ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

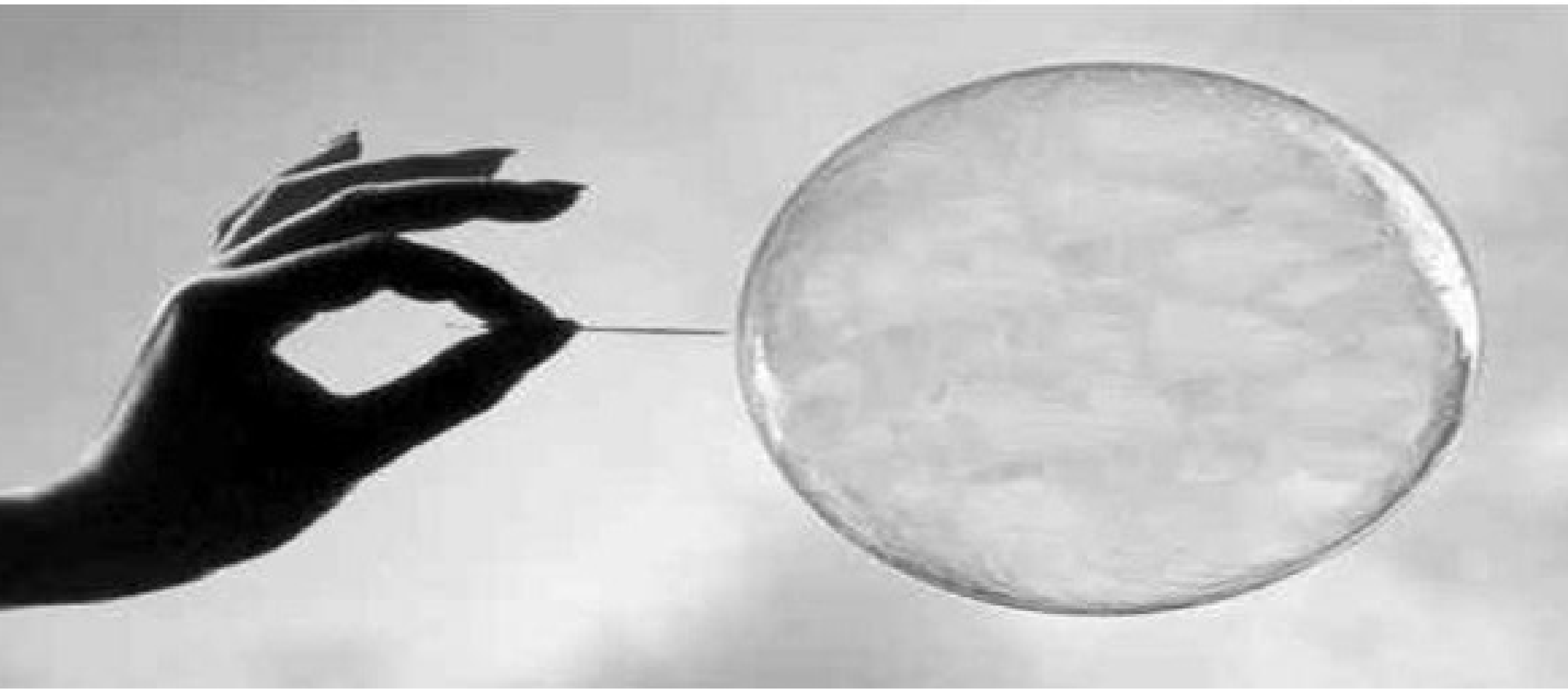
University of Zurich[UZH]

Master thesis

Detection of financial bubbles with the FTS-GARCH model and extensions

Martin Pleischl

May 2013



Master Thesis for the Master of Science in Quantitative Finance.

Supervisor: Prof. Dr. Erich Walter Farkas
Program Director UZH ETH joint degree MSc in Quantitative Finance
Department of Banking and Finance, University of Zurich, Switzerland
Department of Mathematics, ETH Zurich, Switzerland

Co-Supervisor: Prof. Dr. Didier Sornette
Director of the Financial Crisis Observatory
Professor on the Chair of Entrepreneurial Risks
Department of Management, Technology and Economics, ETH Zurich, Switzerland

University of Zurich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Acknowledgements

I would like to express my gratitude to Prof. Didier Sornette for the opportunity to write my Master thesis under his guidance. I thank him for many interesting ideas and suggestions for improvement. In particular, I appreciated his feedback at several meetings we arranged throughout this project. I'm also grateful for the supervision of Prof. Walter Farkas who enabled the thesis in this form. Furthermore, I would like to thank Dr. Mika Kastenholz for helpful inputs at our meetings and especially for establishing the contact between Prof. Sornette and me in terms of this project. Finally, I feel grateful to my parents for providing financial support during my years of study.

The thesis was a great opportunity to get an insight in the research of Prof. Sornette on financial bubbles.

Abstract

The main goal of this Master thesis is to find an appropriate bubble detection method that combines bubble dynamics leading to a finite-time singularity (positive feedback of either price on return or of return on return) with the standard of the financial industry to model volatility clustering and to assess volatiliy risk, namely the $GARCH(1,1)$ model. One form of the resulting stochastic model is the so called **Finite-Time Singularity GARCH model** (invented by Corsi and Sornette (2012)), a $GARCH(1,1)$ enhanced by adding a regression component in the equation for the conditional mean. We demonstrate an algorithm to consistently estimate the coefficients of this kind of model. Then, we test several modified versions of the original $FTS\ GARCH$ model from Corsi and Sornette for their ability to detect speculative bubbles. Finally, we add a second regression component modeling the possibility of positive feedback of volatility on return in addition.
As an alternative class of models we consider **Markov switching models** which characterize financial returns as a process with different regimes (e.g. a regime of slow growth, a bubble regime and a crash regime). Again, we find an algorithm to consistently estimate the coefficients of such a model and consequently apply it for detection of financial bubbles.

Contents

1. Introduction **1**

2. Characteristics of Financial Time Series **3**
2.1. Small linear autocorrelation . . . 3
2.2. Leptokurtic return distribution (fat tails) . . . 4
2.3. Asymmetric return distribution . . . 5
2.4. Standard deviation of returns dominates mean . . . 7
2.5. Volatility clustering and the Taylor effect . . . 7
2.6. Leverage effect . . . 8
2.7. Time varying correlation between assets . . . 10
2.8. Conditional heavy tails . . . 10
2.9. Aggregational Gaussianity . . . 10

3. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model **13**
3.1. Introducing the $GARCH(p,q)$ process . . . 13
3.2. Characteristics of returns in a $GARCH$ model . . . 14
3.3. Maximum likelihood estimation . . . 16

4. Testing for Dependences in Bubble Periods **21**
4.1. Empirical autocorrelation function . . . 21

5. Finite-Time Singularity (FTS) GARCH **25**
5.1. A bubble experiment in the laboratory . . . 25
5.2. Super-exponential growth and positive feedback on return . . . 26
5.3. Finite-time singularities and the log-periodic power law ($LPPL$) . . . 28
5.4. Introducing the FTS $GARCH$ model . . . 30
5.5. Improving the usage of return as regression component . . . 51
5.5.1. Kalman filtered return . . . 51
5.5.2. Average return . . . 58
5.5.3. Exponentially weighted moving average return . . . 68
5.6. Positive feedback of volatility on return . . . 77

6. Markov Switching Model **85**
6.1. Maximum likelihood estimation 86
6.2. Synthetic tests 90
6.3. Detecting bubbles in real data 93
6.3.1. Usage of average returns 93
6.3.2. Usage of EWMA returns 94

7. Conclusion **97**

A. Appendix - The Johansen-Ledoit-Sornette (JLS) model **99**

B. Appendix - Kalman filter: Derivation of the optimal Kalman gain **101**

C. Appendix - MS model: Derivation of the inference algorithm **103**

Bibliography **105**

Software **109**

1 Introduction

Financial bubbles have been analyzed by academics and practitioners for decades. Nevertheless, there have been plenty of financial bubbles in history which most people did not expect to burst. Quite the contrary, people usually extrapolate the past price evolution and continue their expectation of a rising asset price ("trend following"). This leads to a well-known phenomenon of financial markets, the herding behavior of its participants. A rising market attracts more and more buyers, whereas a falling market leads people rather to sell (downward movements often occur even a lot faster than upward movements). This is probably the most obvious mechanism leading to positive feedback on price. By analyzing a bubble simulation experiment, Huesler, Sornette and Hommes (2012) find evidence that financial bubbles are characterized by super-exponential growth (the growth rate is growing itself). This is contradictory to many previously developed models describing bubbles as periods of exponentially growing prices. The positive feedback loops make the price run into a singularity within finite time. This motivates the goal of this Master thesis to develop appropriate mathematical methods that can be used to analyze time series for the presence of a super-exponentially growing bubble. Such tools might be useful for the purpose of risk management of financial institutions since they can serve as early warning systems for financial crashes. Also, they could potentially be used to develop trading strategies generating alpha (risk-adjusted excess return).

The thesis is organized in 7 chapters. After this short introduction, we start with an analysis of the empirical characteristics of typical financial time series where we demonstrate these findings using the time series of NASDAQ Composite. Then, in Chapter 3 we define the $GARCH$ regression model and show how to consistently estimate its parameters. Also, we analyze the characteristics of returns defined as a $GARCH$ process. In Chapter 4 we use the empirical autocorrelation function in order to find time windows of financial bubbles where the correlation of consecutive returns is increased. In Chapter 5 we present in detail the above mentioned bubble experiment and the emerging evidence for super-exponential growth of prices in bubble periods. This is our motivation for a model combining the $GARCH$ process, which is commonly used to model volatility of financial returns, and bubble dynamics incorporating positive feedback effects. We start with the basic version of the so called "Finite-Time Singularity $GARCH$" model developed by Corsi and Sornette (2012) and test various enhanced versions for their ability to detect financial bubbles. As examples for historical bubbles we choose the DAX, NASDAQ and NIKKEI time series. The Gold price serves as a candidate for a real-time bubble. In Chapter 6 we propose a method how to use Markov switching models, which incorporate the possiblity of regime shifts, to detect periods of super-exponentially growing prices. We conclude the thesis with a short summary of our findings.

2 Characteristics of Financial Time Series

As a first step, we would like to do an empirical study on the characteristics of a typical financial time series $P_0, \dots, P_n$. Since the price process is usually non-stationary (it does not have any long-run price level), one analyzes the corresponding series of continuous returns calculated over a certain horizon τ:

$$r_t^\tau := \log\left(\frac{P_{t\cdot\tau}}{P_{(t-1)\cdot\tau}}\right), \text{ where } t = 1, \dots, \left[\frac{n}{\tau}\right] =: T. \tag{2.1}$$

Typically, the return series is implicitly assumed to be a trajectory from a **(weak)-stationary** process, which we also call r_t^τ (for simplicity). This process fulfills the following properties (weak- or covariance-stationarity):

1. $\mathbb{E}[r_t^\tau] = \mu \in \mathbb{R}, \quad \forall\, t = 1, \dots, T$
2. $\mathrm{Cov}(r_{t-k}^\tau, r_t^\tau) = \mathbb{E}[(r_{t-k}^\tau - \mu)(r_t^\tau - \mu)] = \gamma_k \in \mathbb{R}, \quad \forall\, t = k+1, \dots, T, \text{ and } k = 1, \dots, K < T$

Stationarity is a necessary condition in order to estimate the moments of the return distribution with data from different time periods. But stationarity is not sufficient. We have to keep in mind that only for an **ergodic** process the empirical time average converges to the theoretical moment of r_t. This is another implicit assumption usually made when analyzing time series data. Nevertheless, in many cases the concepts of stationarity and ergodicity turn out to require the same conditions. (Cont (2001) [C], Hamilton (1994) [H])
In the empirical analysis below we use the daily return series $r_t := r_t^1,\ t = 1, \dots, T$, of the NASDAQ Composite Index from 5^{th} February 1971 to 26^{th} October 2012 in order to illustrate several **"stylized facts"** listed by Perner and Schlener (2011) [PS]. Sewell (2011) [S1] nicely explains a stylized fact as "a term used in economics to refer to empirical findings that are so consistent (for example, across a wide range of instruments, markets and time periods) that they are accepted as truth. Due to their generality, they are often qualitative."

Stylized fact 1: Small linear autocorrelation

Typically, price changes of financial assets do have little or even insignificant autocorrelation. The **autocorrelation function** of returns defined as

$$C(lag, \tau) := Corr(r_t^\tau, r_{t-lag}^\tau), \text{ where } t = lag + 1, \dots, T, \tag{2.2}$$

(*Corr* being the sample correlation) is usually at low levels for all lags ($lag < T$). Using intraday data one can find a decline to levels around 0 within a few minutes ($lag \geq 15$ min). (Cont (2001) [C])

There are many studies on the autocorrelation of returns in the literature. Sewell (2011) [S1] finds quite a nice summary: "Weekly and monthly stock returns are weakly negatively correlated, whilst daily, weekly and monthly index returns are positively correlated. Campbell et al. (1996) (p. 74) point out that this somewhat paradoxical result can mean only one thing: large positive cross-autocorrelations across individual securities across time. High frequency market returns exhibit negative autocorrelation."

In our empirical analysis we found a few lags (using daily steps) where one can reject $H_0 : Corr(r_t, r_{t-lag}) = 0$ at the 95% confidence level (see Figure 2.1). As we will notice later in the thesis, the NASDAQ time series includes periods of strong autocorrelation at certain lags (especially at a *lag* of 1 day). This is one of the reasons why we will include a regression component of lag 1 into the return equation of the standard GARCH model.

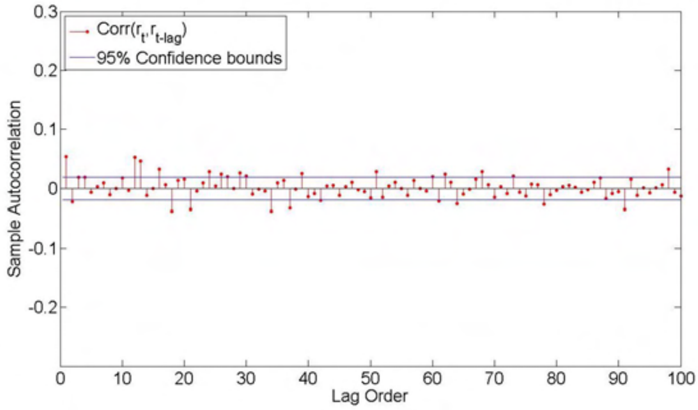


Figure 2.1.: NASDAQ Composite (1971-2012): Autocorrelation of daily returns

Cont (2001) [C] explains the intuition behind the small linear autocorrelation exhibited by financial price movements as follows. If returns are significantly correlated, arbitrage strategies yielding positive expected earnings can be developed. These strategies in turn reduce the autocorrelation of returns except for the reaction time of the market to new information (which is usually very short). Further, he cites Mandelbrot who describes this as "arbitrage tends to whiten the spectrum of price changes". Therefore, traditional tools based on second-order properties, for instance, ACF analysis or ARMA modeling, cannot distinguish between asset returns and white noise. Hence, nonlinear measures to analyze the dependence of financial returns are necessary.

Stylized fact 2: Leptokurtic return distribution (fat tails)

The unconditional distribution of daily returns has heavier tails than a normal distribution. Notice how the histogram in Figure 2.2 has longer and fatter tails and how it is more peaked around zero than the normal distribution. One way to quantify the deviation from the normal distribution is by using the **sample excess kurtosis** of the return series defined as

$$K := \frac{m_4}{m_2^2} - 3, \tag{2.3}$$

where $m_p,\ p \in \mathbb{N}$, is the p-th sample central moment

$$m_p := \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r}_t)^p. \tag{2.4}$$

The excess kurtosis is defined such that $K = 0$ for a Gaussian distribution (mesokurtic). A distribution with $K > 0$ is called leptokurtic and with $K < 0$ it is called platykurtic, respectively. We calculated quite a high sample excess kurtosis of 9.71 for the NASDAQ time series. Hence, the return distribution is leptokurtic.

Stylized fact 3: Asymmetric return distribution

Looking at stock market indizes, the unconditional distribution of daily returns is asymmetric or **negatively skewed**. This means there occur more large drops than large gains (see again Figure 2.2) or, in other words, downward movements occur faster than upward movements. A common measure for the **gain/loss asymmetry** is the **sample skewness**

$$S := \frac{m_3}{m_2^{3/2}}. \tag{2.5}$$

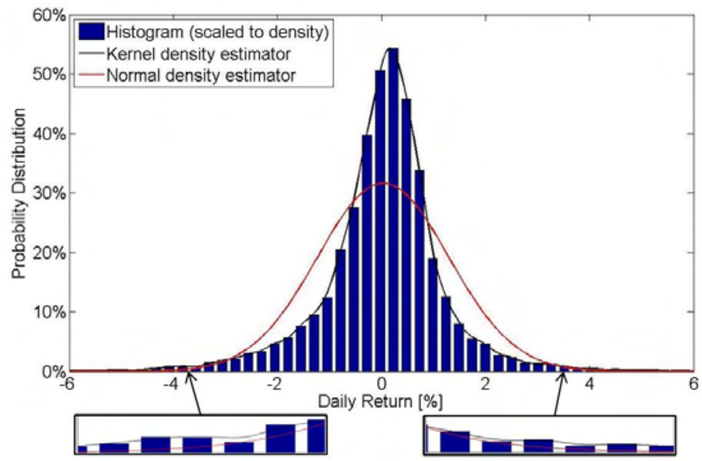


Figure 2.2.: NASDAQ Composite (1971-2012): Histogram of daily returns with kernel and normal density estimator

The NASDAQ return series has a sample skewness of about -0.29. In other markets such as that for foreign exchange one often finds a smaller gain/loss asymmetry. Interestingly, Johansen et al. (2006) [JSJ] show that although one usually finds a gain/loss asymmetry in the stock index, this effect is not present in the time series of the individual stocks nor their average. They take a closer look at the Dow Jones Industrial Average. To find the reason for the negative skewness in stock indizes, Johansen et al. simply compare the gain and loss distributions of the DJIA with the corresponding distributions for single stocks in the DJIA as well as their average. For all 21 stocks in the DJIA by the year 2004 they find the same waiting time distributions for both a positive and a negative return level of 5%. Furthermore, they average the gain and loss distributions separately in order to obtain an average behavior for the stocks in the DJIA and detect an almost perfect agreement of both distributions. Finally, they assess it is twice as slow to move the DJIA down and four times as slow to move it up compared to the average time to move one of its stocks up or down.
Consecutive similar movements of stocks of different economic areas (as there are included in the DJIA) can have hardly any economically fundamental relation, which means there need to be significant correlations between these stocks. Thus, the authors conclude that the gain/loss asymmetry in the market must also be attributed to human psychology. People are in general much more risk averse than risk taking. The implication is that a movement in the DJIA can be initiated by a movement in some particular stock of a certain economic sector. This is often followed by a movement in economically related stocks resulting in a cascade of consecutive movements in all the sectors covered by the DJIA and hence in the DJIA itself ("**feedback loop**").

In order to quantify the deviation of the unconditional return distribution from a normal distribution one can perform the **Jarque-Bera test** (we follow the way of Buening and Thadewald (2004)

[BT]). Under the null hypothesis $H_0 : \; r_t \sim N(\mu, \sigma^2), \; t = 1, ..., T$, the Jarque-Bera test statistic JB fulfills

$$JB := \frac{T}{6}(S^2 + \frac{K^2}{4}) \xrightarrow{D} \chi^2_2. \tag{2.6}$$

Therefore, H_0 has to be rejected at level α if $JB \geq \chi^2_{1-\alpha,2}$. Any deviation from the normal distribution, which has zero skewness and zero excess kurtosis, increases JB. The asymptotic χ^2-distribution has two degrees of freedom because the test statistic is a sum of two asymptotically independent standard normal distributions (see Bowman and Shenton (1975)).
As anticipated, the Jarque-Bera test finds an extremely high test statistic of $JB = 41539$ for the NASDAQ Composite return series between 1971 and 2012. Hence, the p-value is almost zero even for an α of only 0.1% or smaller and we can reject the null hypothesis of a normal distribution. This definitely relies more on the high excess kurtosis of 9.71 than on the relatively small sample skewness of -0.29 we found. In a nutshell, the return distribution is **highly leptokurtic** and has **light negative skewness**.

It is worthwile to make a short excursion to the field of Extreme Value Theory in order to find out more about the estimation of heavy tails (following Winker and Jeleskovic (2006) [WJ]). In general, a fat tail distribution is defined as a distribution whose tails follow a power law (slower than the decline in the tails of a normal distribution). Looking for example at the right tail, the distribution is of **Pareto-type with tail index** $\alpha > 0$ if

$$\lim_{x \to \infty} (1 - F(x))x^\alpha = \beta, \quad \text{with} \quad \beta > 0. \tag{2.7}$$

Note that for a heavy-tailed Pareto-type distribution the higher order moments $\mathbb{E}[r_t^p]$ do only exist for $p < \alpha$ (Embrechts et al., 1999). Hence, the estimation of the tail index α informs us on the one hand about the heaviness of the tails and on the second hand about the existence of moments. For $\alpha < 4$ the kurtosis of the return distribution does not exist and for $\alpha < 2$ not even the variance does exist. As the case may be, sample estimates of variance, skewness or kurtosis will not converge. From a theoretical point of view they should even increase as the sample size increases, but this behavior is not found empirically. Hence, in case $\alpha < 4$ we should use the sample kurtosis and the Jarque-Bera test statistic with additional care.
The most popular estimator for α is the **Hill estimator** developed by Hill (1975). We describe the method for the right tail, the estimator for the left tail is obtained analogously. The Hill estimator is a conditional maximum likelihood estimator based on the order statistics exceeding a high threshold $m > 0$. The crucial assumption is that m is known. Thus, let us define the order statistics of a return time series $\{r_t\}_{t=1}^T$ as $r_{1,T} \leq r_{2,T} \leq \ldots \leq r_{T,T}$. If $k = k(T)$ is the number of order statistics exceeding m, the Hill estimator for $\alpha = \frac{1}{\xi}$ is

$$\frac{1}{\hat{\alpha}_k} = \hat{\xi}_k = \frac{1}{k} \sum_{j=1}^{k} \log(r_{T-j+1,T}) - \log(r_{T-k,T}). \tag{2.8}$$

ξ is the shape parameter in a generalized extreme value (GEV) distribution (see Embrechts et al. (1999) or Winker and Jeleskovic (2006) for details). Wagner and Marsh (2003) postulated m to be chosen such that

$$k(T) \longrightarrow \infty, \quad \frac{k(T)}{T} \longrightarrow 0 \quad \text{as} \quad T \longrightarrow \infty. \tag{2.9}$$

Under these conditions the Hill estimator is consistent, asymptotically normal distributed, asymptotically quite robust with respect to deviations of the return series from independence and converges optimally.

The key issue when we estimate α by the Hill estimator is the choice of the threshold m. It is common practice to analyse the **Hill plot** defined as

$$\{(k, \hat{\alpha}_k), k = 1, \ldots, K\}, \quad (2.10)$$

where one usually chooses α from a stable region. Nevertheless, we would like to have a look at a slightly different version, we plot the Hill estimator against the threshold value m (see Figure (2.3)). Firstly, this enables us to directly compare the graph with the histogram in Figure (2.2), and secondly, the comparison of the tails is more meaningful.

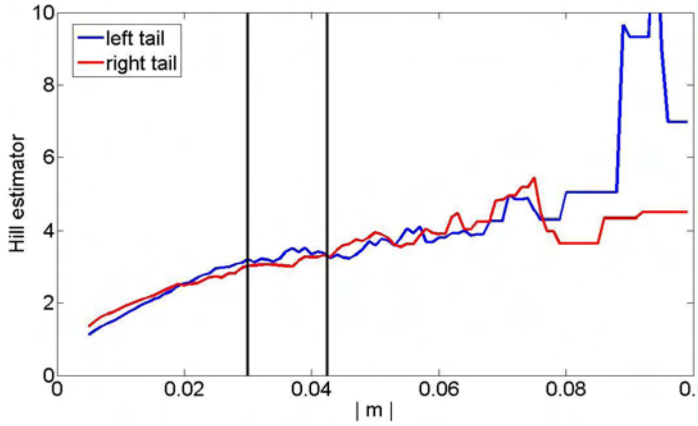


Figure 2.3.: NASDAQ Composite (1971-2012): Hill estimator for tail index α against absolute value of threshold m

Looking at the stable region around levels of $|m| = 3 - 4\%$, we found great evidence of $\alpha < 4$ for the NASDAQ Composite return series. α is even around levels of $3 - 3.5$, which means the distribution is heavy-tailed and does not have a finite forth moment. At least the second moment of r_t does exist, which is a condition for the return series to be stationary. Moreover, we do not find a significant difference in the decline of the left and right tail. This corresponds to the above finding of only a light skewness.

However, the Hill estimator does have many drawbacks. This is why it has been critized a lot in the finance literature. More recently developed estimation methods use the generalized Pareto distribution (additional location and scale parameters) and estimate its parameters either with the "Block Maxima Method" or the "Peak-over-Threshold Method". The interested reader is referred to the studies of P. Embrechts et al.

Stylized fact 4: Standard deviation of returns dominates mean

The standard deviation of daily returns clearly dominates their mean. Typically, it is not possibile to statistically reject a zero mean return. We calculated a mean of 0.032% and a standard deviation of 1.26% for the NASDAQ return series between 1971 and 2012.

Stylized fact 5: Volatility clustering and the Taylor effect

Due to the small linear autocorrelation of returns, random walk models were often considered to model financial prices. But these models assume that returns are independent which is definitely not the case.

As we will see soon (following the ideas of Cont (2001) [C]), there is significant positive serial correlation in the series of absolute as well as squared returns which indicates there occurs **volatility clustering** in the return series. If the increments were independent, these nonlinear functions of the return series would neither exhibit any autocorrelation. As Mandelbrot noted already back in 1963, "large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes" [M1]. The NASDAQ return series in Figure 2.4 shows multiple volatility clusters.

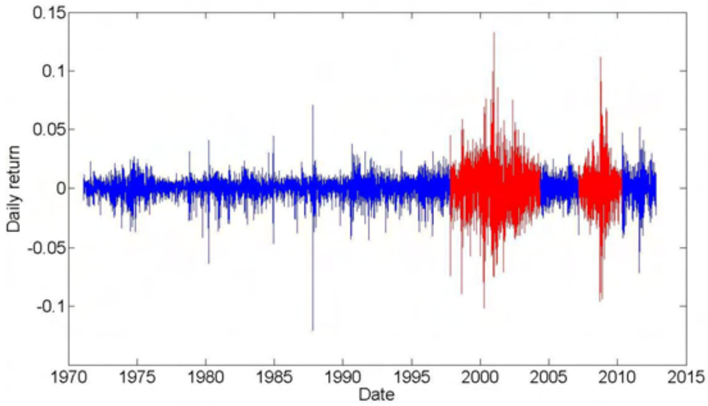


Figure 2.4.: NASDAQ Composite (1971-2012): Daily return series (2 typical volatility clusters marked in red)

A commonly used measure for this effect is the autocorrelation of **squared returns**:

$$C_2(lag,\tau) := Corr(|r_t^\tau|^2, |r_{t-lag}^\tau|^2), \text{ where } t = lag+1,\dots,T. \tag{2.11}$$

The autocorrelations are significantly positive and slowly decaying for increasing lag (see Figure 2.5). This phenomenon makes at least the amplitude of returns to some degree predictable. Sometimes volatility clustering is also called the "ARCH effect" since it is a feature of (G)ARCH models (see Chapter 3).

Similarly one can generalize the above measure for volatility clustering to arbitrary powers of returns ($\alpha > 0$):

$$C_\alpha(lag,\tau) := Corr(|r_t^\tau|^\alpha, |r_{t-lag}^\tau|^\alpha), \text{ where } t = lag+1,\dots,T. \tag{2.12}$$

Comparing the decay of C_α, Ding et al. (1993) remarked that, for fixed lag and τ, this correlation is highest for $\alpha = 1$, which means that **absolute returns** are more predictable than other powers of returns. According to Mora-Galan, Perez and Ruiz (2004) [MPR], Granger and Ding (1995) call this empirical property **Taylor effect**. They find (1994, 1996) the maximum autocorrelation to be obtained rather for values of $\alpha < 1$ than for $\alpha = 1$. Finally, they also point out that autocorrelations of absolute returns are always larger than autocorrelations of squared returns.

Furthermore, for fixed τ, the decay of the autocorrelation function $C_\alpha(lag,\tau)$ as lag increases is well reproduced by a power law (Cont (2001) [C])

$$C_\alpha(lag,\tau) \sim \frac{A}{lag^\beta}, \tag{2.13}$$

with a coefficient $\beta \in [0.2, 0.4]$ for absolute or squared returns respectively (see Figure 2.5). This characteristic of a slowly decaying serial correlation is a sign for long term persistence in volatility. There are models for financial returns that incorporate this effect (e.g. the GARCH model in Chapter 3).

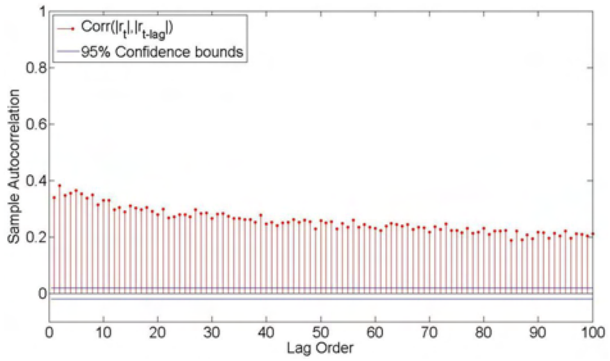


Figure 2.5.: NASDAQ Composite (1971-2012): Autocorrelation of absolute and squared daily returns

Stylized fact 6: Leverage effect

Cont (2001) [C] finds another measure of nonlinear dependence in returns measuring the **"leverage effect"**. For fixed τ the correlation of returns with subsequent squared returns

$$L(lag, \tau) := Corr(r^{\tau}_{t-lag}, |r^{\tau}_{t}|^2), \text{ where } t = lag + 1, \dots, T, \tag{2.14}$$

starts from a negative value and declines to 0 (see Figure 2.6), suggesting that negative returns lead to a rise in volatility and positive returns lead to a fall in volatiliy.

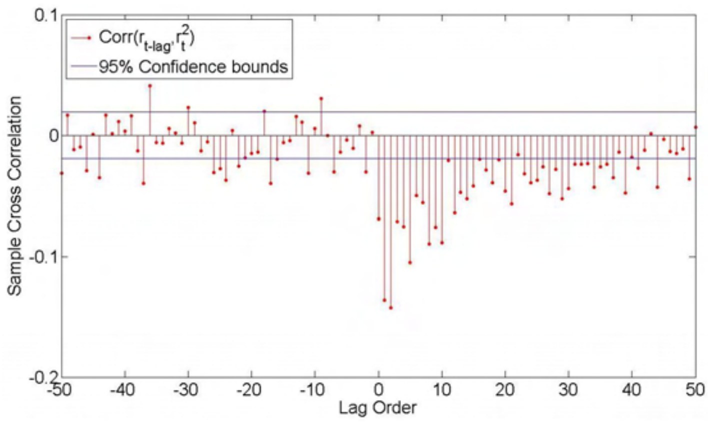


Figure 2.6.: NASDAQ Composite (1971-2012): Leverage effect of daily returns ($\tau = 1$)

According to Ait-Sahalia, Fan and Li (2013) [AFL], Black (1976) and Christie (1982) found an economic interpretation for the "leverage effect": If a company's stock decreases, the ratio of debt to equity (its "leverage") is increased. Consequently, the business becomes riskier and hence the stock's volatility rises. Nevertheless, this explanation is quite controversial in the literature and several alternatives can be found. The inversion, meaning that a movement in volatility influences the direction of subsequent returns, seems negligible ($L(lag, \tau) \approx 0$, for $lag < 0$, see again Figure 2.6). Moreover, according to Reigneron et al. (2011) [RAB], for a stock index the leverage effect can be decomposed into a volatility and a correlation effect. They confirm that downward index trends increase the average correlation between stocks, which explains why the index leverage effect is stronger than for single stocks.

The fact of almost no linear autocorrelation of returns but significant correlation in their volatility motivated a decomposition of the return as a product (Cont (2001) [C])

$$r^{\tau}_{t} = \sqrt{h^{\tau}_{t}} \cdot u_t, \tag{2.15}$$

where u_t is white noise and uncorrelated in time, and h^{τ}_{t} a conditional variance factor whose dynamics should consider the volatility clustering effect. Examples of such models are GARCH (see

Chapter 3) and long-memory stochastic volatility models.

Stylized fact 7: Time varying correlation between assets

Correlations between different financial assets tend to vary within time. Especially in highly volatile bear markets they appear to increase. During market crashes they might even tend to one. There has been developed a large amount of literature on Extreme Value Theory and Copulas during the last few years extending this topic, but in this thesis we do not want to go into more detail on it. The interested reader is again referred to the studies of P. Embrechts et al.

Stylized fact 8: Conditional heavy tails

An interesting question might be if standardizing returns by a time-varying volatility measure (e.g. via GARCH-type models) can lead to a normal conditional distribution of returns. Hence, we fit a $GARCH(1,1)$ model (see Chapter 3 for more details) to the NASDAQ daily return series r_t

$$r_t = \mu + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \quad (2.16)$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \; t = 1, \ldots, T, \quad (2.17)$$

and calculate standardized daily returns

$$r_t^{GARCH} := \frac{r_t}{\sqrt{h_t}}, \; t = 1, \ldots, T. \quad (2.18)$$

As one can see in Figure 2.7, the standardized returns still show fatter than normal tails. However, they are less heavy than in the unconditional case.

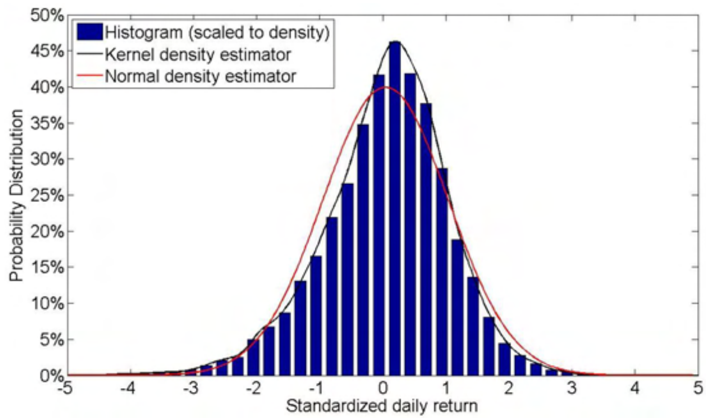


Figure 2.7.: NASDAQ Composite (1971-2012): $GARCH(1,1)$-standardized daily returns

Stylized fact 9: Aggregational Gaussianity

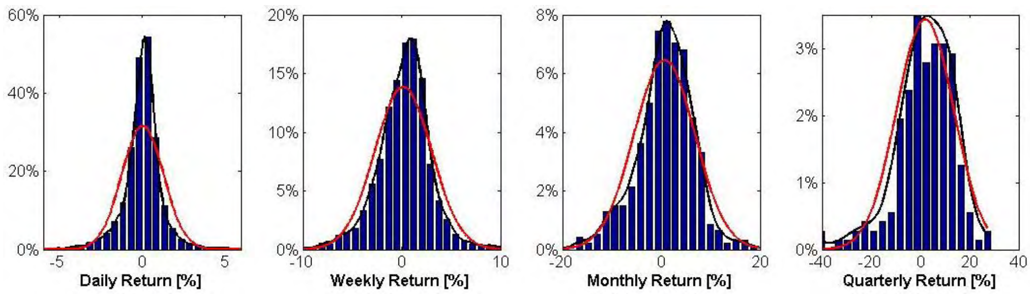


Figure 2.8.: NASDAQ Composite (1971-2012): Histogram (scaled to density), kernel density estimator and normal density estimator for daily, weekly, monthly and quarterly returns.

