Overlaying Time Scales and Persistence Estimation in GARCH(1,1) Models

Eric Hillebrand

Stanford University, Department of Mathematics*

CeVis – Center for Complex Systems and Visualization, Universität Bremen;

erhil@math.stanford.edu, erhil@cevis.uni-bremen.de

January 23, 2003

Abstract

A common finding in the empirical literature is that financial volatility exhibits high persistence, or slow mean reversion of the order of months. We present evidence that financial volatility data contains more than a single time scale. After showing that the expectation of the sum of the estimates of the autoregressive coefficients of a GARCH(1,1) model is one when there are unknown parameter changes, we explore the phenomenon in simulations. For parameter changes within realistic ranges for stock-price volatility we obtain global estimates close to integration while the average data-generating mean reversion is of the order of a few days. Spectral analysis of the Dow Jones Industrial Average and the S&P500 index between 1985 and 2001 reveals a short time scale of the magnitude of 5-10 days present in the data. Thus, two different time scales exist in the data, one of the order of months corresponding to different volatility regimes, and one of the order of days corresponding to the average mean reversion within regimes.

*It is a pleasure to thank George Papanicolaou for his invitation to Stanford and many discussions. I am grateful to Caio Almeida, Arnold Kim, Mordecai Kurz, Jonathan Mattingly, and Knut Sølna for many helpful comments. Any remaining errors are mine.

Contents

1 Time Scales and Persistence in Financial Volatility Data

There are at least two different ways to interpret volatility clustering. An investor with a long-term horizon will see relatively short periods of high or low volatility as persistent jumps in the fluctuation level. He perceives a long term mean to which volatility reverts only slowly after a deviation. An investor with a short-term horizon will hold a different view. If the periods of high or low volatility last longer than his investment horizon, he will see them as different states of the level of fluctuation. His idea of a mean level is short-lived and within each state volatility tends to revert fast to this level. The states are changing though, and they tend to be persistent.

The changing states can be understood as the moves of a second process with a much longer time scale than the one governing the moves within the states. That is, contrary to only a single, long-range time scale we have to deal with two overlaying time scales. For modelling volatility it is thus desirable to have a method that can capture more than one time scale of the process under examination. When only one time scale can be modelled, one has to make a choice. The global estimation of such stationary processes will have to accommodate the jumps by assuming high persistence and this will mask the

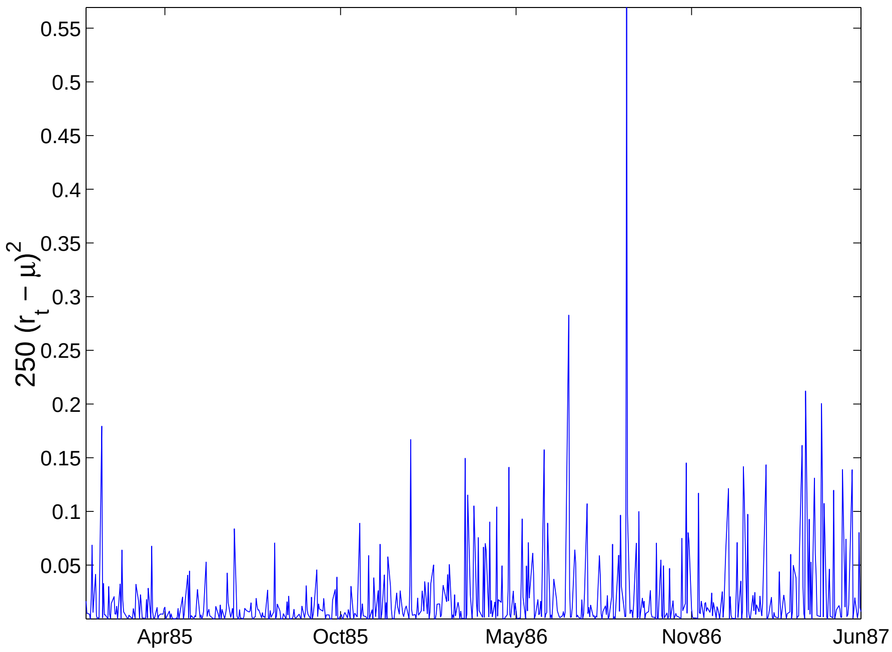

Figure 1: Annualized volatility series $250 \cdot (r_t - \mu)^2$ of the daily Dow Jones Industrial Average between January 3, 1985, and June 5, 1987. Clearly, there is a change in the volatility level in early 1986. An investor with a short investment horizon will be exposed to the small, fast mean reverting fluctuations within segments. He will perceive the change in early 1986 as a shift of the volatility mean. An investor with a long horizon of the order of months or years will not be exposed to the small fluctuations and perceive a global mean. Then, the process moved below the mean in the first segment and above it in the second segment. The investor will interpret the change in the level as a jump that persisted.

short time scale. Allowing for changes will capture the short-run dynamics of the volatility process better at the disadvantage that the long-term scale will be hidden and that the changepoints will have to be identified.

The most commonly employed time-series model for volatility estimation is the generalized model of autoregressive conditional heteroskedasticity, or GARCH. The stylized fact that volatility exhibits long memory is reflected in that the sum of the autoregressive parameters is almost unity. We will refer to this phenomenon as "almost-integration".

We will show that the duality of persistence of jumps and persistence of states translates fully into the GARCH estimation. Analytically and numerically we will demonstrate that parameter changes that are not accounted for in global GARCH estimations lead to high estimated persistence close to integration. This is regardless of the data-generating persistence. We find that a single changepoint between realistic values for stock-market volatility can be sufficient for this effect to occur.

The notion of two overlaying time scales seems to explain these findings reasonably well, so we turn to methods of spectral analysis that allow to detect time scales independently of the model formulation. We clear the volatility time series from the long time scale that was detected by GARCH and estimate the power spectrum of this properly defined residual. This method reveals a short time scale of the magnitude of 5 to 10 days present in the Dow Jones and in the S&P500.

2 Persistence Estimation with GARCH Models

2.1 The Model Formulation

Engle (1982) and Bollerslev (1986) suggested the following approach. The return r_t from a stock with price S_t at time t is modeled as

$$r_t := \log(S_{t+1}) - \log(S_t) = \mathbb{E}(r_t|\mathcal{F}_{t-1}) + \varepsilon_t = \mu(b) + \varepsilon_t. \tag{1}$$

$\mathcal{F}_t$ denotes the filtration modeling the information set. μ is the conditional mean function with argument b, for example a regression $\mu(b) = X_t^T b$, where X_t denotes a set of independent variables. We assume the disturbance ε_t to be normally distributed, conditional on the information available at time $t-1$:

$$\varepsilon_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t), \tag{2}$$

i.e. $\varepsilon_t = \eta_t\sqrt{h_t}$, $\eta_t \sim \mathcal{N}(0,1)$, where h_t denotes the conditional variance. The latter is described by the difference equation

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{3}$$

with $\omega, \alpha_i, \beta_i \geq 0 \,\forall\, i$. This is the GARCH(p,q) model for the conditional variance. To obtain the unconditional expected variance, assume that the process $\{\varepsilon_t\}$ is covariance-stationary. Then $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ holds and

$$\mathbb{E}h = \frac{w}{1 - \sum \alpha_i - \sum \beta_i}. \tag{4}$$

For the sake of simplicity, we will restrict the arguments to the GARCH(1,1) specification

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{5}$$

with $\varepsilon_t = r_t - \mu$, $\mu \in \mathbb{R}$ fixed and $\varepsilon_t|_{\mathcal{F}_{t-1}} \sim \mathcal{N}(0, h_t)$. [1]

[1] We will not report estimates of the constant mean return for the sake of brevity.

