Logarithmic Stock Returns: leptokurtosis, heteroskedasticity and change-points

Christian Weiner*
Université Catholique de Louvain, Institut de statistique
Voie du Roman Pays 20
B 1348 Louvain-la-Neuve
Belgique - Belgium
Tel: +32-10-474314   Fax: +32-10-473032
e-mail: weiner@stat.ucl.ac.be

November 1996

Abstract

There are serious reasons to assume that logarithmic stock returns $\log S_{t_i} - \log S_{t_{i-1}}$ are normally distributed, however, it is frequently denoted that the empirical distribution of stock returns characteristically deviates form a normal distribution: It is leptokurtic, it is peaked and it has thick tails.

But it is not necessary to reject normality as a model for changing volatility, i. e. heteroskedasticity of logarithmic returns, may explain the phenomenon. The paper deals with a change of volatility in the long term, in this case volatility can be estimated nonparametrically as function in time. Particularly it seems to be fruitful to consider this function as a piecewise constant line, in this model one has to estimate the points where volatility changes.

This model for volatility differs form the two common approaches, Stochastic Volatility (SV) and Autoregressive Conditional Heteroskedasticity (ARCH); both put more emphasis on changes of volatility in the short term.

German stock data from 1990 to 1992 are studied as examples.

Acknowledgment: The author is grateful to Pierre Ars, Alois Kneip, Etienne Marceau, Valentin Patilea, Jim Ramsay, Léopold Simar, and Jean-Pierre Vandeuren, who read this paper at different stages, for their helpful comments and corrections.

*Supported by "Projet d'Action de Recherche Concentrées" (PARC 93/98-164) of the Belgian Government

1 Introduction

One of the first references for a stochastic model for stock prices is L. Bachelier's thesis *"Théorie de la speculation"* from 1900 [1]. As stock prices sum up random events Bachelier assumed that changes of stock prices $S_{t_i} - S_{t_{i-1}}$ in distinct intervals $(t_{i-1}, t_i]$ are infinitely divisible and independent. Since a normal distribution is infinitely divisible, Bachelier concluded that stock price changes must follow a normal distribution.

Yet infinite divisibility is a necessary condition for a normal law, but not a sufficient one. Normal distributions are uniquely determined, if we assume the distribution to have finite variance and to be stable; infinite divisibility only implies a generalized Poisson distribution[1]. Infinitely divisible changes $S_{t_i} - S_{t_{i-1}}$ aggregating a large number of events may also look normal, since a Poisson distribution with a large parameter approximates a normal distribution; see Gnedenko and Kolmogorov [12] for the definition and characterization of infinite divisibility and stability.

Much more important is another critic against Bachelier's model: It allows stock prices to become negative. Owing to it P. Samuelson [20], [21] suggested considering relative changes $\frac{S_{t_i}}{S_{t_{i-1}}}$ instead of absolute changes. Moreover, from a speculators point of view we may prefer relative changes because they relate the result to the investment. Canonically Bachelier's condition and its strengthening is converted into a condition on $\log S_t$ instead of S_t. Therefore we consider *logarithmic returns*

$$X_i = \log S_{t_{i+1}} - \log S_{t_i};$$

as above we may assume logarithmic returns to be normally distributed.

Most empirical studies deal with daily, weekly, or monthly observations. Historically the most influential publications seem to be B. Mandelbrot's "The Variation of Certain Speculative Prices"[15] (1963) and E. Fama's "The Behavior of Stock Market Prices"[9](1965). Their result is: The mean is close to zero, the distribution is symmetric, but comparing with a normal distribution there are too much observations around zero ("peak at zero") and in the tails ("thick tails"), and the observed kurtosis is significantly too large, i. e. the distribution is *leptokurtic*. These observations have been confirmed by various researchers, e. g. Officer [18] (1972), Blattberg and Gonedes [3] (1974), or Mittnik and Rachev [17] (1993).

As the mean of a stock should develop like a "riskless asset" plus risk premium, we might expect the mean of X_i to be positive, however, if it were significantly larger than zero, interest and/or risk premium would have to be phenomenal. Therefore the zero mean is not surprising, but leptokurtosis, the peak and thick tails have to be explained.

[1] i. e. the distribution of a weighted sum of independent Poisson distributed random variables or the limit of such sequences

Mandelbrot interpreted these phenomena as follows: Logarithmic returns may follow any stable law. Except for the normal distribution, stable laws have infinite variance, therefore it is too restrictive to assume that the variance of X_i is finite. As the sample size increases, the pth sample moments, $p \geq 2$, go to infinity and there will be a large number of tail observations; this explains excess kurtosis, thick tails and why a normal curve with sample variance is too flat. Moreover Fama [9] showed that stable distributions fit the data quite well.

Another aspect is heteroskedasticity: We assume the stock price $S_{t_{i+1}}$ to be the product of S_{t_i} and a large number of independent elementary actions of different size. What does it imply if activity (i. e. number or size of actions) changes? - We have to care about the variance of X_i. Let us therefore decompose logarithmic returns into a positive factor $v_i > 0$ expressing the activity at the market and into a normalized random variable $Z_i = \frac{1}{v_i} X_i$, e. g. $Z_i \sim N(0,1)$; the factor v_i is the so-called *volatility*.

We shall see in section 2 that several kinds of heteroskedasticity imply leptokurtosis, a peak at zero and thick tails. This has been remarked by different authors dealing with heteroskedasticity. Furthermore Mandelbrot and Fama did not consider heteroskedasticity in their articles referred above. Thus heteroskedastic but normal X_i are an alternative explanation for their observations. Nevertheless, we should not neglect that the X_i might be non-normal.

It is frequently noted, that high volatility tends to follow high volatility and low volatility tends to follow low volatility, cf. Black [2]; this is called *volatility clustering*. The common econometric model for this kind of observations is a time series with a positive autocorrelation. But we can also take into account volatility clustering if we fit a moderately changing function in time nonparametrically. Let us call this function *volatility function*. As we make no parametric assumption on the structure of this volatility function, we can't use it for forecasts; instead it is appropriate for an ex ante analysis of empirical data.

To estimate a regression function we usually assume the function to be sufficiently smooth, e. g. twice continuously differentiable. Several estimators are known, see e. g. the book of J. Fan and I. Gijbels [10] or W. Härdle [13] for a survey. But there is no estimator known which is favorable for all problems, instead we have to choose an estimator depending on the characteristics of the data set. Having no idea about these characteristics, our first choice could be the *Nadaraya Watson* estimator, since it is easy to compute and since we have good knowledge about its properties.

The smoothness assumption, however is too restrictive for a volatility function in the present data set. Consider that there are surprises for the market on which speculators must react immediately. Alternatively to a smooth curve, volatility may be piecewise constant with breaks or *change-points* . The latter approach is inspired by quotations of German stocks from 1990 to 1992, see figure 1. There was Iraq's attack on Kuwait around $t = 150$ and the unsuccessful coup in Moscow around $t = 400$; we can see the changes at these dates in the data.