As one can see in Figure 2.8, increasing the horizon τ over which the returns are calculated leads the unconditional distribution to look more and more like a normal distribution.

In this chapter we have tried to give the reader a nice overview of empirical properties emerging from the statistical analysis of financial returns. These characteristics should be viewed as constraints a stochastic process has to verify in order to reproduce a typical return series accurately (Cont (2001) [C]). But finding and calibrating a model that incorporates all stylized facts at once is quite a difficult task. Nevertheless, there are models available that incorporate some of the most fundamental properties. Moreover, usually the purpose of the model determines which features need to be included inevitably to have an appropriate model and which do not.

3

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model

3.1. Introducing the $GARCH(p,q)$ process

Investors want to be rewarded with a premium for investing in risky assets. Since volatility is considered as a measure of risk, modeling and forecasting volatility, respectively, the covariance structure of asset returns (e.g. for risk management applications as Value-at-Risk) is an important task (Andersen, Davis, Kreiss and Mikosch (2009) [ADKM]). The standard model used in the financial industry to assess volatility risk is the so called **Generalized Autoregressive Conditional Heteroskedasticity** model. In 1986, Tim Bollerslev [B] developed the GARCH model as a generalization of Engle's original ARCH volatility model from 1982 [E]. Bollerslev designed GARCH to use fewer parameters than the ARCH model. Thereby, the computational burden can be reduced. The name can be decomposed in two terms: *Autoregressive* describes the dependence of the process on its own past and *conditional heteroskedasticity* means time-varying conditional variance (as we will see, the unconditional variance does not vary) (The MathWorks (2000) [M2]).
In a GARCH model (following the ideas of its inventor, Bollerslev (1986) [B]), the conditional variance of the innovations depends on the past of the process: Let $\{\epsilon_t\}_{t=1}^T$ denote a real-valued discrete-time stochastic process, and $\{\mathcal{F}_t\}_{t=0}^T$ a σ-algebra such that $\mathcal{F}_t$ includes exactly all available information until time t. Then, a $GARCH(p,q)$ **process with Gaussian innovations** is given by

$$\epsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t), \tag{3.1}$$

where

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} = \alpha_0 + \sum_{i=1}^{q} \alpha_i L^i \epsilon_t^2 + \sum_{i=1}^{p} \beta_i L^i h_t =: \alpha_0 + A(L)\epsilon_t^2 + B(L)h_t. \tag{3.2}$$

L^i is the backshift- or lag-operator defined as $L^i(x_t) := x_{t-i}$ for given time series x_t. $A(L)$ and $B(L)$ are the corresponding lag-polynomials. For the model orders p,q we require $p,q \in \mathbb{N}_0$. For $p = 0$ the process reduces to the $ARCH(q)$ process, and for $p = q = 0$ ϵ_t is simply white noise. In an $ARCH(q)$ process the conditional variance is computed from past innovations only, whereas a $GARCH(p,q)$ process uses also past conditional variances. We impose the constraints $\alpha_0 > 0$, $\{\alpha_i\}_{i=1}^q \geq 0$ and $\{\beta_i\}_{i=1}^p \geq 0$, to ensure the conditional variance h_t is strictly positive. Additionally, one can require α_q and β_p to be stricly positive as well in order to definitely include the maximum lags in a GARCH process of order p and q.

If all solutions of the polynomial equation $1 - B(z) = 0,\ z \in \mathbb{C}$, lie outside the unit circle ($|z| > 1$), then, the definition of h_t in (3.2) can be rewritten as an $ARCH(\infty)$ process:

$$\begin{aligned} h_t &= \alpha_0(1-B(1))^{-1} + A(L)(1-B(L))^{-1}\epsilon_t^2 = & (3.3)\\ &= \alpha_0\left(1-\sum_{i=1}^{p}\beta_i\right)^{-1} + \sum_{i=1}^{\infty}\delta_i\epsilon_{t-i}^2, & (3.4)\end{aligned}$$

The coefficients δ_i can be obtained from a power series expansion of $D(L) := A(L)(1-B(L))^{-1}$, where

$$\delta_i := \begin{cases} \alpha_i + \sum_{j=1}^{n}\beta_j\delta_{i-j}, & i = 1,\dots,q, \\ \sum_{j=1}^{n}\beta_j\delta_{i-j}, & i > q+1, \end{cases} \tag{3.5}$$

and $n := min(p, i-1)$. If $B(1) < 1$, the δ_i's are decreasing for $i > m := max(p,q)$. Therefore, if $D(1) < 1$, we can find an ARCH process approximating the $GARCH(p,q)$ process arbitrarily close. Nevertheless, GARCH is the more parsimonious model and therefore often preferable (reduced computational burden).
Also, S. Pantula found an $ARMA(m,p)$ representation for the squared $GARCH(p,q)$ process:

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{j=1}^{p}\beta_j\epsilon_{t-j}^2 - \sum_{j=1}^{p}\beta_j\nu_{t-j} + \nu_t, \tag{3.6}$$

where

$$\nu_t = \epsilon_t^2 - h_t = (\eta_t^2 - 1)h_t \quad \text{and} \quad \eta_t \overset{i.i.d.}{\sim} N(0,1). \tag{3.7}$$

3.2. Characteristics of returns in a $GARCH$ model

Bollerslev (1986) [B] finds the following sufficient condition for the $GARCH(p,q)$ process ϵ_t in (3.1) to be weak-stationary:

$$A(1) + B(1) = \sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{p}\beta_i < 1. \tag{3.8}$$

In particular, the unconditional mean and variance are given by

$$\mathbb{E}[\epsilon_t] = 0 \quad \text{and} \quad \text{Var}[\epsilon_t] = \frac{\alpha_0}{1 - A(1) - B(1)}. \tag{3.9}$$

This means the unconditional variance is homoscedastic (i.e. constant long-run variance expectation) (Aradhyula and Holt (1988) [AH]).

Furthermore, the $GARCH(p, q)$ process models the volatility clustering effect: The current volatility level is calculated from past innovations and even from past volatility levels. This introduces some kind of "memory effect". If there occur large movements of either sign, they increase the volatility level and hence further large movements are likely to follow (analog for small movements). Obviously, $GARCH(p, q)$ innovations are not independent, but they are uncorrelated. In fact, they can be decomposed in two components:

$$\epsilon_t := \sqrt{h_t} u_t, \tag{3.10}$$

where h_t is the conditional variance defined in (3.2) and $u_t \sim$ *i.i.d.* $N(0, 1)$ with u_t independent of $\mathcal{F}_{t-1}$. Hence, a GARCH process simply rescales an i.i.d. process where the conditional variance introduces serial dependence (The MathWorks (2000) [M2]).

From (3.6) Bollerslev (1986) [B] finds for the squared $GARCH(p, q)$ process ϵ_t^2 the following analogue to the Yule-Walker equations of an ARMA process ($m = max(p, q)$):

$$\gamma_n = \gamma_{-n} = \mathrm{Cov}(\epsilon_{t-n}^2, \epsilon_t^2) = \sum_{i=1}^{q} \alpha_i \gamma_{n-i} + \sum_{i=1}^{p} \beta_i \gamma_{n-i} = \sum_{i=1}^{m} \phi_i \gamma_{n-i}, \quad \text{for } \ n \geq p+1, \tag{3.11}$$

with

$$\phi_i = \alpha_i + \beta_i, \ \text{ for } \ i = 1, \ldots, q, \qquad \alpha_i = 0, \ \text{ for } \ i > q \quad \text{and} \quad \beta_i = 0 \ \text{ for } \ i > p. \tag{3.12}$$

Thus, we end up with

$$\rho_n = \frac{\gamma_n}{\gamma_0} = \sum_{i=1}^{m} \phi_i \rho_{n-i}, \quad \text{for } \ n \geq p+1. \tag{3.13}$$

Given $\rho_{p+1-m}, \ldots, \rho_p$ one can use these equations to uniquely determine ρ_n for $n \geq p+1$. In practice, one usually consistently estimates them by the corresponding sample analogues $\hat{\rho}_{p+1-m}, \ldots, \hat{\rho}_p$. Furthermore, Bollerslev also finds a similar analogue for the partial autocorrelation. These two together can be used for identification and diagnostic checking of the appropriate lag structure (p, q).

There exist many different variations of the standard $GARCH(p, q)$ model, e.g. absolute GARCH, exponential GARCH, integrated GARCH and so forth. Also, many different distributions have been applied to modeling the innovations. But the $GARCH(1, 1)$ with Gaussian innovations is still the most used in practice. For $p, q = 1$ the GARCH process gets the following form:

$$\epsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t), \ \text{ where } \ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}. \tag{3.14}$$

For simplicity, we calculate the following properties only for the standard $GARCH(1, 1)$ process. The skewness of ϵ_t is zero since $\mathbb{E}[\epsilon_t^3] = \mathbb{E}[h_t^{\frac{3}{2}}] \cdot \mathbb{E}[u_t^3] = 0$. Assuming $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 < 1$, the forth moment $\mathbb{E}[\epsilon_t^4]$ exists and one can compute the excess kurtosis to be

$$\frac{\mathbb{E}[\epsilon_t^4]}{\mathbb{E}[\epsilon_t^2]^2} - 3 = \frac{6\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 0. \tag{3.15}$$

Therefore, the unconditional distribution of ϵ_t is leptokurtic (see Bollerslev (1986) [B] for details). Moreover, the GARCH model does not account for the leverage effect. For arbitrary $lag < T$ one finds:

$$\begin{aligned}
\mathbb{E}[\epsilon_{t-lag}\epsilon_t^2] &= \mathbb{E}[\epsilon_{t-lag}h_t u_t^2] = \mathbb{E}[\epsilon_{t-lag}h_t] = \mathbb{E}[\epsilon_{t-lag}(\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1})] = && (3.16)\\
&= \alpha_1\mathbb{E}[\epsilon_{t-lag}\epsilon_{t-1}^2] + \beta_1\mathbb{E}[\epsilon_{t-lag}h_{t-1}] \;=\; \ldots\ldots \;= && (3.17)\\
&= C_1\mathbb{E}[\epsilon_{t-lag}\epsilon_{t-lag}^2] + C_2\mathbb{E}[\epsilon_{t-lag}h_{t-lag}] = 0, && (3.18)\\
& && (3.19)
\end{aligned}$$

for some constants $C_1, C_2 \in \mathbb{R}$.

$$\Rightarrow L(lag) \;=\; Corr(\epsilon_{t-lag}, \epsilon_t^2) = 0. \tag{3.20}$$

3.3. Maximum likelihood estimation

In this section we present the maximum likelihood procedure for the more general $GARCH(p,q)$ **regression model**, which is obtained by letting the ϵ_t's be innovations in a linear regression

$$r_t = b \cdot x_t + \epsilon_t \tag{3.21}$$

where r_t is the dependent variable (in our case the financial return), x_t a vector of l explanatory variables (e.g. past prices or past returns), and b an l-dimensional row vector of unknown parameters.

We follow the way of T. Bollerslev (1986) [B]: First of all, let us reformulate the considered model. We define $z_t := (1, \epsilon_{t-1}^2, \ldots, \epsilon_{t-q}^2, h_{t-1}, \ldots, h_{t-p})^T$, $\omega := (\alpha_0, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)$ and $\theta \in \Theta$, where $\theta := (b, \omega)$ and Θ is a compact subspace of an Euclidean space such that ϵ_t possesses finite second moments. Further, we call the true parameters θ_0, where $\theta_0 \in \text{int}\,\Theta$. Using these definitions the model has the following form:

$$r_t = b \cdot x_t + \epsilon_t, \quad \text{where } \epsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t) \quad \text{with } h_t = \omega \cdot z_t. \tag{3.22}$$

According to Hamilton (1994) [H], for given parameter values θ, estimates for the sequence of innovations $\{\epsilon_t\}_{t=1}^T$ can be inferred from the regression for the return in (3.22) immediately. In order to find with help of the estimated sequence of innovations the sequence of conditional variances $\{h_t\}_{t=1}^T$ from the second equation in (3.22), one needs pre-sample values for $h_{-p+1}, \ldots, h_0$ and $\epsilon_{-q+1}^2, \ldots, \epsilon_0^2$. Therefore, it is convenient to condition on $m = max(p,q)$ pre-sample values for both h_t and ϵ_t. Since we have observations on r_t and x_t for $t = 1, \ldots, T$, Bollerslev (1986) [B] suggested setting

$$h_j = \epsilon_j^2 = \hat{\sigma}^2 := \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2, \quad \text{for } j = -m+1, \ldots, 0. \tag{3.23}$$

The conditional distribution of r_t is Gaussian with mean $b \cdot x_t$ and variance h_t:

$$f(r_t | x_t, \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \cdot \exp\left(\frac{-(r_t - b \cdot x_t)^2}{2h_t}\right). \tag{3.24}$$

Hence, conditioning on m pre-sample values, the **log likelihood function** for a sample of T observations is

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log f(r_t|x_t, \mathcal{F}_{t-1};\theta) = \tag{3.25}$$

$$= -\frac{1}{2T}\sum_{t=1}^{T}\left[\log(2\pi) + \log(h_t) + \frac{\epsilon_t^2}{h_t}\right]. \tag{3.26}$$

Differentiating $l_t(\theta)$ with respect to the variance parameters ω yields

$$\frac{\partial l_t}{\partial \omega} = \frac{1}{2h_t}\frac{\partial h_t}{\partial \omega}\left(\frac{\epsilon_t^2}{h_t} - 1\right), \tag{3.27}$$

$$\frac{\partial^2 l_t}{\partial \omega\, \partial \omega^T} = \left(\frac{\epsilon_t^2}{h_t} - 1\right)\frac{\partial}{\partial \omega^T}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \omega}\right) - \frac{1}{2h_t^2}\frac{\partial h_t}{\partial \omega}\frac{\partial h_t}{\partial \omega^T}\frac{\epsilon_t^2}{h_t}, \tag{3.28}$$

where

$$\frac{\partial h_t}{\partial \omega} = z_t + \sum_{i=1}^{p}\beta_i \frac{\partial h_{t-i}}{\partial \omega}. \tag{3.29}$$

Differentiating $l_t(\theta)$ with respect to the mean parameters b yields

$$\frac{\partial l_t}{\partial b} = \frac{\epsilon_t x_t}{h_t} + \frac{1}{2h_t}\frac{\partial h_t}{\partial b}\left(\frac{\epsilon_t^2}{h_t} - 1\right), \tag{3.30}$$

$$\frac{\partial^2 l_t}{\partial b\, \partial b^T} = -\frac{x_t x_t^T}{h_t} - \frac{1}{2h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial b^T}\frac{\epsilon_t^2}{h_t} - \frac{\epsilon_t x_t}{h_t^2}\frac{\partial h_t}{\partial b} + \left(\frac{\epsilon_t^2}{h_t} - 1\right)\frac{\partial}{\partial b^T}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial b}\right), \tag{3.31}$$

where

$$\frac{\partial h_t}{\partial b} = -2\sum_{j=1}^{q}\alpha_j x_{t-j}\epsilon_{t-j} + \sum_{j=1}^{p}\beta_j \frac{\partial h_{t-j}}{\partial b}. \tag{3.32}$$

In order to have all derivatives of the Hessian to calculate the **Fisher-Information matrix**

$$\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_{bb} & \mathcal{I}_{b\omega} \\ \mathcal{I}_{\omega b} & \mathcal{I}_{\omega\omega} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial l_t(\theta)}{\partial \theta}\frac{\partial l_t(\theta)}{\partial \theta^T}\right] = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial^2 l_t(\theta)}{\partial \theta\, \partial \theta^T}\right] \tag{3.33}$$

we finally need to compute the mixed derivative of $l_t(\theta)$:

$$\frac{\partial^2 l_t}{\partial b\, \partial \omega^T} = -\frac{\epsilon_t x_t}{h_t^2}\frac{\partial h_t}{\partial \omega^T} + \frac{\partial}{\partial \omega^T}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial b}\right)\left(\frac{\epsilon_t^2}{h_t} - 1\right) - \frac{1}{2h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial \omega^T}\frac{\epsilon_t^2}{h_t} \tag{3.34}$$

Taking conditional expectations of the Hessian $\left(\mathbb{E}\left[\frac{\partial^2 l_t(\theta)}{\partial \theta\, \partial \theta^T}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\partial^2 l_t(\theta)}{\partial \theta\, \partial \theta^T}|\mathcal{F}_{t-1}\right]\right]\right)$ and using that h_t is $\mathcal{F}_{t-1}$-measureable, $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\epsilon_t^2|\mathcal{F}_{t-1}] = h_t$, we find the following blocks in the

Fisher-Information:

$$\mathcal{I}_{bb} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{x_t x_t^T}{h_t} + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial b^T}\right] \tag{3.35}$$

$$\mathcal{I}_{\omega\omega} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \omega}\frac{\partial h_t}{\partial \omega^T}\right] \tag{3.36}$$

$$\mathcal{I}_{b\omega} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial b}\frac{\partial h_t}{\partial \omega^T}\right] \tag{3.37}$$

Consistent estimates for $\mathcal{I}_{bb}$ and $\mathcal{I}_{\omega\omega}$ are given by the sample analogues. These estimates involve first derivatives only. Furthermore, the elements in the off-diagonal block $\mathcal{I}_{b\omega}$ are zero (see Engle (1982) [E] for a proof). This finding is of great importance. It means that b and ω are asymptotically independent and thus, they can be estimated separately without loss of asymptotic efficiency. Their variances can be calculated separately as well. Nevertheless, either can be estimated with full efficiency based only on a consistent estimate of the other.

Note, the only difference to the estimation of an $ARCH(q)$ regression model is an additional recursive term appearing in equation (3.29) and (3.32) (compare Engle (1982) [E]). But this complicates the procedure. For the $GARCH(p,q)$ regression model the **Berndt, Hall, Hall and Hausman (1974) algorithm** [BHHH] is an appropriate method to find second-order efficient maximum likelihood estimates:

Let us call the estimates for the parameters θ after i iterations $\theta^{(i)}$. Then, one can iteratively compute the subsequent estimate $\theta^{(i+1)}$ by

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t}{\partial \theta}\frac{\partial l_t}{\partial \theta^T}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t}{\partial \theta}\right), \tag{3.38}$$

where $\frac{\partial l_t}{\partial \theta}$ is evaluated at $\theta^{(i)}$. The parameter λ_i is a variable step length chosen to maximize the likelihood function in the given direction. The direction vector can easily be calculated from a regression of a $T \times 1$ vector of ones on $\frac{\partial l_t}{\partial \theta}$.

As already mentioned above, we perform the iterations for $\omega^{(i)}$ and $b^{(i)}$ separately due to the block diagonality in the Fisher-information. It follows from Weiss (1982) that (under certain regularity conditions) the maximum likelihood estimates $\hat{b}_T$ and $\hat{\omega}_T$ are strongly consistent for b_0 and ω_0 and asymptotically normal with limiting distributions

$$\sqrt{T}(\hat{b}_T - b_0) \xrightarrow{D} N_l(0, \mathcal{I}_{bb}^{-1}), \tag{3.39}$$

$$\sqrt{T}(\hat{\omega}_T - \omega_0) \xrightarrow{D} N_{p+q+1}(0, \mathcal{I}_{\omega\omega}^{-1}). \tag{3.40}$$

Consistent estimates for the asymptotic covariance matrices can be found by $\hat{\mathcal{I}}_{bb}^{-1} = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t}{\partial b}\frac{\partial l_t}{\partial b^T}\right)^{-1}$ and $\hat{\mathcal{I}}_{\omega\omega}^{-1} = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t}{\partial \omega}\frac{\partial l_t}{\partial \omega^T}\right)^{-1}$ from the last iteration of the BHHH algorithm.

As we have seen, GARCH is a time series modeling technique that uses past innovations and past variances to forecast future variances. It considers heavy tails and volatility clustering, which are two of the most important characteristics of financial time series (see Chapter 2). It provides accurate forecasts of variances and covariances of asset returns through its ability to model time-varying conditional variances.

Despite the broad applications of GARCH in many fields of finance (e.g. risk management, portfolio management, foreign exchange), it does have limitations: Although it is explicitly designed to model time-varying conditional volatility, it works best in relatively stable markets. Irregular phenomena, meaning wild market fluctuations and other highly unanticipated events (e.g. bubbles or crashes), are even with GARCH hardly predictable. Moreover, GARCH often fails to fully model the heavy tails of the unconditional return distribution. The included phenomenon of heteroskedasticity explains part of the fat tail behavior, but typically not all of it. Leptokurtic distributions, such as student-t, have been used to model GARCH innovations, but often the choice of distribution is a matter of trial and error (The MathWorks (2000) [M2]).

4 Testing for Dependences in Bubble Periods

4.1. Empirical autocorrelation function

As we have seen in the analysis of typical financial time series in Chapter 2, daily returns do exhibit little or even insignificant autocorrelation in general. In this section we analyze some historical asset bubbles for their autocorrelation at lag 1 during times of large gains. In contrast to our analysis in Chapter 2, we apply an exponentially weighted moving average to the returns. This leads to significantly higher autocorrelation at lag 1 during times of well-known bubble periods.

To start off, let us consider a time series of daily prices $P_0, \ldots, P_n$. The corresponding daily returns can be computed as

$$r_t := \log\left(\frac{P_t}{P_{t-1}}\right), \quad \text{for } t = 1, \ldots, n. \tag{4.1}$$

Using the daily returns, one usually estimates the autocorrelation of returns with lag k as

$$\hat{\rho}_k := \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, \quad \text{with } \hat{\gamma}_k := \frac{1}{n}\sum_{t=1}^{n-k}(r_{t+k} - \bar{r})(r_t - \bar{r}) \quad \text{and} \quad \bar{r} := \frac{1}{n}\sum_{t=1}^{n} r_t. \tag{4.2}$$

In contrast to the common procedure, we average the daily returns over a horizon τ as follows:

$$r_t^{EWMA} := EWMA(r, \tau, \alpha)_t := \frac{1-\delta}{1-\delta^\tau} \cdot \sum_{k=0}^{\tau-1} \delta^k \cdot r_{t\cdot\tau-k}, \quad \text{for } t = 1, \ldots, \left[\frac{n}{\tau}\right] =: T, \tag{4.3}$$

where $\delta = e^{-\alpha}$ and $\alpha > 0$ constant.

Now, we are able to estimate for various time windows $(t_{start}, \ldots, t_{end})$, with $t_{start},\ t_{end} \in \{1, \ldots, T\}$ and $t_{start} < t_{end}$, the **autocorrelation of EWMA returns**:

$$\hat{\rho}_k^{EWMA} := \frac{\hat{\gamma}_k^{EWMA}}{\hat{\gamma}_0^{EWMA}}, \quad \text{with} \quad \hat{\gamma}_k^{EWMA} := \frac{1}{T}\sum_{t=1}^{T-k}(r_{t+k}^{EWMA} - \bar{r}^{EWMA})(r_t^{EWMA} - \bar{r}^{EWMA}) \tag{4.4}$$

$$\text{and} \quad \bar{r}^{EWMA} := \frac{1}{T}\sum_{t=1}^{T} r_t^{EWMA}. \tag{4.5}$$

Let us have a look at some empirical results for the DAX, Gold and the NASDAQ Composite (Figures 4.1-4.3).

We detect quite a high empirical autocorrelation at lag 1 of 0.3-0.4 for the DAX time series if we choose a window starting in 1995 and ending either in 1998 or 2000. At these two end dates there are exactly the sharpest peaks in the DAX price evolution. The bubble in 2000 lets us find such a high empirical ACF even for multiple windows starting between 1995 and 1997. Furthermore, the crash in 2011 was preceded by an ACF of 0.4-0.5 over a window starting at the trough in 2009.

Especially in the years 2009-2012 Gold exhibited a high ACF at lag 1 which obviously corresponds to the continuous rising of the price during the aftermath of the crisis. For time windows starting in 2007 one can find values of 0.3-0.4 for the empirical ACF. Interestingly, looking at the peak in 2006, we computed the highest values for windows ending a few months before it. The reason is a period of stable prices shortly before the bubble peak. Also, around the peak in 2008 the autocorrelation was relatively low compared to the other maxima.

Analyzing the NASDAQ Composite time series, we find a strong empirical autocorrelation at lag 1 of even 0.4-0.5 for windows ending around the bubble peak in 2000. Moreover, we detect this high ACF for the majority of starting dates between the crash in 1987 and the year 1998 which emphasizes the strength of the signal even further. Also, we find quite a large empirical ACF of 0.3-0.4 for a window of 1 year starting before the crash in 1987.

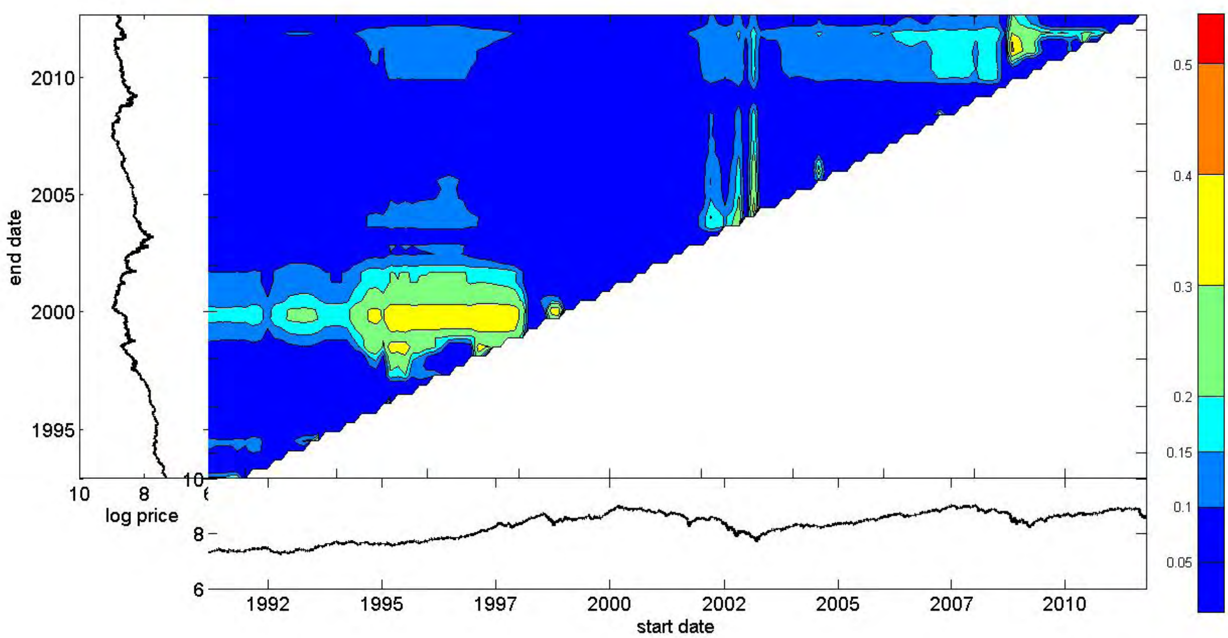


Figure 4.1.: DAX 1991-2011: Empirical ACF at lag 1 of EWMA returns over a horizon of $\tau = 40$ days computed over different time windows

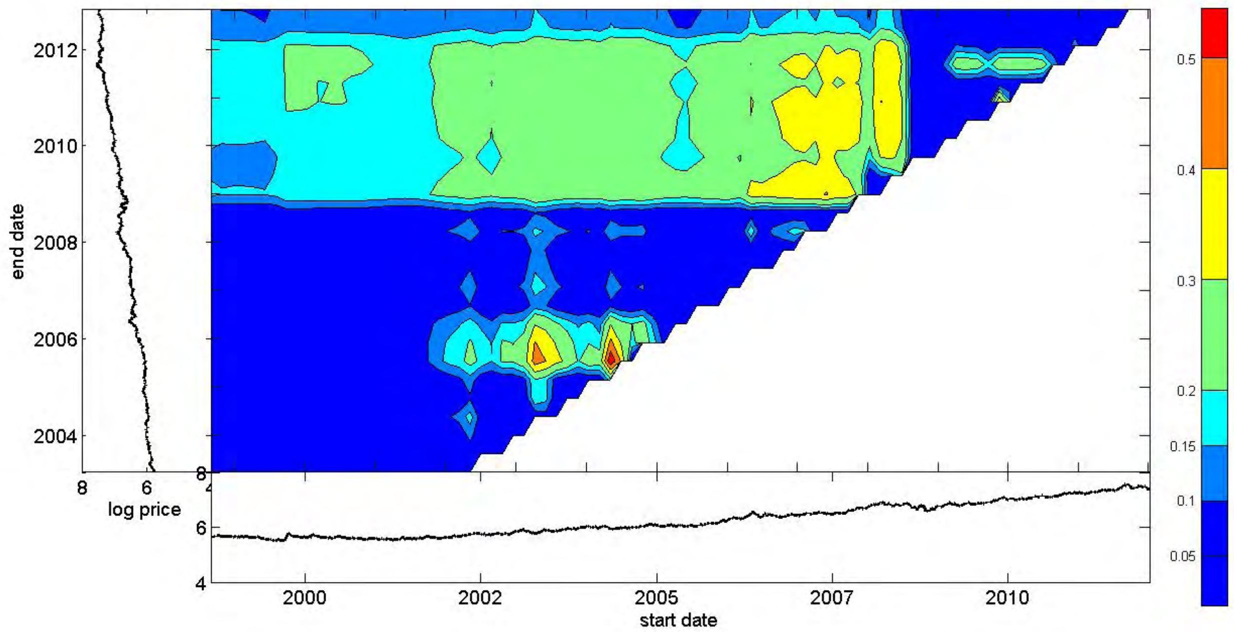


Figure 4.2.: Gold 1999-2012: Empirical ACF at lag 1 of EWMA returns over a horizon of $\tau = 40$ days computed over different time windows

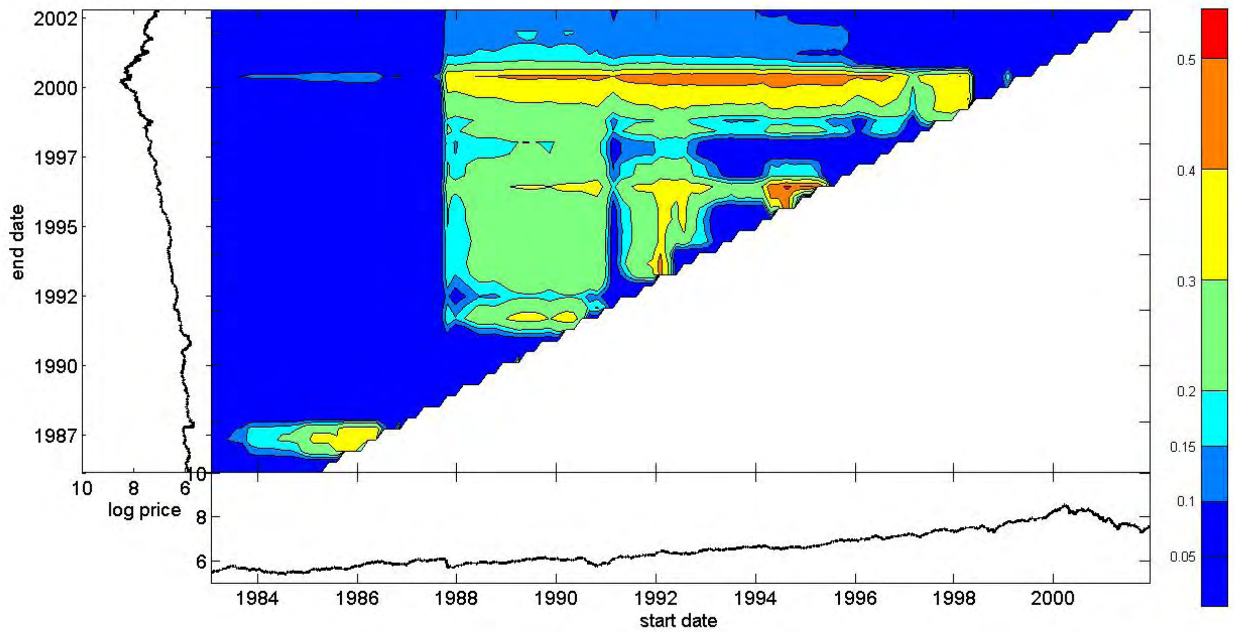


Figure 4.3.: NASDAQ Composite 1983-2002: Empirical ACF at lag 1 of EWMA returns over a horizon of $\tau = 40$ days computed over different time windows

5 Finite-Time Singularity (FTS) GARCH

5.1. A bubble experiment in the laboratory

Our starting point is the paper "Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price" (2012) [HSH] where Huesler, Sornette and Hommes analyze the dynamics of bubbles generated in a controlled price formation experiment in the laboratory by Hommes et al. in 2008 [HSTV]. In this bubble experiment 36 participants were split into 6 different groups trading on 6 different assets. Each "trader" had to forecast the price of an asset in every round. The market price P_t was then formed as an average of all traders' discounted price expectations ($H = 6$ participants per group):

$$P_t = \frac{1}{1+r}\left(\frac{1}{H}\sum_{h=1}^{H} P_{t+1}^{h} + D\right), \tag{5.1}$$

with an interest rate of $r = 5\%$, a dividend $D = 3$ and from each participant a price forecast P_{t+1}^{h} for time $t+1$ based only on information until time $t-1$. Traders were able to calculate the fundamental price $P^f = \frac{D}{r} = 60$ since they knew the interest and the dividend. They were rewarded in dependence on the precision of their price forecasts, but did not know how the market price was formed. Therefore, the reward depended more on one's forecast of the other participant's forecasts than on a precise estimate of the fundamental value which made the market price P_t loosely tied to the fundamental price P^f (expectations-driven price formation). In 5 of the 6 groups the market price built a bubble, it approached 1000 and subsequently dropped (recall $P^f = 60$!). In 4 of them this bubble seemed to burst due to the reason that price forecasts were restricted to be below 1000. Next, the

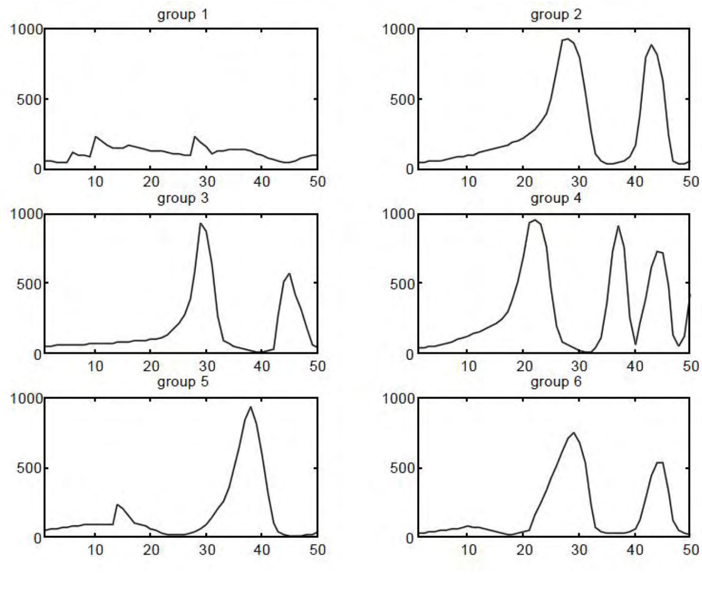


Figure 5.1.: Realized market prices from the bubble experiment for different groups (Figure from Hommes et al. (2008) [HSTV])

price fell even to the lowest value since the start, but then increased again and reached another peak. The authors interpret this as an indication that the dynamics in most of the groups were driven by the interaction between participants trying to extrapolate trends and the restrictions on the price forecasts (0 and 1000). Furthermore, they conclude that the bubbles in the experiment do not correspond to rational bubbles, they rather seem to be speculative bubbles driven by the predicition strategies of the traders (by analyzing the individual strategies within a group, one can see that traders coordinated on a common prediciton strategy). According to the ideas of Blanchard and Watson (1982) [BW], in a rational expectations bubble the price can deviate significantly from its fundamental value, however, it fulfills the condition of rational expectations (the best estimation for P_{t+1} at time t is $\mathbb{E}_t[P_{t+1}]$) and the no-arbitrage condition that the expected return of the asset is equal to the risk-free rate r. Any price process of the form ($R := 1 + r = 1.05$)

$$P_t = P^f + cR^t, \quad \text{with a constant } c \geq 0, \tag{5.2}$$

is a rational bubble satisfying

$$\mathbb{E}_t\left[\frac{P_{t+1} + D}{P_t}\right] = \frac{P^f + cR^{t+1} + rP^f}{P_t} = 1 + r \quad \forall t \in \mathbb{N}_0. \tag{5.3}$$

Hommes et al. (2008) [HSTV] found implied growth rates $\hat{R}$ that are significantly higher than 1.05 in 4 of the 6 groups (except for group 1 and 5). Hence, the trader's expectations are no longer rational (participants know r and P^f!). They seem to apply trend following investment strategies. When they observe small gains, they forecast the price to grow further which in turn makes this prediction to become self-fulfilling ("positive feedback expectations"). There is substantial evidence that real investors act similar as the participants in the experiment. Many professional traders use trend following strategies extrapolating current trends in market prices into the future.