2.2 Measures of Persistence and Mean Reversion

Consider the conditional variance at time $t+k$, $k \geq 2$, and take expectations conditional on $\mathcal{F}_t$:

$$\begin{aligned} \mathbb{E}_t h_{t+k} &= \omega + \alpha \mathbb{E}_t \varepsilon_{t+k-1}^2 + \beta \mathbb{E}_t h_{t+k-1} \\ &= \omega + (\alpha+\beta)\mathbb{E}_t h_{t+k-1}. \end{aligned} \tag{6}$$

Thus, the k-period forecast of the conditional variance according to the GARCH-(1,1)-model is a first-order linear difference equation with autoregressive parameter

$$\lambda := \alpha + \beta. \tag{7}$$

The closer λ is to unity, the more persistent the effect of a change in $\mathbb{E}_t(h_{t+k})$ will be. The parameter λ is the fraction of the forecast that is carried forward per unit of time, so $(1-\lambda)$ is the fraction that is washed out per unit of time. Hence $1/(1-\lambda)$ is the average time needed to return to the mean when the time increment equals one. To formalize this, denote x_t as the distance of $\mathbb{E}_t h_{t+\Delta t}$ from its mean $\mathbb{E}h_t$, that is

$$\mathbb{E}_t h_{t+\Delta t} = \mathbb{E}h_t + x_t.$$

Then, from (6) we have for the case of $\Delta t = 1$ that

$$\mathbb{E}h_t + x_t = \omega + \lambda(\mathbb{E}h_t + x_{t-1}).$$

As $\mathbb{E}h_t = \omega + \lambda \mathbb{E}h_t$, we obtain

$$x_t = \lambda x_{t-1}$$

or

$$x_{t+1} - x_t = (\lambda - 1)x_t \tag{8}$$

for the distance of the forecast from the unconditional mean. We may model x by a decreasing function y defined by the differential equation

$$\lim_{\Delta t \to 0}(y_{t+\Delta t} - y_t) = \lim_{\Delta t \to 0}(-\kappa y_t \Delta t). \tag{9}$$

This differential equation is solved by

$$\frac{y_t}{y_0} = e^{-\kappa t}, \quad \kappa > 0,$$

and the so-called e-folding time

$$\left(t_e \,\middle|\, \frac{y_t}{y_0} = e^{-1} \right)$$

is given by $t_e = 1/\kappa$. Comparing the coefficients in (8) and (9), we see that

$$\kappa = 1 - \lambda$$

and the e-folding time of the distance x of the forecast from the unconditional mean is

$$t_e = \frac{1}{\kappa} = \frac{1}{1-\lambda}. \tag{10}$$

There are other ways to define and measure persistence in a discrete GARCH framework (discussed e.g. in Engle and Patton 2001). Nelson (1990a) uses a variety of persistence definitions and shows that whether or not shocks are persistent depends crucially on the definition chosen.

2.3 Maximum Likelihood Estimation

The most common way to estimate a Gaussian GARCH(1,1) model with constant mean return given a sequence $\{S_t\}_{t\in\mathbb{N}}$ of prices is by maximum likelihood derived from equation (2). Let $\varepsilon_t(\mu) = r_t - \mu$, $\mu \in \mathbb{R}$ fix. Denote the parameter vector by $\theta = (\mu, \omega, \alpha, \beta)$. The log-likelihood function is given by

$$L_N(\theta, \{\varepsilon_t\}_t) := -\frac{1}{2N}\left[N\log(2\pi) + \sum_{t=1}^{N}\left(\log h_t(\theta, \varepsilon_{t-1}) + \frac{\varepsilon_t^2(\mu)}{h_t(\theta, \varepsilon_{t-1})}\right)\right]. \tag{11}$$

The GARCH model is not restricted to the conditionally normal case. Bollerslev (1987) suggests using the t-distribution and treating the number of degrees of freedom as additional parameter.

Consistency and asymptotic normality of the maximum likelihood estimator could only be proven in the conditionally Gaussian GARCH(1,1) case so far. The main results can be found in the papers by Weiss (1986), Bollerslev and Wooldridge (1992), and Lumsdaine (1996). There are no closed analytical expressions for the estimators.

In practice, the likelihood is maximized by numerical optimization methods. Most software packages implement a quasi-Newton method using linesearch and Hessian update algorithms. There are alternatives to this approach, like generalized least squares estimators (Gouriéroux 1997) or scoring methods (Harvey 1976, Greene 2000). We maximize (11) using code written in MATLAB and C++. The MATLAB code uses the ‘fmincon’ routine from the optimization toolbox which implements a quasi-Newton method. The C++ code uses the ‘dfpmin’ routine from the “Numerical Recipes” (Press et al. 2002) which also implements a quasi-Newton method. The gradients are computed using analytical expressions, the Hessians are approximated by finite differencing.

2.4 GARCH(1,1) and Market Data: Long Memory in the Volatility of the Dow Jones and S&P500

We use daily closings of the Dow Jones Industrial Average and the S&P500 ranging from January 2nd, 1985, to January 2nd, 2001. The overall length of the series is 4,031 observations for the Dow Jones and 4,038 for the S&P500.[2] The Dow Jones data was kindly provided by Dow Jones & Company, the S&P500 was downloaded from Datastream. When we globally estimate a Gaussian GARCH(1,1) model for the annualized daily returns of the 16 years that our series cover, we obtain

$$h_t = \underset{(0.00018)}{0.00049} + \underset{(0.0464)}{0.0872}\,\varepsilon_{t-1}^2 + \underset{(0.0393)}{0.8991}\,h_{t-1},$$

for the Dow Jones series. This implies a $\hat{\lambda}$ of 0.9863 ($1/(1-\hat{\lambda}) = 73$ days). For the S&P500 series, we get

$$h_t = \underset{(0.00012)}{0.00037} + \underset{(0.0374)}{0.0888}\,\varepsilon_{t-1}^2 + \underset{(0.0316)}{0.9024}\,h_{t-1},$$

which implies a $\hat{\lambda}$ of 0.9912 ($1/(1-\hat{\lambda}) = 114$ days). The numbers in brackets are heteroskedasticity-robust standard errors according to Bollerslev and Wooldridge (1992). We observe roots close to the unit circle for both indices. These estimations pick up a time scale of the order of months.

We examined model selection criteria for GARCH(p,q) models of the class $p \in \{1,2,3\},\ q \in \{1,2,3\}$. The Akaike and Schwarz information criteria favored GARCH(2,·) and GARCH(3,·) if any, but the margins were of the magnitude of half a per cent or less. For higher order models there was mostly only one β_i significant and it was β_1 in most of the cases. The exceptions were GARCH(3,3) for the Dow Jones (all three β_i's significant) and GARCH(1,3) and GARCH(3,2) for the S&P500 (β_1, β_3 and β_1, β_2 significant). Higher order GARCH models are able to capture more than a single time scale. This might be the reason why we observe the slight advantage of higher order models according to the Akaike and Schwarz criteria. The margins are not very conclusive, though.

2.5 High Persistence as a Stylized Fact

Global estimations of GARCH models usually indicate high persistence or slow mean reversion. For the GARCH(1,1) model, many studies report $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ close to unity, the majority of which base this observation on global estimations of long-range data sets.[3] The conclusion that λ is indeed equal to one and that the constraint $\alpha + \beta < 1$ is active suggests itself. This gave rise to the formulations

[2] We deleted all holidays with zero returns.

[3] Engle/Bollerslev (1986): weekly returns on exchange rates over 12 years; Baillie/DeGennaro (1990): daily returns on stock index over 18 years; Bollerslev/Engle (1993): daily returns on exchange rates over 5 years; Baillie et al. (1996): daily returns on exchange rates over 13 years;

of Integrated GARCH (IGARCH, Engle and Bollerslev 1986) and Fractionally Integrated GARCH (FIGARCH, Baillie et al. 1996), which assume an indefinite memory.

The concern that the apparently high persistence in the observed data may be caused by structural changes was raised early. In a comment to the original IGARCH paper by Engle and Bollerslev (1986), Francis Diebold mentioned with regard to interest rate data that not accommodating shifts in monetary policy regimes, reflected in changes of the constant term ω in (5), might lead to an apparently integrated series of squared disturbances (Diebold 1986). Lamoureux and Lastrapes (1990) showed Diebold's conjecture to be right for stock data, obtaining their results by including dummy variables that indicate different states of the GARCH(1,1) constant ω, equidistant in time. Earlier work of Granger (1980) showed that aggregation over processes with different autoregressive parameters induced long memory properties. Diebold and Inoue (2001) show that stochastic permanent break models and Markov-switching models display behavior consistent with long memory. Granger and Teräsvirta (2001) show that a simple nonlinear model that displays regime switching behavior also leads to long memory properties. Hamilton and Susmel (1994) used the regime switching model to improve volatility forecasts of ARCH models by incorporating state changes. Gray (1996) extended the regime switching approach to GARCH. These locally stationary approaches that segment the data obtained significantly lower estimates of the order of days.

Using high-frequency data of 5 minute returns estimated at different frequencies, Andersen and Bollerslev (1997) give a concise overview of the irregular picture. They obtain estimates ranging from two hours to 7 days for the half-life of the S&P500 1986–1989. Authors who are not using GARCH get still different results: Fouque et al. (2002) obtain about 1.5 days for the average mean reversion length from high-frequency S&P500 data 1994–1998 by spectral methods. Drăgulescu and Yakovenko (2002) obtain 22 days for daily returns of the Dow Jones 1982–2001 by estimating an explicit probability density.

