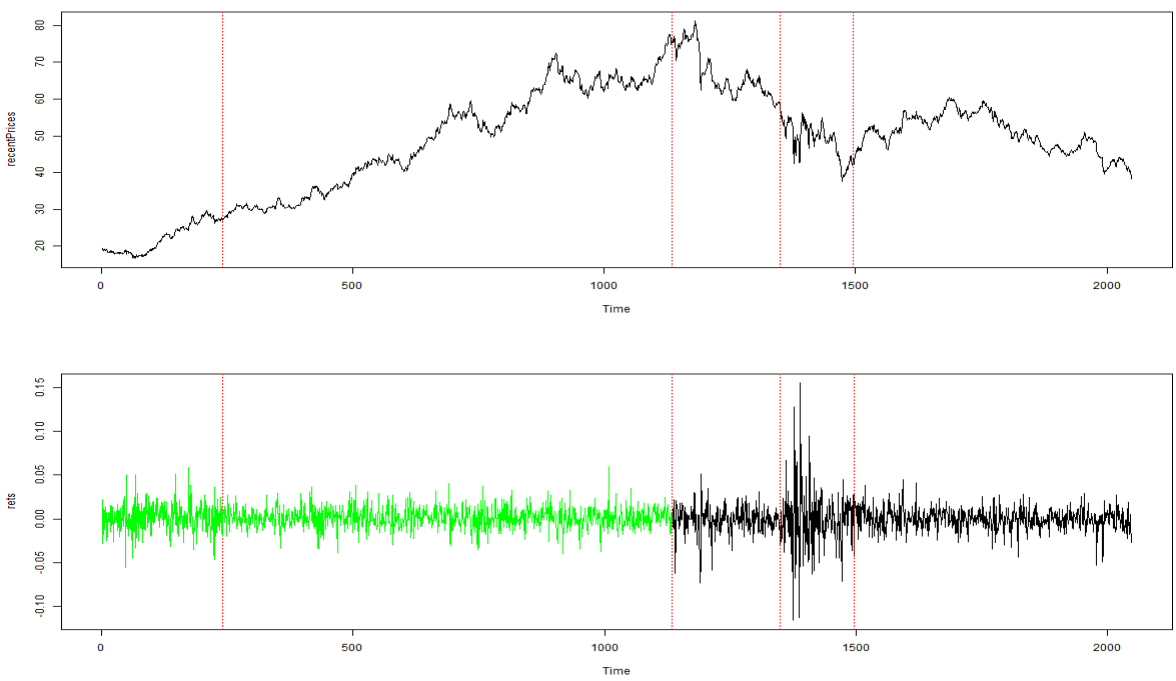



**Figure 6**: Segmentation results for Volkswagen(top) and RWE (below). The periods on which the returns are piecewise homoscedastic are plotted in green.
Short-term jumps (even huge) - as in case of Volkswagen are not necessarily the breakpoints (since, from technical point of view their influence decays immediately).
On the other hand the volatility clusters - as in case of RWE - are clearly separated from each other.

After the segmentation is done, I would like to test on which segments the time series are stationary or at least homoscedastic. But surprisingly, I failed to find *any* implementation of a [nonparametric] stationarity test, which would assess directly the covariance function or the spectrum[14]. So I merely test for the homoscedasticity, which is enough for the traders.

To test the homoscedasticity I proceed as follows: starting from a breakpoint $b_t$ I go $120$ days further. If the next breakpoint occurs, I ignore this segment otherwise I calculate the variance of the subsample $[b_t, b_t + 120]$ (insampleVariance) and of the subsample $[b_t + 121, b_{t+1})$ (outOfSampleVariance), where $b_{t+1}$ is the next breakpoint after $b_t$.

This test imitates the following trading idea: usually we know from the news that there is a "breakpoint" (i.e. disappointing annual report, product innovation and so on). The hype begins and volatility grows. A patient trader waits and let the hype calm down.
Why then wait exactly $120$ days? Well, I took this number because 120 business days is approximately 6 months and so long lasted the acutest phase of the Crisis'2008. However, the auxiliary analysis (which I do not report here) shows that the results are insensitive to the choice of the number of "waiting days".
Then I apply Bartlett test, Levene test *and* check that the
relative error = (insampleVariance - outOfSampleVariance)/inSampleVariance
does not exceed 20%. The verification of the relative error is important because when one

[14] There is a function stationarity(x) in package {fractal}, which implements the Priesley-Subba Rao(1969) test. But the implementation is buggy, at least R 2.13. 2 and additionally it is unclear how to extract the p-values from the returned object.
Several tests for stationarity are mentioned in Dahlhaus(2012), p.53

relies merely on Bartlett and Levene tests, they let sometimes the severe cases pass (as Figure 8 shows).