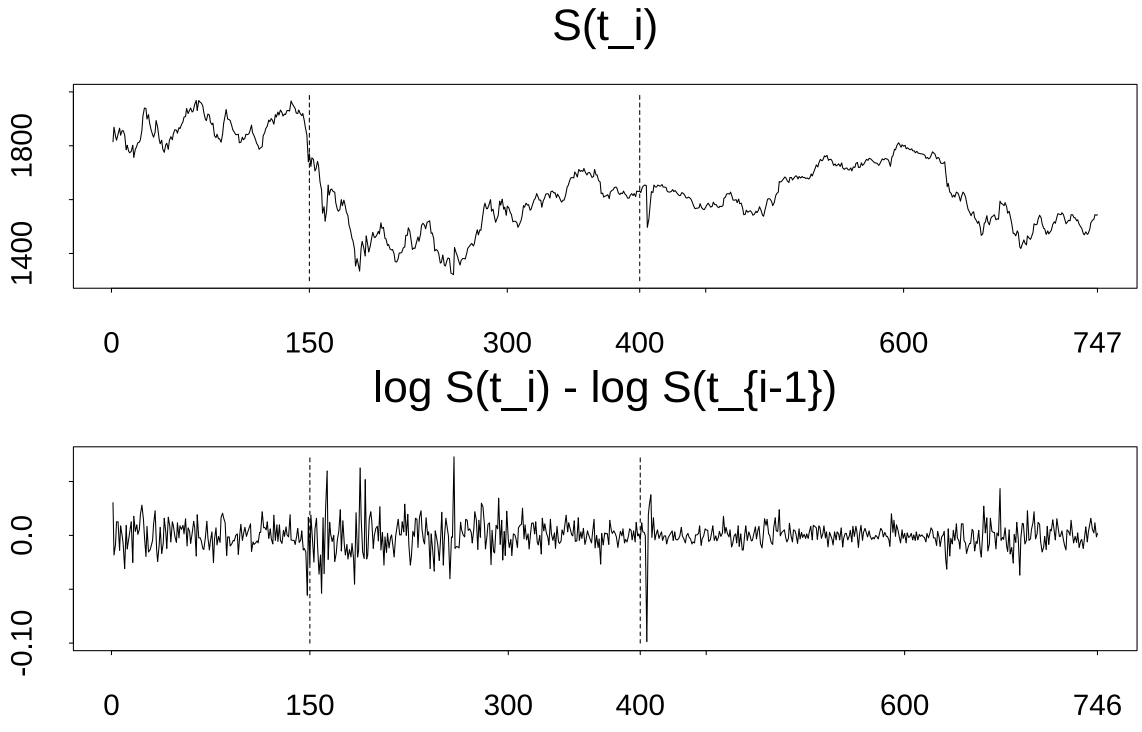

Figure 1: DAX, 2.1.1990 - 30.12.1992

Two other approaches modeling changing volatility are widely used in econometrics: Stochastic Volatility (SV), introduced by Mandelbrot and Taylor [16], and Clark [5]; and Autoregressive Conditional Heteroskedasticity (ARCH), suggested by Engle [7]. The SV model states that volatility is a random variable which is independent of the Z_i; see Ghysels et al. [11] for a survey of SV models. ARCH models state that volatility is a function of the history of the process; modeling volatility as time series with a positive autocorrelation (see above) is a special case of the ARCH model. As in other branches of time series analysis, applications mainly concern ergodic ARCH processes, this is important for estimating volatility and it also allows forecasts; see Bollerslev et al. [4] for a survey of ARCH models.

Comparing the volatility function approach of this paper with the SV and the ARCH approach, the former takes into account changes in the long term, SV and ARCH models put more emphasis on changes in the short term. These approaches are non complementary, instead it is possible to combine them: we could assume volatility to be the sum or the product of a volatility function with a white noise or a time series.

The paper is organized as follows: Section 2 shows that several kinds of heteroskedasticity can explain leptokurtosis, the peak at zero and thick tails; it is a generalization of this well known property of SV and ergodic ARCH models. Section 3 provides the setup for estimating a volatility function and a short description of the Nadaraya Watson estimator. Section 4 suggests a method how to estimate change-points in a piecewise constant line and section 5 presents the results of this method applied to German stock data from 1990 to 1992.

2 Heteroskedasticity

Using descriptive statistics like histograms, nonparametric density estimators, sample moments or quantiles, we characterize the empirical distribution of the observations

$$\rho_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$$

(δ_x denotes the Dirac measure, which assigns one to each set containing x and zero else). For ergodic X_i, especially for iid observations, this empirical measure approximates $\mathcal{L}(X_i)$ ($\mathcal{L}(X_i)$ denotes the marginal distribution of the X_i), i. e. $P[\rho_n \xrightarrow{w} \mathcal{L}(X_i)] = 1$. This allows us to infer on the distribution of X_i, however, in case of heteroskedasticity this inference fails. In the following two subsections we shall see, that ρ_n approximates a mixed normal law, and (under mild conditions) that a mixed normal law is leptokurtic and has a peaked density with thick tails.

2.1 The impact of heteroskedasticity

Let us first consider an easy model: Assume the X_i to be independent, $\mathcal{L}(X_i) = N_{(0,v_i^2)}$ (i. e. X_i normally distributed with mean zero and variance v_i^2) and deterministic v_i's. Then the empirical distribution ρ_n approximates a mixed normal distribution $\frac{1}{n}\sum_{i=1}^{n} N_{(0,v_i^2)}[\cdot]$ in the sense of

$$P\Big[\rho_n[\cdot] - \frac{1}{n}\sum_{i=1}^{n} N_{(0,v_i^2)}[\cdot] \xrightarrow{w} 0\Big] \;=\; 1$$

We may write the mixed normal law as $\frac{1}{n}\sum_{i=1}^{n} N_{(0,v_i^2)}[\cdot] = \int N_{(0,v^2)}[A]Q_n(dv)$ where $Q_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{v_i}$ denotes the mixing distribution.

Heteroskedasticity also induces a mixed normal law for more sophisticated models: Consider a SV model, where the volatilities V_i do not depend on the normalized random variables Z_i. Here ρ_n approximates a mixed normal distribution conditionally on $V_i = v_i$, $i = 1,\ldots n$. If the V_i are ergodic then the mixing distribution $Q_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{v_i}$ approximates the marginal distribution Q of the volatility V_i. This implies[2] $P[\rho_n[\cdot] \xrightarrow{w} \int N_{(0,v^2)}[\cdot]\, Q(dv)] = 1$.

Supposing an ergodic ARCH model, where volatility is a function of the "history of the path", with $X_1 = V_0 Z_1$, $\mathcal{L}(Z_1) = N_{(0,1)}$ and $V_0 > 0$, the empirical

[2] Assume $\int N(0, v^2)[A]Q(dv) < \infty$ for every measurable set A

distribution also approximates a mixed normal law. Consider $\rho_n[A] \approx \mathcal{L}(X_1)[A] = E\Big[E[P_{N_{(0,V_0^2)}}[A]|V_0 = v]\Big] = \int N_{(0,v^2)}[A]Q(dv)$ where Q denotes the distribution of V_0.

2.2 Mixed normal distributions

Let $MN_{(0,Q)}[\cdot] := \int N_{(0,v^2)}[\cdot]Q(dv)$ denote a mixed normal measure, where Q denotes the mixing distribution. Assuming $E_Q[\frac{1}{V}] < \infty$ and $Q[V > 0] = 1$, it has the following characteristics: it has a density

$$f_{MN_{(0,Q)}}(x) = \int f_{v^2}(x)\, Q(dv) \tag{1}$$

where $f_{\sigma^2}(\cdot)$ denotes the density of a $N_{(0,\sigma^2)}$ distribution. The moments are

$$\int x^p f_{MN_{(0,Q)}}(x)dx = \int\int x^p f_{v^2}(x)dx\, Q(dv) = m_p E_Q[V^p]$$

where m_p denotes the p th moment of a $N_{(0,1)}$ distribution. Due to it the mean is zero, it is symmetric, and it is leptokurtic:

$$3\frac{E_Q[V^4]}{E_Q[V^2]^2} - 3 = 3\frac{\mathrm{var}_Q[V^2]}{E_Q[V^2]^2} > 0$$

Finally let us compare the density of a mixed normal distribution (1) with the density $f_{E[V^2]}(x)$, which is approximately the density of a normal distribution with zero mean and sample variance. The second derivative of $f_{\sigma^2}(x)$ with respect to σ^2 (x constant) is

$$\frac{\partial^2}{\partial^2\sigma^2} f_{\sigma^2}(x) = \frac{1}{4\sqrt{2\pi}}\sigma^{-9}e^{-\frac{x^2}{2\alpha}}[(x^2 - 3\sigma^2)^2 - 6\sigma^4] \tag{2}$$

For all σ^2, this is positive at $x = 0$. Therefore $f_{\sigma^2}(0)$ is a convex function in σ^2 and by Jensen's inequality follows $f_{E[V^2]}(0) < \int f_{v^2}(0)Q(dv)$. Because both densities are continuous in $x = 0$, this inequality also holds for x in a small region around zero, thus the density of a mixed normal law is "peaked at zero". Assuming V^2 to be bounded above, i. e. $\exists \bar{v}^2 < \infty : Q[V^2 < \bar{v}^2] = 1$, we also get a positive second derivative (2) for $x > (3 + \sqrt{6})\bar{v}^2$; this implies "thick tails". Note that the upper bound $\bar{v}^2$ for V^2 is a sufficient condition, but it is not necessary.

3 Fitting a volatility function

As stated in the introduction we emphasize on the change of volatility on the long term. Due to it we consider independent normally distributed random variables where the variance is given by a volatility function

$$X_i \sim N_{(0,v(t_i)^2)} \quad \textit{independent};\ v : I\!R_0^+ \to I\!R^+$$

This is the "easy model" for volatility of the previous section, were we have already seen, that it explains leptokurtosis, a peaked density and thick tails.

To estimate the volatility function $v(\cdot)$ we may consider two data transformations: First the squared X_i:

$$X_i^2 \;=\; v(t_i)^2 + \varepsilon_i \tag{3}$$

We have: ε_i independent, $E[\varepsilon_i] = 0$ and $\mathrm{var}[\varepsilon_i] = 2v(t_i)^4$. Alternatively we may consider the logarithm of the squared observations

$$\log X_i^2 \;=\; \log v(t_i)^2 + C_1 + \eta_i \tag{4}$$

where $C_1 := E[\log Z_i^2]$ and $\eta_i := \log Z_i^2 - C_1$ (remember $Z_i = \frac{X_i}{v(t_i)} \sim N_{(0,1)}$); we have η_i iid, $E[\eta_i] = 0$ and $\mathrm{var}[\eta_i] = C_2$. The constants C_1 and C_2 are approximatively[3] $C_1 \approx -1.27$ and $C_2 \approx 4.93$. We shall use these constants below.