The presence of multiple time scales is proposed recently in the context of stochastic volatility models. Fouque et al. (2002) suggest a multi-scale stochastic volatility model. They show that the estimation of such a model on the short time scale is not affected by the long-run dynamics. This corresponds to the observation that local GARCH estimations of properly segmented data do not reveal the long-term high persistence. On the other hand, LeBaron (2001) shows in a similar model with three time scales that the three factors induce long memory properties. This corresponds to the occurrence of an almost unit root in global

Ding/Granger (1996): daily returns on stock index over 63 years; Andersen/Bollerslev (1997): 5 minute returns on exchange rates over *one* year and on stock index future prices over 4 years; Engle/Patton (2001): daily returns on stock index over 12 years.

GARCH estimations. Chernov et al. (2002) also discuss multi-driver stochastic volatility models. Gallant and Tauchen (2001) estimate a two-scale volatility model and find a long and a short correlation structure in daily returns on the Microsoft stock.

The two concepts, overlaying long range processes and parameter switches, differ only in the continuity of their influence. A jump that occurs once every n units of time adds the time-scale of n to the process. Francq et al. (2001) examine GARCH processes which are subject to Markov-switching parameters. They show in simulations that as a result of the stochastic nature of the Markov-switching process the ARCH parameters will be estimated in the neighborhood of integration. Mikosch and Starica (2000) show that the Whittle estimate of ARMA(1,1) parameters will imply almost-integration when there are changepoints.

3 Parameter Changes and Global GARCH(1,1) Estimations

3.1 The Cause of Almost-Integration

We will show why global estimations of GARCH(1,1) models that do not account for a single changepoint in the constant ω will result in almost-integration. Hence, it is not necessary that the long scale process has a specific stochastic structure. This is regardless of the estimation method.

Lemma 1. *Denote by $\mathbb{E}_0 h_t$ the expected value of a stationary Gaussian GARCH(1,1) model conditional on the start value $h_0 \in \mathbb{R}$. Then, the relation*

$$\mathbb{E}_0 h_t = \mathbb{E}h + O(\lambda^t), \tag{12}$$

holds for $t \in \{1, \dots, N\}$.

Proof. The expected value conditional on the start value is given by

$$\mathbb{E}_0 h_t = \omega + \mathbb{E}_0(\alpha\eta_{t-1}^2 + \beta)\mathbb{E}_0 h_{t-1} = \omega + \lambda\mathbb{E}_0 h_{t-1} = \omega\frac{1-\lambda^t}{1-\lambda} + \lambda^t h_0,$$

as $\mathbb{E}_0\eta_t^2 = 1$ for all t and the η_t are independent. Thus, substituting from equation (4)

$$|\mathbb{E}_0 h_t - \mathbb{E}h| = \left|\omega\frac{1-\lambda^t}{1-\lambda} + \lambda^t h_0 - \frac{\omega}{1-\lambda}\right| = \lambda^t\left|h_0 - \frac{\omega}{1-\lambda}\right| = O(\lambda^t). \qquad \square$$

Assumption 2. *We will assume that the processes $\{h_t\}$ and $\{\varepsilon_t\}$ are observable without measurement error, or at least with a measurement error that is independent of the parameter estimates $(\hat{\mu}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$.*

This assumption is, of course, unrealistic. The process h_t is not observable and in real estimation problems h_t is estimated by $\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ and ε_t by $\hat{\varepsilon}_t(\hat{\mu})$. The conjecture is, however, that if we can show that almost-integration would occur if h_t were observable, it will also occur when we have less information. For the sake of notational brevity, we will assume that the measurement is error-free. The case of an error that is independent of the parameter estimates would add a correction term that vanishes with growing sample size, without changing the argument.

Now, let $\{h_t\}$ be generated by

$$h_t = \begin{cases} \omega_1 + (\alpha\eta_{t-1}^2 + \beta)h_{t-1}, & t \in \{1, \ldots, N_1\}, \\ \omega_2 + (\alpha\eta_{t-1}^2 + \beta)h_{t-1}, & t \in \{N_1 + 1, \ldots, N\}. \end{cases} \tag{13}$$

This fact is unknown to the econometrician. The estimated model equation is

$$h_t = \hat{\omega} + \hat{\lambda}^* h_{t-1}, \tag{14}$$

where $\hat{\lambda}^* = \hat{\alpha}\eta_{t-1}^2 + \hat{\beta}$. One might argue that having exact measurement of h_t, it would be an obvious approach to just back out the parameters, thereby finding that there was a jump in ω. However, as our ultimate interest is the case of h_t being unobservable, where this cannot be done, we will nevertheless proceed with the estimation of (14).

Subtract the mean from (14):

$$h_t - \bar{h} = \hat{\alpha}(\varepsilon_{t-1}^2 - \overline{\varepsilon^2}) + \hat{\beta}(h_{t-1} - \bar{h}) \tag{15}$$

If the segmentation were known, the econometrician would insert a term for the difference in ω.

Let $\mathbb{E}_{(i)}h_t$ denote the expected values with respect to the start value in segment i, where i is 1 for $t \in \{1, \ldots, N_1\}$ and i is 2 for $t \in \{N_1 + 1, \ldots, N\}$. In other words,

$$\mathbb{E}_{(1)}h_t = \mathbb{E}(h_t|\mathcal{F}_0)$$
$$\mathbb{E}_{(2)}h_t = \mathbb{E}(h_t|\mathcal{F}_{N_1}).$$

Lemma 3. *Let $\mathbb{E}h^{(i)}$ denote the expected value of a process h_t generated by $\theta = (\mu, \omega_i, \alpha, \beta)$ with respect to the stationary measure. Let h_t be generated according to* (13)*. Then,*

$$\bar{h} = \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1},$$

$$\overline{\varepsilon^2} = \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N - N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1},$$

where $o(1)_{N_1} \to 0$ as $N_1 \to \infty$ and $o(1)_{N-N_1} \to 0$ as $N - N_1 \to \infty$.

Proof. Write $h_t = \mathbb{E}_{(i)} h_t + x_t$, x_t being the deviation from the expectation conditional on the start values within segments such that

$$\frac{1}{N_1}\sum_{t=1}^{N_1} x_t = o(1)_{N_1}$$
$$\frac{1}{N-N_1}\sum_{t=N_1+1}^{N} x_t = o(1)_{N-N_1}.$$

From this and Lemma 1 we obtain

$$\begin{aligned}
\bar{h} &= \frac{1}{N}\sum_{t=1}^{N} h_t,\\
&= \frac{1}{N}\sum_{t=1}^{N_1}\mathbb{E}_{(1)}h_t + \frac{1}{N}\sum_{t=N_1+1}^{N}\mathbb{E}_{(2)}h_t + \frac{1}{N_1}\sum_{t=1}^{N_1}x_t + \frac{1}{N-N_1}\sum_{t=N_1+1}^{N}x_t,\\
&= \frac{1}{N}\sum_{t=1}^{N_1}\mathbb{E}h^{(1)} + \frac{1}{N}\sum_{t=N_1+1}^{N}\mathbb{E}h^{(2)} + o(1)_{N_1} + o(1)_{N-N_1}\\
&\quad + \frac{1}{N}\sum_{t=1}^{N_1}O(\lambda^t) + \frac{1}{N}\sum_{t=N_1+1}^{N}O(\lambda^{t-N_1})\\
&= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}.
\end{aligned}$$

In the same manner, write $\varepsilon_t^2 = \mathbb{E}_{(i)}\varepsilon_t^2 + y_t = \mathbb{E}_{(i)}h_t + y_t$ by the distribution assumption (2). Then,

$$\begin{aligned}
\overline{\varepsilon^2} &= \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2,\\
&= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}. \qquad \square
\end{aligned}$$

Proposition 4. *If there is an unknown switch in the data-generating constant of the conditional variance equation of a Gaussian GARCH(1,1) model, as specified in equation* (13)*, and the model is estimated on the entire series, then, under Assumption 2 the condition*

$$\mathbb{E}_{(i)}\hat{\lambda} = \mathbb{E}_{(i)}(\hat{\alpha} + \hat{\beta}) = 1$$

must hold in both segments i, up to terms that vanish with growing length of the segments.