5.2. Super-exponential growth and positive feedback on return

Huesler, Sornette and Hommes (2012) [HSH] show that the explosive increase in price during a bubble can not be described by an exponential growth, rather one needs to incorporate a positive feedback mechanism corresponding to even faster than exponential growth. There had already been papers providing empirical evidence that prices grow super-exponentially during bubble periods, but this paper was actually the first demonstrating it unambiguously in a controlled laboratory experiment. Huesler, Sornette and Hommes find strong evidence for super-exponential growth during bubble periods by fitting two models (in which traders anchor their price predictions either on price or on return) to the bubbles built in the experiment ($\bar{P}_t := P_t - P^f$ is the excess price):

(1) Anchoring on price:
The growth rate $r_t := \log(\frac{\bar{P}_t}{\bar{P}_{t-1}})$ is equal to a constant a plus b times the last excess price:

$$r_t = a + b \cdot \bar{P}_{t-1} \tag{5.4}$$

For $b = 0$ the corresponding price grows exponentially (rational bubble), but if $a > 0$ and $b > 0$ it grows even super-exponentially (positive feedback of price on return):

$$\bar{P}_t = \bar{P}_{t-1} \cdot e^{a+b\cdot\bar{P}_{t-1}} \tag{5.5}$$

(2) Anchoring on return:
Alternatively, one can also test for positive feedback of return on return:

$$r_t = a + b \cdot r_{t-1} \tag{5.6}$$

Again, if $a > 0$ and $b > 0$ there is faster than exponential growth in the excess price.

In addition, we would like to have a closer look at the explicit form for r_t in (5.6) and its asymptotics:

$$r_t = a + br_{t-1} = a + ab + b^2 r_{t-2} = ... = a(1 + b + b^2 + ... + b^{t-2}) + b^{t-1} r_1. \tag{5.7}$$

Thus, in case $a > 0$ and $b > 1$ r_t grows exponentially and the corresponding price P_t grows super-exponentially. By applying the formula for a finite geometric series we find for the case $a > 0$ and $0 < b < 1$

$$r_t = a\frac{1 - b^{t-1}}{1 - b} + b^{t-1} r_1 \quad \longrightarrow \quad \frac{a}{1-b} \quad \text{as} \quad t \longrightarrow \infty. \tag{5.8}$$

Thus, the return is asymptotically constant. Nevertheless, one can have faster than expontential growth even for $0 < b < 1$. Looking at the differences in return from (5.7):

$$r_t - r_{t-1} = b^{t-2}(r_2 - r_1) = \begin{cases} > 0, & \text{if } r_2 > r_1, \\ \leq 0, & \text{if } r_2 \leq r_1. \end{cases} \tag{5.9}$$

This implies the price process P_t grows super-exponentially (although the returns grow in a decelerating fashion) for $a > 0$ and $0 < b < 1$ if $r_2 > r_1$.
Now, we come back to the results of Huesler, Sornette and Hommes (2012) [HSH] in their analysis of the bubble experiment. They estimate the coefficients a and b in (5.4) and (5.6) over different windows of the bubbles in group 2, 3, and 4. Then, they compare the lower 95% confidence intervals for the null hypothesis that a, respectively, b is zero.
To sum up, they find evidence for positive feedback on price (equation 5.4) for windows ending around the bubble peak or shortly before. For smaller end dates they detect b around 0, implying exponential growth in the initial phase of the bubble. a is significantly greater than 0 for almost all time windows.
For the analysis of positive feedback on return (equation 5.6) the picture is not that clear. The windows where $a > 0$ and $b > 0$ are only those with the earliest starting dates (largest time windows) and, interestingly, all possible end dates, not only the ones around the bubble peak (we will find similar results for the coefficient of the regression component in an *FTS GARCH* model later in the thesis). They also detect high significance for $b > 0$ for windows starting later and ending around the bubble peak, but for these windows a is around 0. They conclude that traders seem to anchor their expectations more on price rather than on return (in the latter case the signal for jointly positive a and b is relatively small).

5.3. Finite-time singularities and the log-periodic power law ($LPPL$)

As we have seen in the analysis of the laboratory experiment, bubbles can be characterized by super-exponential growth in market price due to positive feedback effects. Generalizing the standard mathematical models by incorporating positive feedback loops let's them exhibit a **finite-time singularity**. Therefore, the solutions of these models exist only over a finite time horizon. As the singularity is approached, new mechanisms come into play leading to new price dynamics (e.g. a crash), a change of regime occurs (Corsi and Sornette (2012) [CS], Kaizoji and Sornette (2009) [KS]). Typically, economists avoid to work with models where there does not exist a solution for all times, but actually this captures quite well the transient nature of bubbles. According to Johansen and Sornette (1999) [JS], the simplest formula to describe the price in a super-exponential bubble is a power law with finite-time singularity at some critical time $t_c > 0$:

$$P_t = A + B(t_c - t)^z, \quad \text{where } A, B \text{ are constant and } z < 0. \tag{5.10}$$

Such an expression can be obtained as solution of models incorporating positive feedback, namely the typical population growth models. Cohen and others idealized the logistic equation

$$\frac{dP(t)}{dt} = R \cdot P(t)(K - P(t)), \tag{5.11}$$

where R is a constant and K is a finite carrying capacity, by letting K be time-dependent instead: $K(t)$ is assumed to increase with $P(t)$ due to technological progress, which means that for $K(t) > P(t)$ the limiting factor $-P(t)$ can be dropped. Under the assumption of a power law relation $K \propto P^\delta$, for a $\delta > 1$, their model has the following form:

$$\frac{dP(t)}{dt} = R \cdot P(t)^{1+\delta}. \tag{5.12}$$

Due to the acceleration in growth according to a power law (corresponding to super-exponential growth), the solution exhibits a finite-time singularity (Corsi and Sornette (2012) [CS]):

$$P(t) = P(0)(1 - \frac{t}{t_c})^{-1/\delta}, \tag{5.13}$$

where one can find the critical time t_c to be

$$t_c = \frac{1}{\delta R} P(0)^{-1/\delta}. \tag{5.14}$$

One obtains the standard exponential solution $P(t) = P(0)e^{Rt}$ as $\delta \to 0^+$, which can be proved by a Taylor approximation:

$$(1 - \frac{t}{t_c})^{-1/\delta} = e^{-1/\delta \, \cdot \, \log(1-t/t_c)} \approx e^{1/\delta \, \cdot \, t/t_c} \overset{\delta \to 0^+}{\longrightarrow} e^{Rt} \quad \text{for } \; t \ll t_c \; \to \infty. \tag{5.15}$$

Johansen and Sornette (2008) [JS] fit formula (5.10) to empirical data for super-exponential bubbles. They find the acceleration in growth to be well described, but the simple power fit fails to determine the critical time t_c precisely. There exists great variability around the average power law behavior, which brings them to the idea of a generalisation to power laws with complex exponents $z := \beta + i\omega$. The real part of $(t_c - t)^{\beta + i\omega}$ has the form

$$\mathrm{Re}((t_c - t)^{\beta + i\omega}) = \mathrm{Re}(e^{(\beta + i\omega)\log(t_c - t)}) = (t_c - t)^\beta \cos(\omega \log(t_c - t)). \tag{5.16}$$

This leads to the so-called **log-periodic power law (LPPL)** modeling faster-than-exponential price growth:

$$P_t \approx A + B(t_c - t)^{\beta} + C(t_c - t)^{\beta}\cos(\omega \log(t_c - t) + \phi), \tag{5.17}$$

with an additional constant C and another phase constant ϕ. The cosine leads to log-periodic oscillations (periodic in $\log(t_c - t)$ instead of t) around the average power law behavior that accelerate before the critical date. The time intervals between consecutive local maxima of the price tend to zero as the critical time is approached (with a constant ratio of consecutive intervals $\lambda = e^{2\pi/\omega}$). The authors find that these oscillations succeed to capture a large part of the variability about the power law growth in developing bubbles in a variety of markets. Including log-periodic corrections provides a better parametrization of the data and better constraints on β and t_c.

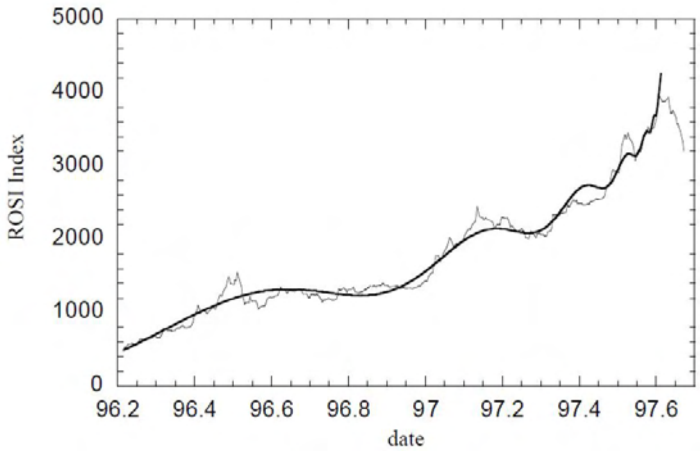


Figure 5.2.: LPPL model fit for the ROSI index: $A \approx 4254$, $B \approx -3166$, $C \approx 246$, $\beta \approx 0.4$, $t_c = 97.61$, $\phi \approx 0.44$ and $\omega \approx 7.7$) (Example and Figure from Johansen, Sornette and Ledoit (1999) [JSL]).

Johansen and Sornette (2008) [JS] propose also a nonlinear log-periodic model which expands to the next order of corrections to the power law. The interested reader is referred to the studies of D. Sornette et al. (e.g. in [JS] or [S2]). Furthermore, Johansen, Sornette and Ledoit (1999) [JSL] propose to model the price evolution as a jump diffusion process, embedding the $LPPL$ model in a stochastic framework (**JLS model**). This goes beyond the objective of the current chapter, however, we would like to give an interesting summary in Appendix A.

5.4. Introducing the *FTS GARCH* model

In this section our goal is to develop a suitable tool that analyzes financial time series for the presence of bubble-like dynamics. As we have seen so far, financial bubbles exhibit high autocorrelation of returns (compare Chapter 4) and grow even faster than exponential. This can be captured in a mathematical model by incorporating positive feedback of either price on return or return on return leading to a finite-time singularity. Inspired by the ideas of F. Corsi and D. Sornette (2012) [CS], we would like to combine this bubble dynamics with the standard of the financial industry to model volatility clustering and to assess volatiliy risk, namely the $GARCH(1,1)$ model. The resulting stochastic model is the so called **Finite-Time Singularity GARCH** model. This is the standard $GARCH(1,1)$ enhanced by adding a regression component y_t (e.g. last price or last return) in the conditional mean modeling positive feedback:

$$r_t = \mu + \gamma \cdot y_{t-1} + \epsilon_t, \quad \text{where } \epsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t), \tag{5.18}$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \tag{5.19}$$

where the daily continuous return r_t is computed from a time series of daily prices $P_0, \dots, P_n$ as

$$r_t := \log\left(\frac{P_t}{P_{t-1}}\right), \quad \text{for } t = 1, \dots, n =: T. \tag{5.20}$$

Note that this is a $GARCH(1,1)$ regression model as defined in (3.22), where $b := (\mu, \gamma)$ and $x_t := (1, y_{t-1})^T$. In Section 3.3 we already presented the full maximum likelihood procedure to consistently estimate the parameters of this model. In its original form the $FTS\ GARCH$ model is proposed with regression component $y_t := P_t$ (last price). Recall that according to Huesler, Sornette and Hommes (2012) [HSH], traders seem to anchor their expectations more on price rather than on return. Nevertheless, the price series is highly non-stationary which is definitely an undesirable property as we already pointed out at the beginning of Chapter 2. On account of this we propose to use the logarithm of price instead. This should manage to reduce the non-stationarity problem, although the log price is in general non-stationary as well. We will also try to find improved versions of using daily returns as regression component and test all these different regression components. However, the two main models we will have a closer look at are:

(1) Using the logarithm of last price as regression component ($y_t = \log(P_t)$):

$$r_t = \mu + \gamma \cdot \log(P_{t-1}) + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \tag{5.21}$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \quad t = 1, \dots, T. \tag{5.22}$$

(2) Using last return as regression component ($y_t = r_t$):

$$r_t = \mu + \gamma \cdot r_{t-1} + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \tag{5.23}$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \quad t = 2, \dots, T. \tag{5.24}$$

Our main interest lies in the parameter γ since it represents the strength of the feedback. We propose a standard one-sided t-test of

$$H_0 : \gamma \leq 0 \text{ against } H_1 : \gamma > 0 \tag{5.25}$$

and interpret the rejection of the null hypothesis as evidence for an FTS dynamics (with positive γ), and hence, for the presence of a bubble (compare the analysis of the bubble experiment in Section 5.2). Actually, one could also test for $H_0 : \gamma = 0$, but we would like to have as alternative hypothesis only positive feedback (and no negative feedback).

Before we apply the *FTS GARCH* model to real data, we would like to analyze it in a few synthetic tests where we work with simulated data. The idea of these experiments is to get a feeling for the model itself and its ability to detect regimes of positive feedback of both log price on return and return on return.

Quality of the maximum likelihood estimation

In order to assess the quality of the maximum likelihood procedure for the estimation of the coefficients in an *FTS GARCH* model we propose the following synthetic test. We first of all estimate realistic parameters of an *FTS GARCH* model from empirical data (Nasdaq Composite). Then, we simulate an *FTS GARCH* process with the estimated coefficients (γ_{input} is the estimated γ) 1000 times and infer the coefficient of the regression component from the simulated data. Finally, we calculate the t-statistics of the test $H_0 :\ \gamma = \gamma_{input}$ against $H_1 :\ \gamma \neq \gamma_{input}$. If the ML estimation works appropriately, there should be at most 50 exceptions lying either below the 2.5% quantile or above the 97.5% quantile as well as only 10 outside the 0.5% and 99.5% quantile (due to the large sample size the t-quantile is almost the same as the normal quantile). Using log price as regression component, we find 45, respectively, 10 exceptions (see Figure 5.3). Using return as regression component, we obtain almost the same result (Figure 5.4: 46 t-statistics below the 2.5% or above the 97.5% quantile and 10 outside the 0.5% and 99.5% quantile).
Note, this test does not say that the *FTS GARCH* model describes the empirical data well, neither that the parameters inferred from empirical data are estimated appropriately. But what it does tell us is that at least for simulated data the used ML procedure estimates the coefficient γ properly (consistent with the theoretical result from Section 3.3).

How do simulated trajectories from an *FTS GARCH* model look like?

Another preliminary step to get a feeling for the *FTS GARCH* model is to look at some simulated trajectories and compare them to bubbles in real data. For this reason we estimate appropriate coefficients for a $GARCH(1,1)$ from the first 6000 prices of NASDAQ Composite between 1971 and 2000 as well as for an *FTS GARCH* model from the subsequent 1350 prices. Then, we simulate $GARCH(1,1)$ for 6000 days and finally, *FTS GARCH* dynamics until the price exceeds the maximum price of the NASDAQ time series. As regression component we use (1) log price and (2) daily returns for both estimation and simulation of the *FTS GARCH*. We find the simulated prices in Figure 5.5 and 5.6 (the corresponding log prices are plotted in 5.7 and 5.8) to look quite similar to the real time series of NASDAQ.

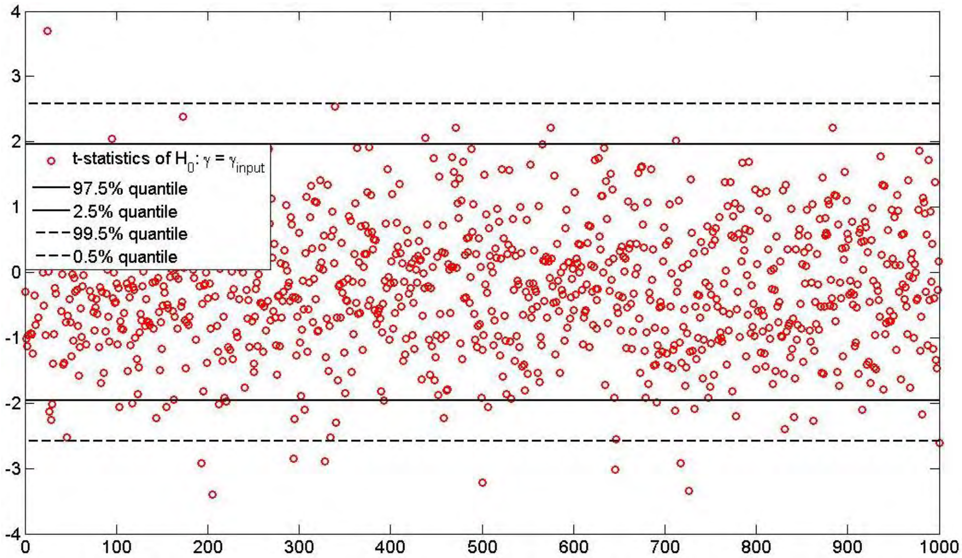


Figure 5.3.: *FTS GARCH* model ($y_t = \log(P_t)$): Goodness-of-fit test for synthetic data simulated with coefficients estimated from NASDAQ Composite (1982-2000) ($\gamma = 4 \cdot 10^{-4}$)

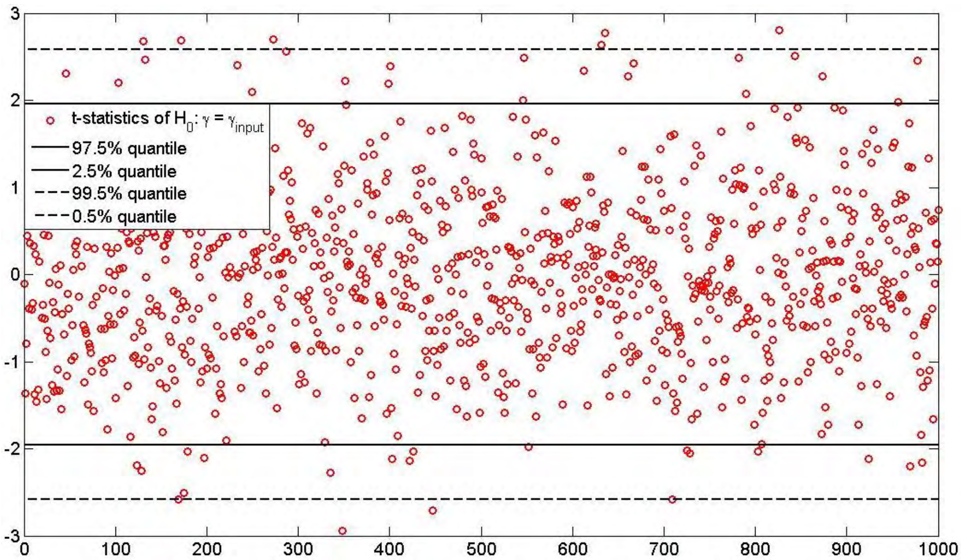


Figure 5.4.: *FTS GARCH* model ($y_t = r_t$): Goodness-of-fit test for synthetic data simulated with coefficients estimated from NASDAQ Composite (1982-2000) ($\gamma = 0.2$)

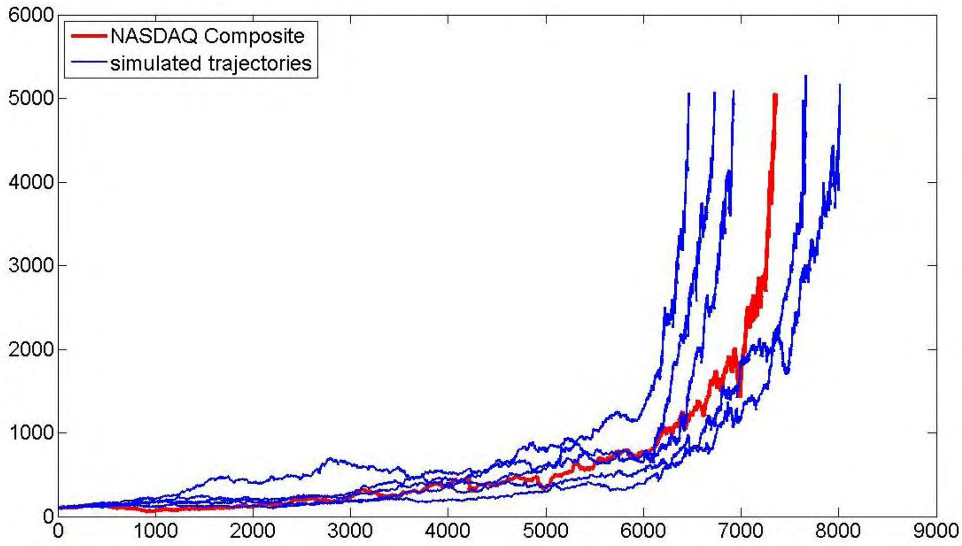


Figure 5.5.: Simulated trajectories of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ Composite, $\gamma = 2 \cdot 10^{-4}$) until the price level exceeds the maximum of NASDAQ

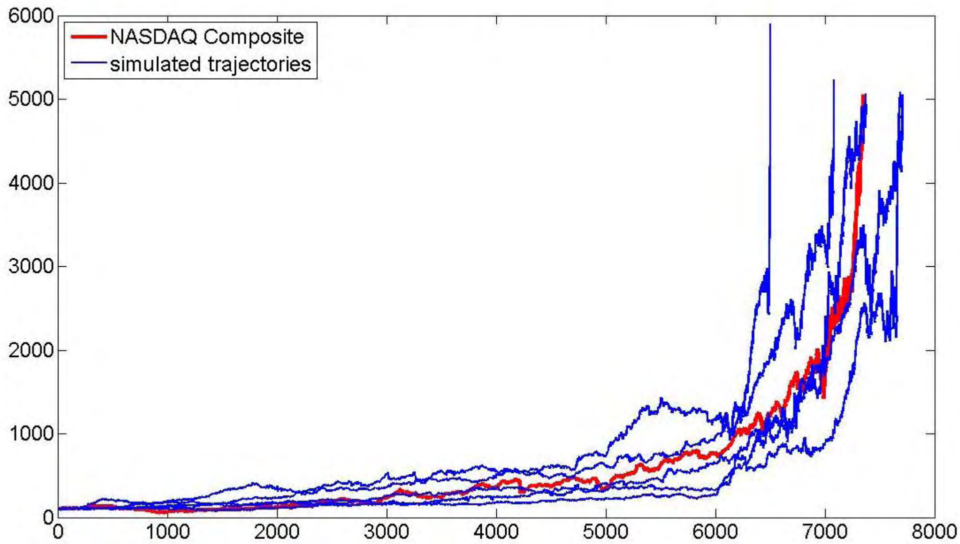


Figure 5.6.: Simulated trajectories of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics ($y_t = r_t$, coefficients estimated from NASDAQ Composite, $\gamma = 0.1$) until the price level exceeds the maximum of NASDAQ

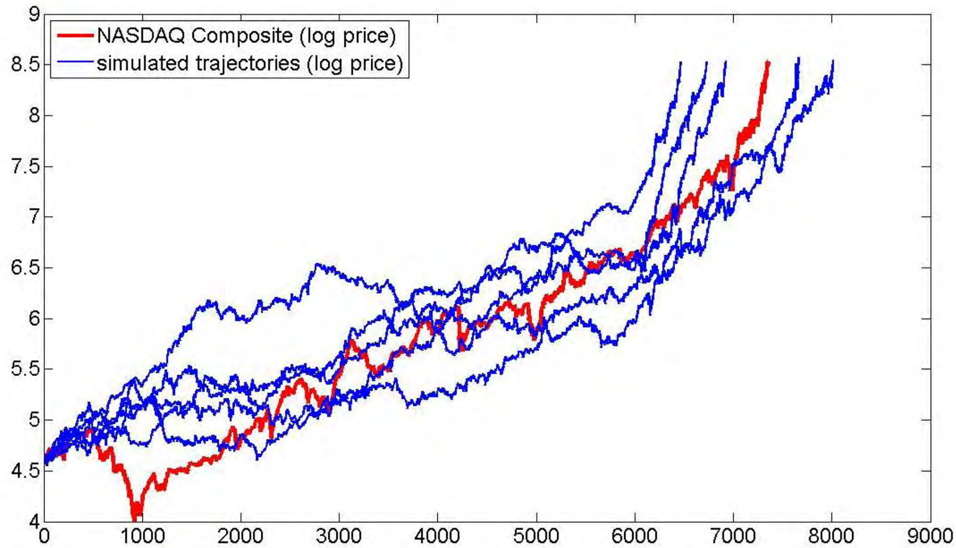


Figure 5.7.: Simulated trajectories (in log price) of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ Composite, $\gamma = 2 \cdot 10^{-4}$) until the price level exceeds the maximum of NASDAQ

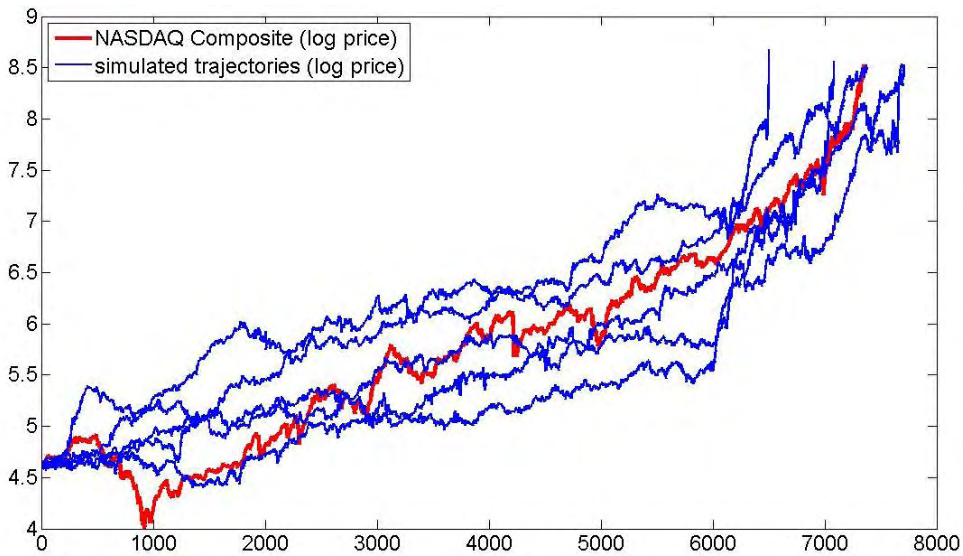


Figure 5.8.: Simulated trajectories (in log price) of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics ($y_t = r_t$, coefficients estimated from NASDAQ Composite, $\gamma = 0.1$) until the price level exceeds the maximum of NASDAQ

Can we detect a simulated bubble regime with $\gamma > 0$?

The main idea of the *FTS GARCH* model is to have a tool that enables us to detect positive feedback regimes where γ is greater than 0. Therefore, a basic simulation experiment that should definitely be done is to simulate two regimes, one where $\gamma = 0$ and another one where $\gamma > 0$, and look if the *FTS GARCH* model is able to detect the different regimes. So we simulate again trajectories of $GARCH(1,1)$ for 6000 days and then, *FTS GARCH* for another 1350 days (the fixed end date is the single difference to the simulation procedure from above) with (1) log price and (2) daily returns as regression component.
Now, we fit an *FTS GARCH* model to the daily return series calculated from various windows of each of the simulated price trajectories. We plot a rescaled version of the NASDAQ Composite price series between 1971 and 2000, the average of the 50 simulated trajectories at each point in time and finally, the average of the 50 t-statistics of the coefficient γ for each window size and various points in time, where we shift the window by 100 days each time. In order to reduce the computational effort the parameters are estimated only every 100 days and not each day.
In both cases (Figure 5.9 and 5.10) the t-statistic reacts immediately to the bubble regime and increases. If we use log prices as regression component, the t-statistic at the end of the simulation (where the price is at its highest level) is the higher the larger the chosen time window is. Interestingly, the t-statistic of γ decreases during the last days of the bubble period for windows of size 800 and 1350 days. This shows that the window needs to include part of the $GARCH(1,1)$ regime to yield high significance for a positive γ. Thus, this method is influenced mainly by the change of regime to a higher slope in log price around point 6000 but rarely by the positive feedback simulated in the last 1350 days per se (otherwise the t-statistic at the end would be higher for a window of 1350 days).
In this experiment the method of using return as regression component performs better. When $\gamma > 0$ the t-statistic is continuously increasing for all windows and reaches its highest value for the window equal to the length of the bubble period (1350 days). Hence, the *FTS GARCH* model (with $y_t = r_t$) reacts to a regime shift from $\gamma = 0$ to $\gamma > 0$ in simulated data.

How does the *FTS GARCH* model react to a regime with high μ but $\gamma = 0$ instead of $\gamma > 0$?

Let us further investigate the reasons for the evolution of the t-statistics in the last two simulation experiments. We propose to do another synthetic test where we simulate trajectories of 6000 days $GARCH(1,1)$ and then, again $GARCH(1,1)$ dynamics but with different coefficients (instead of *FTS GARCH* as usual) for another 1350 days. Subsequently, we estimate an *FTS GARCH* model from the simulated return series with (1) $y_t = \log(P_t)$ and (2) $y_t = r_t$. The results for the corresponding average t-statistics of γ are plotted in Figure 5.11 and 5.12.
This synthetic test confirms our recent finding for the usage of log price as regression component: The *FTS GARCH* model shows the same evidence for a bubble regime as before (compare Figures 5.9 and 5.11) although we simulated only exponential price growth in both regimes ($\gamma = 0$). Hence, the *FTS GARCH* model using log price as regression component signals super-exponential behavior as soon as there is an increase in the average return.
On the contrary, the *FTS GARCH* model with $y_t = r_t$ strictly distinguishes between a switch from $\gamma = 0$ to $\gamma > 0$ and an increase in μ. It does not signal evidence for faster than exponential growth in this simulation experiment, although there is a jump in the average return μ at point 6000.