Both equations, (3) and (4), are nonparametric regressions on a fixed design of the form

$$Y_i \;=\; f(t_i) + u_i \tag{5}$$

where $f()$ is a function and u_i denotes an error term. For the present regressions (3) and (4) f is the squared volatility function respectively the logarithm of the squared volatility function plus C_1

As pointed in the introduction we have to be aware that logarithmic stock returns are not normal. In this case the two transformations (3) and (4) are still useful. Assuming X_i to have finite fourth moments, we can estimate the squared volatility function from the squared observations; non-normality only affects the residual variance. Assuming $E[\log X_i^2] < \infty$ allows to estimate $\log v(\cdot)^2$ up to an unknown constant; consider that non-normality affects both constants C_1 and C_2. Nevertheless $\log X_i^2$ provides information on the change of the volatility function.

3.1 Nadaraya Watson estimator

The Nadaraya Watson estimator for the nonparametric regression (5) is a locally weighted average

$$\hat{f}_h(t_i) \;=\; \sum_{j=1}^{n} w_j(h,t_i)Y_j; \quad w_j(h,t_i) \;:=\; \frac{K(\frac{t_i - t_j}{h})}{\sum_{k=1}^{n} K(\frac{t_i - t_k}{h})},$$

where the weights $w_j(h,t_i)$ are given by a *kernel* function K, i. e. a probability density, and the *bandwidth* h. A good choice for the kernel is the Epanechnikov kernel $K(x) = \frac{3}{4}(1-x^2)_+$. (The subscript + denotes the "positive part of a function", i. e. $K(x)$ is $\frac{3}{4}(1-x^2)$ for $|x| < 1$ and zero else.)

[3] These constants can be obtained by calculating the first and second derivative of the moment generating function; on needs numerical evaluations of the Γ function and its derivative.

Using this kernel we take a weighted average of observations in $(t_i - h, t_i + h)$ where the weights decrease as the distance $|t_j - t_i|$ increases. If the bandwidth is small and if the function doesn't vary too much the bias will be moderate. On the other hand the larger the bandwidth, the smaller the variance, because we take the sum of a larger number of residuals.

If f is continuous in t_i it becomes asymptotically constant on $(t_i - h, t_i + h)$ as h decreases to zero. This implies that the estimator is asymptotically unbiased as h goes to zero. Choosing a sequence h_n that decreases to zero, such that the number of observations in $(t_i - h_n, t_i + h_n)$ goes to infinity we obtain a consistent estimator. If f is not continuous in t_i, the estimator is generally not consistent; see Fan and Gijbels [10] or Härdle [13] for a discussion of this estimator and for methods how to choose the bandwidth h.

Summarizing its properties, the Nadaraya Watson estimator detects variation of a function in the long term, however, information about variation in the short term and discontinuities is lost. We shall see in the empirical results - chapter 5, figure 3 - that the Nadaraya Watson estimator is a poor estimator around the "surprises"; nevertheless it shows the change of volatility in the long term.

4 Change-points

Suppose the volatility function $v(\cdot)$ is piecewise constant. Then $v(t_i)^2$ and $\log v(t_i)^2 + C_1$ are also piecewise constant and we may search the change-points in the X_i^2 or in the $\log X_i^2$ data. We prefer the latter as the residuals are iid and as we know the residual variance, $\mathrm{var}[u_i] = C_2$. Moreover in case of non-normal X_i we need not to worry: the change-points of a function f do not depend on the additive constant C_1 and as we shall see below the constant C_2 doesn't affect the location of our estimates for the change-points.

For more convenience let us assume that the change-points are at observation times t_{k_j} only. Consider the following three properties of least squares approximations:

Property A *Let f be an arbitrary function and $\{t_{k_j} : 1 \le j \le M-1,\ 0 =: k_0 < k_1 < \ldots < k_{M-1} < k_M := n\}$ a given set of intersections. The solution of the minimizing problem*

$$(\hat{a}_1, \ldots \hat{a}_M) : \sum_{j=1}^{M} \sum_{i=k_{j-1}+1}^{k_j} (f(t_i) - \hat{a}_j)^2 = \min_{(a_1, \ldots a_M)} \sum_{j=1}^{M} \sum_{i=k_{j-1}+1}^{k_j} (f(t_i) - a_j)^2$$

is given by the means between the intersections

$$\hat{a_j} = \overline{f(t.)}^{k_{j-1}+1, k_j} := \frac{1}{k_j - k_{j-1}} \sum_{i=k_{j-1}+1}^{k_j} f(t_i),\ 1 \le j \le M.$$

Property B *Let f be an arbitrary function. Then we have for all $k = 1, \dots n$ the following inequality*

$$\sum_{i=1}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2 \geq \sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,k})^2 + \sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{k+1,n})^2$$

and: f is constant if and only if there is equality for every k.

Property C *Let f be piecewise constant with change-points at*

$$\{t_{k_j} \,:\, 1 \leq j \leq M-1,\ 0 =: k_0 < k_1 < \ldots < k_{M-1} < k_M := n\}$$

and let t_{k^*} be the "best choice for the first change-point", i. e.

$$\begin{aligned} k^* \quad : \quad & \sum_{i=1}^{k^*}(f(t_i) - \overline{f(t.)}^{1,k^*}))^2 + \sum_{i=k^*+1}^{n}(f(t_i) - \overline{f(t.)}^{k^*+1,n})^2 \\ & = \min_{1\leq k\leq n} \sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,k}))^2 + \sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{k+1,n})^2, \end{aligned}$$

then t_{k^*} is one of the change-points, i. e. $t_{k^*} \in \{t_{k_1}, \dots, t_{k_{M-1}}\}$

A proof of these properties is given in appendix A

The crucial point is property C. It allows to search the change-points sequentially. This has two advantages: a sequential search is much faster than a joint search of M change-points, and we don't need to know the number M.

To find the first change-point, we have to minimize in k

$$S(k) = \sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,k}))^2 + \sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{k+1,n})^2$$

Yet we can't do it exactly, because we don't know the real values of $f(t_i)$. But we can estimate $S(k)$, consider

$$\sum_{i=1}^{k}((f(t_i) - \overline{f(t.)}^{1,k})^2 = \sum_{i=1}^{k}(Y_i - \overline{Y.}^{1,k})^2 + \underbrace{\sum_{i=1}^{k}(u_i - \overline{u.}^{1,k})^2 - 2\sum_{i=1}^{k}(Y_i - \overline{Y.}^{1,k})(u_i - \overline{u.}^{1,k})}_{(*)};$$

$(*)$ is a random variable which we do not observe. But for large k it has a small variance and we may replace $(*)$ by its mean

$$E[(*)] = (k-1)\mathrm{var}[u_i] - 2(k-1)\mathrm{var}[u_i] = -(k-1)C_2$$

($C_2 \approx 4.93$ is the constant defined above in section 3) and get an unbiased estimator for $S(k)$:

$$\hat{S}(k) := \sum_{i=1}^{k}(Y_i - \overline{Y.}^{1,k})^2 + \sum_{i=k+1}^{n}(Y_i - \overline{Y.}^{k+1,n})^2 - (n-2)C_2;$$

here we take advantage of the iid residuals and the knowledge of $\mathrm{var}[u_i] = C_2$.

The variance of this estimator is large for small k or $n-k$, and it is comparatively small for k around $\frac{n}{2}$. Therefore we have to be careful if the minimizing value k of $\hat{S}(k)$ is close to 1 or n, it might be due to random. We reduce this problem by assuming a minimal distance $\delta > 0$, i. e. we search the minimum for $k \in (\delta, n-\delta]$ only. However, if there is a change-point outside $(\delta, n-\delta]$ we shall not find it, or we shall find δ, respectively $n-\delta$, instead. For the particular choice of δ we must weigh these two types of error.

Comparing the estimator $\hat{S}(k)$ with $S(k)$, we replace the real value $f(t_i)$ by the noisy observation Y_i and subtract $(n-2)C_2$ for correction. As this term is the same for all $S(k)$ it has no influence on the location of the point k^* that minimizes $\hat{S}(k)$, but we may not leave it out. Consider there is a change-point in $(0, n]$, only if $S(k)$ is smaller than $S := \sum_{i=1}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2$ (property B); the latter is estimated by $\hat{S} := \sum_{i=1}^{n}(Y_i - \overline{Y.}^{1,n})^2 - (n-1)C_2$. This estimator has a larger correction term than $\hat{S}(k)$, $(n-1)C_2 > (n-2)C_2$. So we choose k^* as estimate for a change-point, if it minimizes $\hat{S}(k), k \in (\delta, n-\delta]$, and if $\hat{S}(k^*) < \hat{S}$, otherwise we stop the search. We also stop, if $\delta \geq n - \delta$.