Proof. Take expectations of (15) conditional on the start value within segments:

$$\mathbb{E}_{(i)}h_t - \mathbb{E}_{(i)}\bar{h} = \mathbb{E}_{(i)}\hat{\alpha}(\mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}\overline{\varepsilon^2}) + \mathbb{E}_{(i)}\hat{\beta}(\mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}\bar{h}). \qquad (16)$$

Here, we use Assumption 2 and equation (2):

$$\begin{aligned}
\mathbb{E}_{(i)}(\hat{\alpha}\varepsilon_{t-1}^2) &= \mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}_{(i)}\varepsilon_{t-1}^2 = \mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}_{(i)}h_{t-1},\\
\mathbb{E}_{(i)}(\hat{\alpha}\overline{\varepsilon^2}) &= \mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}_{(i)}\overline{\varepsilon^2},\\
\mathbb{E}_{(i)}(\hat{\beta}h_{t-1}) &= \mathbb{E}_{(i)}\hat{\beta}\mathbb{E}_{(i)}h_{t-1},\\
\mathbb{E}_{(i)}(\hat{\beta}\bar{h}) &= \mathbb{E}_{(i)}\hat{\beta}\mathbb{E}_{(i)}\bar{h}.
\end{aligned}$$

From Lemma 3 we have that

$$\begin{aligned}
\mathbb{E}_{(i)}\bar{h} &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1},\\
\mathbb{E}_{(i)}\overline{\varepsilon^2} &= \frac{N_1}{N}\mathbb{E}h^{(1)} + \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}
\end{aligned}$$

where only the $O(\cdot)$ terms may have changed from (4). From this and from Lemma 1

$$\begin{aligned}
\mathbb{E}_{(1)}h_t - \mathbb{E}_{(1)}\bar{h} &= \mathbb{E}h^{(1)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1}\\
&\quad + o(1)_{N-N_1} + O(\lambda^t),\\
\mathbb{E}_{(2)}h_t - \mathbb{E}_{(2)}\bar{h} &= \mathbb{E}h^{(2)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1}\\
&\quad + o(1)_{N-N_1} + O(\lambda^{t-N_1})\\
\mathbb{E}_{(1)}h_t - \mathbb{E}_{(1)}\overline{\varepsilon^2} &= \mathbb{E}h^{(1)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1}\\
&\quad + o(1)_{N-N_1} + O(\lambda^t),\\
\mathbb{E}_{(2)}h_t - \mathbb{E}_{(2)}\overline{\varepsilon^2} &= \mathbb{E}h^{(2)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} + O(1/N) + o(1)_{N_1}\\
&\quad + o(1)_{N-N_1} + O(\lambda^{t-N_1}).
\end{aligned}$$

Plugging this into (16) and arranging terms, we obtain

$$\begin{aligned}
&\mathbb{E}h^{(1)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} =\\
&\quad \mathbb{E}_{(1)}(\hat{\alpha}+\hat{\beta})\left(\mathbb{E}h^{(1)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)}\right)\\
&\quad + O(\lambda^{t-1}) + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}\\
&\mathbb{E}h^{(2)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)} =\\
&\quad \mathbb{E}_{(2)}(\hat{\alpha}+\hat{\beta})\left(\mathbb{E}h^{(2)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N-N_1}{N}\mathbb{E}h^{(2)}\right)\\
&\quad + O(\lambda^{t-1-N_1}) + O(1/N) + o(1)_{N_1} + o(1)_{N-N_1}. \qquad (17)
\end{aligned}$$

That is, with growing N_1 and growing $N - N_1$, the expected value of the sum of the estimators of the autoregressive parameters conditional on the start value in each segment i must fulfil

$$\mathbb{E}_{(i)}(\hat{\alpha} + \hat{\beta}) = 1$$

up to vanishing terms, in order to satisfy condition (17). Observe that the difference

$$\mathbb{E}h^{(i)} - \frac{N_1}{N}\mathbb{E}h^{(1)} - \frac{N - N_1}{N}\mathbb{E}h^{(2)} \neq 0$$

in both segments i if $\omega_1 \neq \omega_2$. Thus, condition (17) is not trivial. □

The fact that condition (17) is not trivial is the main difference to the stationary GARCH analysis without parameter switches. In that case, the condition reads zero equals zero with respect to the stationary measure.

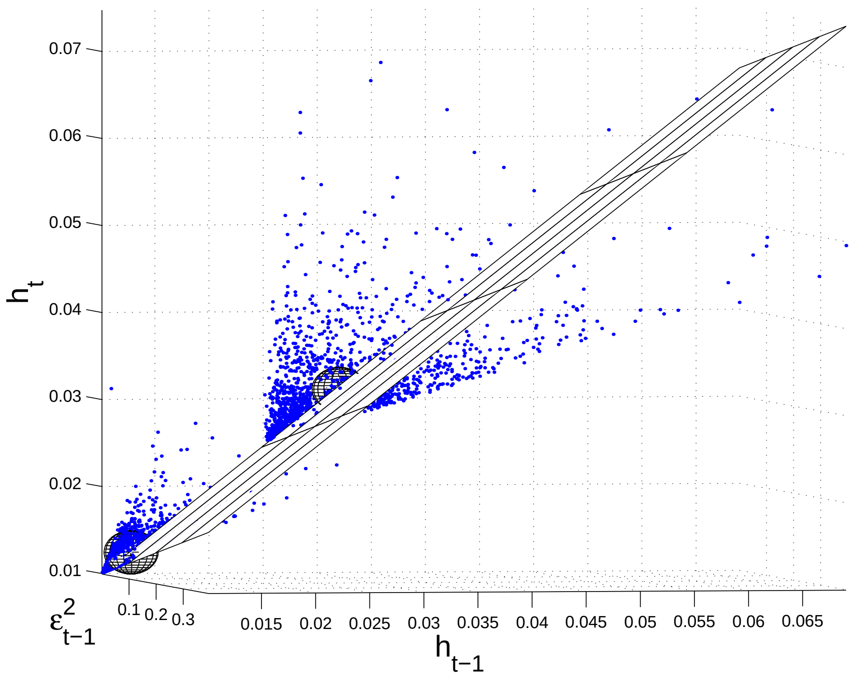

Figure 2: Plot of the least squares problem (14) for a synthetic data series. The $\{\varepsilon_t\}$ and $\{h_t\}$ were generated by the parameters ω_1 =2e-5 and ω_2 =5e-5, $\alpha = 0.10$ and $\beta = 0.50$. The length of the entire series was $N = 4200$ and the changepoint N_1 was set at one half of N. The spheres are centered at the unconditional, stationary expected values $\mathbb{E}h^{(1)} = 250 * 2e - 5/(1 - 0.1 - 0.5) = 0.0125$ and $\mathbb{E}h^{(2)} = 250 * 5e - 5/(1 - 0.1 - 0.5) = 0.03125$. The fact that a single hyperplane is fitted through both segments, reflected in the two clusters, leads to almost-integration. The slope of the clusters with respect to the (h_t, h_{t-1})-subspace, which is $\beta = 0.5$, is largely overestimated. The slope of the clusters with respect to the $(h_t, \varepsilon^2_{t-1})$-subspace, which is $\alpha = 0.1$, is underestimated. The estimated parameters are $\hat{\omega} = 2.6$e-5, $\hat{\alpha} = 0.018$, and $\hat{\beta} = 0.981$.

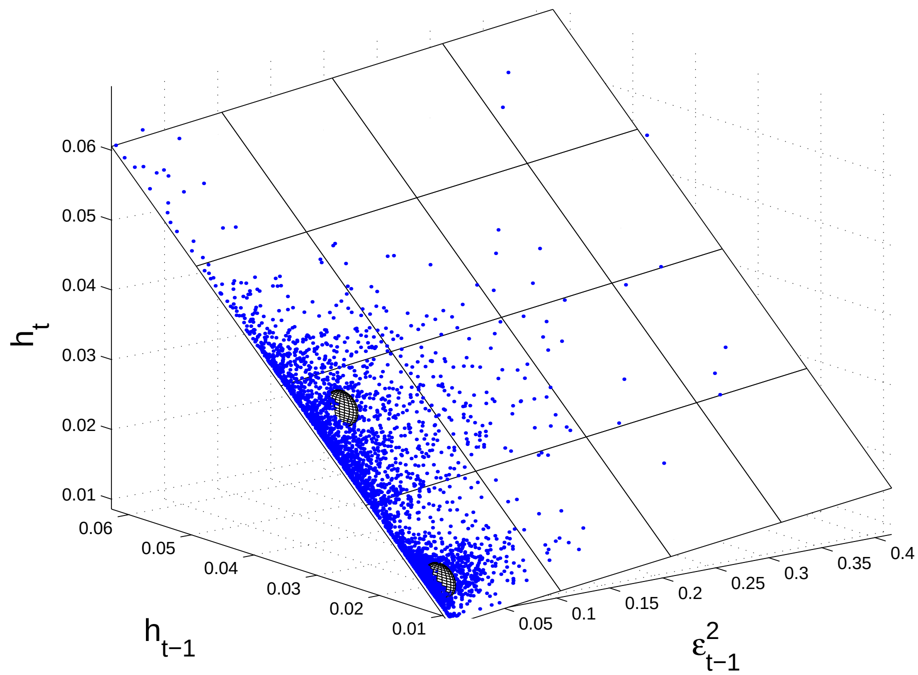

Figure 3: Plot of the least squares problem (14) for the same synthetic data series considered in Figure 2. Here, the estimated series $\{\hat{\varepsilon}_t(\hat{\mu})\}$ and $\{\hat{h}_t(\hat{\theta})\}$ are plotted. All the points are lying on the hyperplane according to the estimates $\hat{\omega} = 2.6\text{e-}5$, $\hat{\alpha} = 0.018$, and $\hat{\beta} = 0.981$. However, the two-cluster structure is still visible. The viewpoint is chosen differently from Figure 2.