In total there are 8191 segments on which the homoscedasticity holds true. They are plotted in green, as at Figure 6. In total there are 20131 segments (including those that are shorter than 120 business days). **This means that the trader, who carefully watches the structural breaks in dynamics of stock returns can hope in *40%* cases be able to infer from the past volatility the future volatility (up to the next structural break). This is pretty encouraging!**

Surprisingly, there are nearly as many homoscedastic segments during the crisis times as during the calm times (Figure 7), which means that after the crisis the volatility jumps but remains at the new higher level approximately constant.
It is also encouraging, since the traders yields at most during the turbulent times. However, it may be not very relevant for a private investor, who is risk-averse.

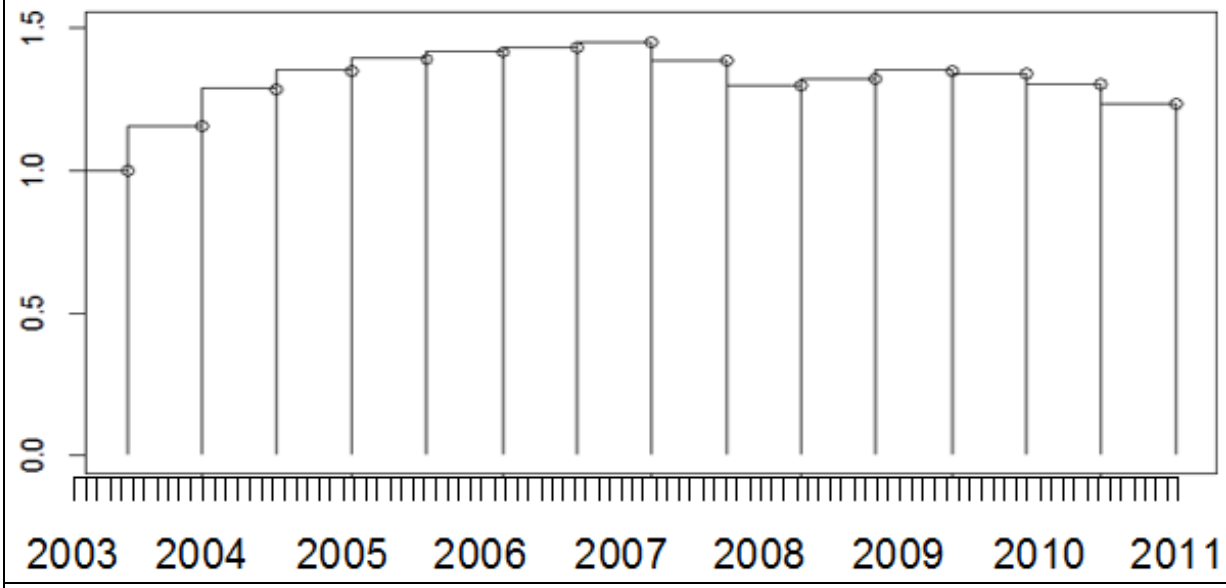



**Figure 7**: Temporal distribution of the number of homoscedastic segments over all stocks. (The distribution is normalized: the height of the first bin is set to 1.0)