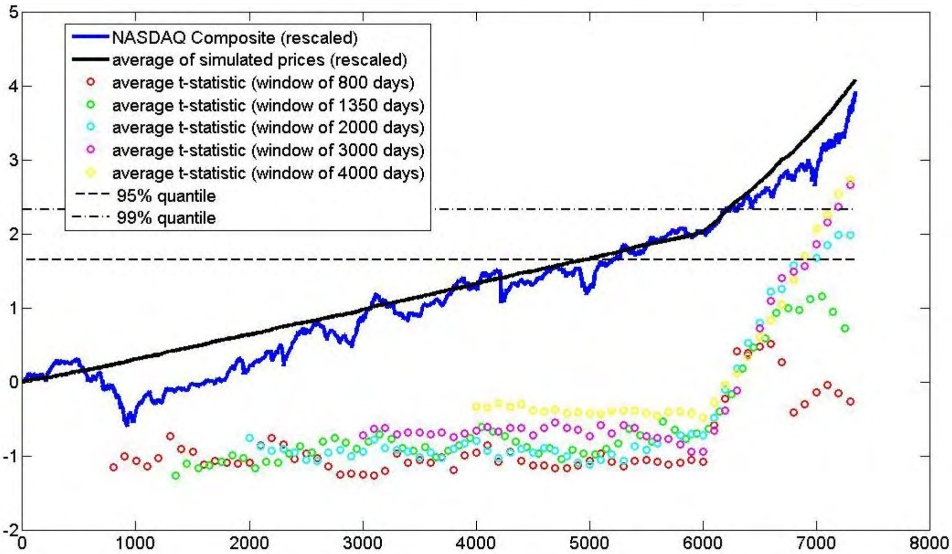


Figure 5.9.: MC Simulation (50 simulations): Sim. trajectories of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics with log price as regression component (coefficients estimated from NASDAQ, $\gamma = 6 \cdot 10^{-4}$) for another 1350 days. Estimation of an $FTS\ GARCH$ model ($y_t = \log(P_t)$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

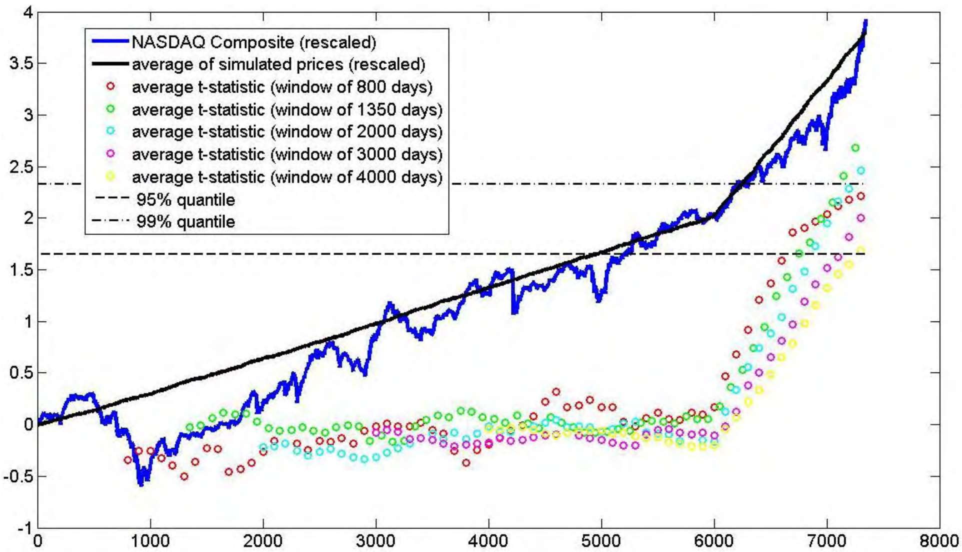


Figure 5.10.: MC Simulation (50 simulations): Sim. trajectories of 6000 days $GARCH(1,1)$ and then, $FTS\ GARCH$ dynamics with return as regression component (coefficients estimated from NASDAQ, $\gamma = 0.7$) for another 1350 days. Estimation of an $FTS\ GARCH$ model ($y_t = r_t$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

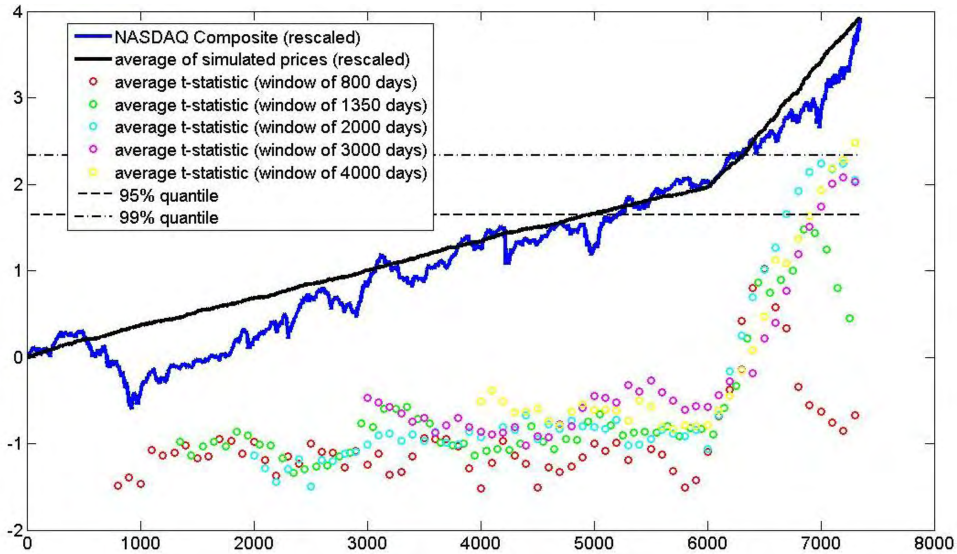


Figure 5.11.: MC Simulation (50 simulations): Sim. trajectories of 6000 days $GARCH(1,1)$ and then, again $GARCH(1,1)$ dynamics with different coefficients (both estimated from NASDAQ) for another 1350 days. Estimation of an $FTS\ GARCH$ model ($y_t = \log(P_t)$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

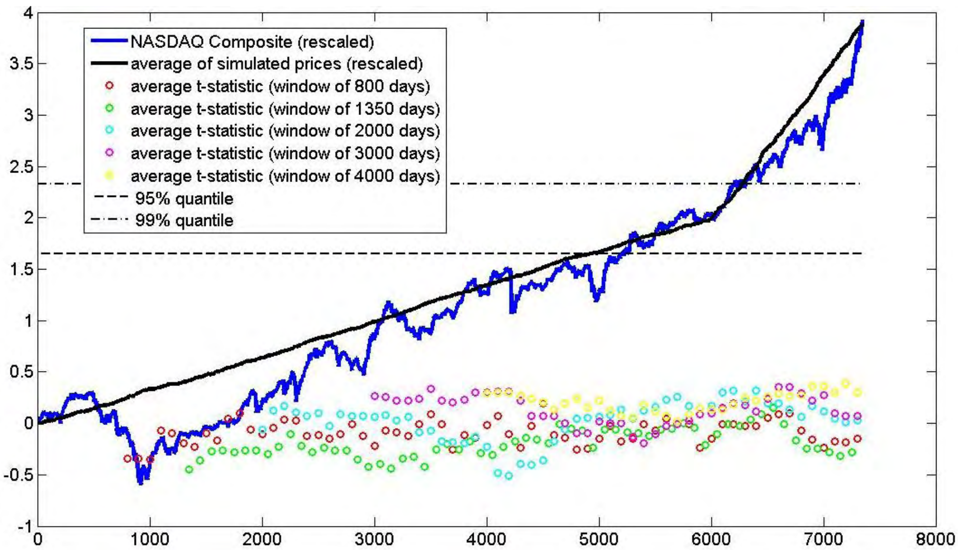


Figure 5.12.: MC Simulation (50 simulations): Sim. trajectories of 6000 days $GARCH(1,1)$ and then, again $GARCH(1,1)$ dynamics with different coefficients (both estimated from NASDAQ) for another 1350 days. Estimation of an $FTS\ GARCH$ model ($y_t = r_t$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

How does the t-statistic of the regression component in an $FTS\ GARCH$ model evolve if we include a period of alternating regimes ($GARCH(1,1)$ and $FTS\ GARCH$)?

So far we have always simulated 6000 days of a non-bubble regime followed by 1350 days of a bubble regime. Now, let us include a period of alternating bubble and non-bubble regimes where each of them lasts for a quarter (60 days). This means we simulate 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics for 3000 days and finally, a pure $FTS\ GARCH$ regime (where γ is chosen even higher than in the alternating regime) for another 350 days. We would like to know at which point in time the $FTS\ GARCH$ gives a signal for a bubble regime. Concerning the regression component, we do not only use (1) log price and (2) return for both simulation and estimation, but also mix the regression components (simulation and estimation are done with different regression components). The results of these four simulation experiments can be found in Figures 5.13 - 5.16.
Independent of the chosen regression component for simulation, if we use log price as regression component in the $FTS\ GARCH$ model used for estimation, we find the t-statistic to increase a bit in the period of alternating regimes and to increase rapidly as soon as the strong bubble regime at the end of the simulation starts. At first glance this evolution of the t-statistic looks quite good, but we should keep in mind we could find high significance for a positive γ estimated over windows reaching to the end of the simulated series also in case of exponential growth with high μ (compare with the synthetic test in Figure 5.11).
We also estimate an $FTS\ GARCH$ model from the same bubbles by using return as regression component and find quite different results depending on the chosen regression component for simulation. If we choose log price in the simulation step, we find only a pretty small increase in the evolution of the t-statistic of the coefficient of $y_t = r_t$ at the end of the simulation period. We do not detect a significantly positive γ (Figure 5.14). On the contrary, if we choose return in the simulation step, we detect high significance for a positive γ already in the alternating period (Figure 5.16). On the one hand this means the model shows a nice consistency, but on the other hand it gives us a bubble signal already in rather calm periods. We would prefer a lower signal / lower t-statistic in the alternating period becoming higher for the pure bubble regime at the end of the simulation, which then could serve as an early warning signal at the right time. Hence, depending on the type of simulated bubble (positive feedback of log price on return or return on return) we get either almost no signal or a very strong (in our case even undesirably strong) signal. This is one of the main drawbacks of using return as regression component in an $FTS\ GARCH$ model (we will find similar results for empirical data below).

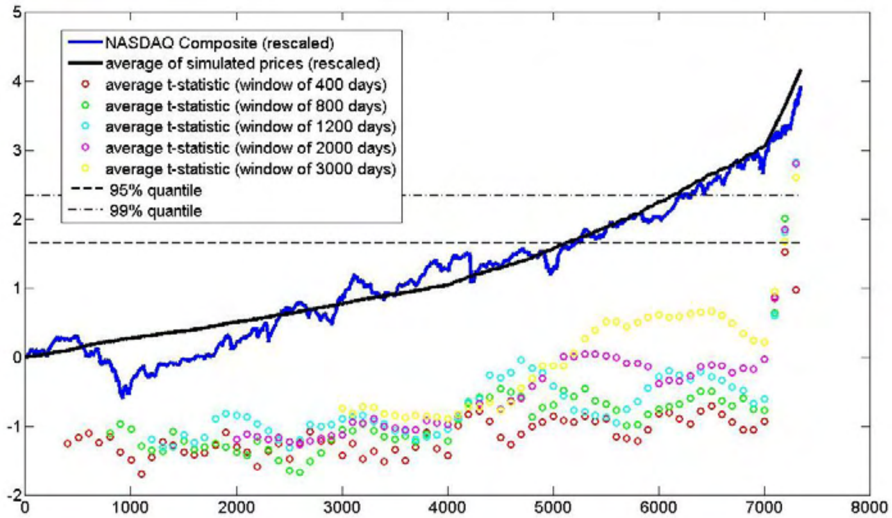


Figure 5.13.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an *FTS GARCH* model ($y_t = \log(P_t)$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

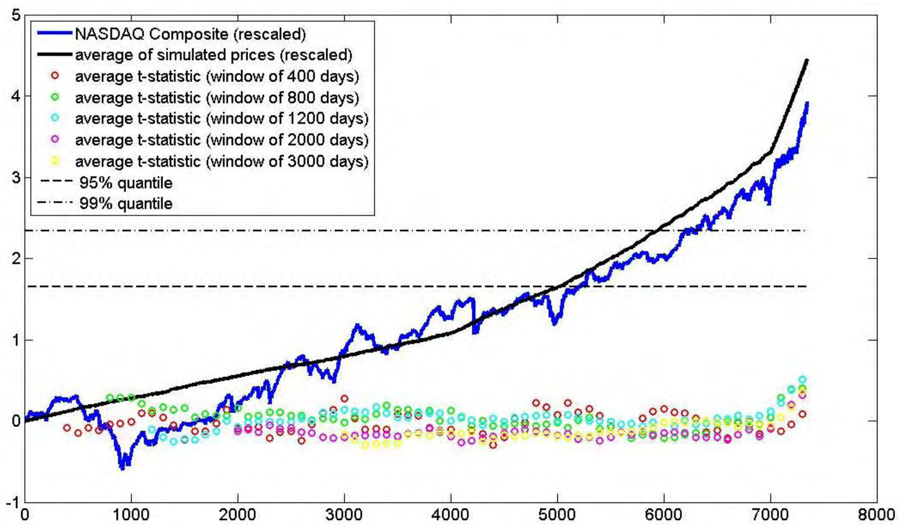


Figure 5.14.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an *FTS GARCH* model ($y_t = r_t$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

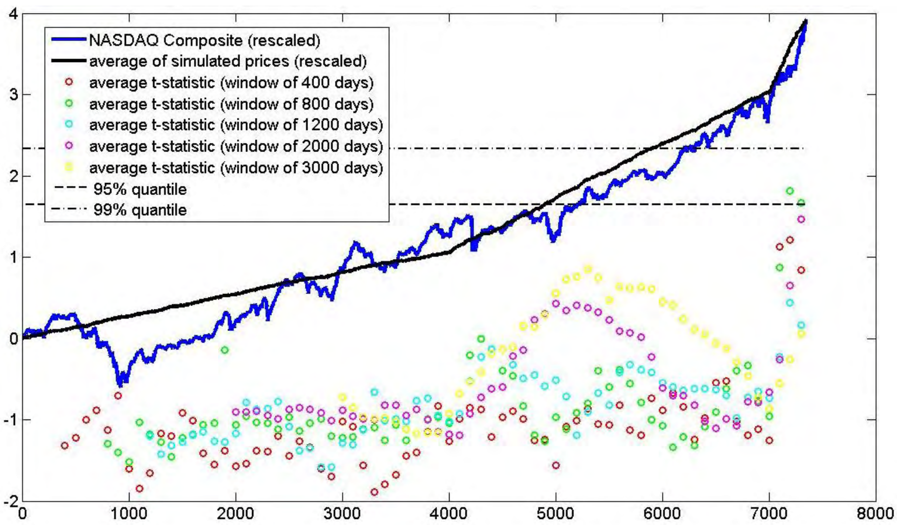


Figure 5.15.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an *FTS GARCH* model ($y_t = \log(P_t)$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

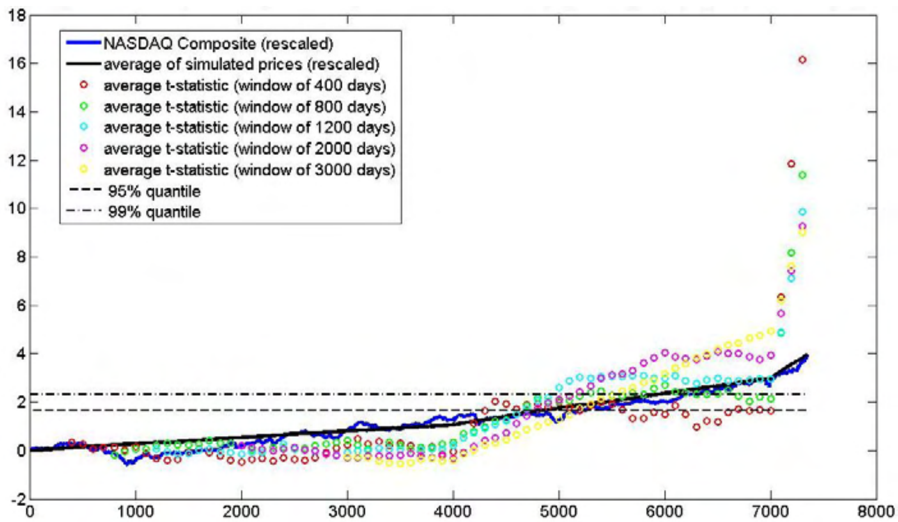


Figure 5.16.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an *FTS GARCH* model ($y_t = r_t$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

Empirical results:

We propose the following time window analysis to find windows where γ is significantly greater than 0 (evidence for a bubble regime):
Choose various time windows $(t_{start}, \dots, t_{end})$, with $t_{start},\ t_{end} \in \{1, \dots, T\}$ and $t_{start} < t_{end}$, and fit an *FTS GARCH* model to the corresponding return series $r_{t_{start}}, ..., r_{t_{end}}$:

$$r_t = \mu + \gamma \cdot y_{t-1} + \epsilon_t, \quad \text{where } \epsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t), \tag{5.26}$$
$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \quad t = t_{start} + 1, \dots, t_{end}. \tag{5.27}$$

Then, we plot the parameter γ and its t-statistic for the chosen windows in a heat color plot. On the x-axis we have the start date t_{start} and on the y-axis the end date t_{end} of the corresponding time window. Also, we add the logarithmic price series for both the start and the end dates. The numbers 1.28, 1.64, 1.96 and 2.32 are the 90%, 95%, 97.5% and 99% quantiles of a standard normal distribution. We analyze the time series of DAX, Gold, NASDAQ Composite and NIKKEI, where we use both log price and return as regression component.

Using $y_t = \log(P_t)$, the result for the DAX time series is quite nice (Figure 5.17). There is very high significance for a positive γ for windows ending around the bubble peaks in 1997, 1998 and 2000. The start dates leading to the highest t-statistics are in 1991 and 1994 where one can find local maxima in the price series.
Looking at the analysis of the Gold price series (Figure 5.18), we see that a start date around 1998 enables us to identify the maxima in 2006 and 2008 as bubbles. Also, there has been a strong signal for a positive feedback of log price on return for the last three years. Therefore, the *FTS GARCH* model with log price as regression component has suggested a bubble in the Gold price.
For NASDAQ Composite we find many different start dates indicating a bubble regime around the peak in 2000 (Figure 5.19). If one uses windows starting at the local maxima in 1981 or 1983, one finds evidence for a positive γ even for windows ending already in 1997.
Fitting an *FTS GARCH* model with log price as regression component to various time windows of the NIKKEI time series (Figure 5.20), one detects a positive feedback regime for windows ending between 1986 and 1992. This means the NIKKEI price series with its peak in 1990 leads to the strongest bubble signal compared to the other three assets we analyzed.

The results using an *FTS GARCH* model with return as regression component for the same time series (DAX, Gold, NASDAQ Composite and NIKKEI) are not that clear.
Let's start again with the analysis of the results for the DAX (Figure 5.21). For the peaks in 1997 and 1998 we can not find evidence for a bubble regime. At least for windows ending in 1999 and around the peak in 2000 there are bubble signals. The windows leading to the highest t-statistics are those starting either in 1991 or at the peaks in 1997 and 1998.
For the Gold price (Figure 5.22) we see even lower significance for a positive γ, where the window with the highest t-statistic is from 2004 to 2008.
The analysis of the NASDAQ Composite (Figure 5.23) and the NIKKEI time series (Figure 5.24) face a different problem. We observe the t-statistic of γ depending heavily on the size of the chosen window. We find very high significance for a positive coefficient for the majority of windows we are looking at. This is even a too strong signal we certainly do not want to have because we can not use it as an early warning signal at the right time. One idea to handle this strong signal is to modify

the test and analyze

$$H_0 : \gamma \leq 1 \text{ against } H_1 : \gamma > 1. \tag{5.28}$$

The rejection of H_0 would then signal a very strong explosive bubble since the return itself grows exponentially (recall the result from equation 5.7). But even for the time windows with the highest γ's we have only values around 0.3 at most (Figures 5.23 and 5.24 for γ). Hence, in general we will reject H_0 in 5.28 for all considered time windows.

In summary, we found positive feedback of log price on return for all four assets analyzed, whereas for many time windows of DAX and Gold the positive feedback of return on return is only low and very high for NASDAQ Composite and NIKKEI. Recall we found similar results for the estimation of an *FTS GARCH* model with return as regression component in the simulation experiments in Figures 5.14 and 5.16. Fitting an *FTS GARCH* model with $y_t = r_t$ to a simulated bubble where $y_t = \log(P_t)$ led to almost no significance for a positive γ, however, by using $y_t = r_t$ for the simulation we found strong evidence for a bubble already during the alternating period.
In the next section we try to overcome the above problem with three different methods that reduce the volatility of the return series. First of all, we apply a Kalman filter to the daily return series that finds an optimal estimate with minimal variance for the true state of the return process. Then, we also estimate the *FTS GARCH* model for returns averaged over longer than daily horizons. And finally, we use an exponentially weighted moving average of the returns as regression component.

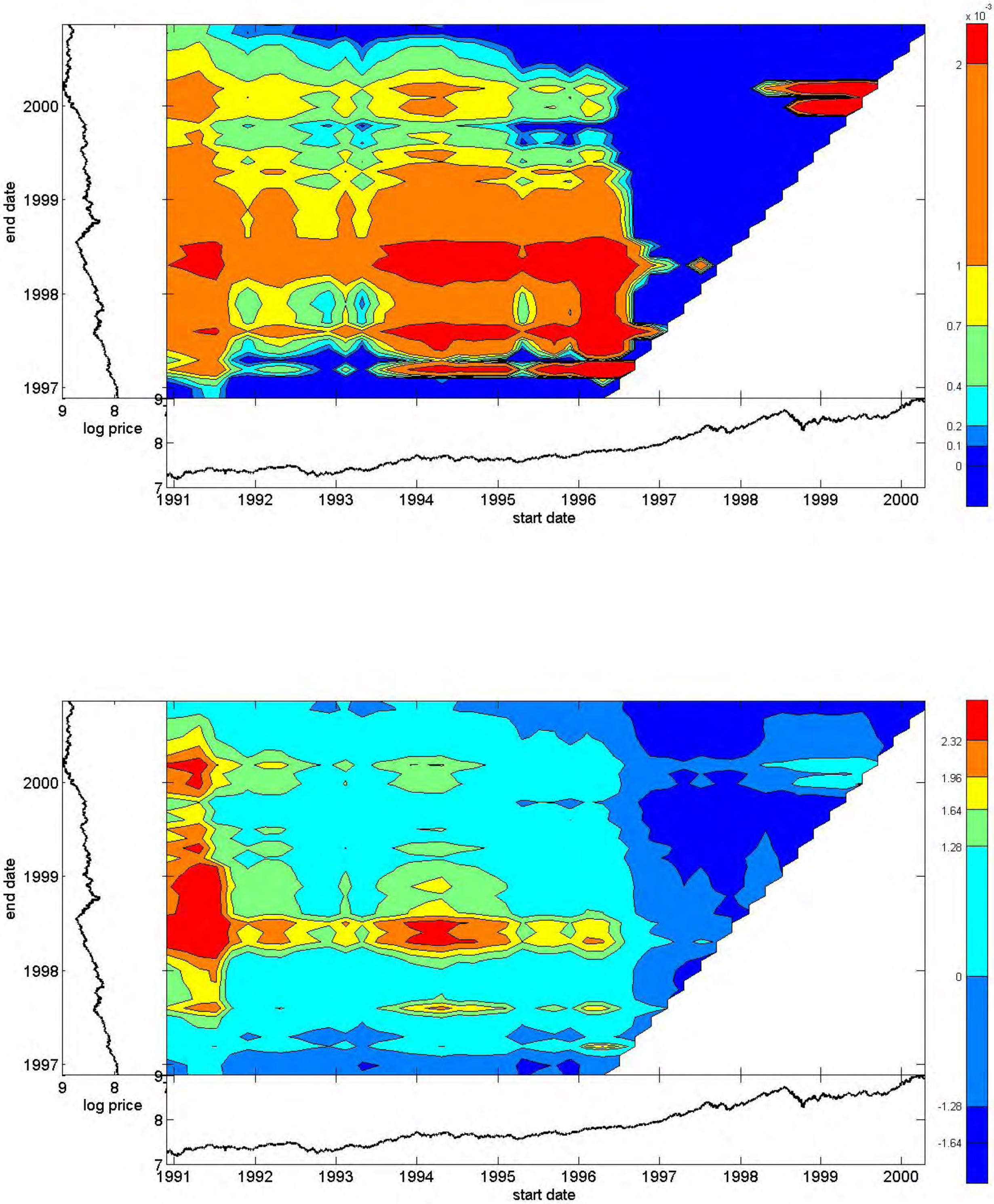


Figure 5.17.: DAX 1991-2000: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = \log(P_t)$) fitted over different time windows

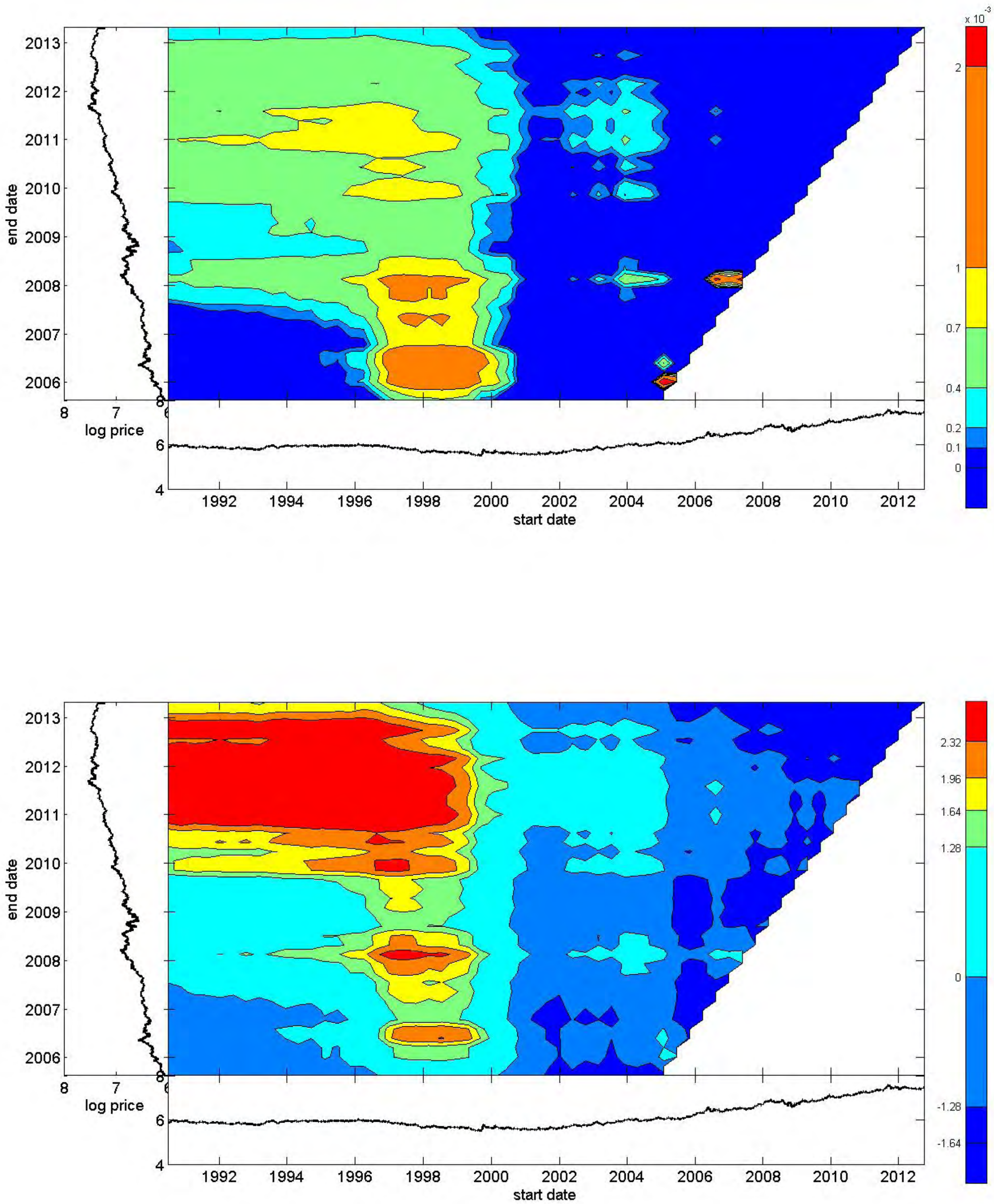


Figure 5.18.: Gold 1991-2013: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = \log(P_t)$) fitted over different time windows

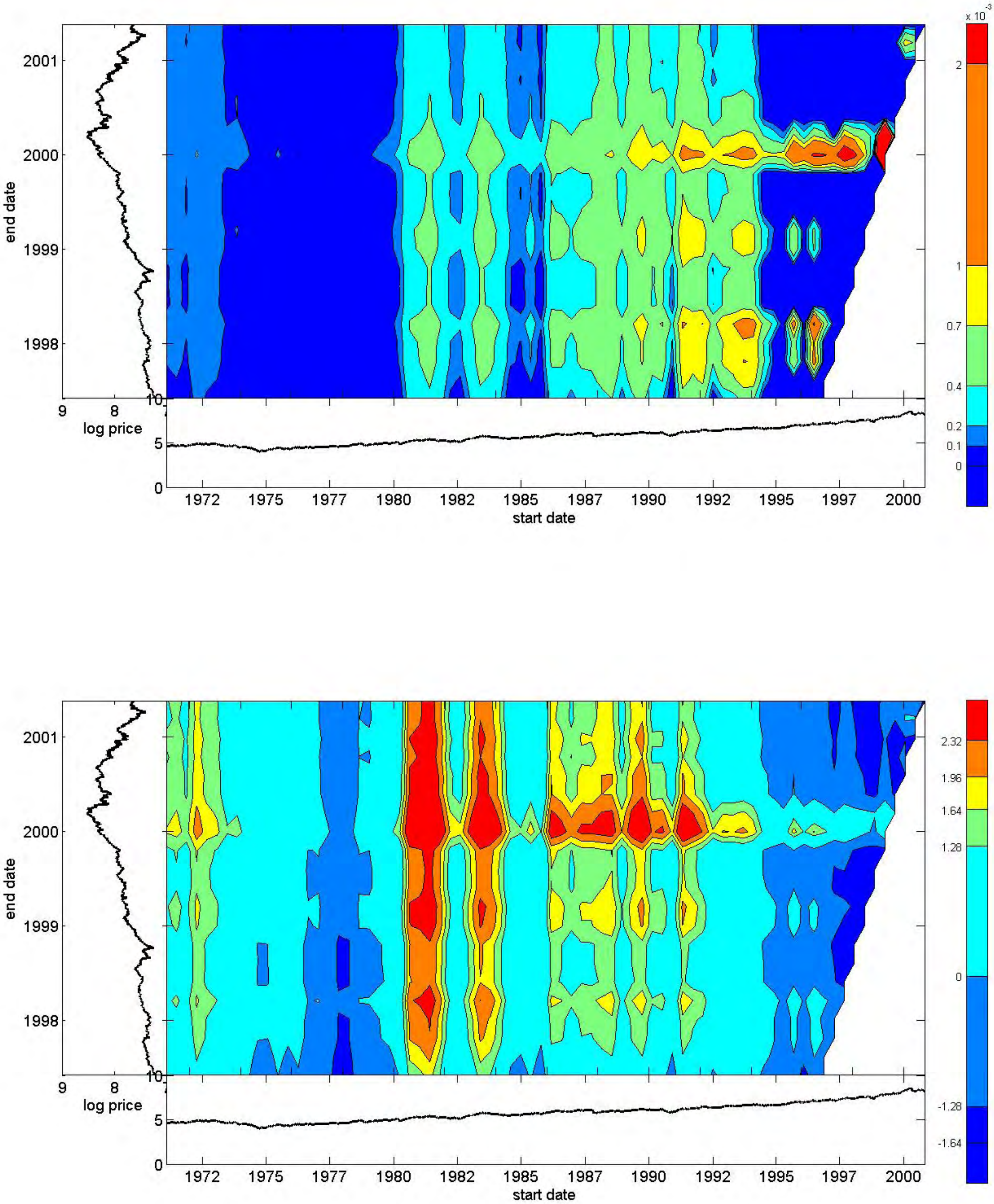


Figure 5.19.: NASDAQ Composite 1971-2001: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = \log(P_t)$) fitted over different time windows

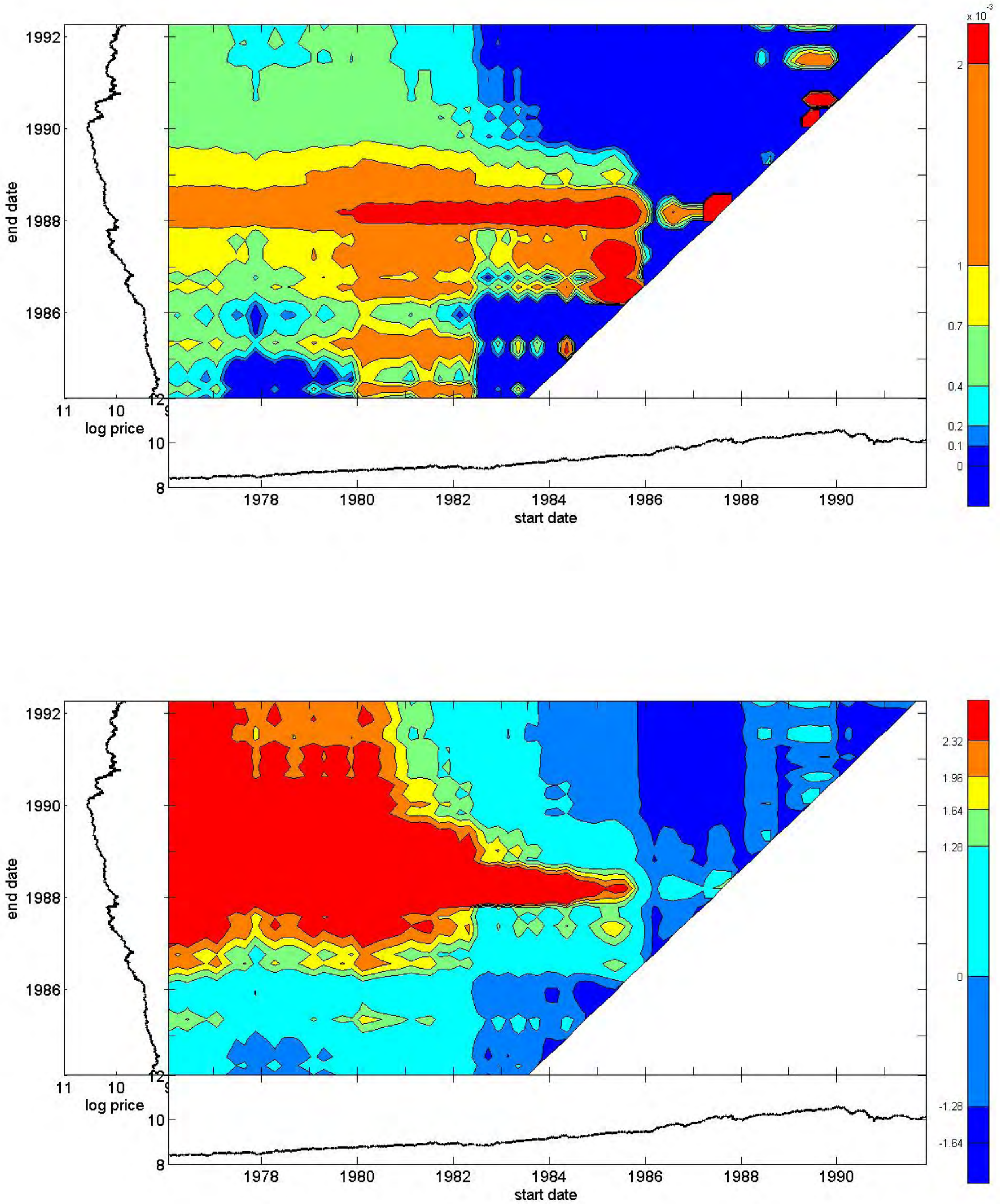
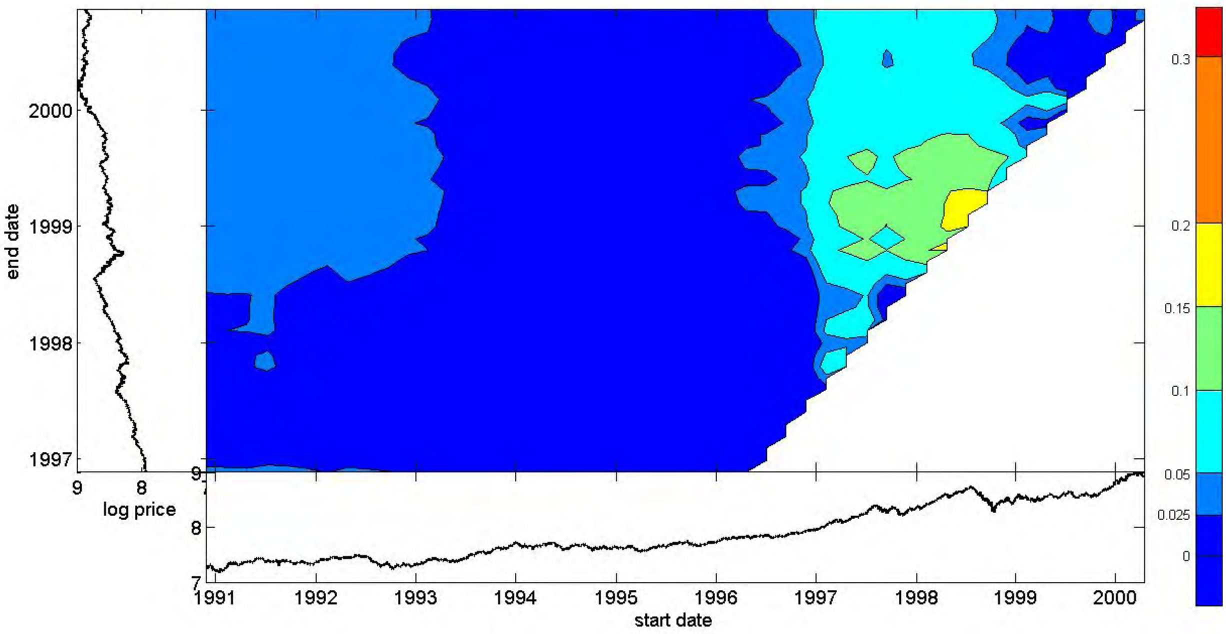