Equation (14) can be interpreted as a least squares problem when the $\{\varepsilon_t\}$ and $\{h_t\}$ are known. This allows a geometric interpretation. Figure 2 shows the problem for a synthetically generated series with a single changepoint in ω. The two different data-generating parameters induce two distinct expected values $\mathbb{E}h^{(1)}$ and $\mathbb{E}h^{(2)}$. The spheres in Figure 2 are centered at these expected values. The $\{\varepsilon_t\}$ and $\{h_t\}$ values of each segment are centered at these expectations, as shown in Lemma 1, so that two distinct clusters can be observed. The effect of almost-integration occurs as the data are not properly segmented. If they were, a best fit hyperplane would be estimated within each cluster. This would capture the slopes of the clusters. As a single hyperplane is fitted to the entire series, the relative position of the clusters dominates the estimation. The single hyperplane has to go through $\mathbb{E}h^{(1)}$ and $\mathbb{E}h^{(2)}$. This is the geometric interpretation of the fact that condition (17) is not trivial. Here, $\bar{h}$ is a weighted average of the two expectations, so that $\mathbb{E}h^{(i)} - \bar{h}$ is not zero, contrary to the case where there is only a single cluster.

To support the conjecture proposed in the context of Assumption 2, Figure 3 depicts the plot of the $\{\hat{\varepsilon}_t(\hat{\mu})\}$ and $\{\hat{h}_t(\hat{\theta})\}$ together with the estimation hyperplane. Of course, all the points lie on the estimation hyperplane. For this reason, the viewpoint is chosen differently. The two clusters that reflect the data-generating structure are still distinguishable.

3.2 Simulations

We expect changes in the mean reversion parameters α and β to have a similar effect. To explore this, we consider global GARCH(1,1) estimates of a synthetic series constructed in three segments of length 1400 in four mean-reversion scenarios:

Table 1: GARCH(1,1) segment parameters of artificial series.

segment	1	2	3	average
length	1400	1400	1400	
Scenario 1.				
ω	1e-5	1e-5	2.5e-5	1.5e-5
α	0.10	0.10	0.10	0.10
β	0.75	0.75	0.75	0.75
Scenario 2.				
ω	1e-5	1e-5	1e-5	1e-5
α	0.10	0.10	0.10	0.10
β	0.65	0.80	0.65	0.70
Scenario 3.				
ω	1e-5	1e-5	1e-5	1e-5
α	0.65	0.80	0.65	0.70
β	0.10	0.10	0.10	0.10
Scenario 4.				
ω	1e-5	2e-5	1e-5	1.3e-3
α	0.10	0.05	0.10	0.083
β	0.65	0.80	0.65	0.70

1. ω switches at the point 2800 from 1e-5 to 2.5e-5 while $\alpha = 0.10$ and $\beta = 0.75$ are constant. In terms of annualized standard deviations this is a jump from 13 to 20 per cent volatility.

2. β switches from 0.65 to 0.80 and back, $\alpha \leq \beta$ holds. $\omega =$ 1e-5 and $\alpha = 0.10$ are constant. This corresponds to changes between 10 and 16 per cent volatility.

3. α switches from 0.65 to 0.80 and back, $\alpha \geq \beta$ holds. $\omega =$ 1e-5 and $\beta = 0.10$ are constant.

4. All parameters change: ω from 1e-5 to 2e-5 and back, α from 0.10 to 0.05 and back, and β from 0.65 to 0.80 and back. The annualized volatility switches between 10 and 18 per cent.

Table 1 shows the specification of the parameters in the four scenarios over the three segments.

We generated 10,000 series for each scenario. On each series we estimated a global Gaussian GARCH(1,1) model with constant mean return. Figure 4 shows the histograms of the estimations of ω, α, β, and λ for each scenario. Table 2 shows the moments statistics.

Table 2: Moments statistics of the estimates of ω, α, β, and λ from the GARCH(1,1) estimation of 10,000 artificial series for every scenario according to Table 1.

	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
	Scenario 1.				**Scenario 2.**			
mean	3e-6	0.0794	0.8926	0.9720	2e-6	0.0765	0.8925	0.9690
std.dev.	1e-6	0.0208	0.0325	0.0123	9e-7	0.0223	0.0366	0.0149
skewness	0.5005	-0.2548	0.0299	-0.5020	0.4957	-0.1554	-0.0559	-0.5195
kurtosis	3.4224	3.1788	3.0779	3.4609	3.2266	2.8505	2.8331	3.3147
	Scenario 3.				**Scenario 4.**			
mean	1e-5	0.6844	0.1022	0.7867	4e-7	0.0402	0.9545	0.9947
std.dev.	6e-7	0.0510	0.0204	0.0488	2e-7	0.0121	0.0149	0.0029
skewness	0.0357	0.9084	0.1094	0.9214	1.5536	0.9214	-1.0281	-1.5523
kurtosis	2.9608	4.5097	3.0380	4.5514	6.3984	3.8801	4.2242	6.4092

The global estimations differ widely from the average of the parameters in every scenario. Most pronounced is the effect in Scenarios 1, 2, and 4. Here, λ is estimated close to one, regardless of the different data generating parameters. In Scenario 3 where α and β take the values of Scenario 2 in reverse order, the effect is not observed. In Scenarios 1, 2, and 4, the global estimate of β is close to or above 0.90 and that of α is less than 0.10.

In Scenarios 1, 2, and 4 the transitions between parameter states cause the estimator $\hat{\beta}$ to take a large value in order to accommodate the large and apparently persistent "jumps". The influence of the $\{\varepsilon_t^2\}$ series on the process appears to be diminished by choosing $\hat{\alpha}$ residually. In these simulation setups the estimation of ω was much lower than the segment's average. Intuitively, this is not very surprising as the high $\hat{\lambda}$ takes much of the variation of the series. The estimator $\hat{\sigma}^2$ of $\mathbb{E}\varepsilon_t^2$ is given by the series (and $\hat{\mu}$) and as $\mathbb{E}\varepsilon_t^2 = \mathbb{E}h_t = \omega/(1-\lambda)$ in the estimated model, a high estimation of λ must be compensated by a low estimation of ω.

We conclude that the global estimation of GARCH(1,1) parameters is highly sensitive to the presence of parameter changes. A single deterministic changepoint can add a long-term time scale to the volatility process and push the estimate of λ close to one. Changes in ω and in β can cause this effect.

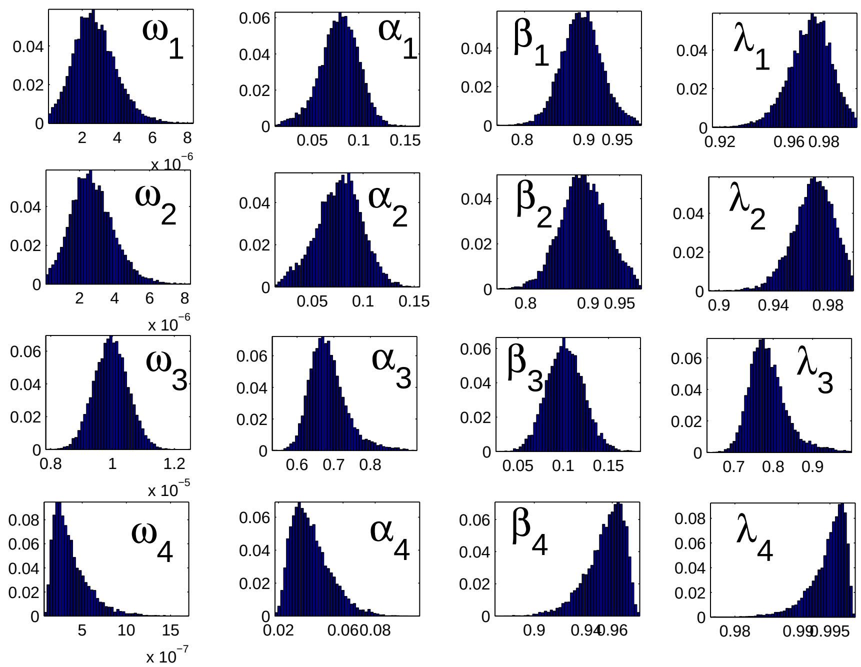

Figure 4: Histograms of the GARCH(1,1) estimations of ω, α, β, and λ of 10,000 artificial series for each scenario, constructed according to Table 1. The subscripts denote the scenarios.