If we don't stop we have to start an induction to search further change points: Having estimated $m-1$ change-points $0 = k_0 < k_1 \ldots < k_{m-1} < k_m = n$, we have m intervals $(k_{j-1}, k_j]$, in which we have to find a change-point, respectively where we have to state, that there is no further change. This is the same problem as finding the first change-point; see appendix B for the complete algorithm.

Finally, if the X_i are non-normal, then the variance $\mathrm{var}[u_i]$ can be different form C_2. This affects the number of estimated change-points but not the location.

5 Data and Empirical Results

We use daily quotations (closing quotes) of 15 german stocks at the stock exchange in Frankfurt a. M. and the german stock index DAX[4] from January 1990 to December 1992. The data are available at `savage@dg-rtp.dg.com` via FTP.

Weekends and holidays are counted as no days, so we have 747 days. There are observations at all days for 12 of the stocks and the DAX. Allianz quotations start at March 19th 1991, RWE quotations start at May 2nd 1990. Furthermore there is no quotation for Bayer at January 24th, 1990. This day is counted as no day.

Figure 1 in the introduction (section 1) shows the closing quotation of the DAX and the logarithmic returns. On the x axis there are numbers from 1 to 747, respectively 746. In the logarithmic returns we see a period with higher oscillation from 150 to 250, some extreme values around 400 and a higher oscillation at the end of

[4] **D**eutscher **A**ktien Inde**x**

logarithmic returns $X_i = \log S_{t_{i+1}} - \log S_{t_i}$							
	n	diff. states	*zeros*		n	diff. states	zeros
DAX	746	745	2	Dresdner Bank	746	688	35
Allianz	444	421	22	Hoechst	746	725	19
BASF	746	719	25	Mannesmann	746	693	34
Bayer	745	729	12	RWE	663	636	19
BMW	746	697	36	Siemens	746	731	13
Commerzbank	746	698	29	Thyssen	746	703	30
Daimler Benz	746	725	18	VEBA	746	723	17
Deutsche Bank	746	719	21	VW	746	722	21

Table 1: tied observations and zeros

the period. The first is due to the Kuwait crisis, the second is during the coup in Moscow and the last can be explained by troubles in the European Currency System after speculations against the British Pound.

5.1 Empirical distributions

Stocks are usually quoted with one digit after the decimal point, the DAX is published with two digits. We have for all stock ties and zeros. Most of the ties are lost because of the logarithmic transformation, so the ties in the X_i are mainly zeros; table 1 summarizes them. However we have still a large number of different states, therefore we may approximately assume an absolute continuous distribution.

Figure 2 provides a view on the empirical distribution for the DAX and Siemens. The solid line is a nonparametric density estimate; it is estimated with a kernel estimator using an Epanechnikov kernel and a bandwidth according to Silverman's rule of thumb; see Silverman [22]. For comparison the dotted line is a normal curve with sample mean and sample variance plus a correction term for the bias[5]. The dashed lines are the kernels with the actual bandwidth scaled by $\frac{1}{10}$; these lines are drawn to give an impression of the degree of smoothing.

The peak around zero is easily visible, thick tails are comparatively hard to see; this is due to the fact that the number of observations in the tails is anytime small. Zero mean and symmetry is also supported by these estimates. Estimates for the other stocks, which are not included in the paper, show the same properties.

[5]The kernel estimator is biased, i. e. its mean is the real density convoluted with the smoothing kernel. Therfore the normal curve is convoluted with this kernel

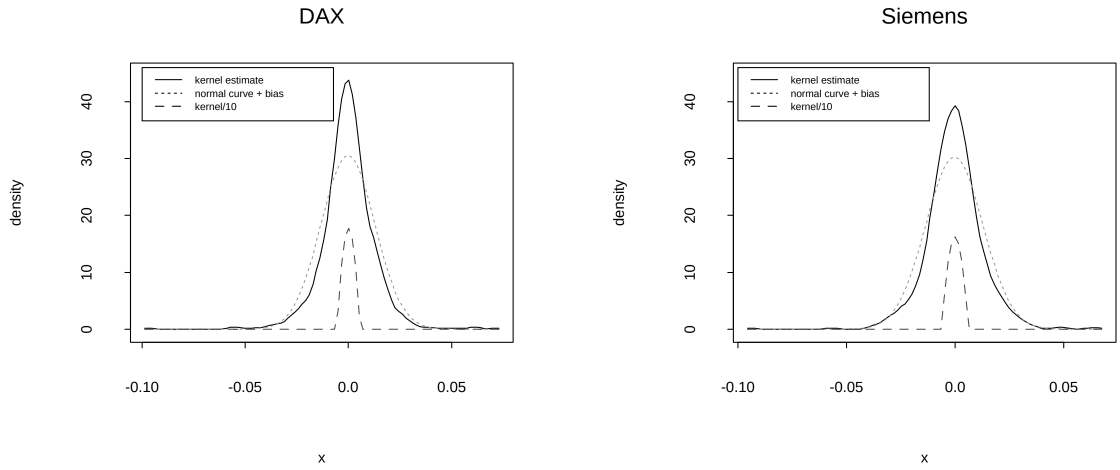

Figure 2: estimates for the density

We can also apply some tests: table 2 shows the kurtosis G_2

$$G_2 := \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^4}{(\frac{1}{n}\sum_{i=1}^{n} X_i^2)^2} - 3$$

and the p value for a two sided alternative. G_2 is asymptotically normal with zero mean and variance $\frac{24}{n}$; e. g. see Lomnicki [14]. All p values are lower than 0.0005. This implies a very high level of certainty for a rejection of normality.

	G_2	$(p_{2\text{-}side})$		G_2	$p_{2\text{-}side}$
DAX	8.190	(0.000)	Dresdner Bank	9.106	(0.000)
Allianz	10.024	(0.000)	Hoechst	7.352	(0.000)
BASF	8.633	(0.000)	Mannesmann	10.588	(0.000)
Bayer	5.572	(0.000)	RWE	7.409	(0.000)
BMW	6.350	(0.000)	Siemens	6.224	(0.000)
Commerzbank	9.559	(0.000)	Thyssen	5.383	(0.000)
Daimler Benz	4.655	(0.000)	VEBA	8.426	(0.000)
Deutsche Bank	7.140	(0.000)	VW	7.058	(0.000)

Table 2: kurtosis

	T^{sig}	($p_{2\text{-}side}$)	T^{Wil}	($p_{2\text{-}side}$)	T^{tri}	($p_{2\text{-}side}$)
DAX	-0.807	(0.420)	-0.519	(0.604)	0.119	(0.905)
Allianz	-1.266	(0.206)	-0.368	(0.713)	1.547	(0.122)
BASF	-1.452	(0.146)	-0.724	(0.469)	0.238	(0.812)
Bayer	-1.293	(0.196)	-0.363	(0.717)	-0.064	(0.949)
BMW	-0.676	(0.499)	-0.094	(0.925)	0.391	(0.696)
Commerzbank	-1.307	(0.191)	-0.314	(0.754)	0.333	(0.739)
Daimler Benz	-1.408	(0.159)	-1.058	(0.290)	0.195	(0.845)
Deutsche Bank	-0.780	(0.435)	-0.661	(0.508)	-0.222	(0.825)
Dresdner Bank	-0.563	(0.574)	-0.403	(0.687)	-0.204	(0.838)
Hoechst	-1.446	(0.148)	-0.700	(0.484(	0.896	(0.370)
Mannesmann	-0.525	(0.600)	-0.902	(0.367)	0.035	(0.972)
RWE	-0.079	(0.937)	-0.197	(0.844)	-0.562	(0.574)
Siemens	-0.702	(0.483)	-0.869	(0.385)	0.598	(0.550)
Thyssen	-1.046	(0.295)	-0.828	(0.408)	-1.059	(0.289)
VEBA	0.333	(0.739)	0.172	(0.863)	-0.810	(0.418)
Volkswagen	**-2.265**	**(0.023)**	-1.755	(0.079)	-0.060	(0.952)

Table 3: tests for the median and skewness; rejections are bold fat (only one rejection)

Table 3 shows the following test:

	Test	Hypothesis H_0	Alternatives H_1
1	sign Test T^{sig}	$median(X_i) = 0$	$median(X_i) \neq 0$
2	Wilcoxon Test T^{Wil}	$median(X_i) = 0$ and X_i symmetric	$median(X_i) \neq 0$ or X_i not symmetric
3	triple Test T^{tri}	X_i symmetric	X_i not symmetric

These tests are preferred over the first and the third sample moment, because we don't need to assume the variance of X_i to be finite. Thus they are also valid, if we have doubt in the finite variance, such as Mandelbrot [15] and Fama [9]. The triple test is due to Randles et al. [19]. They show in a Monte Carlo study, that the triple test is superior a test based on the sample skewness (i. e. the third moment). Assuming the distribution is symmetric, the Wilcoxon test is as efficient as the mean test, and the sign test is always a good test for the median. Therefore these three tests are powerful competitors for tests based on sample moments. Zero mean and symmetry are confirmed by these tests, only the sign test for Volkswagen shows a median significantly (95%) different from zero.