Figure 5.20.: NIKKEI 1976-1992: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = \log(P_t)$) fitted over different time windows



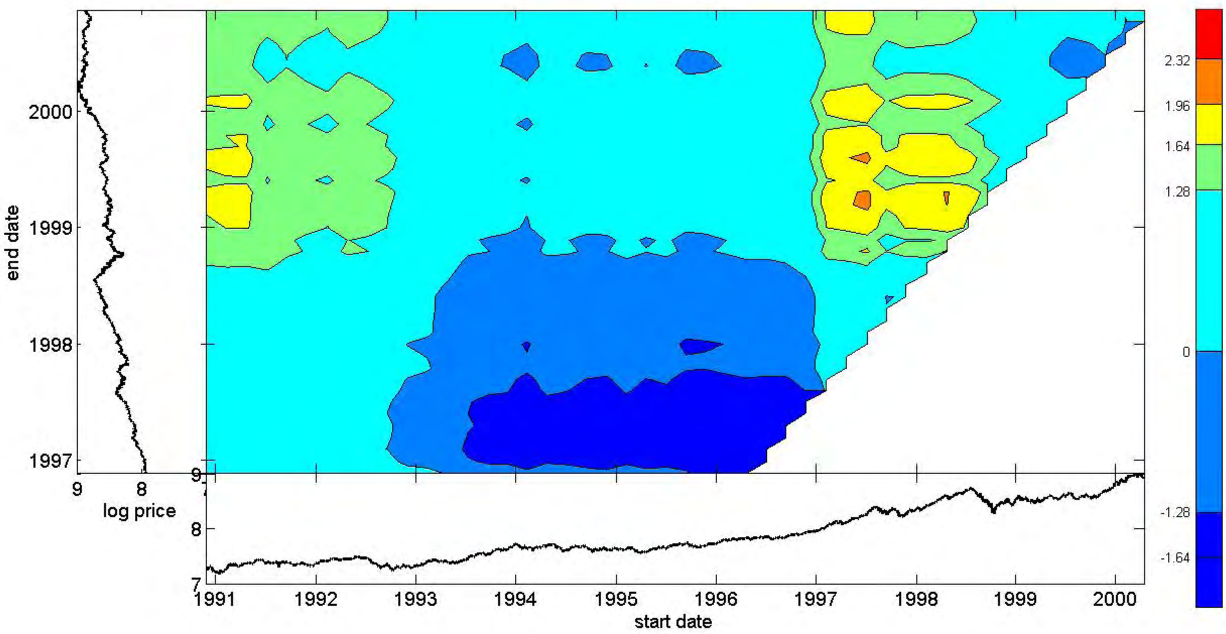


Figure 5.21.: DAX 1991-2000: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = r_t$) fitted over different time windows

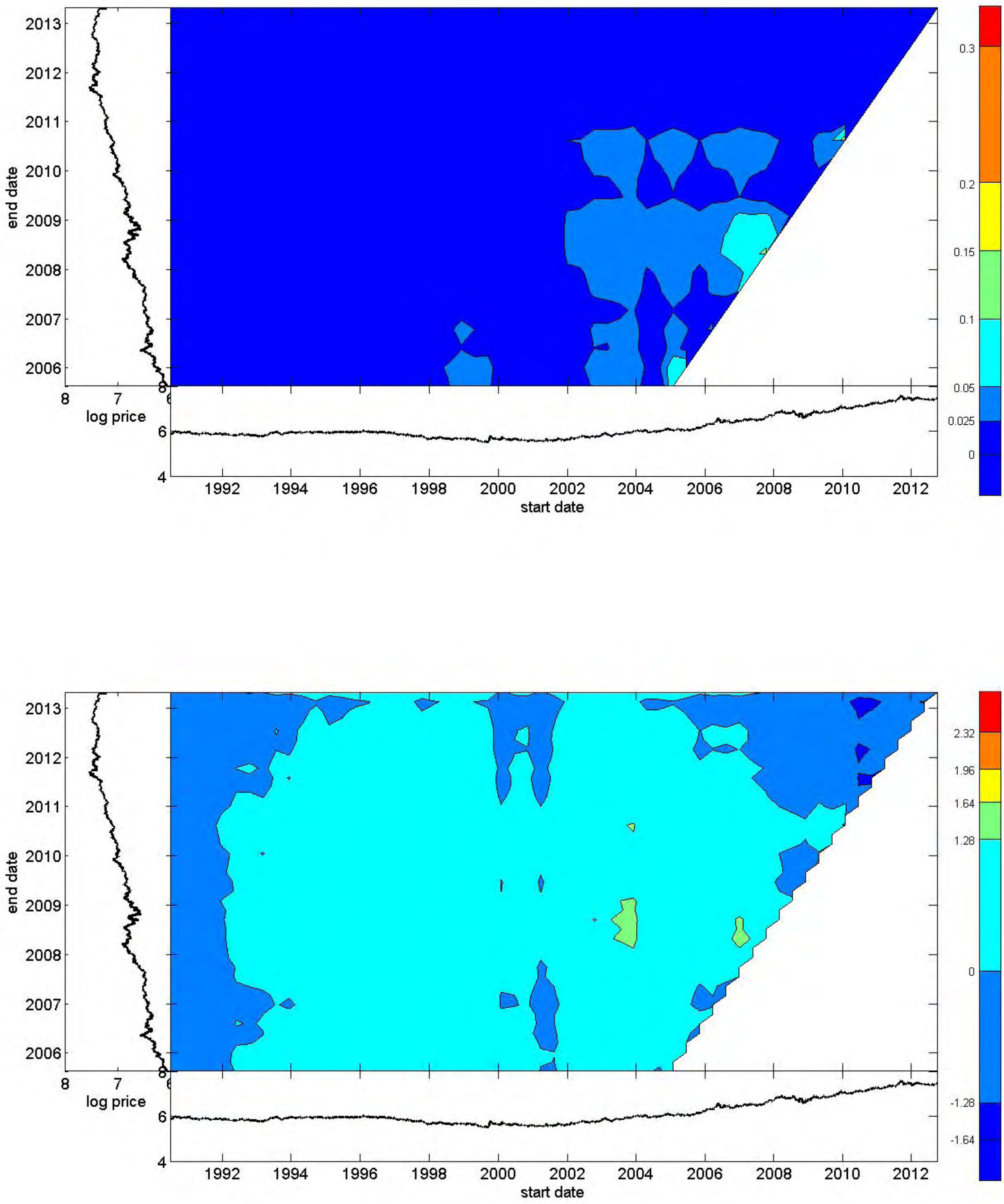
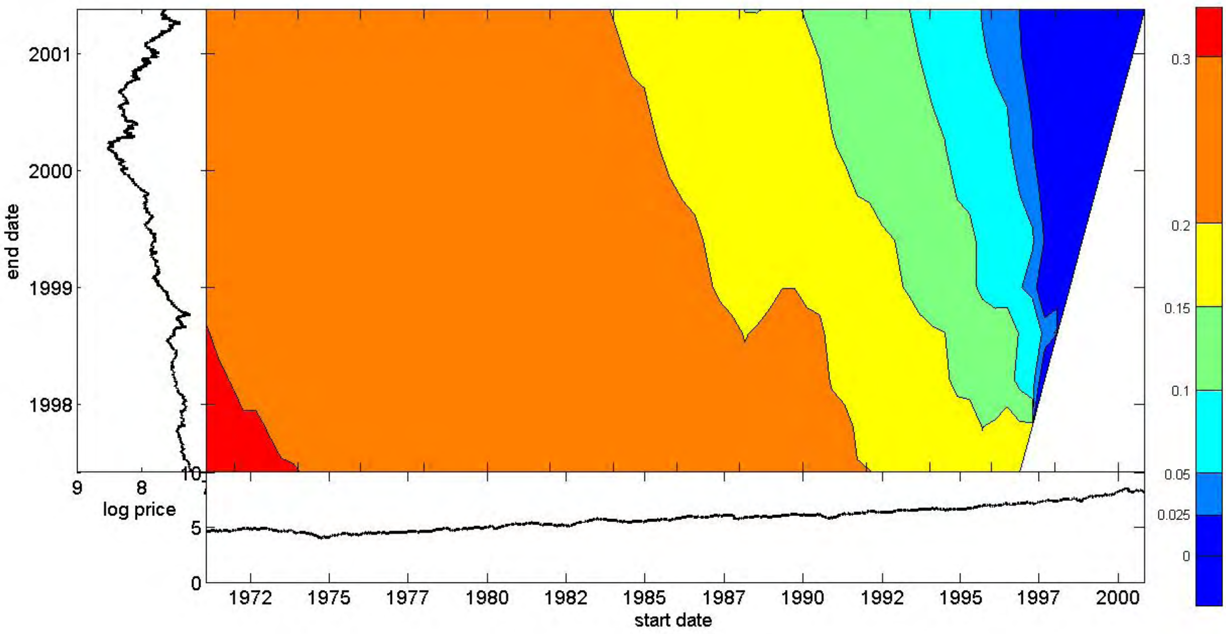


Figure 5.22.: Gold 1991-2012: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = r_t$) fitted over different time windows



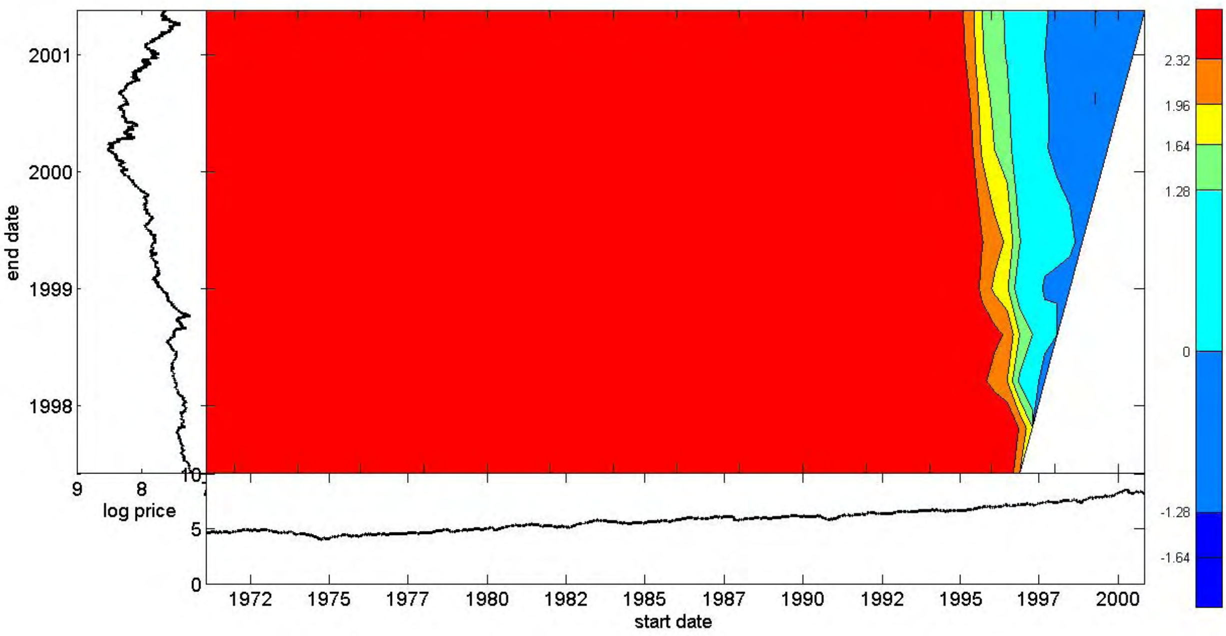


Figure 5.23.: NASDAQ Composite 1971-2000: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an $FTS\ GARCH$ model ($y_t = r_t$) fitted over different time windows

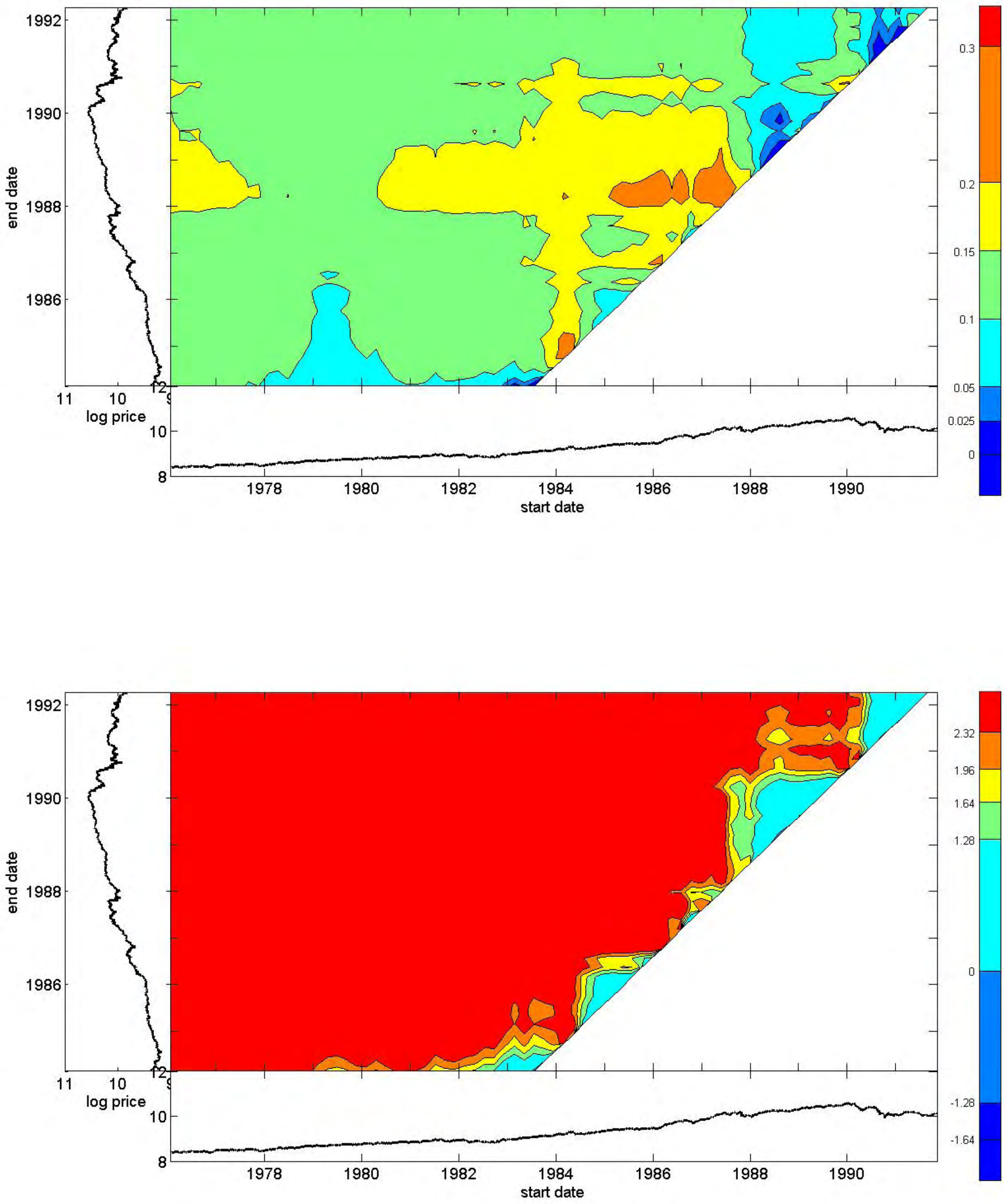


Figure 5.24.: NIKKEI 1976-1994: estimated γ (upper panel) and corresponding t-statistic (lower panel) in an *FTS GARCH* model ($y_t = r_t$) fitted over different time windows

5.5. Improving the usage of return as regression component

5.5.1. Kalman filtered return

The Kalman filter is a remarkable method to reduce the noise in a time series. Hence, we will try to use it in order to improve the usage of return as regression component in an *FTS GARCH* model. We start off with a general description of the discrete Kalman filter following the ideas of Welch and Bishop (1995) [WB]. Subsequently, we look in detail at the implemented Kalman filter method.

The system of equations

The discrete Kalman filter provides an estimate for the **true state** $x \in \mathbb{R}^p$ of a discrete-time process. The estimation is done in a way that minimizes the mean of the squared error. The true state x is assumed to follow the linear stochastic difference equation

$$x_k = A \cdot x_{k-1} + B \cdot u_{k-1} + w_{k-1}, \qquad \text{for } k = 2, \dots, n, \tag{5.29}$$

where $u \in \mathbb{R}^l$ is an optional control input and $A \in \mathbb{R}^{p \times p}$ and $B \in \mathbb{R}^{p \times l}$ are constant coefficient matrices. The **noisy measurement** $z \in \mathbb{R}^q$ of the true state should satisfy

$$z_k = H \cdot x_k + v_k, \qquad \text{for } k = 1, \dots, n, \tag{5.30}$$

where $H \in \mathbb{R}^{q \times p}$ is another constant coefficient matrix. w_k and v_k are the **process and** the **measurement noise**, respectively. They are assumed to be white noise, independent of each other and with normal probability distributions

$$w_k \sim N(0, Q_k) \text{ and } v_k \sim N(0, R_k). \tag{5.31}$$

The process noise covariance Q_k and the measurement noise covariance R_k change with each time step. In practice, their estimation is quite a challenging task. There is a variety of techniques proposed in the literature. In order to estimate them, for example, by the sample covariance or an exponentially weighted moving average calculated over a certain time window N, one needs a surrogate for the true process state x_k. Even more promising but also more complicated is the Autocovariance Least-Squares method.

The method of estimation

Let $\hat{x}_k^- \in \mathbb{R}^p$ be the **a priori state estimate** for x at step k given knowledge of the process prior to step k, and $\hat{x}_k \in \mathbb{R}^p$ the **a posteriori estimate** at step k given measurement z_k. Then, one can define the **a priori** and **a posteriori estimate errors** by

$$e_k^- := x_k - \hat{x}_k^- \qquad \text{and} \qquad e_k := x_k - \hat{x}_k. \tag{5.32}$$

Hence, the corresponding error covariance matrices can be computed as

$$P_k^- := \mathbb{E}[e_k^- \cdot e_k^{-T}] \qquad \text{and} \qquad P_k := \mathbb{E}[e_k \cdot e_k^T]. \tag{5.33}$$

Now, let us rougly describe how Maybeck (1979) [M3] finds the optimal estimator algorithm. From a Bayesian point of view, one is interested in using the density of the random variable x_{k-1} conditioned on the entire measurement history $Z_{k-1} := (z_1, \ldots, z_{k-1})$ to find the density of x_k conditioned on Z_k. Maybeck starts with the assumption that

$$x_{k-1}|Z_{k-1} \sim N(\hat{x}_{k-1}, P_{k-1}). \tag{5.34}$$

In particular, this means

$$\mathbb{E}[x_{k-1}|Z_{k-1}] = \hat{x}_{k-1} \quad \text{and} \quad P_{k-1} = \mathbb{E}[e_{k-1} \cdot e_{k-1}^T|Z_{k-1}]. \tag{5.35}$$

This assumption is verified in form of an inductive proof type of derivation ($k-1 \Rightarrow k$) to be true for all $k = 1, \ldots, n$ (see Maybeck (1979) [M3] for details). In other words, the optimal estimator is found to coincide with the conditional mean and the covariance recursion does not depend on the measurement history (not only P_k can be shown to be independent of Z_k, but also P_k^- of Z_{k-1}). If Z_k changes, then the conditional mean of x_k changes, but the shape of the density of $x_k|Z_k$ remains unchanged. Maybeck interprets this independence in the following way: "[...] the estimator gleans out as much information from the measurements as possible, and there is nothing left in the measurements that could tell you anything about the error." Hence, we can compute the covariance of the errors by using $\hat{x}_k$ as the optimal estimate of the true state x_k.
These thoughts (and some further arguments that can be found in Maybeck (1979) [M3] as well) initiate the goal to find the a posteriori state estimate $\hat{x}_k$ as a linear combination of the a priori estimate $\hat{x}_k^-$ and a weighted difference between the actual measurement z_k and the measurement prediction $H \cdot \hat{x}_k^-$ (Welch and Bishop (1995) [WB]):

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot (z_k - H \cdot \hat{x}_k^-) \tag{5.36}$$

The difference $z_k - H \cdot \hat{x}_k^-$ is called **measurement innovation** or **residual**. The **Kalman gain** $K_k \in \mathbb{R}^{p \times q}$ is chosen to minimize the a posteriori error covariance P_k. In order to find an explicit form for the minimizer, one substitutes $\hat{x}_k$ from (5.36) into the definition for e_k (5.32) and subsequently puts that into the formula for P_k in (5.33). Then, one takes the derivative of the trace of P_k with respect to K_k and sets the result equal to 0. Finally, one can solve for the optimizing K_k (please see Appendix B for the full derivation of the minimizer K_k and the optimal form for P_k).
Eventually, Kalman found the following two-step procedure as solution for the filtering problem. At first, the optimal state estimate and the error covariance are predicted from measurement time $k-1$ to k by (**time update**)

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}, \tag{5.37}$$
$$P_k^- = AP_{k-1}A^T + Q_k. \tag{5.38}$$

Then, they are updated by computing the Kalman gain

$$K_k = P_k^- H^T (HP_k^- H^T + R_k)^{-1}. \tag{5.39}$$

and incorporating it in both the mean and covariance relations (**measurement update**):

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot (z_k - H \cdot \hat{x}_k^-), \tag{5.40}$$
$$P_k = (I - K_k H)P_k^-. \tag{5.41}$$

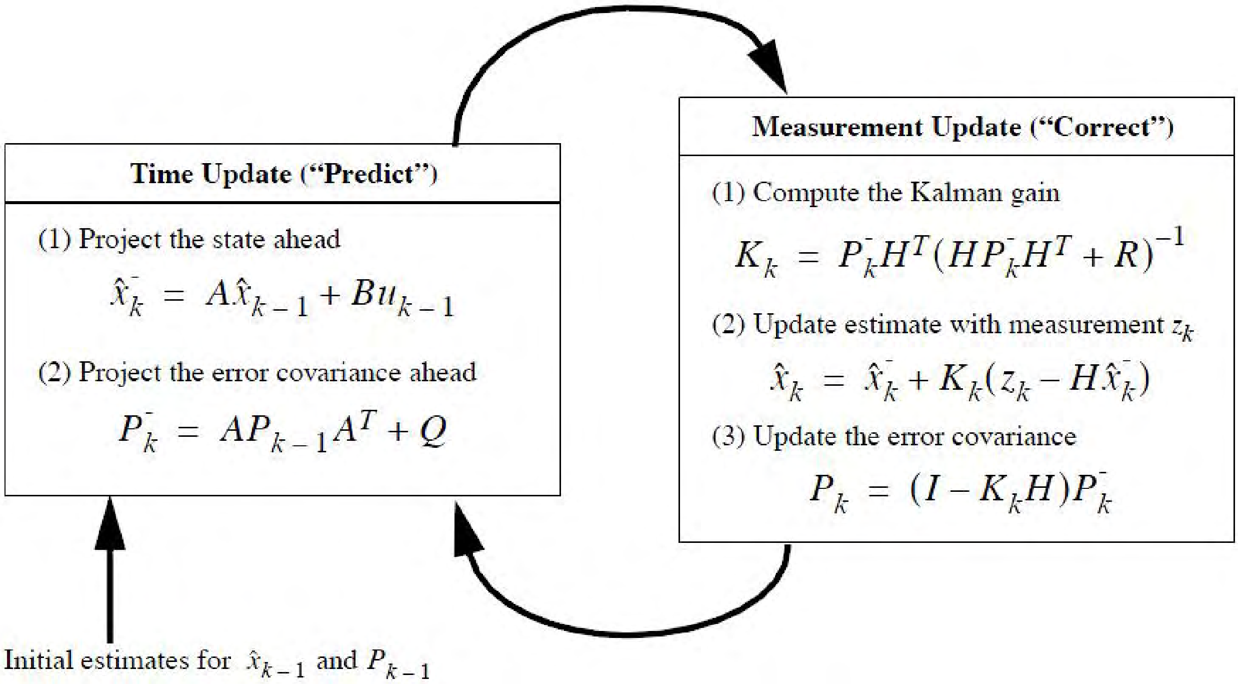


Figure 5.25.: Illustration of the discrete Kalman filter method (Figure from Welch and Bishop (1995) [WB]).

In summary, the Kalman filter operates along some kind of **feedback control**: At first, it estimates the true state of the process and then obtains feedback in form of a measurement. The time update equations forecast the state and error covariance estimates to obtain a priori estimates for the next time step, whereas the measurement update equations are responsible for the feedback. They incorporate the measurement into the a priori estimate to obtain an improved a posteriori estimate.

According to Maybeck (1979) [M3], Kalman originally developed the filter by the geometrical relation that the optimal estimate $\hat{x}_k$ is the orthogonal projection of the true state x_k onto the subspace spanned by the measurement history Z_k. Furthermore, the Kalman filter estimate is optimal with respect to other criteria as well, which makes it a quite powerful tool for different applications. The optimal estimate $\hat{x}_k$ minimizes the mean square error since it is equal to the conditional mean. Also, it is the minimum variance unbiased linear estimate and, if there is no a priori state information, even the maximum likelihood estimate.

The implemented form of the Kalman filter

For the calculation of the Kalman filtered returns r_k^{Kalman}, $k = 1, \ldots, n$, we use a one dimensional discrete Kalman filter ($p = 1$, $q = 1$). Moreover, we simplify the filter by setting $A = 1$, $B = 0$ and $H = 1$. Therefore, we start off with the following system of equations:

$$x_k = x_{k-1} + w_{k-1}, \quad k = 2, \ldots, n, \tag{5.42}$$

$$r_k = x_k + v_k, \quad k = 1, \ldots, n, \tag{5.43}$$

where

$$w_k \sim N(0, Q_k) \quad \text{and} \quad v_k \sim N(0, R_k). \tag{5.44}$$

For the estimation of the process noise variance Q_k and the measurement noise variance R_k in each time step we use the sample variance over a window of length N. An exponentially weighted moving average of the noisy measurements r_k (window length $MAlen = N$) serves as a surrogate for the true process state x_k.

Algorithm:

Smoothing data: $r^{smoothed} = EWMA(r, N)$,

where

$$EWMA(r, N)_t := \lambda \cdot EWMA(r, N)_{t-1} + (1 - \lambda) \cdot r_t, \quad \text{for} \quad t = 2, \dots, T, \tag{5.45}$$

with

$$EWMA(r, N)_1 = 0 \quad \text{and} \quad \lambda := 1 - \frac{1}{N+1}. \tag{5.46}$$

Start values: $\hat{x}_k = r_k^{smoothed}$, for $k = 1, \dots, 2 \cdot N - 2$, $P_{2 \cdot N-2} = 1$

for $k = 2 \cdot N - 1, \dots, T$

1. **A priori state estimate** (project the state ahead):

$$H \cdot \hat{x}_k^- = \hat{x}_k^- = \hat{x}_{k-1},$$

   which means the preceding Kalman filtered return is used as **measurement forecast**.

2. **Estimate process and measurement noise**:

$$\begin{aligned} \hat{Q}_k &= \frac{1}{N-1} \sum_{j=k-N+1}^{k} \left(r_j^{smoothed} - \frac{1}{N} \sum_{j=k-N+1}^{k} r_j^{smoothed} \right)^2 \\ \hat{R}_k &= \frac{1}{N-1} \sum_{j=k-N+1}^{k} \left((r_j - r_j^{smoothed}) - \frac{1}{N} \sum_{j=k-N+1}^{k} (r_j - r_j^{smoothed}) \right)^2 \end{aligned}$$

   Project error variance ahead:

$$P_k^- = P_{k-1} + \hat{Q}_k$$

3. **Compute Kalman gain:**

$$K_k = P_k^- \cdot (P_k^- + \hat{R}_k)^{-1}$$

4. **Combine forecast optimally with measurement r_k:**

$$r_k^{Kalman} := \hat{x}_k = \hat{x}_k^- + K_k \cdot (r_k - \hat{x}_k^-) \text{ (a posteriori state estimate)}$$

5. **Update error variance:**

$$P_k = (1 - K_k) \cdot P_k^-$$

end

Using the filtered returns r_k^{Kalman}, $k = 1, \ldots, n$, we are now able to estimate the *FTS GARCH* model with Kalman filtered regression component for various time windows $(t_{start}, \ldots, t_{end})$, where t_{start}, $t_{end} \in \{1, \ldots, n\}$ and $t_{start} < t_{end}$:

$$r_t = \mu + \gamma \cdot r_{t-1}^{Kalman} + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \tag{5.47}$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \quad t = t_{start} + 1, \ldots, t_{end}, \tag{5.48}$$

and test for

$$H_0 : \gamma \leq 0 \text{ against } H_1 : \gamma > 0. \tag{5.49}$$

Synthetic tests:

How does the t-statistic of the regression component ($y_t = r_t^{Kalman}$) in an *FTS GARCH* model evolve if we simulate a bubble using the *FTS GARCH* model including a period of alternating regimes ($GARCH(1,1)$ and *FTS GARCH*)?

We propose to repeat the simulation experiments from Figures 5.13 - 5.16, but instead of using either return or log price as regression component in the fitted *FTS GARCH* model we use the Kalman filtered returns. For both types of simulated bubbles we find an improvement of the bubble signal in the preferred way (Figures 5.26 and 5.27). Now, we successfully detect a bubble simulated with $y_t = \log(P_t)$ for windows ending at the end of the simulation period. Moreover, in case of $y_t = r_t$ in the simulated *FTS GARCH*, we see a reduction in the strength of the bubble signal. There is high significance for a positive γ during the pure bubble regime but not during the alternating regime. Hence, the applicaton of a Kalman filter to the return series could improve our empirical results as well.

Empirical results:

Summing up the empirical results for DAX and Gold using Kalman filtered returns as regression component, we could not achieve a rise in the strength of the bubble signals (independent of the chosen parameter N). Nevertheless, the filter manages to reduce the strength of the bubble signal in case of the NASDAQ Composite and NIKKEI time series. For the NIKKEI we detect a perfect bubble signal between 1988 and 1990 preceding the peak in 1990 using almost all start dates (Figure 5.29). But the results for NASDAQ show there is no guarantee the filter solves our problem completely. Obviously it reduces the signal strength, however, we still observe increasing significance for larger time windows (Figure 5.28).