4 Estimation of the Short Scale in Stock Volatility

The GARCH model implies correlation structures for the series ε_t^2:

$$\mathbb{E}\varepsilon_t^2\varepsilon_s^2 = \mathbb{E}\eta_t^2\eta_s^2 h_t h_s = \mathbb{E}h_t h_s,$$

and for the residual $\nu_t = \varepsilon_t^2 - h_t$:

$$\mathbb{E}\nu_t\nu_s = \mathbb{E}(\eta_t^2 - 1)(\eta_s^2 - 1)h_t h_s = 0.$$

We will extract the long time scale by estimating GARCH(1,1) with constant mean return on the $\hat{\varepsilon}_t(\hat{\mu})$-series, thereby obtaining the $\hat{h}_t(\hat{\omega}, \hat{\alpha}, \hat{\beta})$-series, and calculate the residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$. If there is a second time scale in the data, it will be visible in the $\hat{\nu}_t$'s.

4.1 Synthetic Data

The averaged periodogram is estimated by subsampling with a Tukey-Hanning window of 256 points length allowing for 64 points overlap. A Lorentzian spectrum model

$$h(w) = a + b/(c^2 + w^2), \tag{18}$$

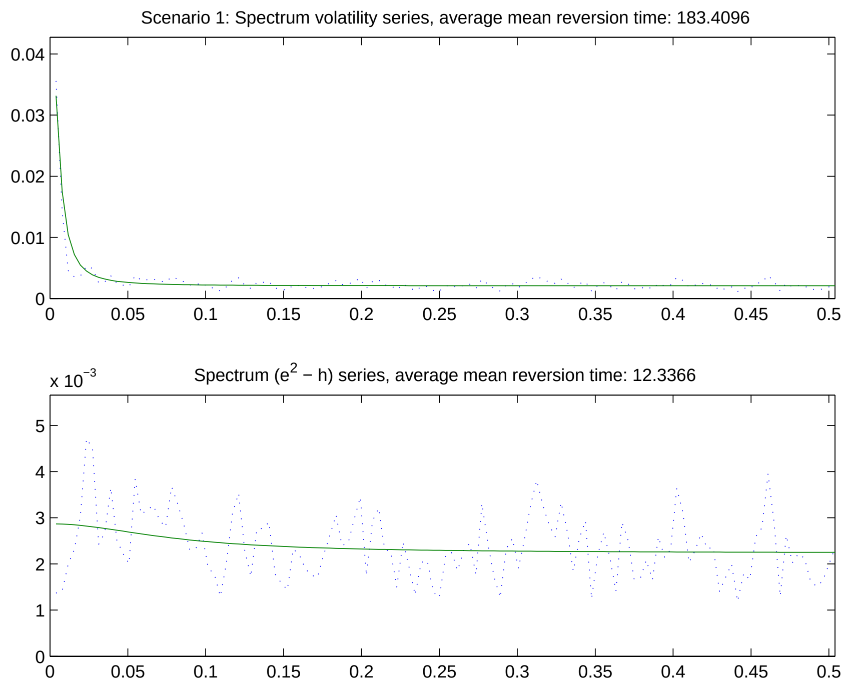

Figure 5: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ constructed according to Scenario 1 from Section 3.2 and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time was computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. Here we see that the GARCH(1,1) estimation indeed "peels off" the long time scale and the short time scale in the residuals $\hat{\nu}$ is revealed.

was fitted to the periodogram. w denotes the frequencies and (a, b, c) are parameters. The average mean reversion time was estimated by $1/c$. The parametrization of the Lorentzian is motivated in the Appendix.

For the series ε_t^2 and ν_t there is no explicit parametrization of the Lorentzian (18) in terms of the parameters of the discrete GARCH(1,1) model available. To establish the correspondence between the mean reversion time $1/c$ from the Lorentzian spectral model and $1/(1-\lambda)$ from the GARCH model, we generated 10,000 synthetic GARCH(1,1) series of 5000 points length ranging from $\lambda = 0.75$ to $\lambda = 0.99$ and estimated $1/c$ by a nonlinear least squares fit of the Lorentzian to the estimate of the power spectrum. (In particular, we set $\alpha \equiv 0.10$ and let β go through the interval [0.65, 0.89] while $\omega \equiv$ 1e-5.) The relation from a linear regression of the resulting series of $\{1/\hat{c}_i\}_i$ on the logs of $\{1/(1-\lambda_i)\}_i$ was obtained as

$$\frac{1}{\hat{c}} = -86.74 + 61.20 \log\left(\frac{1}{1-\lambda}\right), \quad R^2 = 0.93. \tag{19}$$

Note that the $\{1/(1-\lambda_i)\}_i$ series was known from the construction of the artificial

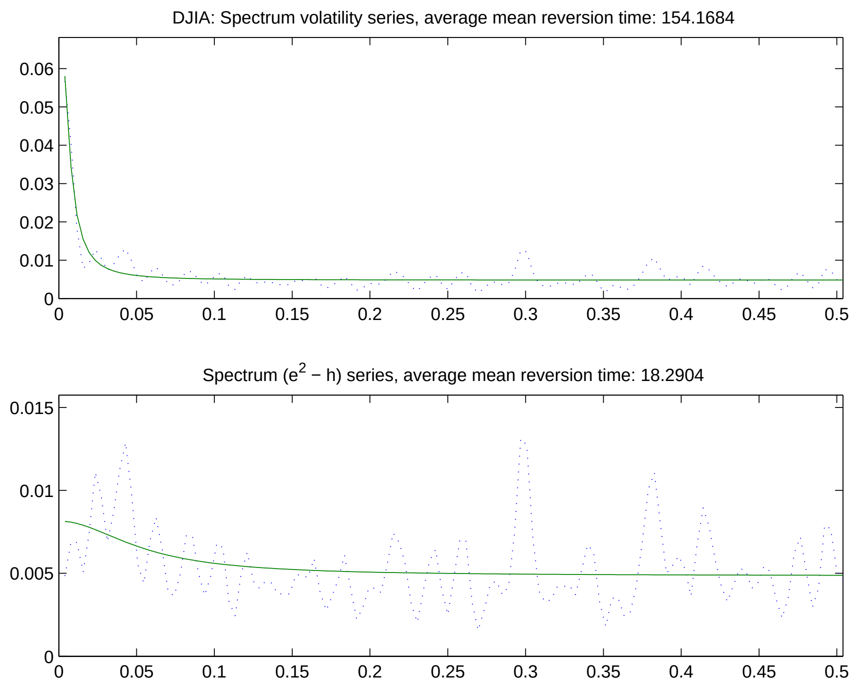

Figure 6: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ of the Dow Jones series and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time was computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. Clearly two distinct time scales can be observed in the Dow Jones series, a slower one of the magnitude of about 154 (51) days and a faster one of the magnitude of 18 (6) days. (The numbers in brackets are the time scales according to equation (19).)

series. .

We estimate the power spectra for series built according to Scenario 1 in Section 3.2. Here, a time scale of 7 days is superposed by a time scale of the complete length of the series as there is a single switch in ω. The power spectra of the residuals exhibit a short time scale as shown in Figure 5 for a typical realization. The GARCH(1,1) maximum likelihood estimates of this realization were $(\hat{\omega}, \hat{\alpha}, \hat{\beta}) = (8\text{e-}7, 0.0384, 0.9556)$. This is a long time scale of $1/(1-\hat{\lambda}) = 167$ days according to the parameter estimates, 180 days according to the Lorentzian model and $1/(1-\lambda) = 82$ days according to (19). A maximum likelihood estimation of a GARCH(1,1) model for the residuals $\hat{\nu}_t$ could not detect the second, short time scale but essentially repeated the estimates of the GARCH(1,1) model for the volatility series $\hat{\varepsilon}_t^2$.