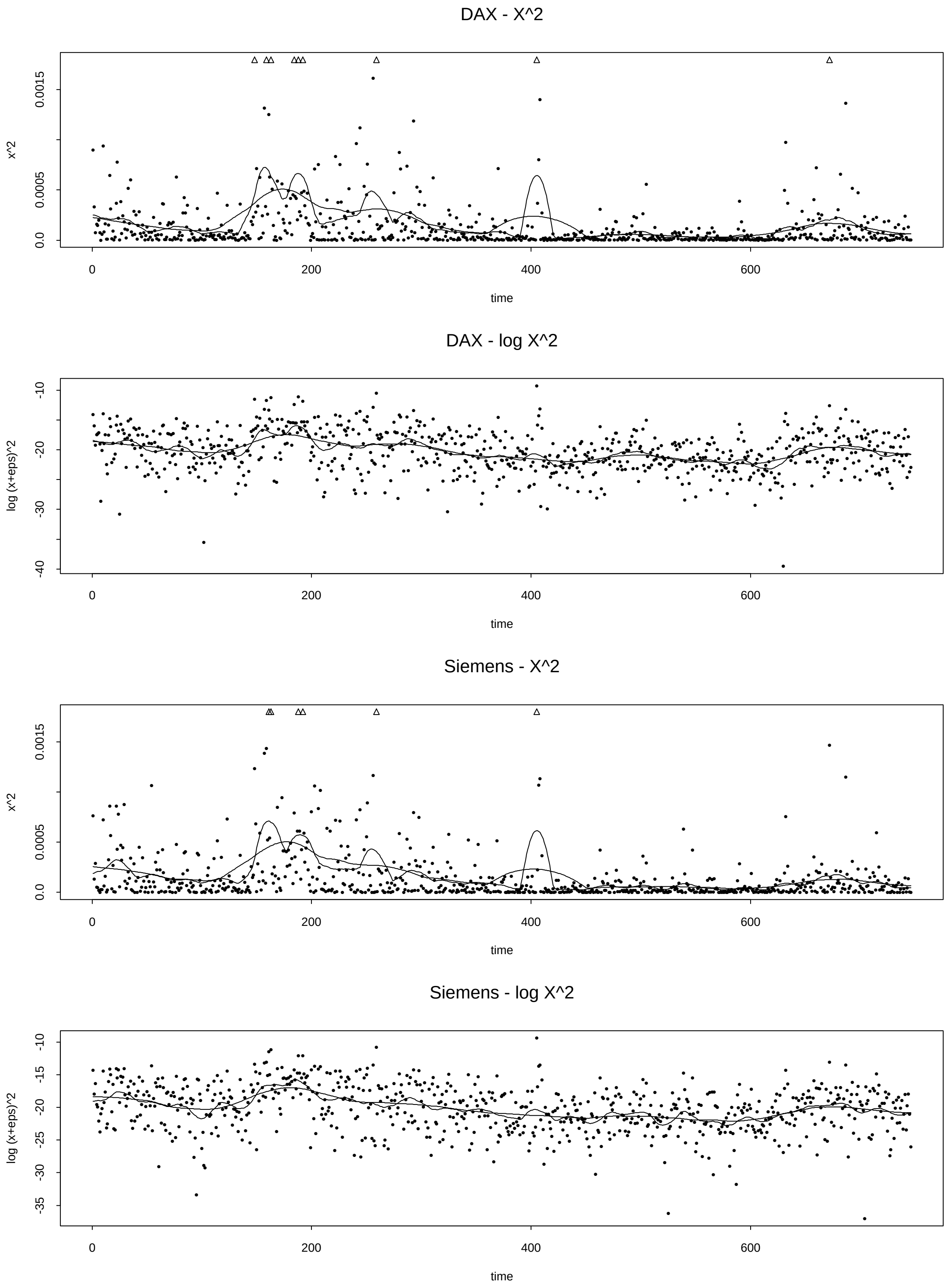

Figure 3: Nadaraya Watson estimates with bandwidth 15 and 45

5.2 Volatility

5.2.1 Nadaraya Watson estimates

Figure 3 shows Nadaraya Watson estimates of the squared volatility (3) and the log squared volatility (4) for the DAX and Siemens; different lines correspond to estimates with different bandwidth (15 and 45). Estimates with the larger bandwidth are more flat and do not reflect the sudden changes, e. g. those around $t = 150$ or $t = 400$; estimates with the lower bandwidth are more oscillating and adapt better to changes, but they also do not show change-points. Moreover the oscillation is due to the residual variance rather than to changes in the function. Estimates for the other stocks, which are not included in the paper, show the same behavior.

Consider that a bandwidth 15 corresponds to an average of three weeks before and after, and a bandwidth 45 averages over 9 weeks before and after. Having in mind, that there are breaks in the volatility, the Nadaraya Watson estimator takes too much old (respectively future) data for estimating the true volatility after (respectively before) the break. Therefore it cannot detect change-points.

5.2.2 Change-points

Figure 4 shows the X_i^2 and the $\log X_i^2$ and a step-line for the DAX and Siemens. Jumps are at the estimated change-points and the level of the steps is the sample mean between these change-points (either of the X_i^2 or the $\log X_i^2$). In order to have a reasonable scale, very large observations in the squared data are not included in the plots, instead of them there are small triangles at the upper margin. The small triangles on the bottom of the $\log X_i^2$ plots denote the zero X_i (cf table 1).

The change-points are derived from the $\log X_i^2$ data with the search algorithm in appendix B. Because of the zero observations we have added a random disturbance ε_i, $\varepsilon_i \sim N(0, \frac{1}{2500}\frac{1}{n}\sum(X_i - \bar{X})^2)$ to the X_i. The minimal distance we used is $\delta = 40$. This distance is to large to detect the beginning and the end of the coup in Moscow in August 1991. Therefore change-points at 404 and 410 are chosen in advance and the algorithm started with searching the third change-point.

Let us compare the empirical distribution of the logarithmic stock returns X_i between the change-points with a normal distribution: table 4 shows kurtosis and nonparametric test for median and symmetry and figures 5 and 6 show nonparametric density estimates and the expected normal density "plus bias".

There are similar dates for change-points at 145 (31. 7. 90), 198/199 (15./16. 10. 90), 303 (19. 3. 91), 451/455 (22./28. 10.91), 566/571 (8./15. 4. 92), and 704/706 (27./29. 10. 92). These dates are also estimated for other stocks. Furthermore there are non-common change-points, this might be due to badly chosen change-points or to events that are only important for one company.