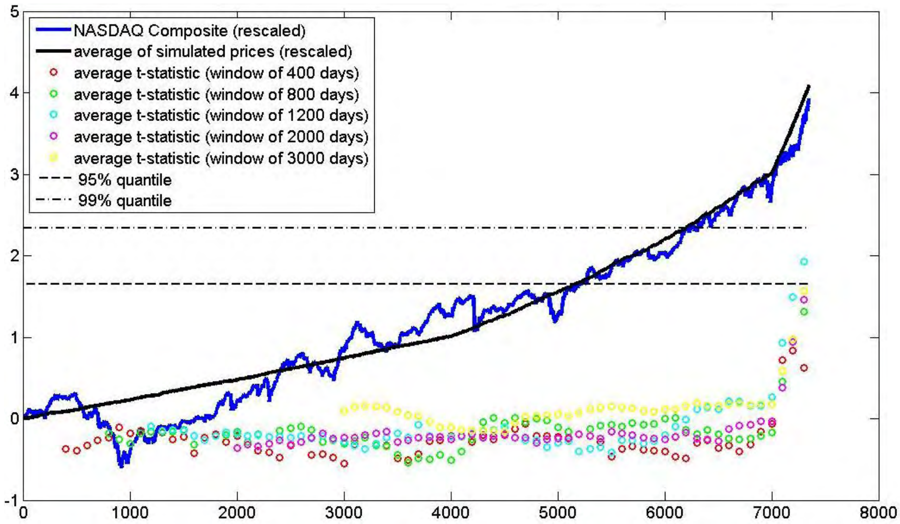


Figure 5.26.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an $FTS\ GARCH$ model ($y_t = r_t^{Kalman}$, $N = 20$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

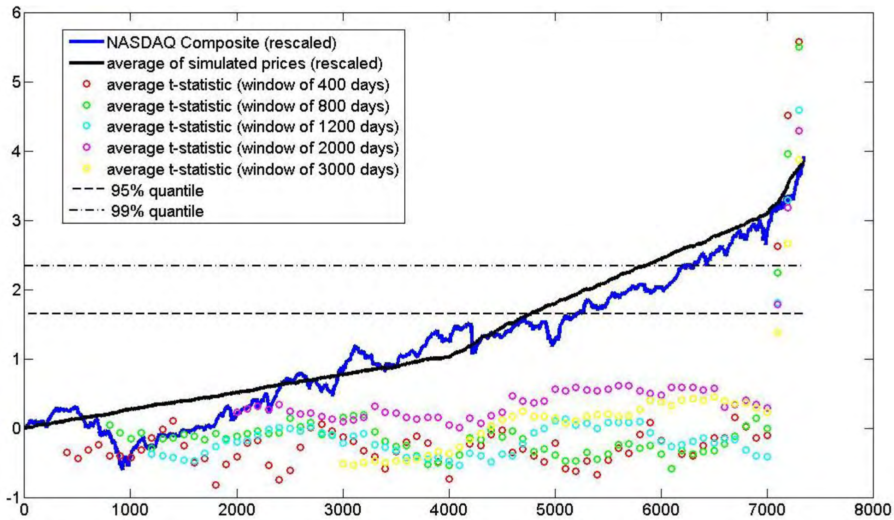


Figure 5.27.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an $FTS\ GARCH$ model ($y_t = r_t^{Kalman}$, $N = 20$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

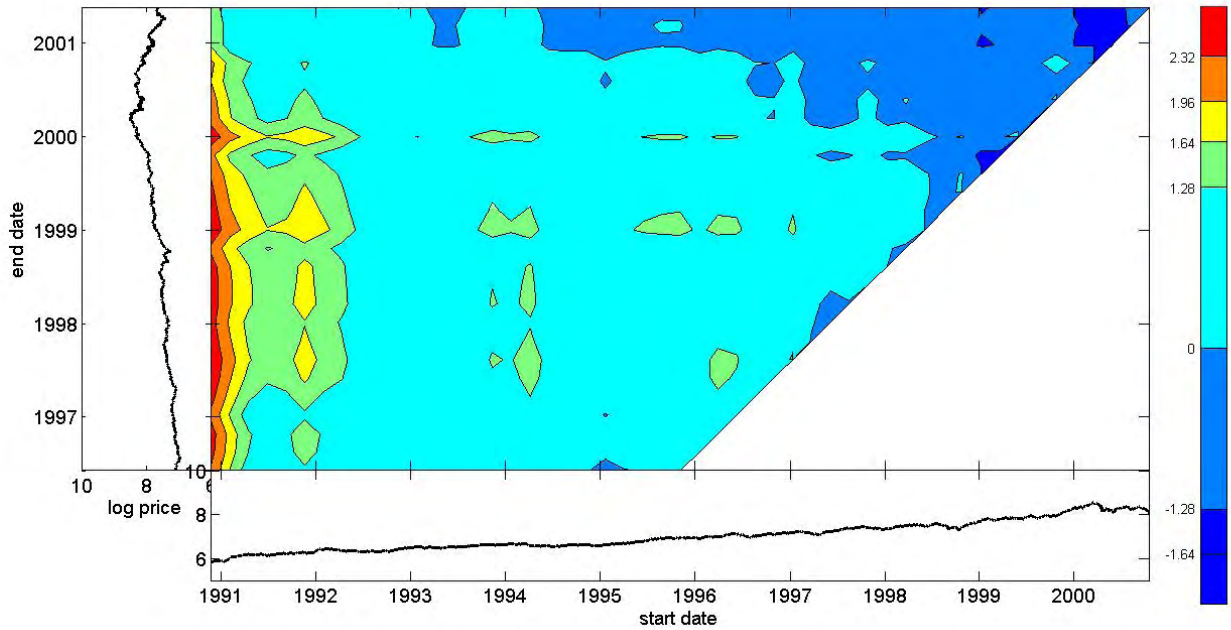


Figure 5.28.: NASDAQ Composite 1991-2000: t-statistic of γ in an FTS-GARCH model ($y_t = r_t^{Kalman}$, $N = 10$) fitted over different time windows

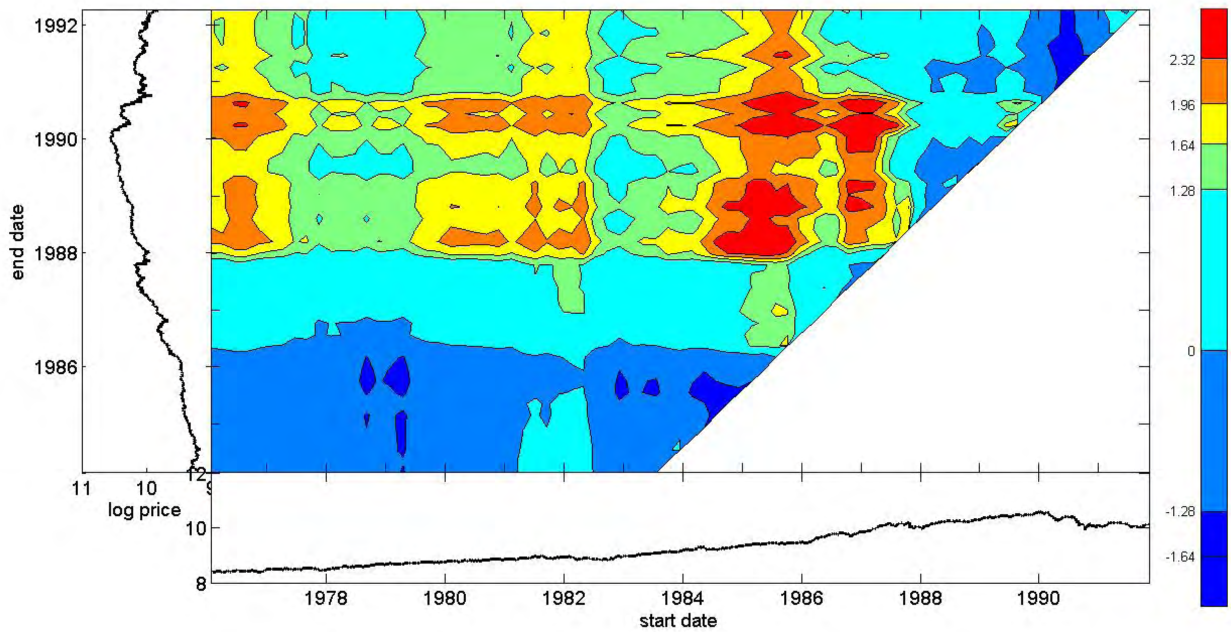


Figure 5.29.: NIKKEI 1976-1994: t-statistic of γ in an FTS-GARCH model ($y_t = r_t^{Kalman}$, $N = 20$) fitted over different time windows

5.5.2. Average return

Another alternative instead of using the daily return series itself for the estimation of an *FTS GARCH* model is to use averages of multiple returns. Hence, let us define the **average return** over a certain horizon of $\tau \in \mathbb{N}$ days as

$$r^{\tau}_{t|s} \;:=\; \log\left(\frac{P_{s+t\cdot\tau}}{P_{s+(t-1)\cdot\tau}}\right) = \sum_{k=s+(t-1)\cdot\tau+1}^{s+t\cdot\tau} \frac{r_k}{\tau}, \quad \text{for } t = 1,\ldots,\left[\frac{n-s}{\tau}\right] =: T, \tag{5.50}$$

where $s \in \{0,\ldots,\tau-1\}$ is the start date in the price series $P_s,\ldots,P_n$ chosen for the calculation. Note that for $s = 0$ we obviously use the whole sample $P_0,\ldots,P_n$ for the analysis. We will now show that the estimation over various time windows $(t_{start},\ldots,t_{end})$, with $t_{start},\ t_{end} \in \{1,\ldots,T\}$ and $t_{start} < t_{end}$, of the following form of an *FTS GARCH* process depends a lot on the chosen start date s:

$$r^{\tau}_{t|s} \;=\; \mu + \gamma \cdot r^{\tau}_{t-1|s} + \epsilon_t, \text{ with } \epsilon_t \sim N(0,h_t), \tag{5.51}$$

$$h_t \;=\; \alpha_0 + \alpha_1 \cdot \epsilon^2_{t-1} + \beta_1 \cdot h_{t-1}, \quad t = t_{start}+1,\ldots,t_{end}. \tag{5.52}$$

For this purpose, let us consider, for example, the NASDAQ Composite time series between December 9th, 1986 and May 18th, 2001 including 3650 prices $P_0,\ldots,P_{3649}$. Using a horizon of $\tau = 40$ days, the number of average returns T is equal to

$$T = \left[\frac{n-s}{\tau}\right] = \begin{cases} 91, & \text{for } s \in \{0,\ldots,9\}, \\ 90, & \text{for } s \in \{10,\ldots,39\}. \end{cases} \tag{5.53}$$

One can see the dependence on s in Figure 5.30 where we plot the t-statistic of γ in 5.51-5.52 for $s \in \{0,\ldots,39\}$, $t_{start} = 1$ and

$$t_{end} \in \begin{cases} \{66,\ldots,91\}, & \text{for } s \in \{0,\ldots,9\}, \\ \{65,\ldots,90\}, & \text{for } s \in \{10,\ldots,39\}, \end{cases} \tag{5.54}$$

We label the y-axis with the date of the last price included in the corresponding window of prices whose index is $n' = s + t_{end} \cdot \tau$. Let us compare, for example, the cases $s = 3$ and $s = 33$ (the corresponding first date of the price series used for calculating the average returns is December 12th, 1986, respectively, January 27th, 1987). For $s = 3$ we find high significance for a positive γ for the time windows ending around the bubble peak, whereas for $s = 33$ the maximum likelihood estimation finds a negative value for γ. The reason is that due to the averaging over a certain horizon the obtained return series look quite different for different start dates s (although the sum of all returns over each series of average returns is approximately the same). Figure 5.31 shows the average return series ($\tau = 40$ days) for NASDAQ Composite between December 9th, 1986 and 18th May, 2001 for $s = 3$ and $s = 33$. Intuitively, for $s = 33$ (in blue color) we find 4 highly negative average returns followed by highly positive ones which leads to a negative coefficient γ, whereas for $s = 3$ (in red color) there are less extreme returns. Alternatively, this can be visualized nicely in a scatterplot with regression line for the following regression

$$r^{40}_{t|s} = \mu + \gamma \cdot r^{40}_{t-1|s} + \epsilon_t, \quad t = 2,\ldots,t_{end} \text{ and } s \in \{3,33\}. \tag{5.55}$$

For t_{end} we choose the end date near the bubble peak where we found the highest significance for a positive γ. ϵ_t are the residuals. In Figure 5.32 we see that in a standard regression analysis one finds a quite similar result for γ as in the *FTS GARCH* model, the slope γ is positive for $s = 3$ and negative for $s = 33$.

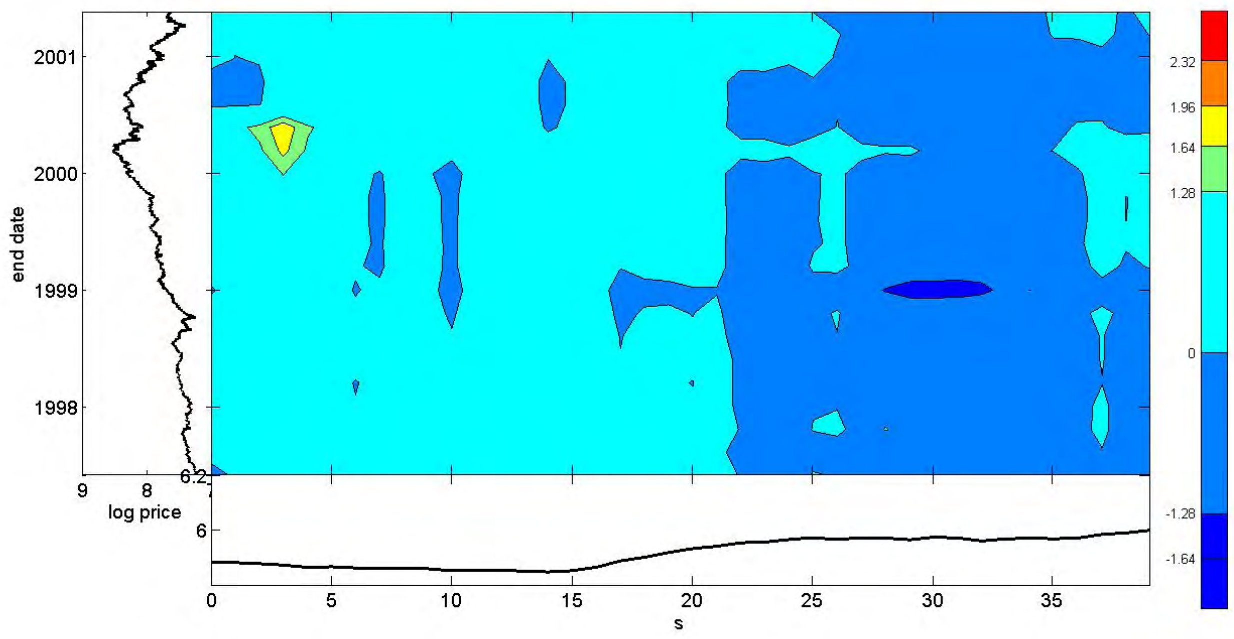


Figure 5.30.: NASDAQ Composite (December 9th, 1986 - May 18th, 2001): t-statistic of γ in an *FTS GARCH* model with regression component $y_t = r^{40}_{t|s}$ fitted for $s \in \{0, \ldots, 39\}$, $t_{start} = 1$ and various t_{end}.

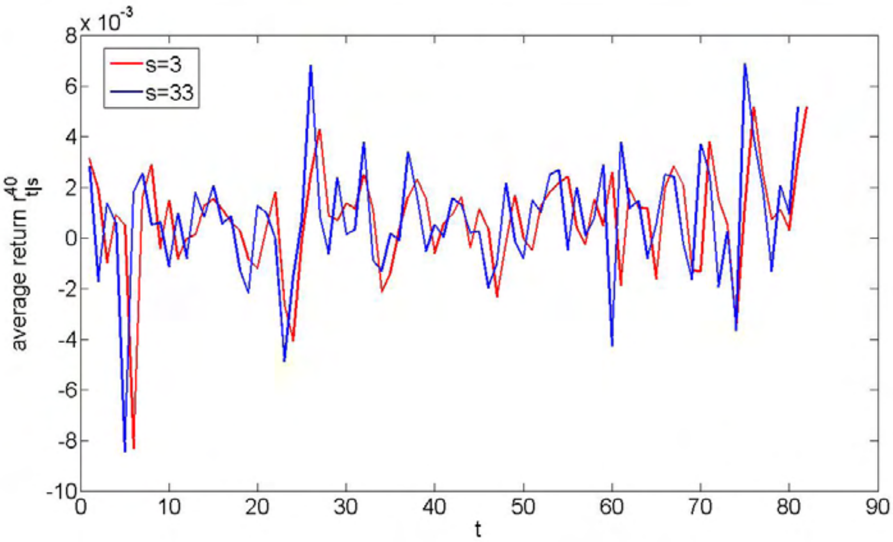


Figure 5.31.: NASDAQ Composite 1986-2001: Comparison of average returns over $\tau = 40$ days calculated for $s = 3$ and $s = 33$ ($r^{40}_{t|3}$ and $r^{40}_{t|33}$).

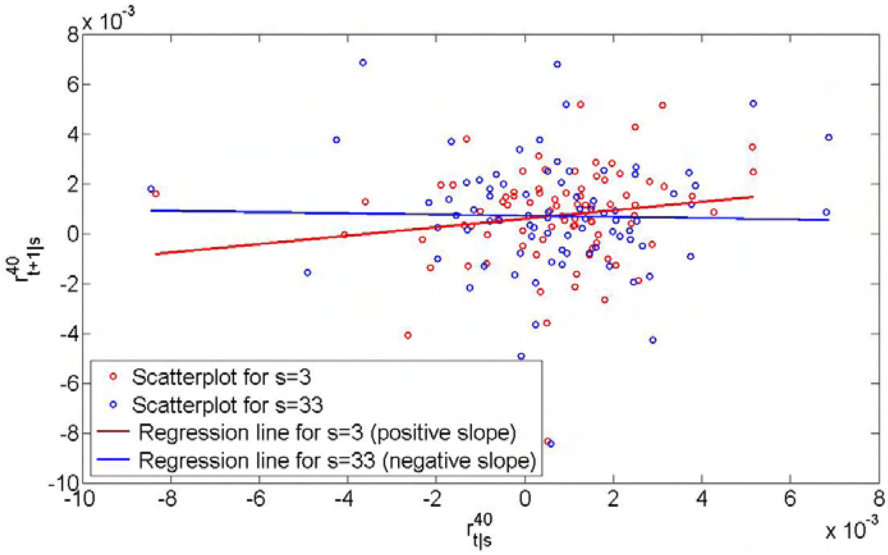


Figure 5.32.: NASDAQ Composite 1986-2001: Scatterplot of $r^{40}_{t+1|s}$ against $r^{40}_{t|s}$ with corresponding regression line for $s = 3$ and $s = 33$.

If we let s' vary even in $\{0, \ldots, 3\tau - 1\}$ we find some kind of periodicity in s' (3 periods). This is due to the fact that $s' \geq \tau$ corresponds to the case $s = s' \bmod \tau \in \{0, \ldots, \tau - 1\}$ with $t_{start} = \left[\frac{s'}{\tau}\right] + 1 > 1$, which means we are analyzing for various t_{end} as defined in equation 5.54 the windows of average returns $\{r^{\tau}_{t'|s'}\ ,\ t' = 1, \ldots, t_{end} - t_{start} + 1\} = \{r^{\tau}_{t|s}\ ,\ t = t_{start}, \ldots, t_{end}\}$, and thus, subsets of the series $\{r^{\tau}_{t|s}\ ,\ t = 1, \ldots, t_{end}\}$. Since the difference between these two sets is only the first few average returns, the maximum likelihood estimation yields quite similar coefficients for fixed s and $t_{start} \in \{1, 2, 3\}$ or equivalently for $s' \in \{s, s + \tau, s + 2\tau\}$ (Figure 5.33).
The upcoming question is how to choose the parameter s. I propose to pick s such that one maximizes the t-statistic of γ, or with other words, one gets the highest significance for a positive γ (in particular for windows ending around the bubble peak). Recall that we are looking for indication of super-exponential growth in price. Now, if we can find a way of averaging the return series such that there is positive feedback in the average returns we found evidence for such a behavior (even if there are other start dates s where we do not get such a strong signal). Basically this is just part of our goal to find a window in the price series with high significance for a positive γ around bubble peaks.
In the example above this means we choose $s = 3$ as start date for the price time series, where P_3 is the price of NASDAQ Composite on December 12$^{\text{th}}$, 1986.

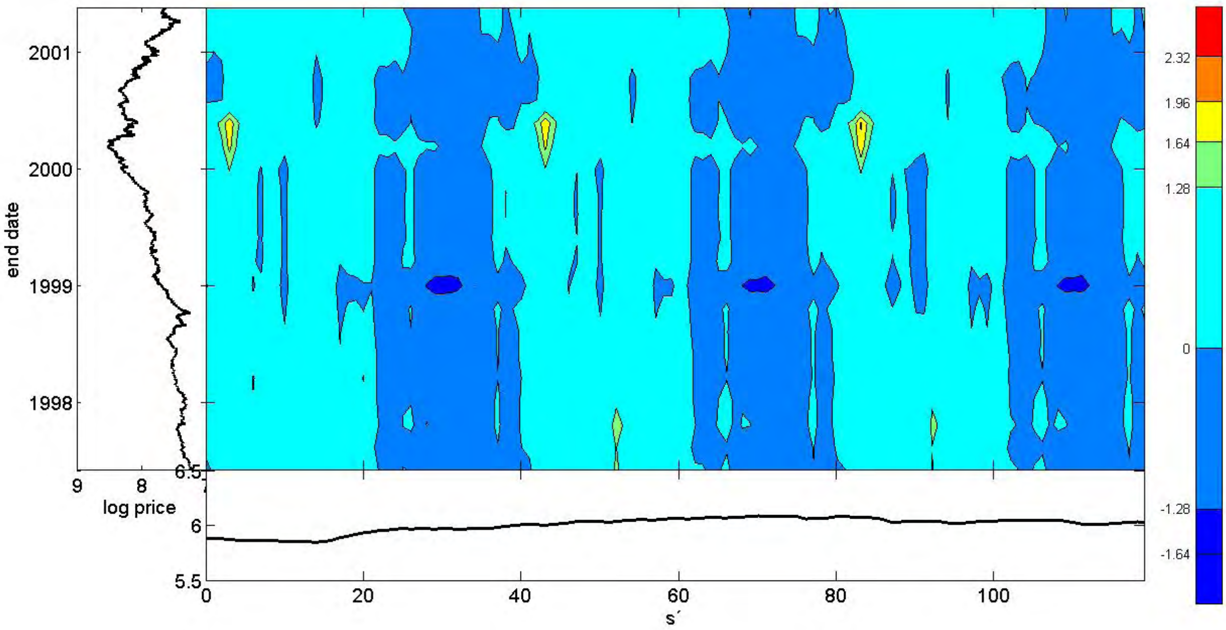


Figure 5.33.: NASDAQ Composite (December 9th, 1986 - May 18th, 2001): t-statistic of γ in an $FTS\ GARCH$ model with regression component $y_t = r^{40}_{t|s}$ fitted for $s \in \{0, \ldots, 39\}$, $t_{start} \in \{1, 2, 3\}$ and various t_{end}.

Synthetic tests:

Now, let us have a look at the performance of the method of using average returns to fit an $FTS\ GARCH$ model in various simulation experiments. We simulate again 50 trajectories of two types of bubbles. We start with the simulation of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics for another 3000 days and finally, a pure $FTS\ GARCH$ regime with γ chosen even higher as in the alternating period for 350 days. The coefficients are estimated from NASDAQ Composite between 1971 and 2000 in order to use realistic values. Note that we could do the same experiment with totally different parameters as well and that the analysis does not depend on the NASDAQ time series itself. As regression component in the $FTS\ GARCH$ model used for simulation we use (1) log prices $y_t = \log(P_t)$ and (2) daily returns $y_t = r_t$.
Next, we fit the $FTS\ GARCH$ model from equations 5.51 - 5.52 to various windows of the average returns calculated as in definition 5.50 from each of the simulated price trajectories with $s = 0$ and two different horizons τ. As usual, we plot a rescaled version of the NASDAQ Composite price series between 1971 and 2000, the average of the 50 simulated trajectories at each point in time and finally, the average of the 50 t-statistics of the coefficient γ for each window size and various points in time, where we shift the window by 100 days each time. For simplicity we assume that $s = 0$. Since we simulate multiple trajectories and average the corresponding t-statistics, the choice of s should not have any influence on our results. Actually, one could also look for an optimal start date s for each trajectory separately such that the t-statistic of γ is maximized. In general this would lead

to higher significance for a positive γ.
The results can be found in Figures 5.34 - 5.37. In Figure 5.34 and 5.35 we consider the cases $y_t = \log(P_t)$ and (1) $\tau = 20$ days and (2) $\tau = 40$ days. Comparing the results with Figure 5.14 where we used daily returns as regression component in the fitted *FTS GARCH* model, we see that the averaging of returns leads to higher significance for a positive γ in the bubble period (a pure *FTS GARCH* regime lasting for 350 days). Moreover, if we compare the cases $y_t = r_t$ and (1) $\tau = 5$ days and (2) $\tau = 20$ days in Figure 5.36 and 5.37 with Figure 5.16 where we used daily returns as regression component we can find an improvement as well.
In a nutshell, the usage of average returns improves the bubble signal found from simulated trajectories in a preferable way for both simulation experiments, using log price and daily returns as regression component.

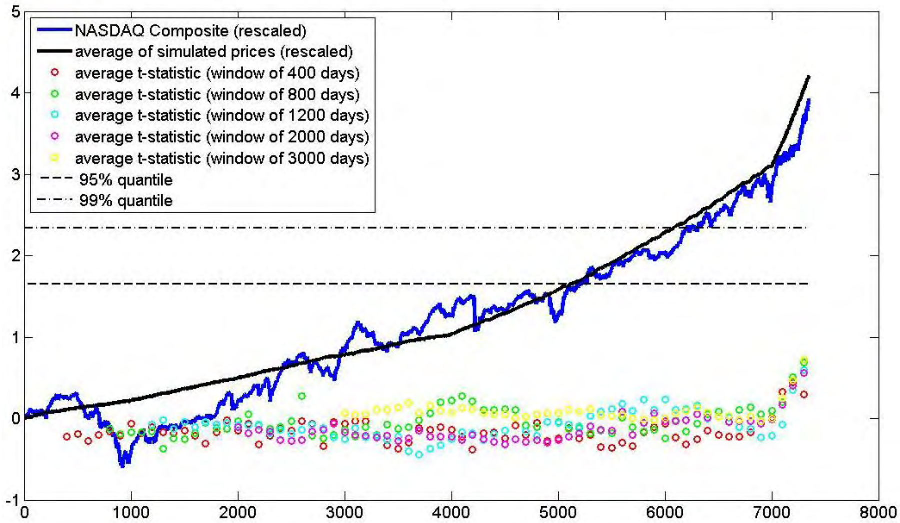


Figure 5.34.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an *FTS GARCH* model (average return, $\tau = 20$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

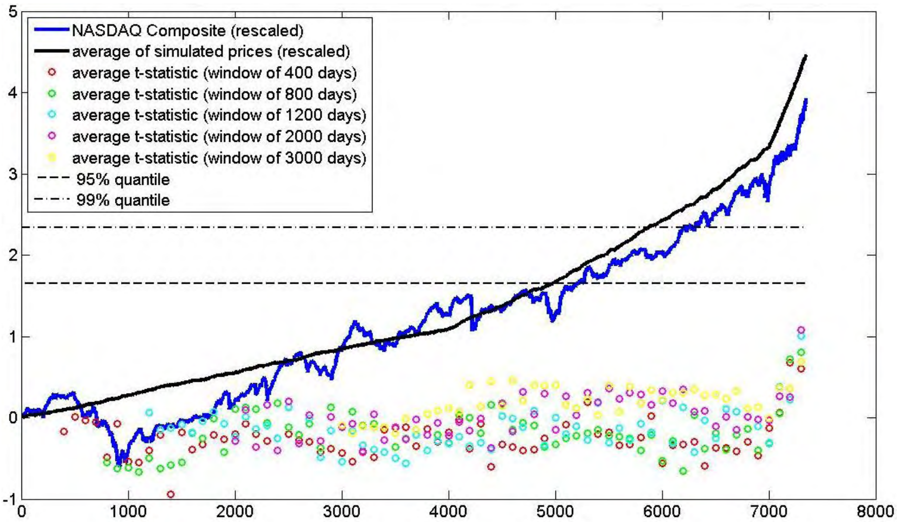


Figure 5.35.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an $FTS\ GARCH$ model (average return, $\tau = 40$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

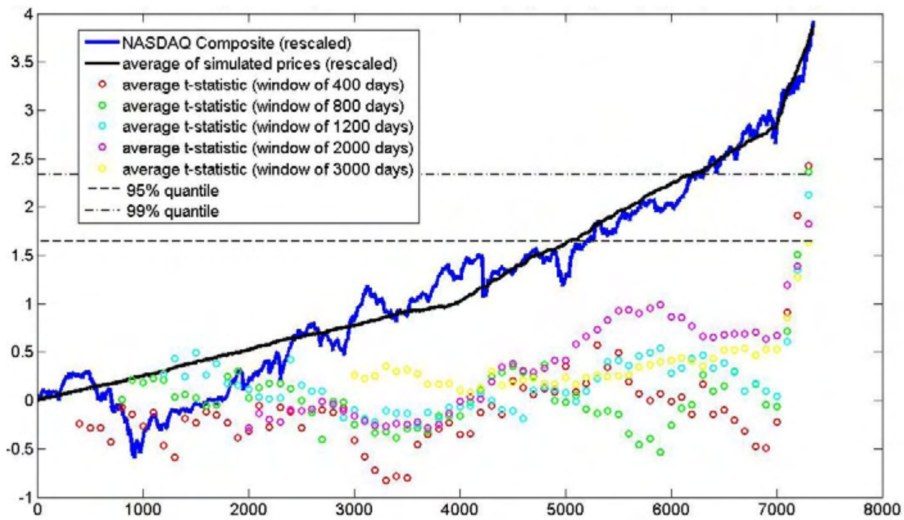


Figure 5.36.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an $FTS\ GARCH$ model (average return, $\tau = 5$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

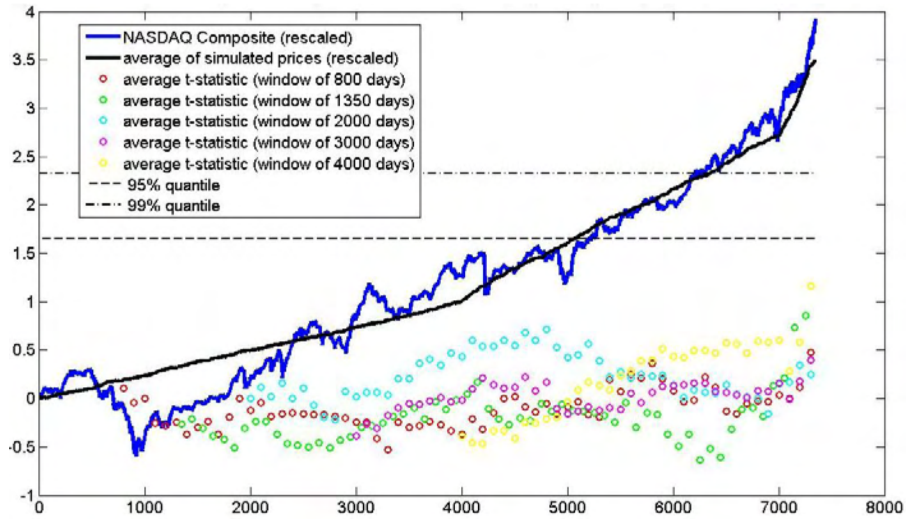


Figure 5.37.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an $FTS\ GARCH$ model (average return, $\tau = 20$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

Empirical results:

In order to get the following empirical results, we have always looked for an optimal parameter s such that one finds the highest t-statistics (as already described above), or equivalently, for an appropriate start date for the price time series. Moreover, the most meaningful results are found for $\tau = 20$, $\tau = 40$ and $\tau = 60$ days (1 month, 2 months and 3 months) which are compared below.
Looking at the results for DAX (Figures 5.38-5.40), there is a strong bubble signal for the peak in 2000 in case $\tau = 20$, for the peak in 1998 in case $\tau = 40$ and for the peak in 1997 in case $\tau = 40$ and $\tau = 60$.
Analyzing the Gold price, we could not detect a strong enough signal for a bubble in case $\tau = 20$ or $\tau = 40$. But for $\tau = 60$ there is a bubble signal for the peaks in 2006 and 2008 (Figure 5.41).
Finally, fitting the $FTS\ GARCH$ with average return as regression component to the NASDAQ Composite, we manage to reduce the strength of the signal. In case $\tau = 20$ (Figure 5.42) the reduction in the strength of the signal is similar to the application of a Kalman filter (compare Figure 5.28). In order to have a meaningful bubble signal we propose to use $\tau = 40$ days (Figure 5.43). Then there is high significance for a positive γ only around the bubble peak in 2000 (and also for the subsequent crash). A horizon of $\tau = 60$ days seems to be already too large, we could not detect a strong enough bubble signal.