The power spectrum estimation reveals a short scale in the residuals that is of the magnitude of $1/\hat{c} = 12$ days or $1/(1-\lambda) = 5$ days according to (19). This

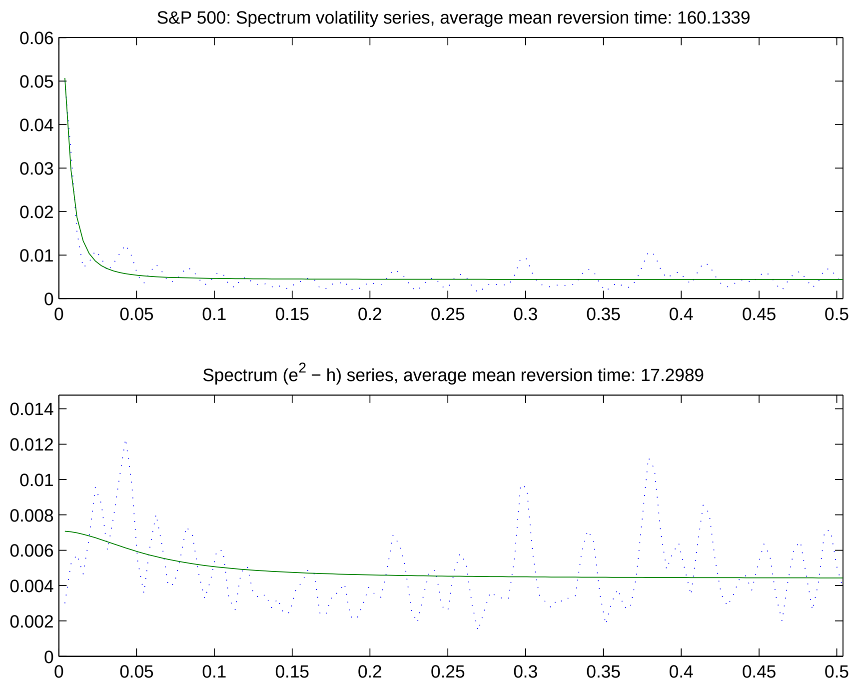

Figure 7: Upper graph: Estimation of the power spectrum (dotted line) of the volatility process $\hat{\varepsilon}_t^2$ of the S&P500 series and nonlinear least squares fit of a Lorentzian spectrum (solid line). The estimate of the average mean reversion time was computed as $1/c$ from the Lorentzian. Lower graph: same analysis for the residuals $\hat{\nu}_t = \hat{\varepsilon}_t - \hat{h}_t$ after a GARCH(1,1) estimation of the series above. The two distinct time scales compare to those of the Dow Jones series in Figure 6.

compares to the data-generating short scale of $1/(1-0.85) \approx 7$ days.

We conclude that by estimating GARCH(1,1) and computing the residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$ using the estimated GARCH parameters, we can eliminate the long time scale from the data. Spectral analysis is capable of measuring the short time scale left in the residual.

4.2 Market Data

Figure 6 shows the spectra of the volatility series $\hat{\varepsilon}_t^2$ of the Dow Jones series (above) and of the residual $\hat{\nu}_t$ (below). Two distinct time scales can be observed, the longer one about 154 days or $1/(1-\lambda) \approx 51$ days according to (19), the faster scale about 18 days or $1/(1-\lambda) \approx 6$ days. Figure 7 shows the spectra of the volatility (above) and the residual (below) of the S&P500 series. Again, two time scales can be observed and their magnitudes compare closely to those of the Dow Jones.

For inference statistics, we turn to the estimation of the autocorrelation func-

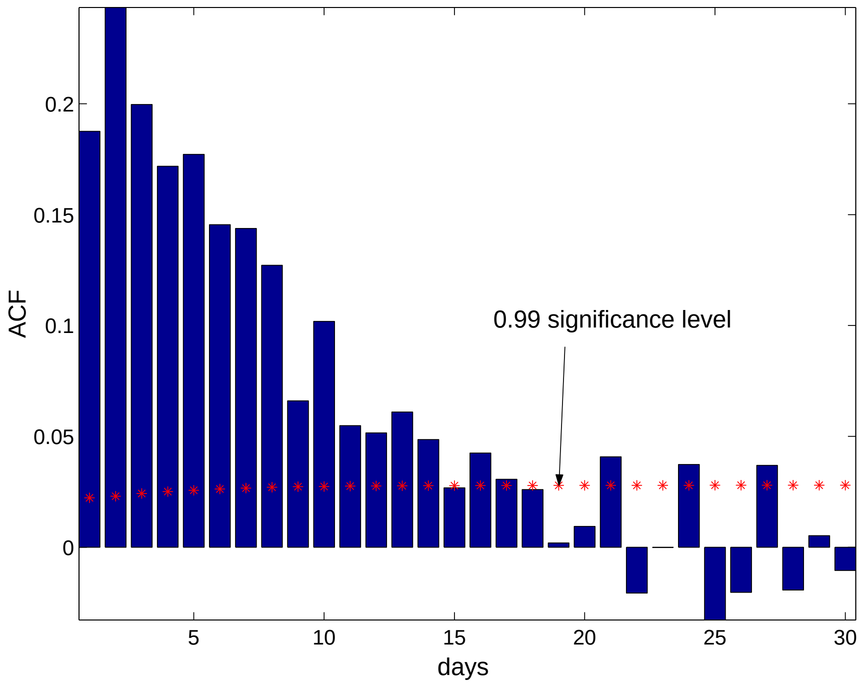

Figure 8: Sample autocorrelation function of the series $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$ of the Dow Jones series. Up to the lag of 14 days, the estimates are clearly significant and the median lag is 6 days. The sample autocorrelation function of the $\hat{\nu}_t$ series of the S&P500 looks essentially the same.

tion, which is equivalent to the estimation of the power spectrum of stationary processes by the Wiener-Khintchine theorem (e.g., Priestley 1981). The results for the Dow Jones and for the S&P500 series look essentially the same, so that we will report only the estimation for the Dow Jones in Figure 8.

5 Summary and Conclusion

Changes in GARCH(1,1) parameters that are not accounted for in global estimations lead to an estimated persistence that is much higher than the average persistence within the regimes. We show that for switches in the constant ω of the conditional variance equation of the GARCH(1,1) model, the sum of estimated autoregressive parameters $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ must be close to one. In simulations we obtain global estimates close to integration for parameter changes within realistic ranges for stock-price volatility. Changes in the GARCH constant ω and changes in the autoregressive parameter β can cause this effect. It is not necessary to have a certain underlying stochastic structure that drives the changes but a single deterministic changepoint is sufficient.

A global GARCH estimation has to take shifts in the unconditional variance induced by the parameter changes as “jumps” that persist as long as the parameter regime lasts. The infrequent regime changes induce a long time scale that dominates the parameter estimation and masks the short correlation structure that governs the process within regimes.

This short time scale can be uncovered in the GARCH(1,1) residual $\hat{\nu}_t = \hat{\varepsilon}_t^2 - \hat{h}_t$, where $\hat{h}_t$ is the estimated conditional volatility. By periodogram estimation of synthetic data we recover the short time scale that we inserted in the data. Applying this method to daily Dow Jones and S&P500 returns ranging from 1985 to 2000, we find a short time scale of the magnitude of 5 to 10 days. We conclude that at least two overlaying time scales are present in the considered series.

In this Appendix we will derive the Lorentzian model

$$h(w) = a + b/(c^2 + w^2)$$

for the power spectrum of the GARCH(1,1) process as used in Section 4. (a, b, c) are parameters and w denotes the frequencies.

Proposition 5. *The power spectrum of the log of ε_t^2 in the continuous time analogue of a Gaussian GARCH(1,1) model with constant mean return can be represented by the function*

$$h(w) = \frac{\gamma^2}{2\pi} + \frac{\alpha^2}{2\pi}\frac{1}{w^2+\vartheta^2}$$

where $\vartheta \approx 1 - \alpha - \beta$, γ^2 is the variance of $\log \eta_t^2$, $\eta_t \sim \mathcal{N}(0,1)$, and w denotes the frequencies.