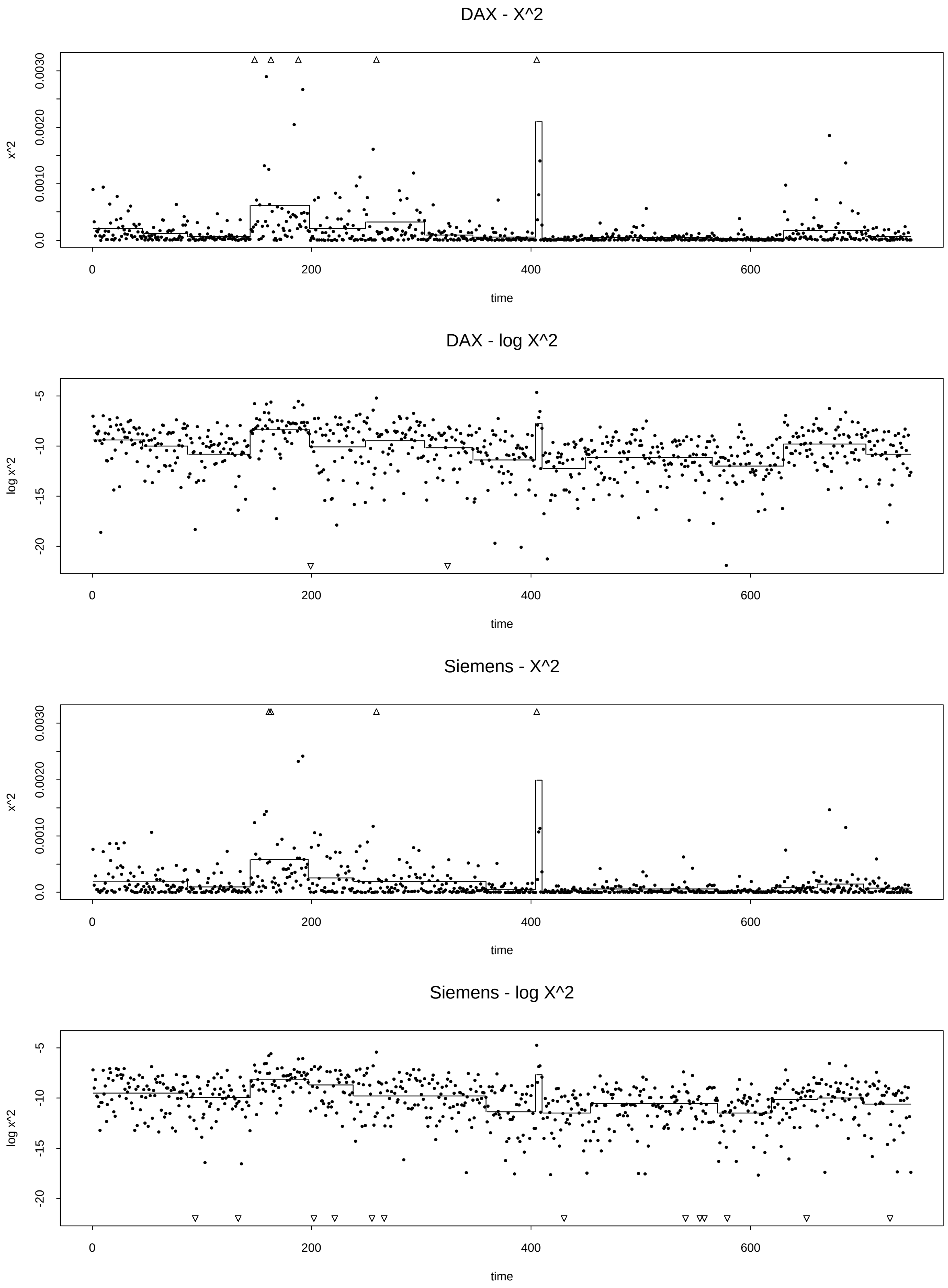

Figure 4: constant volatility within estimated change-points

Between the change-points the empirical distribution is closer to a normal distribution than for the whole period: zero mean and symmetry are mainly accepted. The kurtosis is too large for 6 periods, but we also have 10 negative values.

Consider two periods: First there is a negative mean from 145 to 197 for the DAX and Siemens. This are the first two months of the Kuwait crisis, which may explain a trend downwards in the stock prices; moreover we can see this trend in figure 1. Second the period from 455 to 570 for the DAX, respectively 451 to 565 for Siemens, are very nice examples: the nonparametric density estimate looks normal, the test are not rejected and actually there are more than 100 observations, i. e. 5 months[6].

Comparing with the other stocks, DAX and Siemens are good examples for the detection of change-points. For the other stocks the number of change-points is similar, but there are more rejections of the kurtosis test and less negative values for the kurtosis. It is reasonable to assume, that there is still some heteroskedasticity in the empirical distributions, e. g. remember the assumed minimal distance $\delta = 40$.

6 Conclusions

Leptokurtosis, a peaked density and thick tails occur in the empirical distribution of stock returns, but this does not contradict normality; instead it may indicate heteroskedasticity, i.e. that the volatility of stock returns changes.

The present paper shows changes of volatility in the long term for german stock data from 1990 to 1992. The most striking feature of this change are some change-points. Between two change-points there is less deviation between the empirical distribution and a normal law.

We have not shown that change-points are the only source of heteroskedasticity and we have not shown that logarithmic returns have been anytime normal; there are still periods where the empirical distribution is significantly non-normal. On the other hand we must consider that our search algorithm (appendix B) cannot find the breaks exactly since it is disturbed by random noise and the minimal distance δ. Moreover between the change-points the volatility needs not to be constant, either it could be a smooth function or it could be a SV or an ARCH process.

Somewhat mysterious is the large number of zero observations. We can assume a zero variance at these days or we must pass to non absolute continuous distributions, e. g. generalized Poisson distributions.

Finally, the search algorithm (appendix B) is on a basic level, e. g. there is no rule for the choice of δ. We may search for a rule or for other methods to reduce the variance. Alternatively we may search for more sophisticated nonparametric curve estimators which allow the curve to have discontinuities.

[6]November 1991 to March 1992

DAX									
interval	n	G_2	$(p_{2\text{-}side})$	T^{sig}	$(p_{2\text{-}side})$	T^{Wil}	$(p_{2\text{-}side})$	T^{tri}	$(p_{2\text{-}side})$
all data	746	**8.19**	**(0.00)**	-0.81	(0.42)	-0.52	(0.60)	0.12	(0.91)
1 - 46	46	-0.76	(0.30)	0.30	(0.77)	0.13	(0.90)	-0.40	(0.69)
47 - 87	41	-0.55	(0.47)	-0.16	(0.88)	0.49	(0.63)	0.58	(0.56)
88 - 144	57	0.23	(0.73)	-0.93	(0.35)	-0.20	(0.84)	-0.53	(0.60)
145 - 198	54	0.71	(0.29)	**-2.18**	**(0.03)**	-1.86	(0.06)	0.58	(0.56)
199 - 249	51	-0.17	(0.81)	0.28	(0.78)	-0.37	(0.71)	-0.75	(0.45)
250 - 303	54	**3.08**	**(0.00)**	0.27	(0.79)	0.65	(0.52)	0.61	(0.54)
304 - 347	44	-0.36	(0.62)	1.98	(0.05)	1.55	(0.12)	1.49	(0.14)
348 - 404	57	**1.72**	**(0.01)**	-0.13	(0.90)	-0.35	(0.72)	0.34	(0.74)
405 - 410	6	0.71	(0.72)	0.82	(0.41)	0.73	(0.46)	0.13	(0.90)
411 - 450	40	-0.39	(0.62)	-1.90	(0.06)	-1.75	(0.08)	-1.24	(0.22)
451 - 565	115	0.46	(0.32)	1.03	(0.31)	0.96	(0.34)	1.19	(0.24)
566 - 630	65	**2.35**	**(0.00)**	-0.62	(0.54)	-0.04	(0.97)	-0.83	(0.41)
631 - 705	75	**1.33**	**(0.02)**	-1.50	(0.13)	-1.67	(0.10)	-1.17	(0.24)
706 - 746	41	-0.74	(0.33)	0.16	(0.88)	0.63	(0.53)	0.49	(0.63)

Siemens									
interval	n	G_2	$(p_{2\text{-}side})$	T^{sig}	$(p_{2\text{-}side})$	T^{Wil}	$(p_{2\text{-}side})$	T^{tri}	$(p_{2\text{-}side})$
all data	746	**6.22**	**(0.00)**	-0.70	(0.48)	-0.87	(0.39)	0.60	(0.55)
1 - 87	87	-0.55	(0.30)	0.11	(0.92)	0.35	(0.73)	0.25	(0.81)
88 - 144	57	0.22	(0.74)	-1.21	(0.23)	-0.65	(0.52)	1.87	(0.06)
145 - 197	53	0.37	(0.59)	**-2.06**	**(0.04)**	-1.82	(0.07)	1.51	(0.13)
198 - 238	41	-0.44	(0.56)	1.12	(0.26)	1.14	(0.26)	-0.19	(0.85)
239 - 359	121	**3.73**	**(0.00)**	1.19	(0.23)	0.35	(0.73)	0.13	(0.90)
360 - 404	45	0.55	(0.45)	0.45	(0.66)	0.27	(0.79)	-0.20	(0.84)
405 - 410	6	0.66	(0.74)	0.82	(0.41)	0.73	(0.46)	0.13	(0.90)
411 - 454	44	-0.49	(0.51)	**-2.29**	**(0.02)**	**-2.29**	**(0.02)**	-1.55	(0.12)
455 - 570	116	0.57	(0.21)	0.66	(0.51)	1.18	(0.24)	1.28	(0.20)
571 - 619	49	1.31	(0.06)	-0.87	(0.39)	-0.92	(0.36)	-0.94	(0.35)
620 - 661	42	0.05	(0.95)	-1.72	(0.09)	**-2.04**	**(0.04)**	-1.29	(0.20)
662 - 703	42	**1.67**	**(0.03)**	-0.62	(0.54)	-0.68	(0.50)	-0.56	(0.58)
704 - 746	43	-0.31	(0.68)	1.23	(0.22)	0.74	(0.46)	1.05	(0.30)