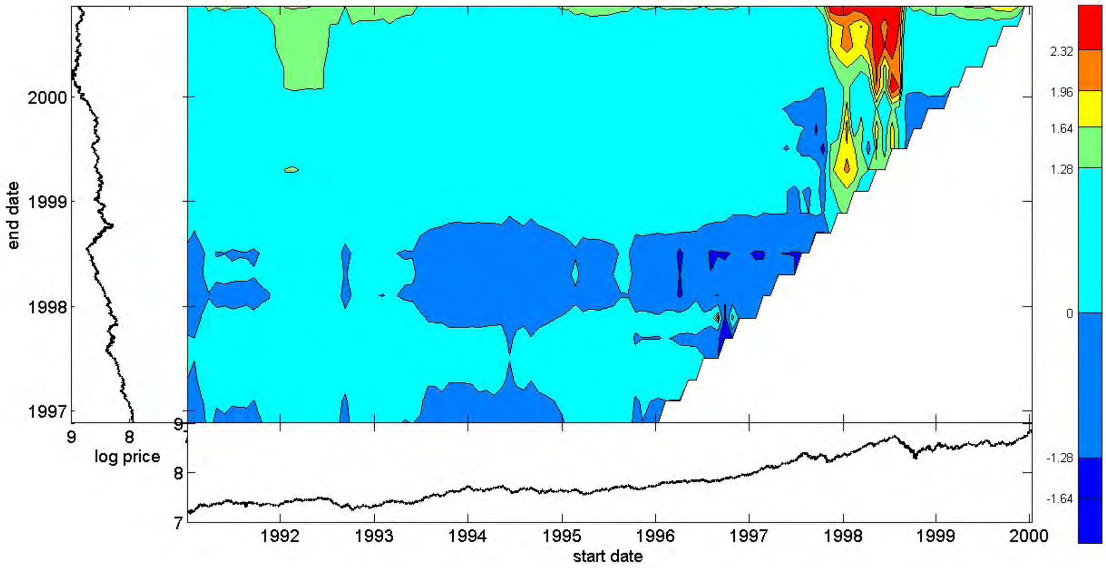


Figure 5.38.: DAX (January 2nd, 1991 - November 10th, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 20$ days) fitted over different time windows

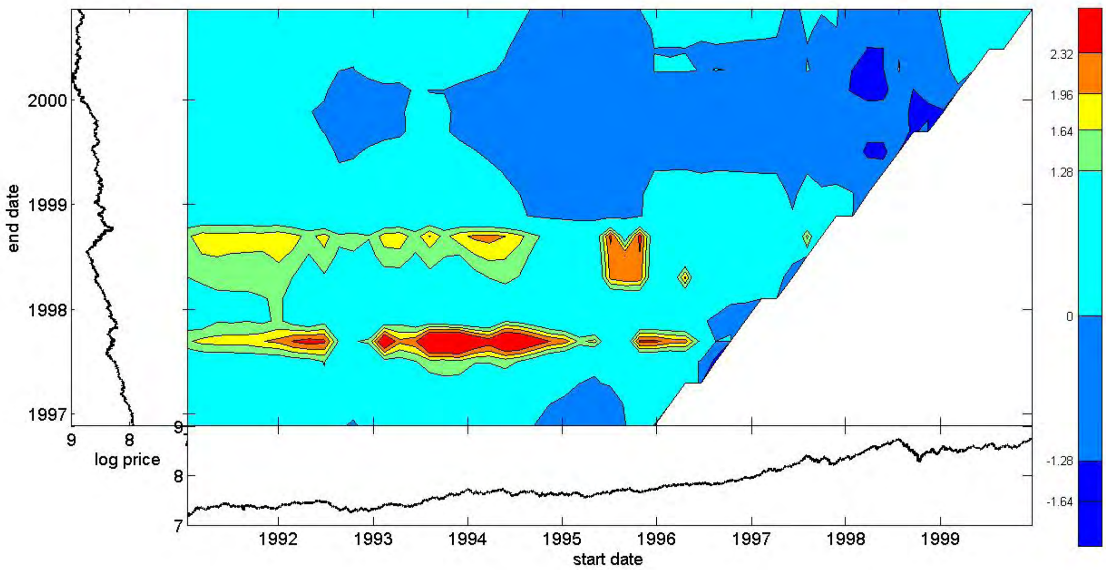


Figure 5.39.: DAX (January 14th, 1991 - November 10th, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 40$ days) fitted over different time windows

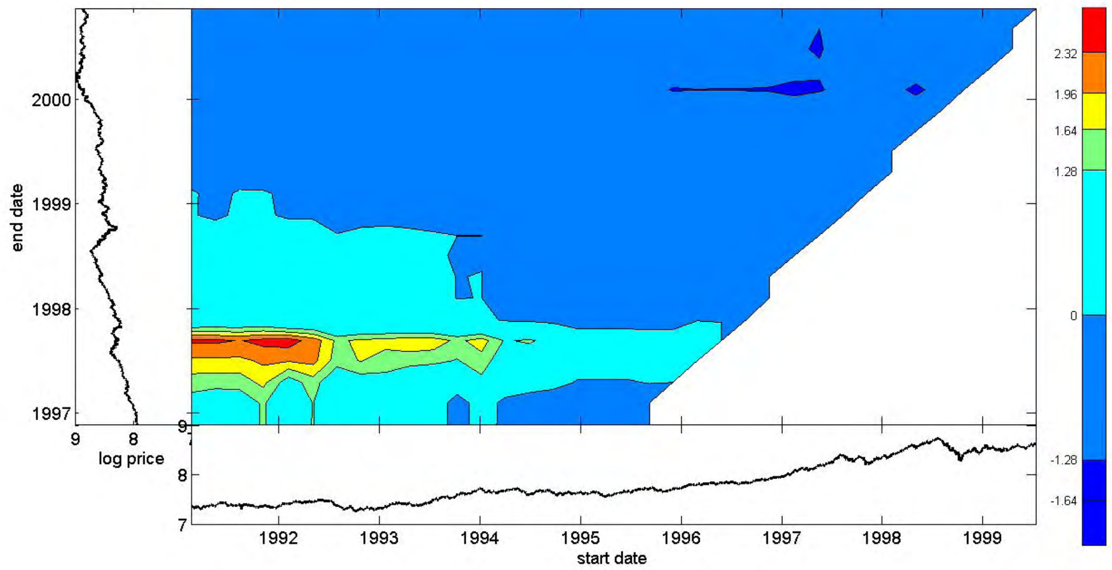


Figure 5.40.: DAX (February 18th, 1991 - November 10th, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 60$ days) fitted over different time windows

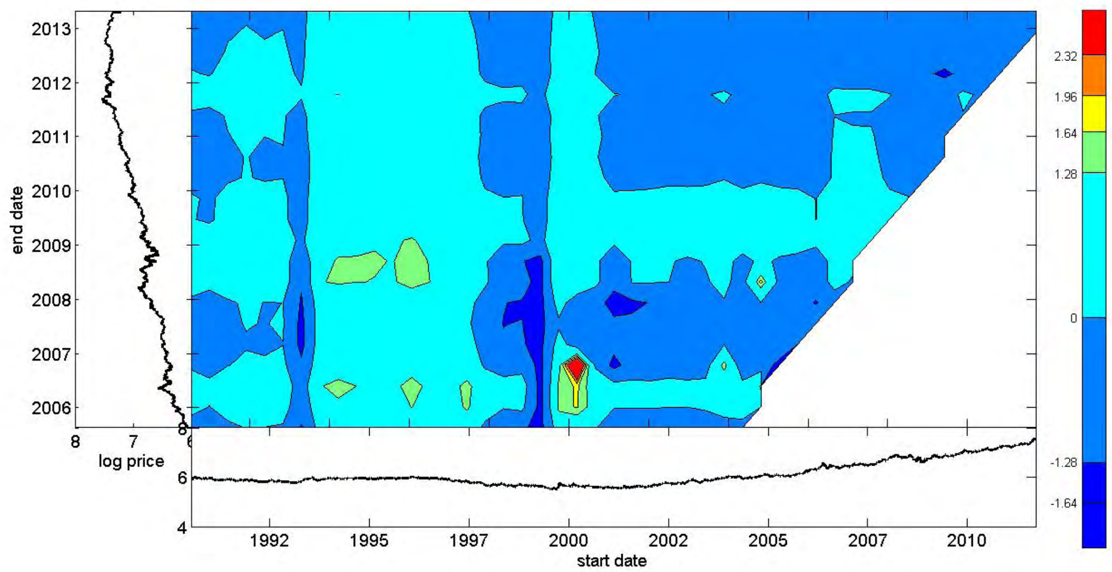


Figure 5.41.: Gold (July 17th, 1990 - April 22nd, 2013): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 60$ days) fitted over different time windows

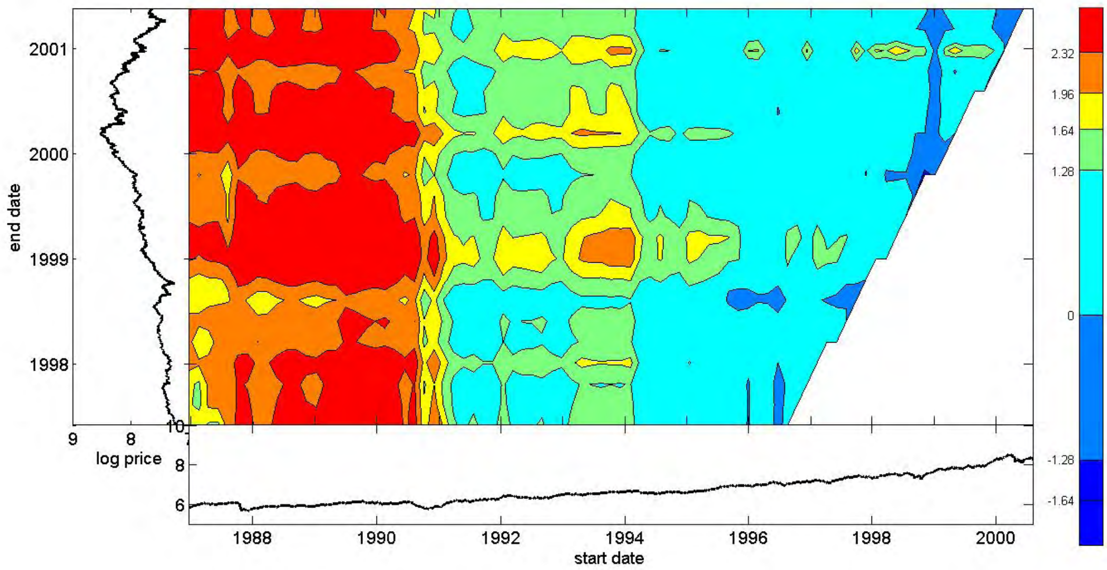


Figure 5.42.: NASDAQ Composite (December 18th, 1986 - May 18th, 2001): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 20$ days) fitted over different time windows

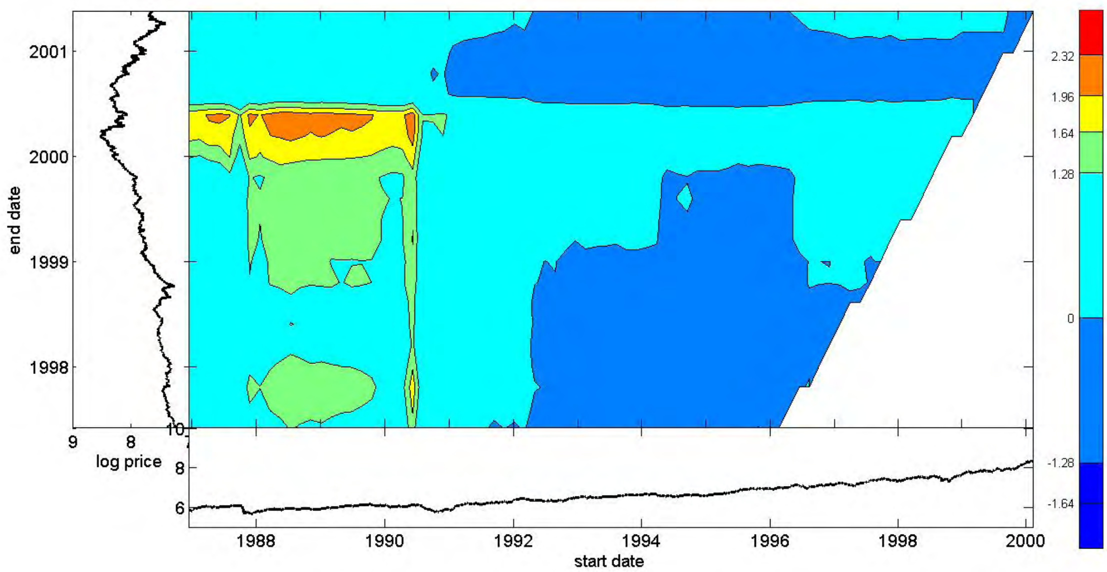


Figure 5.43.: NASDAQ Composite (December 12th, 1986 - May 18th, 2001): t-statistic of γ in an *FTS GARCH* model (regression component = average return over $\tau = 40$ days) fitted over different time windows

5.5.3. Exponentially weighted moving average return

The last method we propose to improve the *FTS GARCH* model with returns as regression component is the usage of exponentially weighted moving average returns instead. The **EWMA return** over a certain horizon of $\tau \in \mathbb{N}$ days is defined as

$$r_{t|s}^{EWMA} := \frac{1-\delta}{1-\delta^{\tau}} \cdot \sum_{k=0}^{\tau-1} \delta^k \cdot r_{s+t\cdot\tau-k}, \qquad \text{for } \; t = 1, \dots, \left\lfloor \frac{n-s}{\tau} \right\rfloor =: T, \tag{5.56}$$

where $s \in \{0, \dots, \tau - 1\}$ is again the start date in the price series $P_s, \dots, P_n$ chosen for the calculation, $\delta = e^{-\alpha}$ and $\alpha > 0$ constant.
Now, we would like to estimate for various time windows $(t_{start}, \dots, t_{end})$, with $t_{start}, \; t_{end} \in \{1, \dots, T\}$ and $t_{start} < t_{end}$, the following *FTS GARCH* process:

$$r_{t|s}^{\tau} \;=\; \mu + \gamma \cdot r_{t-1|s}^{EWMA} + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \tag{5.57}$$

$$h_t \;=\; \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \qquad t = t_{start} + 1, \dots, t_{end}, \tag{5.58}$$

and test again for

$$H_0 : \gamma \leq 0 \;\text{ against }\; H_1 : \gamma > 0. \tag{5.59}$$

Synthetic tests:

As usual, we do some simulation experiments before we apply the method to real world data. We simulate the two types of bubbles ($y_t = \log(P_t)$ and $y_t = r_t$) where we include a period of alternating regimes once again, but now we use the *FTS GARCH* model with exponential weighting of returns from 5.57-5.58 for estimation. The results are quite similar compared to the usage of average returns as regression component (compare Figures 5.34-5.37). They can be found in Figures 5.44-5.47. For both types of simulated bubbles we find a preferable bubble signal compared to the usage daily returns as regression component.

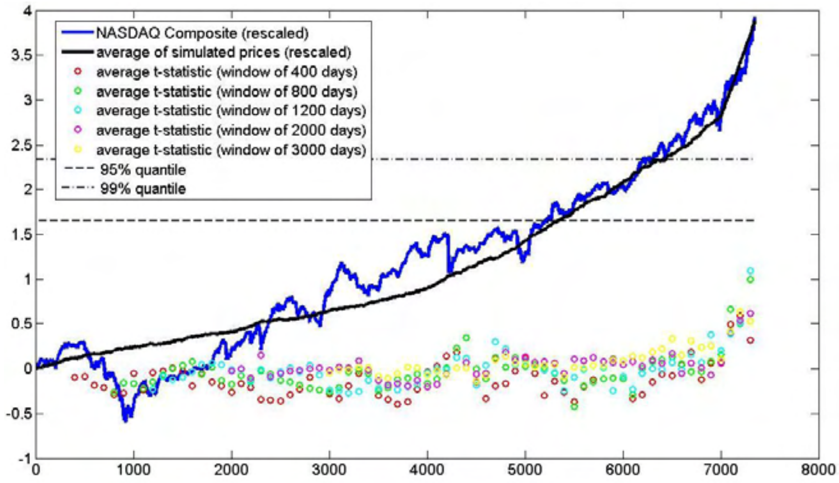


Figure 5.44.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an $FTS\ GARCH$ model (EWMA return, $\tau = 20$, $\alpha = 0.15$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

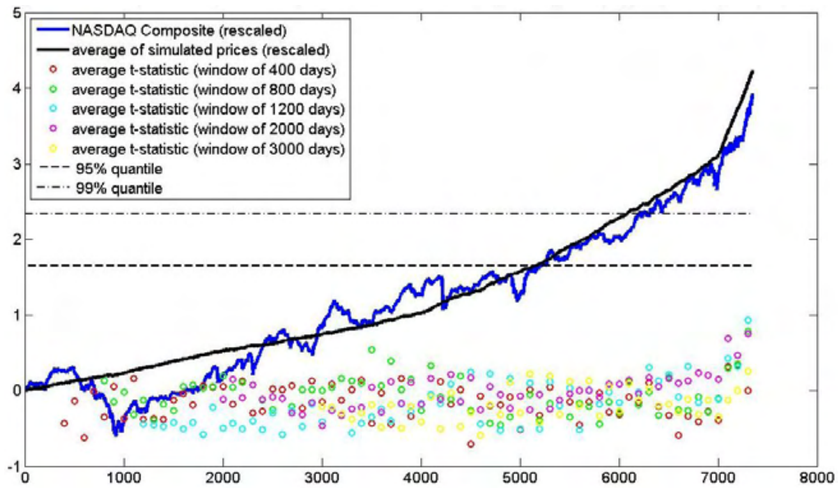


Figure 5.45.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and $FTS\ GARCH$ dynamics ($y_t = \log(P_t)$, coefficients estimated from NASDAQ, $\gamma = 4 \cdot 10^{-4}$) for another 3000 days and finally, a pure $FTS\ GARCH$ regime with even higher $\gamma = 6 \cdot 10^{-4}$ for 350 days. Estimation of an $FTS\ GARCH$ model (EWMA return, $\tau = 40$, $\alpha = 0.1$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

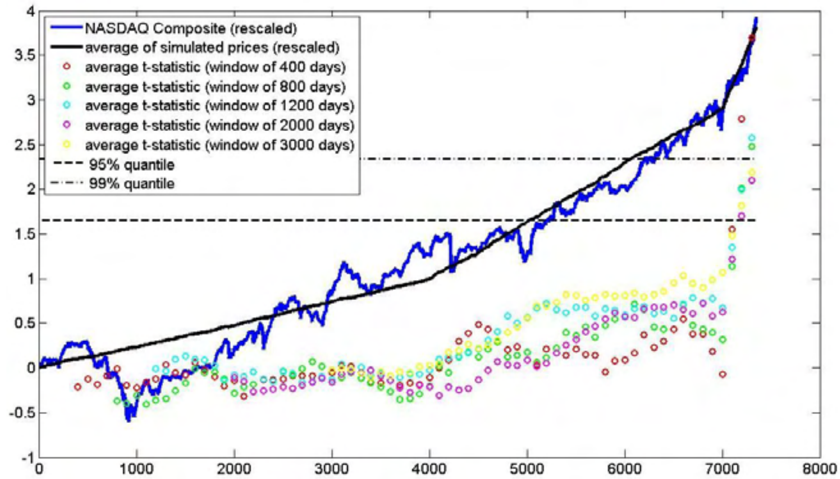


Figure 5.46.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an *FTS GARCH* model (EWMA return, $\tau = 5$, $\alpha = 0.3$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

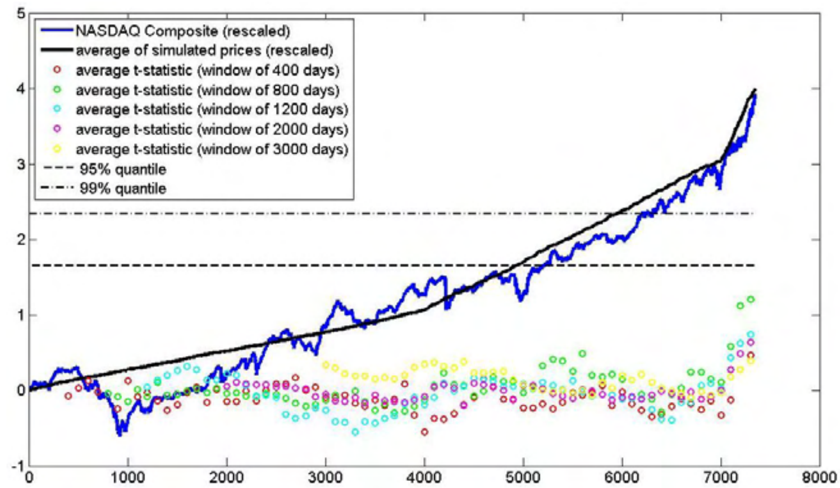


Figure 5.47.: MC Simulation (50 simulations): Sim. trajectories of 4000 days $GARCH(1,1)$, then, alternating with a regime length of 60 days $GARCH(1,1)$ and *FTS GARCH* dynamics ($y_t = r_t$, coefficients estimated from NASDAQ, $\gamma = 0.2$) for another 3000 days and finally, a pure *FTS GARCH* regime with even higher $\gamma = 0.7$ for 350 days. Estimation of an *FTS GARCH* model (EWMA return, $\tau = 20$, $\alpha = 0.15$) from the simulated return series and plot of the average evolution of the t-statistic of γ calculated over different window sizes.

Empirical results:

In order to find plausible results for real data (we choose again the DAX, Gold and NASDAQ time series such that we are able to directly compare the results), we need to look again for an optimal parameter s (an optimal start date for the analyzed time series) in order to maximize the t-statistics of γ. Analog to the usage of average returns, we observe the most meaningful results for $\tau = 20$, $\tau = 40$ and $\tau = 60$ days (1 month, 2 months and 3 months).
As usual, let us start with the analysis of the DAX time series (Figures 5.48-5.50). For $\tau = 20$ days we have a clear bubble signal around the peak in 2000 for various start dates of the time window. Using a horizon of $\tau = 40$, there is very strong positive feedback on return for the whole period 1998-2000. For windows ending at the peak in 2000 the t-statistic of γ is above the 99% quantile for almost all start dates we consider. There is also a strong bubble signal before the peak in 1998. The peak in 1997 is detected best by using $\tau = 60$ days (independent of the chosen start date for the window).
Next, we analyze the t-statistics of γ for the Gold price (Figures 5.51-5.53). In case $\tau = 20$ the t-statistic is above the 90% quantile for many windows ending between 2011 and 2013. For $\tau = 40$ the signal is even stronger, telling us there has been a bubble in Gold since 2009. Enlarging τ to 60 days we detect a Gold bubble already in 2006, where the t-statistic of γ has its peaks around the peaks in the Gold price. The highest t-statistics are found for 2008 (in particular, there is strong positive feedback during the crash period).
Finally, let us consider the NASDAQ Composite time series (Figures 5.54-5.56). Consistent with our previous results for NASDAQ, a horizon of $\tau = 20$ days seems too small to identify the critical time of the bubble (we find high significance for a bubble during the whole period 1997-2001). But if one increases τ to 40, the time around the bubble peak in 2000 can be identified as the most critical period. The 90% quantile is reached already at the beginning of 1999, and in 2000 even the 97.5% quantile is exceeded. Analog to the Gold price, we find also positive feedback during the time of the subsequent crash in 2001. The results in case of $\tau = 60$ days are quite similar, however, there is an even stronger signal for the crash.

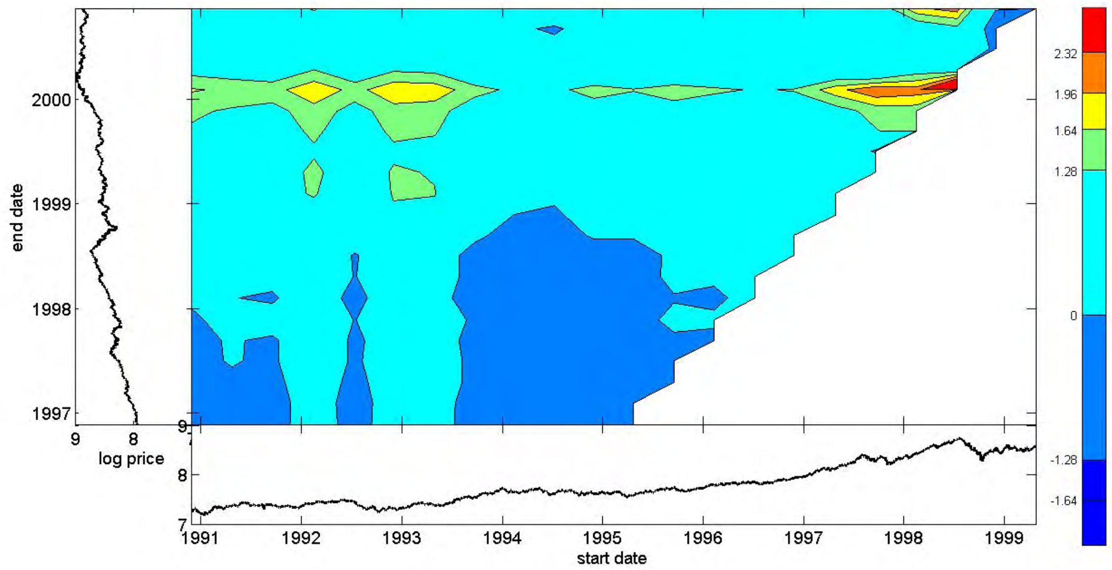


Figure 5.48.: DAX (January 2^{nd}, 1991 - November 10^{th}, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 20$ days, $\alpha = 0.15$) fitted over different time windows

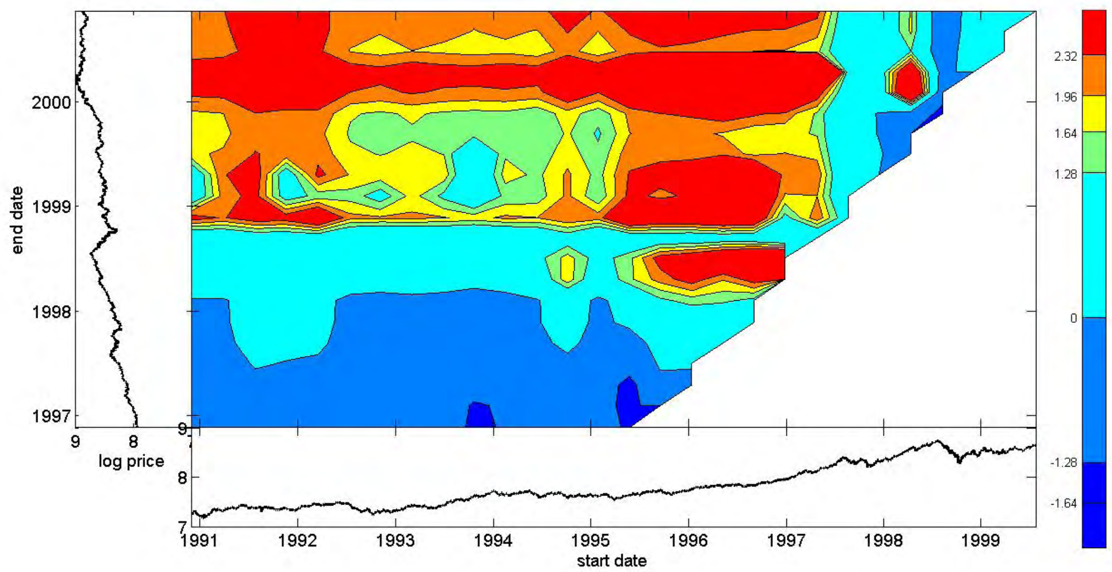


Figure 5.49.: DAX (January 31^{st}, 1991 - November 10^{th}, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

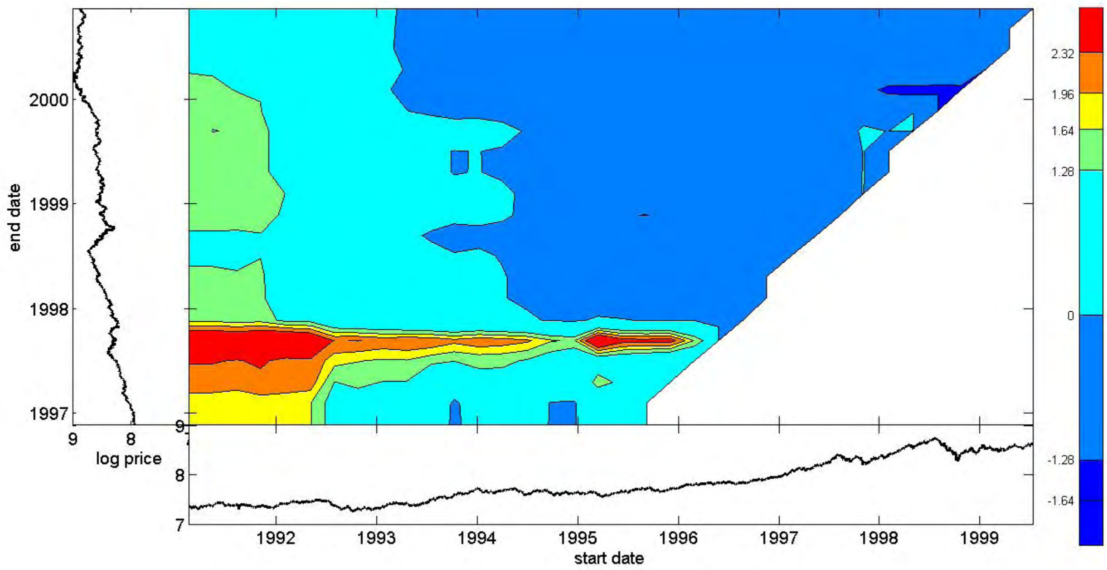


Figure 5.50.: DAX (February 18th, 1991 - November 10th, 2000): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 60$ days, $\alpha = 0.05$) fitted over different time windows

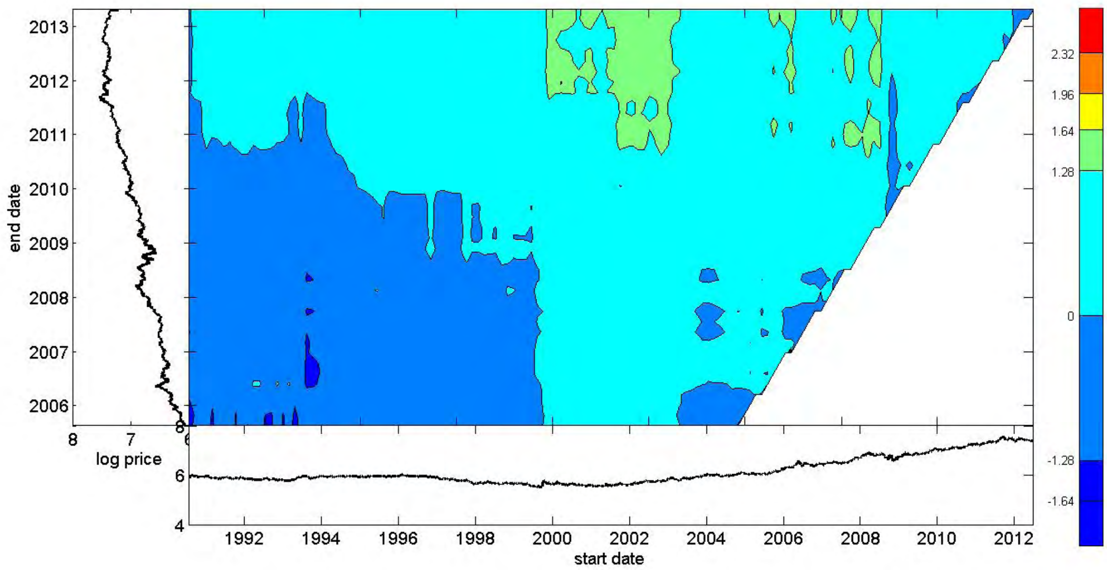


Figure 5.51.: Gold (July 18th, 1990 - April 22nd, 2013): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 20$ days, $\alpha = 0.15$) fitted over different time windows

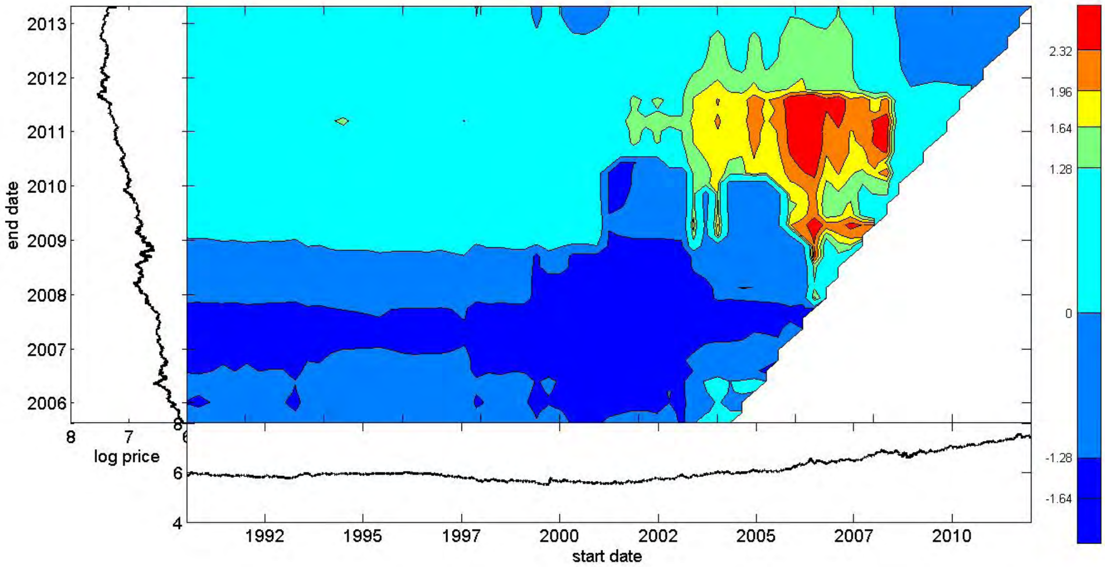


Figure 5.52.: Gold (July 6th, 1990 - April 22nd, 2013): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

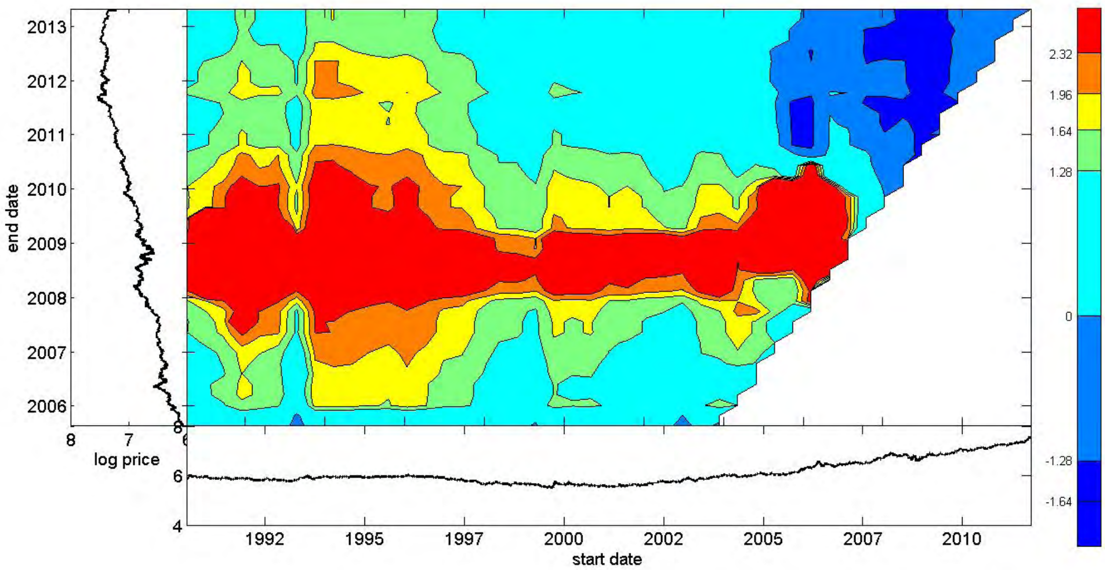


Figure 5.53.: Gold (July 16th, 1990 - April 22nd, 2013): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 60$ days, $\alpha = 0.05$) fitted over different time windows

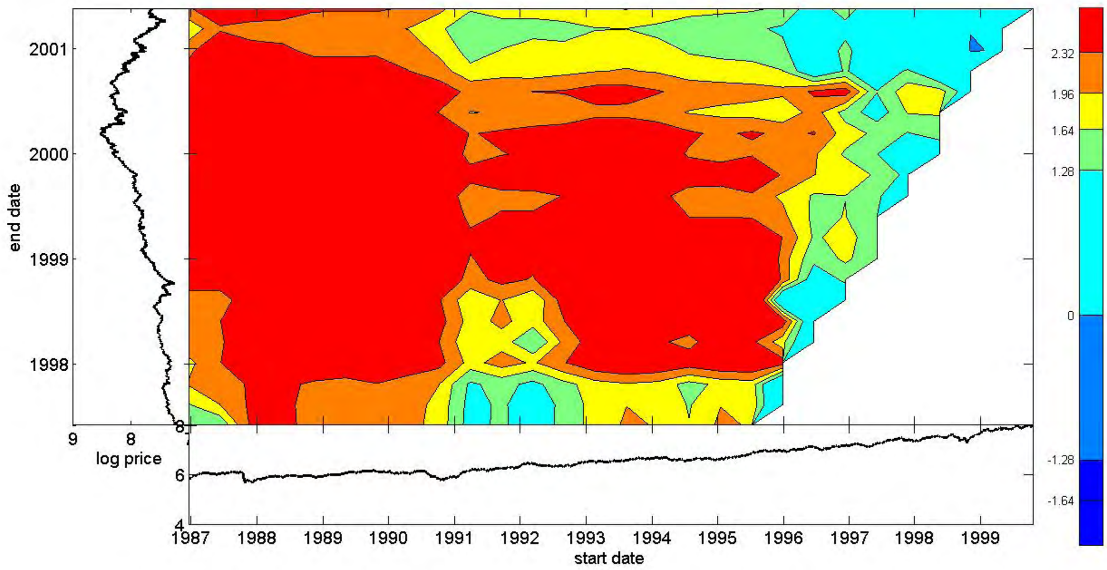


Figure 5.54.: NASDAQ (December 17th, 1986 - May 18th, 2001): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 20$ days, $\alpha = 0.15$) fitted over different time windows