Proof. Nelson (1990) showed that the discrete GARCH(1,1) model with constant mean return converges with $\Delta t \to 0$ in distribution to the system of stochastic differential equations

$$\begin{aligned} dY(t) &= \sigma(t)dW_1(t) \\ d\sigma^2(t) &= (\omega - \vartheta\sigma^2(t))dt + \alpha\sigma^2(t)dW_2(t), \end{aligned}$$

where $Y_t = \sum_{i=0}^{t}(r_i - \mu)$ are the cumulative excess returns, σ_t^2 is the volatility process, ω and α are the discrete GARCH(1,1) parameters and $\vartheta \approx 1 - \alpha - \beta$. $W_1(t)$, $W_2(t)$ are two independent Brownian Motions.

Taking the log of the volatility driver and denoting $V_t = \log \sigma^2(t)$, $f(V_t) = \sqrt{\exp(V_t)}$, and $m = \log(\omega/\vartheta) - \alpha^2/2\vartheta$, we have the first order equivalent

$$\begin{aligned} dY_t &= f(V_t)dW_1(t), \\ dV_t &= \vartheta\,(m - V_t)\,dt + \alpha dW_2(t). \end{aligned}$$

V_t is an Ornstein-Uhlenbeck process with solution

$$V_t = m + (V_0 - m)e^{-\vartheta t} + \alpha\int_0^t e^{-\vartheta(t-s)}dW_2(s).$$

As described for example in Arnold (1973), the correlation with respect to the stationary measure is given by

$$\operatorname{cov}(V_s, V_{s+t}) = \frac{\alpha^2}{2\vartheta}e^{-\vartheta t} \text{ for } s \to \infty \text{ and } t > 0.$$

We discretize the volatility process with $\Delta t = 1$ and obtain

$$Y_t - Y_{t-1} = \varepsilon_t = r_t - \mu = \sqrt{e^{V_t}}\,\eta_t, \quad \eta_t \sim \mathcal{N}(0,1).$$

This motivates the transformation

$$x_t := \log \varepsilon_t^2 = V_t + \log \eta_t^2,$$

$\log \eta_t^2$ being White Noise with mean zero and variance γ^2.

The autocorrelation of x_t is then given by

$$R_x(t) = \frac{\alpha^2}{2\vartheta} e^{-\vartheta t} + \gamma^2 \delta_0(t),$$

where $\delta_0(t)$ is the Dirac-function with unit mass at zero.

According to the Wiener-Khintchine theorem, the power spectrum of the real process x_t is then given by

$$\begin{aligned} h(w) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwt} \left(\frac{\alpha^2}{2\vartheta} e^{-\vartheta t} + \gamma^2 \delta_0(t) \right) dt \\ &= \frac{\gamma^2}{2\pi} + \frac{1}{2\pi} \frac{\alpha^2}{\vartheta} \operatorname{Re} \int_0^{\infty} e^{-(iw+\vartheta)t}\, dt \\ &= \frac{\gamma^2}{2\pi} + \frac{\alpha^2}{2\pi} \frac{1}{w^2 + \vartheta^2} \end{aligned} \qquad \square$$

Simplifying this to $h(w) = a + b/(w^2 + c^2)$, we can recover the e-folding time by $1/c$.

References

[1] ANDERSEN, T. G. / BOLLERSLEV, T. (1997). *Intraday Periodicity and Volatility Persistence in Financial Markets.* Journal of Empirical Finance 4: 115–158.

[2] ARNOLD, L. (1973). **Stochastische Differentialgleichungen: Theorie und Anwendung.** Oldenbourg: München.

[3] BAILLIE, R. T. / DEGENNARO, R. P. (1990). *Stock Returns and Volatility.* Journal of Financial and Quantitative Analysis 25(2): 203–214.

[4] BAILLIE, R. T. / BOLLERSLEV, T. / MIKKELSEN, H. O. (1996). *Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity.* Journal of Econometrics 74: 3–30.

[5] BOLLERSLEV, T. (1986). *Generalized Autoregressive Conditional Heteroskedasticity.* Journal of Econometrics 31: 307–327.

[6] BOLLERSLEV, T. (1987). *A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return.* Review of Economics and Statistics 69: 542–547.

[7] BOLLERSLEV, T. / WOOLDRIDGE, J. M. (1992). *Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances.* Econometric Reviews 11(2): 143–172.

[8] BOLLERSLEV, T. / ENGLE, R. F. (1993). *Common Persistence in Conditional Variances.* Econometrica 61(1): 167–186.

[9] CHERNOV, M. / GALLANT, A.R. / GHYSELS, E. / TAUCHEN, G. (2002). *Alternative Models for Stock-Price Dynamics.* mimeo. http://www.ssc.upenn.edu/∼fdiebold/w2002/program.htm

[10] DIEBOLD, F. X. (1986). *Modeling the Persistence of Conditional Variances: A Comment.* Econometric Reviews 5(1): 51–56.

[11] DIEBOLD, F. X. / INOUE, A. (2001). *Long Memory and Regime Switching.* Journal of Econometrics 105: 131–159.

[12] DING, Z. / GRANGER, C. W. J. (1996). *Modeling Volatility Persistence of Speculative Returns: A New Approach.* Journal of Econometrics 73: 185–215.

[13] DRĂGULESCU, A. A. / YAKOVENKO, V. M. (2002). *Probability Distribution of Returns for a Model with Stochastic Volatility.* mimeo. http://lanl.arXiv.org/abs/cond-mat/0203046.

[14] Engle, R. F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.* Econometrica 50(4): 987–1007.

[15] Engle, R. F. / Bollerslev, T. (1986). *Modelling the Persistence of Conditional Variances.* Econometric Reviews 5(1): 1–50.

[16] Engle, R. F. / Patton, A. J. (2001). *What good is a volatility model?* Quantitative Finance 1(2): 237–245.

[17] Fouque, J. P. / Papanicolaou, G. / Sircar, K. R. / Sølna, K. (2002). *Short time-scale in S&P 500 Volatility.* mimeo. http://georgep.stanford.edu/∼papanico/pubs.html

[18] Francq, C. / Roussignol, M. / Zakoïan, J.-M. (2001). *Conditional Heteroskedasticity Driven by Hidden Markov Chains.* Journal of Time Series Analysis 22(2): 197–220.

[19] Gallant, A. R. / Tauchen, G. (2001). *Efficient Method of Moments.* mimeo. http://www.unc.edu/∼arg.

[20] Gouriéroux, C. (1997). **ARCH Models and Financial Applications.** Springer: New York.

[21] Granger, C. W. J. (1980). *Long Memory Relationships and the Aggregation of Dynamic Models.* Journal of Econometrics 14: 227–238.

[22] Granger, C. W. J. / Teräsvirta, T. (2001). *A Simple Nonlinear Time Series Model with Misleading Linear Properties.* Economics Letters 62: 161–165.

[23] Gray, S. F. (1996). *Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process.* Journal of Financial Economics 42: 27–62.

[24] Greene, W. H. (2000). **Econometric Analysis.** 4th ed., Prentice-Hall: New Jersey.

[25] Hamilton, J. D. / Susmel, R. (1994). *Autoregressive Conditional Heteroskedasticity and Changes in Regime.* Journal of Econometrics 64: 307–333.

[26] Harvey, A. C. (1976). *Estimating Regression Models with Multiplicative Heteroscedasticity.* Econometrica 44(3): 461–465.

[27] Lamoureux, C. G. / Lastrapes, W. D. (1990). *Persistence in Variance, Structural Change, and the GARCH Model.* Journal of Business & Economic Statistics 8(2): 225–234.

[28] LeBaron, B. (2001). *Stochastic Volatility as a Simple Generator of Apparent Financial Power Laws and Long Memory.* Quantitative Finance 1(6): 621–631.

[29] Lumsdaine, R. (1996). *Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models.* Econometrica 64(3): 575–596.

[30] Mikosch, T. / Starica, C. (2000). *Change of Structure in Financial Time Series, Long Range Dependence and the GARCH Model.* University of Aarhus, Aarhus School of Business, Centre for Analytical Finance, Working Paper No. 58.

[31] Nelson, D.B. (1990). *ARCH Models as Diffusion Approximations.* Journal of Econometrics 45: 7–38.

[32] Nelson, D. B. (1990a). *Stationarity and Persistence in the GARCH(1,1) Model.* Econometric Theory 6: 318–334.

[33] Press, W. H. / Teukolsky, S. A. / Vetterling, W. T. / Flannery, B. P. (2002). **Numerical Recipes in C++.** Cambridge University Press: Cambridge, U.K.

[34] Priestley, M. B. (1981). **Spectral Analysis and Time Series.** Academic Press: San Diego.

[35] Weiss, A. A. (1986). *Asymptotic Theory for ARCH Models: Estimation and Testing.* Econometric Theory 2: 107–131.