Table 4: tests for kurtosis, median and skewness; rejections are bold fat

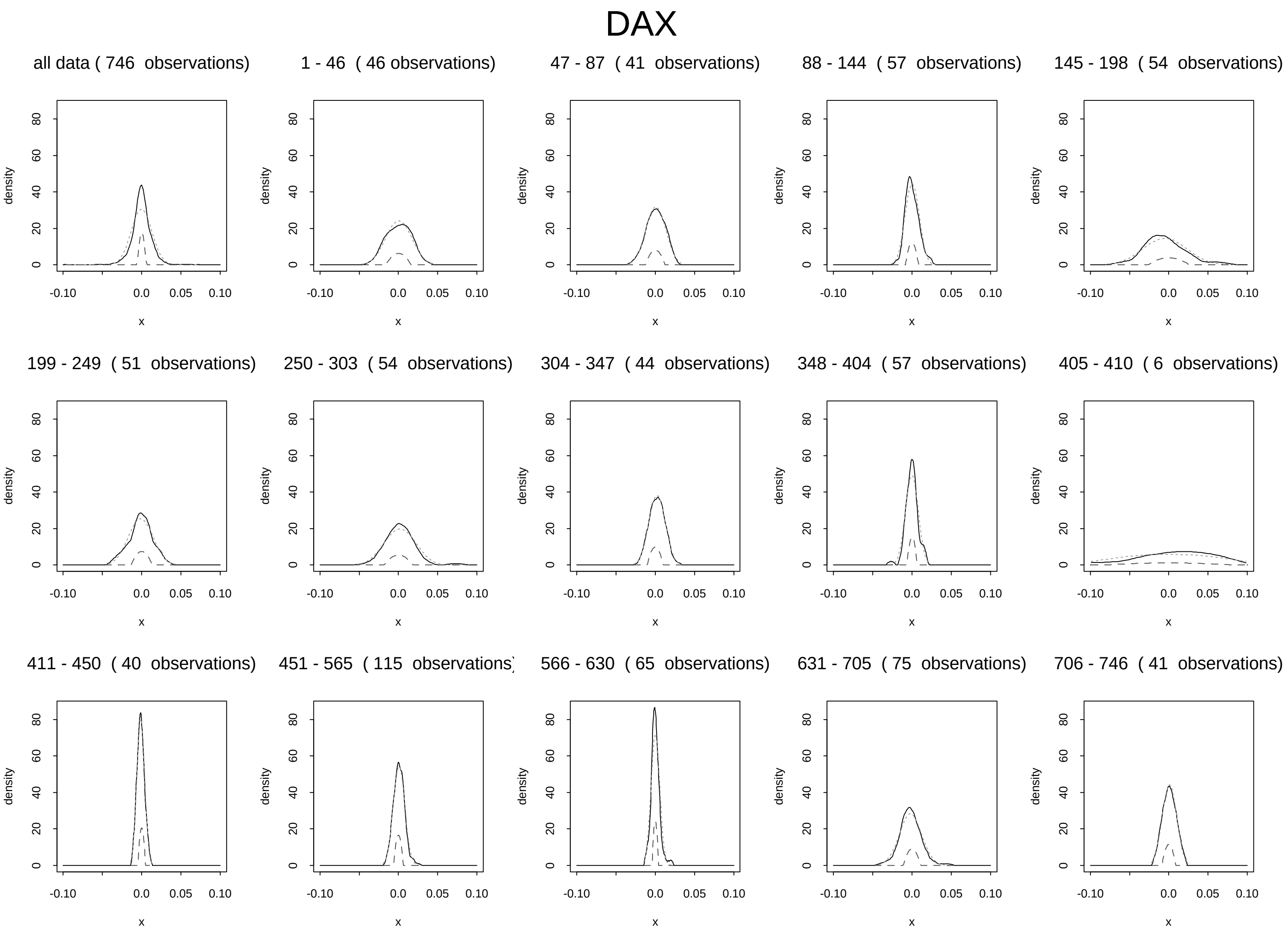

Figure 5: marginal distribution within estimated change-points

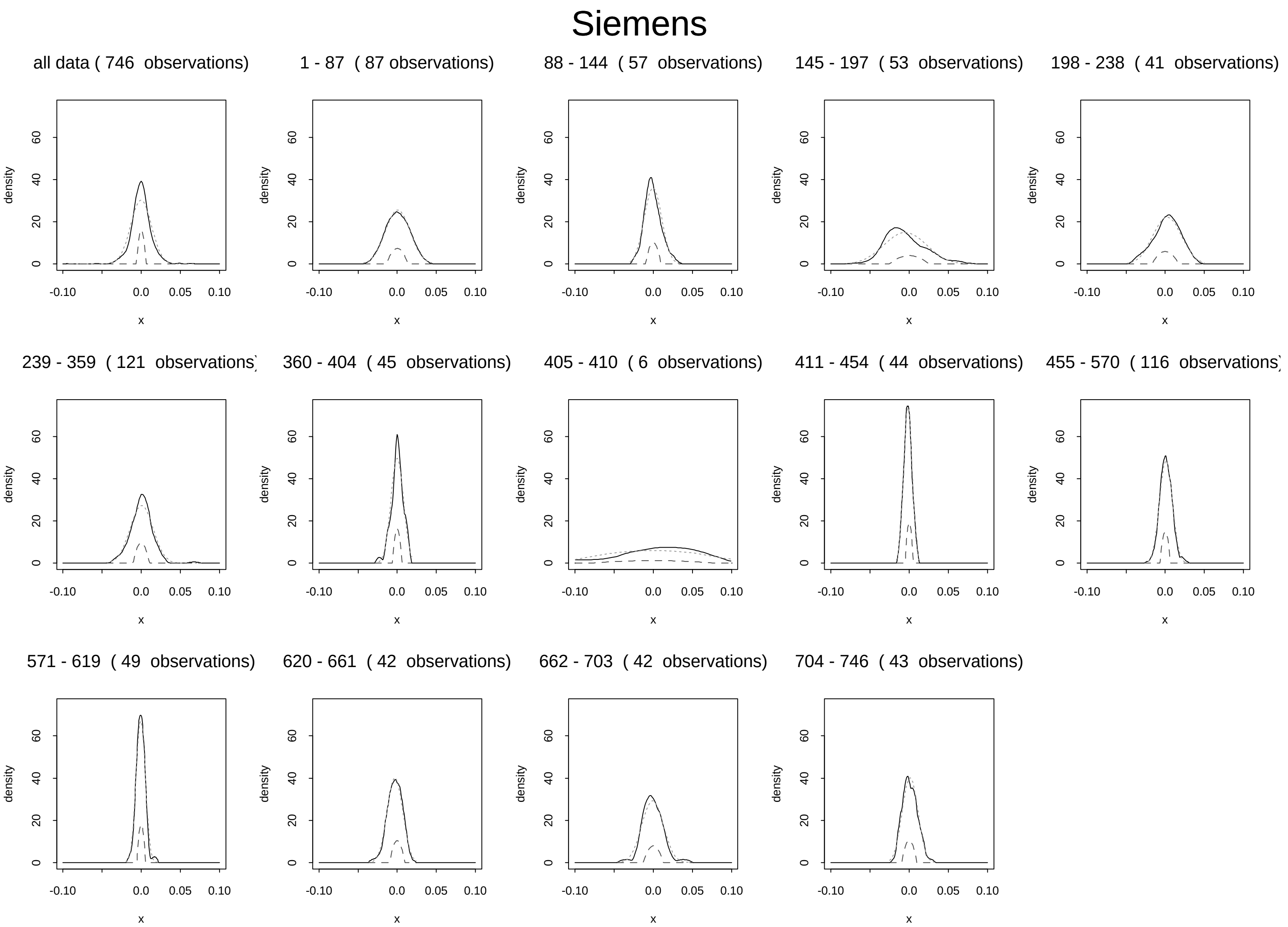

Figure 6: marginal distribution within estimated change-points

A Proof of properties A, B and C

Proof Property A follows form the principle of least squares. For property B consider that

$$\sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,k})^2 \;\leq\; \sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,n})^2$$
$$\sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{k+1,n})^2 \;\leq\; \sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2$$

which is a consequence of property A. The second part ("f is constant if and only if there is equality for every k") is shown by induction. $k=1$:

$$\sum_{i=1}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2 \;=\; 0 + \underbrace{\sum_{i=2}^{n}(f(t_i) - \overline{f(t.)}^{2,n})^2}_{\leq \sum_{i=2}^{n}(f(t_{n,i}) - \overline{f(t.)}^{1,n})^2}$$

This implies $(f(t_1) - \overline{f(t.)}^{1,n})^2 = 0$ and $f(t_1) = \overline{f(t.)}^{1,n}$. For the induction step assume $f(t_1) = \ldots = f(t_k) = \overline{f(t.)}^{1,n}$. We have

$$\sum_{i=1}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2 \;=\; \sum_{i=1}^{k+1}(f(t_i) - \overline{f(t.)}^{1,k+1})^2 + \underbrace{\sum_{i=k+2}^{n}(f(t_i) - \overline{f(t.)}^{k+2,n})^2}_{\leq \sum_{i=k+2}^{n}(f(t_i) - \overline{f(t.)}^{1,n})^2}$$

This implies $\overline{f(t.)}^{1,n} = \overline{f(t.)}^{1,k+1}$ and $f(t_{k+1}) = \overline{f(t.)}^{1,n}$. Thus it is shown that f is constant.