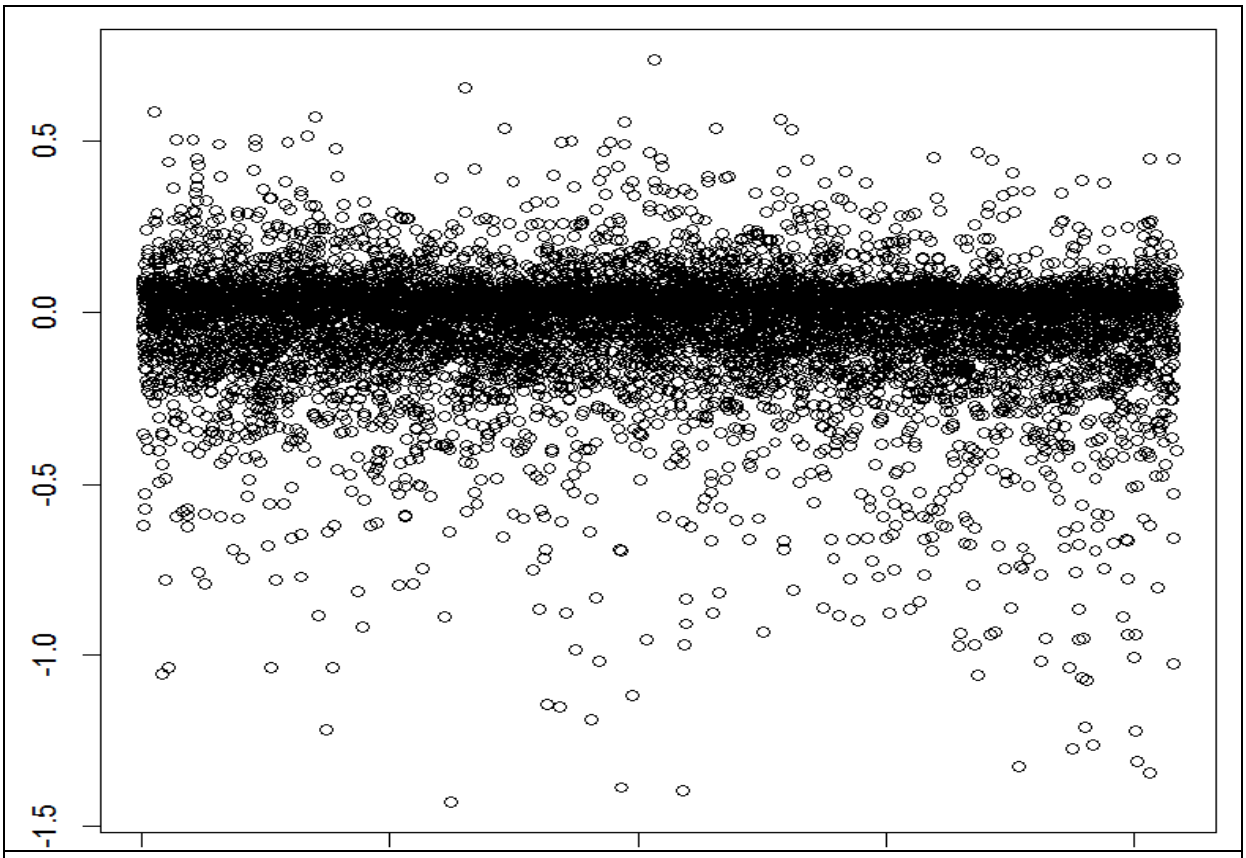



**Figure 8**: Relative variance deviation: (inSampleVola - outOfSampleVola) / inSampleVola for the cases when both Bartlett and Levene test do not reject the homogeneous variance hypothesis.
Though in most cases the deviation is less than 25%, outOfSampleVola may be underestimated - and sometimes severely!

## *References*


- Cho, H. and Fryzlewicz, P. (2012) "Multiscale and multilevel technique for consistent segmentation of nonstationary time series."
  Statistica Sinica, 22, 207-229.
  Available at: http://personal.lse.ac.uk/CHOH1/papers/msml_technique.pdf and http://personal.lse.ac.uk/CHOH1/msml_technique.html

- Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. Ann. Statist., 25, 1–37.

- Dahlhaus, R. (2012) . Locally Stationary Processes. Preprint at http://arxiv.org/pdf/1109.4174v2.pdf

- Nason, G.P. (2008) Wavelet Methods in Statistics with R. Springer

- Nason, G.P., von Sachs, R. and Kroisandt, G. (2000) "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum."
  J. R. Statist. Soc. Series B, 62, 271-292.
  available at http://www.stats.bris.ac.uk/~guy/Research/papers/WavProcEWS1.pdf

- Percival, D.B. and Walden, A.T. (2000) Wavelet Methods for Time Series Analysis. Cambridge University Press

- Schwert, W.G. (1989) " Why Does Stock Volatility Change Over Time?"
  available at http://schwert.ssb.rochester.edu/afa.htm

- "Turtles", "The original turtle trading rules"
  available at http://bigpicture.typepad.com/comments/files/turtlerules.pdf"