For property C consider that

$$\begin{aligned}
& \min_{1\leq k\leq n} \sum_{i=1}^{k}(f(t_i) - \overline{f(t.)}^{1,k})^2 + \sum_{i=k+1}^{n}(f(t_i) - \overline{f(t.)}^{k+1,n})^2 \\
= \;& \max_{1\leq k\leq n} \frac{1}{k}(\sum_{i=1}^{k} f(t_i))^2 + \frac{1}{n-k}(\sum_{i=k+1}^{n} f(t_i))^2
\end{aligned}$$

This is a function in k. We may write for $k \in [k_{j^*-1}, k_{j^*}]$, respectively $k \in (0, k_{j_1}]$ or $k \in [k_{j_{M-1}}, n)$

$$\frac{1}{k}\left(A_{j^*} + a_{j^*}k\right)^2 + \frac{1}{n-k}\left(B_{j^*} + a_{j^*}(n-k)\right)^2$$

where

$$A_{j^*} \;=\; \sum_{j=1}^{j^*-1}(k_j - k_{j-1})(a_j - a_{j^*}) \qquad B_{j^*} \;=\; \sum_{j=j^*+1}^{M}(k_j - k_{j-1})(a_j - a_{j^*})$$

This function is twice differentiable between the change-points. The second derivative is $2(\frac{A_{j*}^2}{k^3} + \frac{B_{j*}^2}{(n-k)^3})$. This is positive unless $A_{j*} = B_{j*} = 0$. If the second derivative is positive, the maximum is taken at the margin, i. e. it is one of the change-points. On the other hand $A_{j*} = B_{j*} = 0$ implies

$$a_{j*} = \frac{1}{k_{j*}} \sum_{j=1}^{j^*-1} (k_j - k_{j-1}) a_j = \frac{1}{n - k_{j*}} \sum_{j=j^*+1}^{M} (k_j - k_{j-1}) a_j$$

and for all $k \in (k_{j^*-1}, k_{j*}]$

$$\sum_{i=1}^{n} (f(t_i) - \overline{f(t.)}^{1,n})^2 = \sum_{i=1}^{k} (f(t_i) - \overline{f(t.)}^{1,k})^2 + \sum_{i=k+1}^{n} (f(t_i) - \overline{f(t...)}^{k+1,n})^2$$

thus the minimum is not taken in $(k_{j*}, k_{j*}]$ unless the function is constant. q. e. d.

B Search Algorithm

1. INPUT $i = 1, \ldots n : Y_i = \log(X_i + \varepsilon_i)^2, \quad \delta > 0$
2. $m = 1$, $k_{1,0} = 0$, $k_{1,1} = n$ and $\hat{S}_{1,1} := \sum_{i=1}^{n} (Y(t_i) - \overline{Y(t.)}^{1,n})^2 - (n-1)C_2$
3. $j = 1, \ldots, m$:
   (a) IF $k_{m,j-1} + \delta > k_{m,j} - \delta$ THEN GOTO OUTPUT
   (b) $k = k_{m,j-1} + \delta + 1, \ldots k_{m,j} - \delta$

$$\begin{aligned} \hat{S}_j(k) \quad := & \sum_{i=k_{m,j-1}+1}^{k} (Y_i - \overline{Y.}^{k_{m,j-1}+1,k})^2 \\ & + \sum_{i=k+1}^{k_{m,j}} (Y_i - \overline{Y.}^{k+1,k_{m,j}})^2 - (k_{m,j} - k_{m,j-1} - 2)\, C_2 \end{aligned}$$

4. IF for $j = 1, \ldots, m$, $k = 1, \ldots, n$: $\hat{S}_j(k) \geq \hat{S}_{m,j}$ THEN GOTO OUTPUT
5. Find j^*, k^* such that

$$\hat{S}_{j*}(k^*) = \min_{1 \leq j \leq m} \min_{1 \leq k \leq n_j} \hat{S}(k_j) - \hat{S}_{m,j}$$

6. $m = m + 1$,
   (a) $k_{m,0} = 0$, $k_{m,m} = n$
   (b) $j = 1, \ldots, j^* - 1$: $k_{m,j} = k_{m-1,j}$, $\hat{S}_{m,j} = \hat{S}_{m-1,j}$

(c)

$$
\begin{aligned}
\hat{S}_{m,j^*} &= \sum_{i=k_{m-1,j^*-1}+1}^{k^*} (Y_i - \overline{Y}_{\cdot}^{k_{m-1,j^*-1}+1,k^*})^2 - (k^* - k_{m-1,j^*-1} - 1)\, C_2 \\
\hat{S}_{m,j^*+1} &= \sum_{i=k^*+1}^{k_{m-1,j^*}} (Y_i - \overline{Y}_{\cdot}^{k^*+1,k_{m-1,j^*}})^2 - (k_{m-1,j^*} - k^* - 1)\, C_2 \\
k_{m,j^*} &= k^*
\end{aligned}
$$

(d) $j = j^* + 1, \ldots, m-1$: $k_{m,j} = k_{m-1,j-1}$, $\hat{S}_{m,j+1} = \hat{S}_{m-1,j}$

7. GOTO 3.

8. OUTPUT

References

[1] Bachelier, L. (1900,1964), "Theory of Speculation", english translation of "Théorie de la speculation" (Gauthier-Villars, Paris, 1900) in [6]

[2] Black, F. (1976) "Studies in Stock Price Volatility Changes", Proceedings on the 1976 Business and Economic Statistic Section, Am. Stat. Assoc., pp 177-181

[3] Blattberg, R. C., Gonedes, N. J. (1974) "A Comparison of the Stable and Student Distribution as Statistical Model for Stock Prices", Journal of Business 47, pp 244 - 280

[4] Bollerslev, T., Engle, R. F., Nelson, D. B. (1994) "ARCH Models" in [8], chapter 49, pp 2961-3031

[5] Clark, P. K. (1973) "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", Econometrica 41, pp 135 - 155

[6] Cootner, P. H., editor (1964) "The Random Character of Stock Market Prices", 1^{st} edition, M.I.T. Press, Cambridge, Massachusetts

[7] Engle, R. F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica 50, pp 987 - 1007

[8] Engle, R. F., McFadden, D. L. (1994) "Handbook of Econometrics, Volume IV", Elsevier Science B.V. Amsterdam

[9] Fama, E. (1965) "The Behavior of Stock Market Prices", Journal of Business 38, pp 34 - 105

[10] Fan, J., Gijbels, I. (1996) "Local Polynomial Modeling and its Applications", Monographs on Statistics and Applied Probability 66, Chapman & Hall, London

[11] Ghysels, E., Harvey, A., Renault, E. (1995) "Stochastic Volatility", CORE Discussion Paper 9569, Center for Operations Research & Econometrics, Université Catholique de Louvain; to appear in "Handbook of Econometrics, Volume V"

[12] Gnedenko, B. V., Kolmogorov, A. N. (1968) "Limit Distributions for Sums of Independent Random Variables" (2^{nd} edition 1^{st} ed.: 1954), translated from the Russian, annotated, and revised by K. L. Chung, Addison - Wesley, Reading, Massachusetts

[13] Härdle, W. (1990) "Applied Nonparametric Regression", Cambridge University Press, Cambridge

[14] Lomnicki, Z. A. (1961) "Tests for Departure from Normality in the Case of Linear Stochastic Processes", Metrika, 4, pp 37 - 62

[15] Mandelbrot, B. (1963) "Variation of certain speculative prices", Journal of Business 36, pp 394 - 419

[16] Mandelbrot, B., Taylor, H. (1967) "On the Distribution of Stock Price Differences", Operations Research 15, pp 1057 - 1062

[17] Mittnik, S., Rachev, S. T. (1993) "Modeling Asset Returns with Alternative Stable Distributions", Econometric Reviews 12, pp 261 - 389

[18] Officer, R. R. (1972) "The Distribution of Stock Returns", Journal of the American Statistical Association 67, No 340, pp 807 - 812

[19] Randles, R. W., Flinger, M. A., Policello, G. E., Wolfe, D. A. (1980) "An asymptotically Distribution-Free Test for Symmetry Versus Asymmetry", Journal of the American Statistical Association, 75 pp 168 - 172

[20] Samuelson, P. A. (1965) "Proof That Properly Anticipated Prices Fluctuate Randomly", Industrial Management Review 6, pp 41 - 50, reprinted in "The Collected Scientific Papers of Paul A. Samuelson, Vol III", R. C. Merton, ed., M.I.T Press, Cambridge, Massachusetts

[21] Samuelson, P. A. (1965) "Rational Theory of Warrant Pricing", Ibid 6, pp 13 - 32, reprinted in "The Collected Scientific Papers of Paul A. Samuelson, Vol III", R. C. Merton, ed., M.I.T Press, Cambridge, Massachusetts

[22] Silverman, B. W. (1986) "Density Estimation for Statistics and Data Analysis", Monographs on Statistics and Applied Probability 26, Chapman & Hall, London