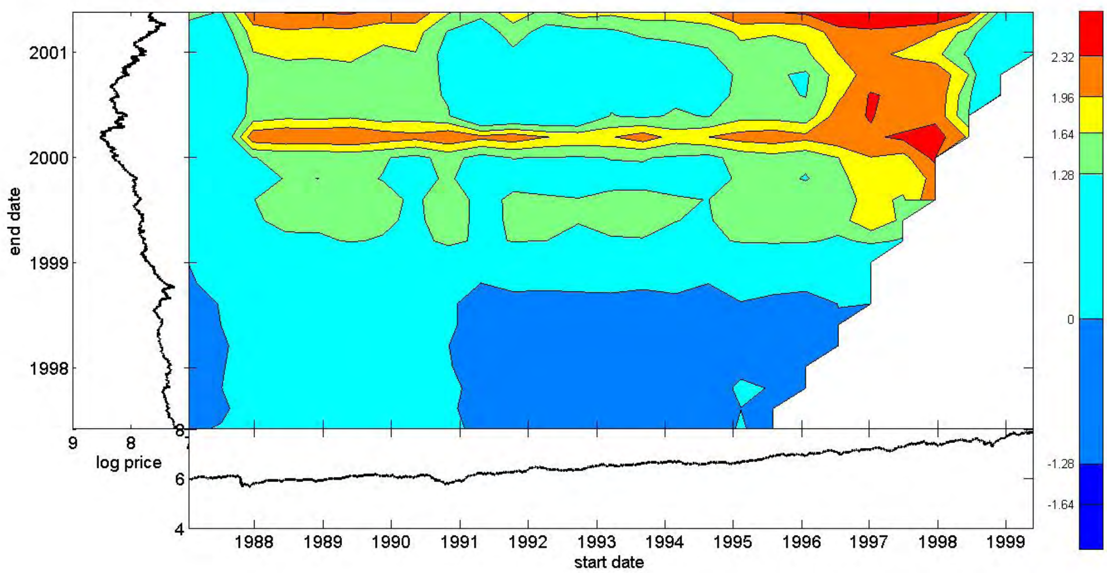


Figure 5.55.: NASDAQ (January 13th, 1987 - May 18th, 2001): t-statistic of γ in an *FTS GARCH* model (regression component = EWMA return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

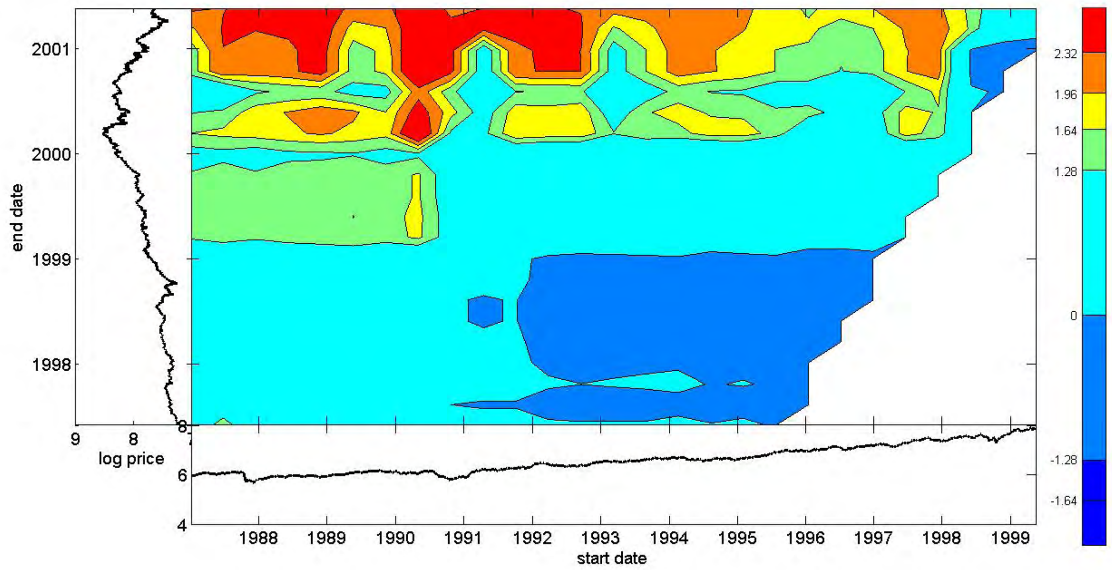


Figure 5.56.: NASDAQ (January 6th, 1987 - May 18th, 2001): t-statistic of γ in an $FTS\ GARCH$ model (regression component = EWMA return over $\tau = 60$ days, $\alpha = 0.05$) fitted over different time windows

5.6. Positive feedback of volatility on return

The goal of this section is to find an appropriate way how to incorporate the ideas of Andersen and Sornette (2004) [AS] in an *FTS GARCH* model. They define the so called **nonlinear rational expectation bubble model** that considers not only positive feedback of return on return, but also positive feedback of volatility on return. In comparison to previous bubble models where the price process P_t is a jump-diffusion process of the form

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t - \kappa dJ_t, \tag{5.60}$$

where $\mu > 0$ is the drift, $\sigma > 0$ the volatility, W_t a standard Brownian motion and J_t a jump process modeling corrections or crashes of size $\kappa \cdot P_t$, they let $\mu = \mu(P_t)$ and $\sigma = \sigma(P_t)$ be nonlinearly dependent on the instantaneous price. More specifically,

$$\mu(P_t)P_t = \frac{m}{2P_t}\left(P_t\sigma(P_t)\right)^2 + \mu_0\left(\frac{P_t}{P_0}\right)^m, \tag{5.61}$$

$$\sigma(P_t)P_t = \sigma_0\left(\frac{P_t}{P_0}\right)^m, \tag{5.62}$$

where $P_0,\ \mu_0,\ \sigma_0,\ m > 0$ are fixed parameters of the model.
Hence, the authors model the possibility that a speculative bubble exhibits two behaviors simultaneously that are usually seen as conflicting: **(1) super-exponential price growth** and **(2) accompanying volatility growth** (which in turn leads to return growth due to equation 5.61 - positive feedback of volatility on return). They call this type of speculative bubble a "**fearful singular bubble**" (risk-aversion increases as the bubble builds up). These two properties seem to disagree due to the leverage effect (compare Chapter 2, stylized fact 6): negative returns lead to a rise in volatility and positive returns lead to a fall in volatiliy. However, the leverage effect describes only a short-term correlation and does not exclude isolated periods of positive correlation between return and future volatility. Of course, there can be market regimes where only one of the two properties is satisfied as well.
As a last step in this chapter, we thus include the possibility of positive feedback of volatility on return (growth of return due to growth of volatility) in our model. Since we found the most meaningful empirical results when we used exponentially weighted moving average returns, we use this type of the *FTS GARCH* model as a starting point and add the exponentially weighted average of the absolute returns as a second regression component. For this purpose let us define the **EWMA absolute return** over a certain horizon of $\tau \in \mathbb{N}$ days as

$$r_{t|s}^{EWMA\ abs} := \frac{1-\delta}{1-\delta^\tau} \cdot \sum_{k=0}^{\tau-1} \delta^k \cdot |r_{s+t\cdot\tau-k}|, \quad \text{for } t = 1, \ldots, \left[\frac{n-s}{\tau}\right] =: T, \tag{5.63}$$

where $s \in \{0, \ldots, \tau - 1\}$ is as usual the start date in the price series $P_s, \ldots, P_n$ chosen for the calculation. Hence, we propose the following form of the *FTS GARCH* to model a "fearful singular bubble" (including the possibility of positive feedback of volatility on return):

$$r_{t|s}^{\tau} = \mu + \gamma_1 \cdot r_{t-1|s}^{EWMA} + \gamma_2 \cdot r_{t-1|s}^{EWMA\ abs} + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \tag{5.64}$$

$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}, \quad t = t_{start} + 1, \ldots, t_{end}, \tag{5.65}$$

where $(t_{start}, \ldots, t_{end})$ are all possible windows with $t_{start},\ t_{end} \in \{1, \ldots, T\}$ and $t_{start} < t_{end}$. Note that this is still a $GARCH(1,1)$ regression model as defined in (3.22), where $b := (\mu, \gamma_1, \gamma_2)$ and $x_t := (1, r_{t-1|s}^{EWMA}, r_{t-1|s}^{EWMA\ abs})^T$, which means we can use the same maximum likelihood procedure to consistently estimate the parameters of this model (compare Section 3.3).
We perform t-tests for γ_1 and γ_2 of the form

$$H_0 : \gamma_1 \leq 0 \ \text{ against } \ H_1 : \gamma_1 > 0 \tag{5.66}$$

as well as

$$H_0 : \gamma_2 \leq 0 \ \text{ against } \ H_1 : \gamma_2 > 0. \tag{5.67}$$

The rejection of H_0 for γ_1 will then indicate positive feedback of return on return and the rejection of H_0 for γ_2 positive feedback of volatility on return. If we can reject both null hypotheses for the same time window, we have indication for a "fearful bubble".

Empirical results:

As usual, we analyze the DAX, Gold and NASDAQ time series and maximize the t-statistics by choosing an appropriate start date for the analyzed time series. Note that we need to analyze the t-statistics of γ_1 and γ_2 separately. In Figures 5.57-5.59 we present the results of each time series for a single choice of τ.
For the DAX one finds strong positive feedback of return on return (γ_1) and also positive feedback of volatility on return (γ_2) around the bubble peak in 2000 (in particular for windows starting in 1991 or 1996). For the maximum value in 1998 there are also start dates with significantly positive values for either γ_1 or γ_2. Moreover, one can detect high t-statistics for γ_2 around the peak in 1997, however, for these windows γ_1 is often negative.
In the Gold time series there is positive feedback of return for the whole period 2006-2013, attaining its maximum in 2008. Around the peak in 2006 and during the period 2009-2013 we find positive feedback of volatility in addition.
The t-statistics obtained from analyzing NASDAQ show a pretty clear pattern. For various start dates we detect strong evidence for growth in return due to growth in both return and volatility at the peak in 2000 and in advance. Also, during the crash period in 2001 there was strong positive feedback of return.

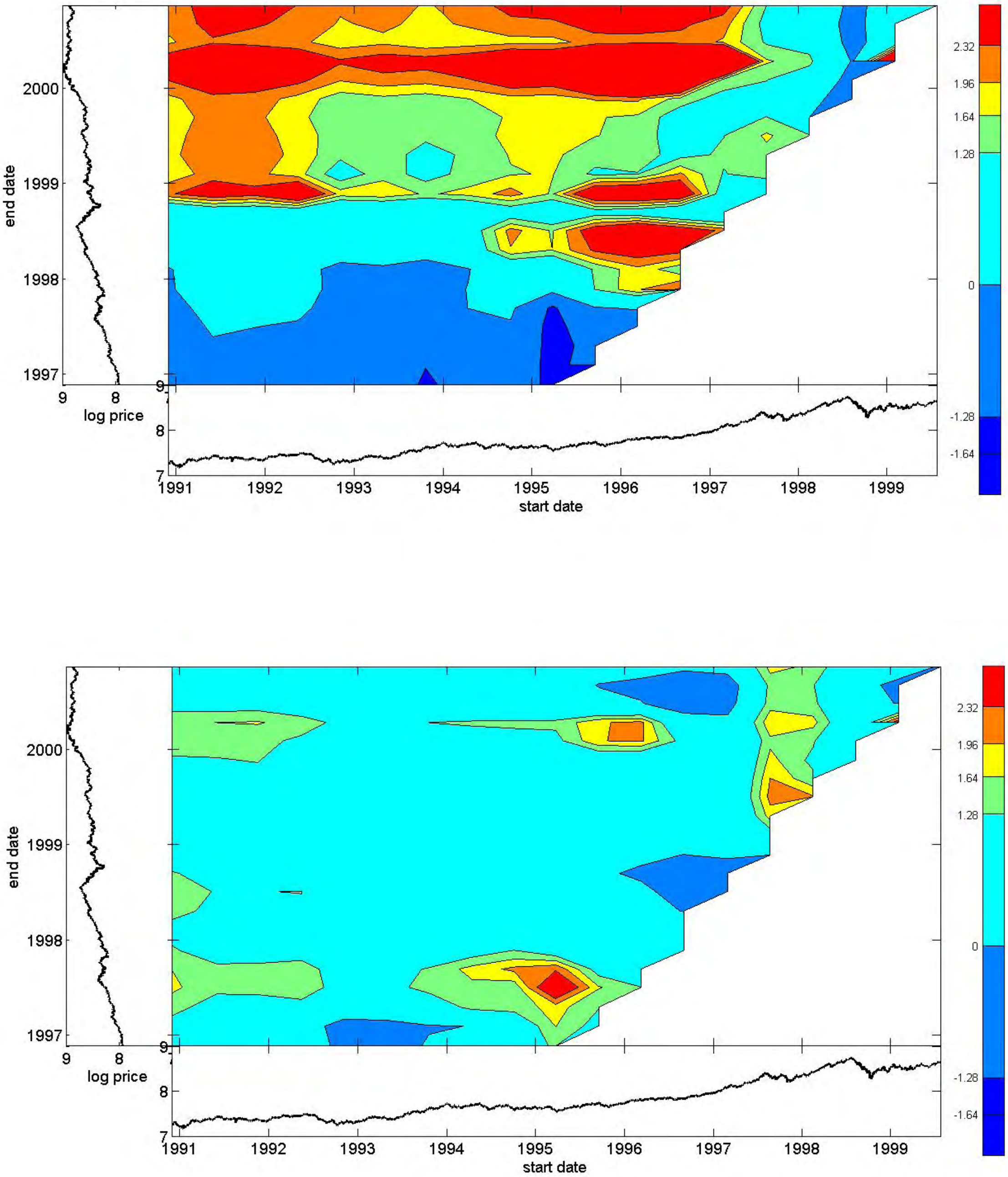


Figure 5.57.: DAX (January 31st, 1991 - November 10th, 2000): t-statistic of γ_1 (upper panel) and γ_2 (lower panel) in an *FTS GARCH* model (regression components = EWMA return and absolute value of EWMA return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

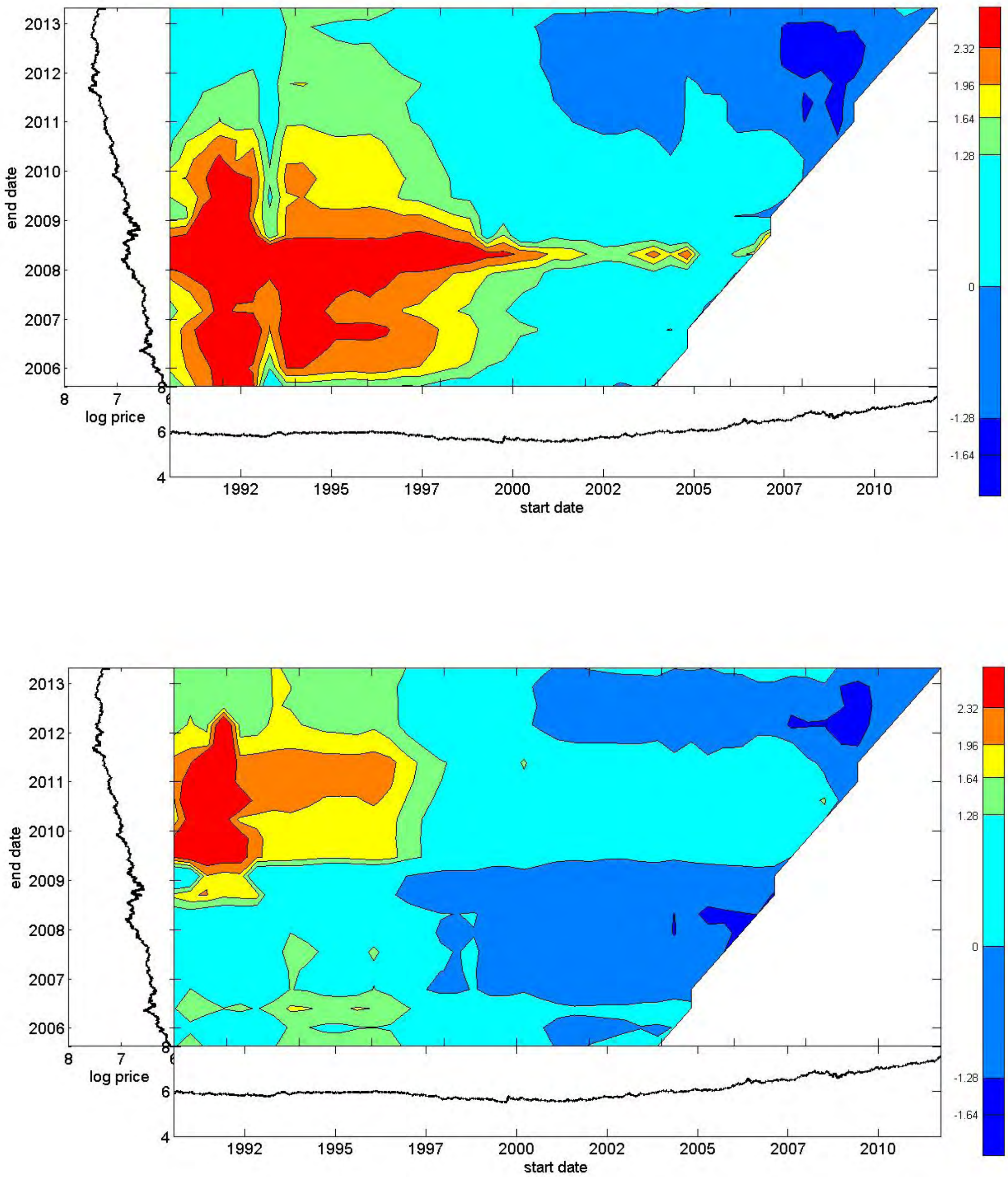


Figure 5.58.: Gold (July 13th, 1990 - April 22nd, 2013): t-statistic of γ_1 (upper panel) and γ_2 (lower panel) in an *FTS GARCH* model (regression components = EWMA return and absolute value of EWMA return over $\tau = 60$ days, $\alpha = 0.05$) fitted over different time windows

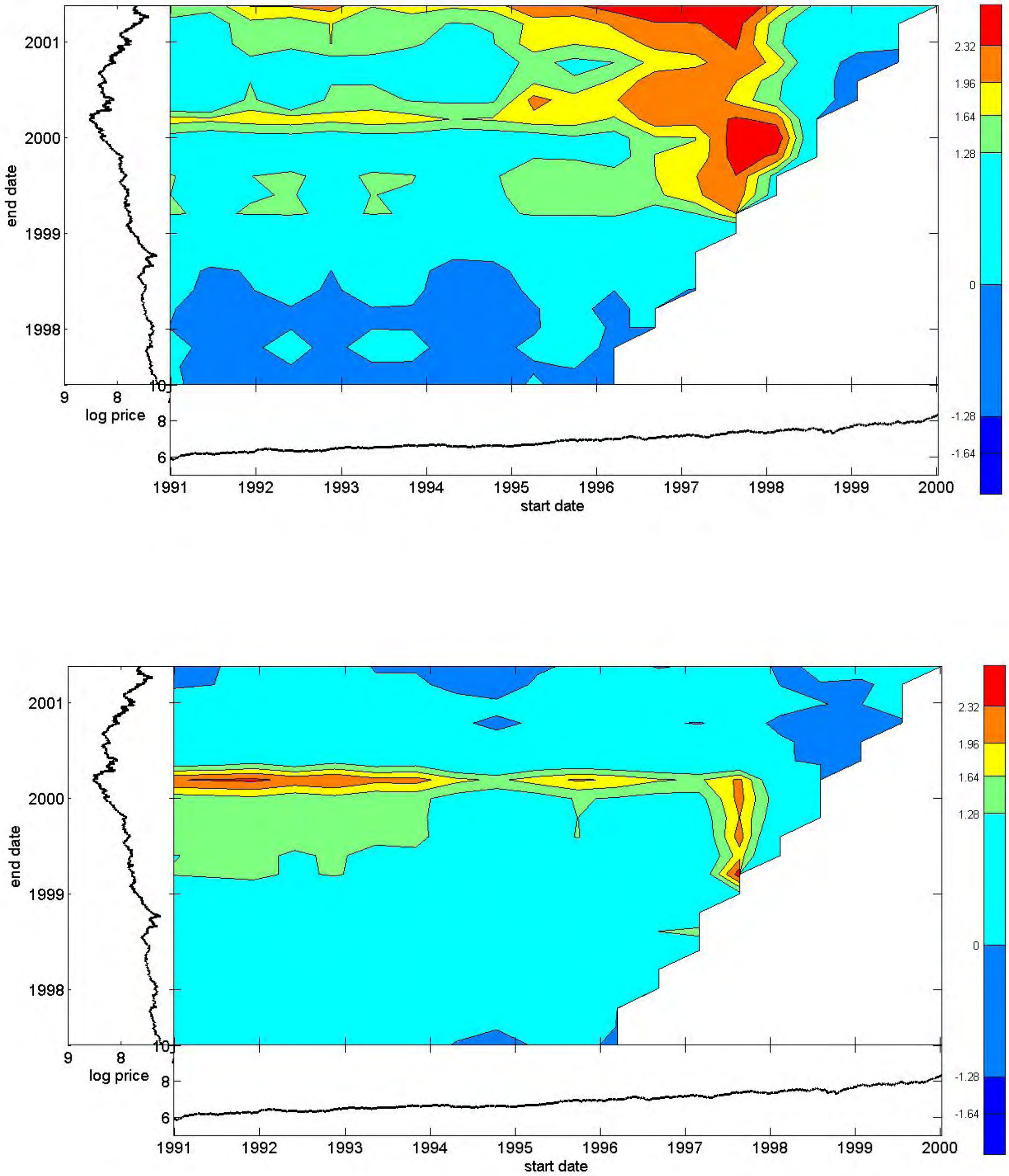


Figure 5.59.: NASDAQ (December 26th, 1990 - May 18th, 2001): t-statistic of γ_1 (upper panel) and γ_2 (lower panel) in an *FTS GARCH* model (regression components = EWMA return and absolute value of EWMA return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

Alternatively, we propose to analyze a reduced form of 5.64 by defining $\gamma_2 := \gamma_1$. This leads

$$r^{\tau}_{t|s} = \mu + \gamma_1 \cdot (r^{EWMA}_{t-1|s} + r^{EWMA\ abs}_{t-1|s}) + \epsilon_t, \text{ with } \epsilon_t \sim N(0, h_t), \quad (5.68)$$
$$h_t = \alpha_0 + \alpha_1 \cdot \epsilon^2_{t-1} + \beta_1 \cdot h_{t-1}, \quad t = t_{start} + 1, \ldots, t_{end}. \quad (5.69)$$

Summing EWMA returns and EWMA absolute returns means we use an exponentially weighted average of only the positive returns. Hence, there could be stronger positive feedback on this combined form of the regression component in times of a bubble. Moreover, this *FTS GARCH* model has the advantage of only one regression component. As usual, we propose the following t-test:

$$H_0 : \gamma_1 \leq 0 \text{ against } H_1 : \gamma_1 > 0. \quad (5.70)$$

Empirical results:

We apply the common procedure to receive empirical results for the t-statistic of γ_1 for DAX, Gold and NASDAQ. In a nutshell, this method yields very nice results, showing clear evidence for strong positive feedback at all bubble peaks we have usually analyzed. Apparently this tool could serve as a good early warning system.

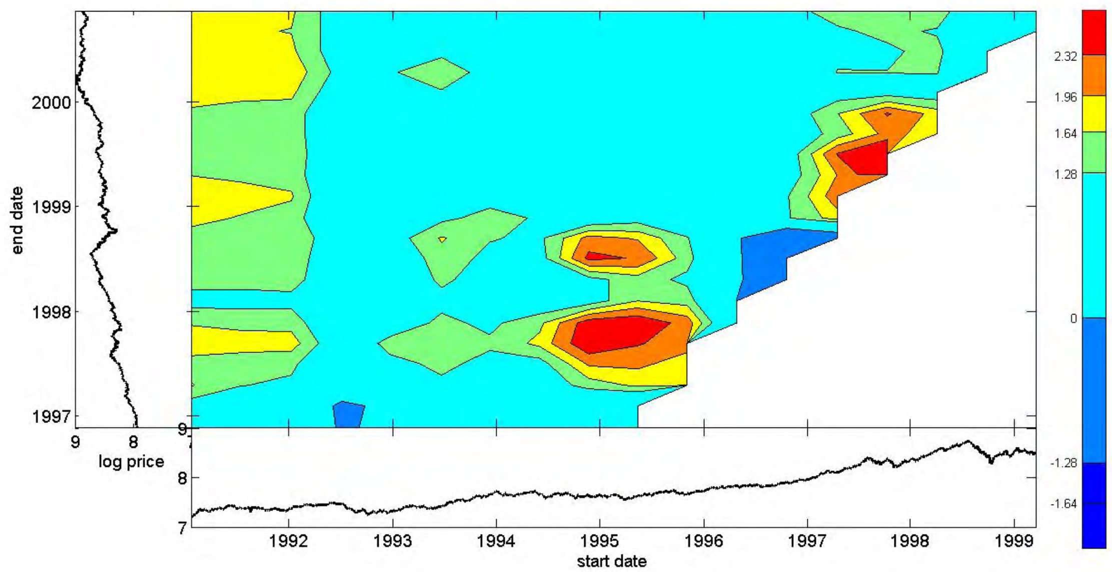


Figure 5.60.: DAX (January 22nd, 1991 - November 10th, 2000): t-statistic of γ_1 in an *FTS GARCH* model (regression component = sum of EWMA return and EWMA absolute return over $\tau = 40$ days, $\alpha = 0.1$) fitted over different time windows

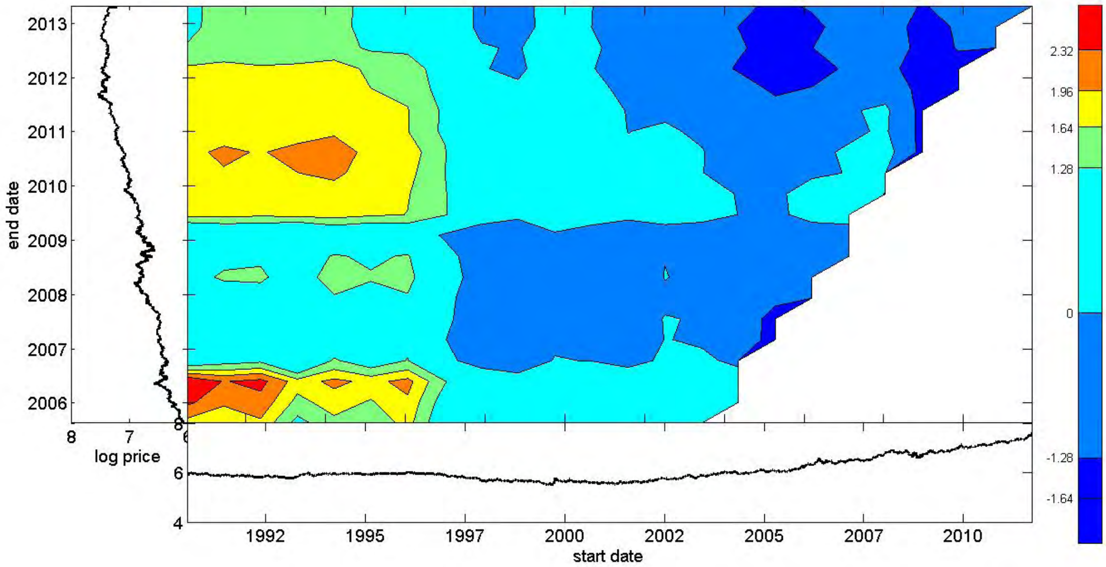


Figure 5.61.: Gold (July 18th, 1990 - April 22nd, 2013): t-statistic of γ_1 in an *FTS GARCH* model (regression component = sum of EWMA return and EWMA absolute return over τ = 60 days, $\alpha = 0.1$) fitted over different time windows

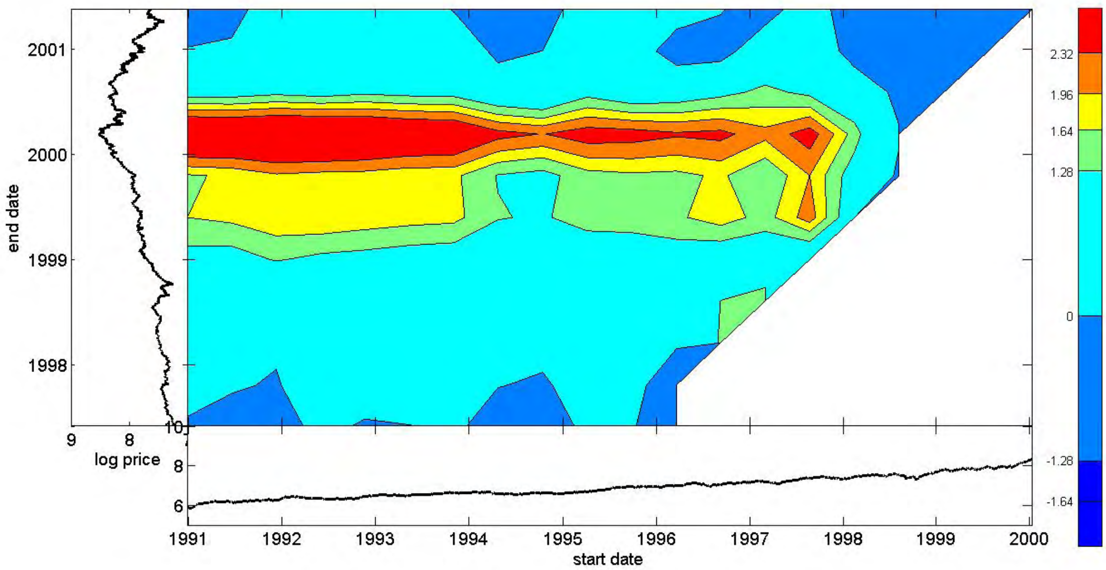


Figure 5.62.: NASDAQ Composite (December 27th, 1990 - May 18th, 2001): t-statistic of γ_1 in an *FTS GARCH* model (regression component = sum of EWMA return and EWMA absolute return over τ = 40 days, $\alpha = 0.1$) fitted over different time windows

6 Markov Switching Model

In order to define the Markov switching model and estimate it's coefficients we follow the way of J.D. Hamilton (1994) [H].
We start off by considering a process s_t that should describe changes in regime. We model s_t as an N-state **Markov chain** assuming values in $\{1, \dots, N\}$ where the probability that s_t equals j depends on the past only through the latest value s_{t-1}:

$$\mathbb{P}[s_t = j | s_{t-1} = i, s_{t-2} = k, \dots] = \mathbb{P}[s_t = j | s_{t-1} = i] =: p_{ij}. \tag{6.1}$$

Therefore, p_{ij} is the probability that state i will be followed by state j. We call $\{p_{ij}\}_{i,j=1,\dots,N}$ **transition probabilities** and collect them in the **transition matrix** P:

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \cdots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{pmatrix} \tag{6.2}$$

Note that the element in row j and column i is p_{ij}, the probablity to switch from state i to j. The transition probabilities satisfy

$$\sum_{j=1}^{N} p_{ij} = 1 \quad \forall i \in \{1, \dots, N\}. \tag{6.3}$$

Hence, the sum of each column in P is 1.
It is quite useful to have a representation of a Markov chain with a vector autoregression. For this reason let us define

$$\xi_t := \begin{cases} (1, 0, 0, \dots, 0)^T, & \text{if } s_t = 1, \\ (0, 1, 0, \dots, 0)^T, & \text{if } s_t = 2, \\ \quad\vdots & \quad\vdots \\ (0, 0, 0, \dots, 1)^T, & \text{if } s_t = N. \end{cases} \tag{6.4}$$

If the process is in state i at time t, then the j-th element of ξ_{t+1} is a random variable that is 1 with probability p_{ij} and 0 otherwise. Thus, its expectation is p_{ij}, and we find

$$\mathbb{E}[\xi_{t+1}|s_t = i] = \begin{pmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iN} \end{pmatrix}, \tag{6.5}$$

the i-th column of the transition matrix P. Recall that ξ_t is simply the i-th column of the identity matrix I_N if $s_t = i$. Thus, equation 6.5 yields

$$\mathbb{E}[\xi_{t+1}|\xi_t] = P \cdot \xi_t \tag{6.6}$$

and using the Markov property from equation 6.1 one finds

$$\mathbb{E}[\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots] = P \cdot \xi_t. \tag{6.7}$$

This result let's us express a Markov chain in form of a **first-order vector autoregression** for ξ_t:

$$\xi_{t+1} = P \cdot \xi_t + v_{t+1}, \tag{6.8}$$

where

$$v_{t+1} := \xi_{t+1} - \mathbb{E}[\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots]. \tag{6.9}$$

Note that the innovation process v_t is a martingale difference sequence that is on average 0 and impossible to forecast on the basis of previous states of the process.
From 6.8 we can also find ***m*-period-ahead forecasts** by using

$$\xi_{t+m} = v_{t+m} + P \cdot v_{t+m-1} + \ldots + P^{m-1} \cdot v_{t+1} + P^m \cdot \xi_t, \tag{6.10}$$

where P^m is the matrix multiplication of P by itself m times, and taking the conditional expectation

$$\mathbb{E}[\xi_{t+m}|\xi_t, \xi_{t-1}, \ldots] = P^m \cdot \xi_t. \tag{6.11}$$

Note that the element in the j-th row and i-th column of P^m is $\mathbb{P}[s_{t+m} = j|s_t = i]$, the probability that an observation from regime i will be followed m periods later by an observation from regime j.

6.1. Maximum likelihood estimation

Let us denote by y_t a vector of observed endogenous variables and by x_t a vector of observed exogenous variables. Then, $\mathcal{Y}_t := (y_t^T, y_{t-1}^T, \ldots, y_{-m}^T, x_t^T, x_{t-1}^T, \ldots, x_{-m}^T)^T$ is the vector of all observations until time t. We assume the conditional density of y_t in regime j to be given as

$$f(y_t|s_t = j, x_t, \mathcal{Y}_{t-1}; \alpha), \tag{6.12}$$

where α is a parameter vector characterizing the density. We collect the N different densities in a vector η_t. Note that we assume in 6.12 that the conditional density depends only on the current state:

$$f(y_t|s_t = j, x_t, \mathcal{Y}_{t-1}; \alpha) = f(y_t|s_t = j, s_{t-1} = i, s_{t-2} = k, \ldots, x_t, \mathcal{Y}_{t-1}; \alpha). \tag{6.13}$$

In terms of the Markov chain s_t we assume independence of past x_t's and y_t's in addition

$$\mathbb{P}[s_t = j | s_{t-1} = i, s_{t-2} = k, \dots, x_t, \mathcal{Y}_{t-1}] = \mathbb{P}[s_t = j | s_{t-1} = i] = p_{ij} \tag{6.14}$$

and that x_t is strictly exogenous, meaning it is independent of s_τ for all t and τ.
Next, we would like to collect all parameters characterizing y_t in a vector θ, i.e. α and all transition probabilities p_{ij}. The main goal of this section is to find an appropriate algorithm to estimate θ based on all observations in $\mathcal{Y}_T$. However, let us assume for the moment we know the parameter vector θ. Then, one can find an optimal inference about the regime at each time t based on θ and all observations up to time t:

$$\mathbb{P}[s_t = j | \mathcal{Y}_t; \theta] \quad \forall\, j = 1, \dots, N, \tag{6.15}$$

which are collected in a vector denoted $\hat{\xi}_{t|t}$. Moreover, consider also forecasts of how likely the process is to be in regime j at time $t+1$ based on θ and observations until t:

$$\mathbb{P}[s_{t+1} = j | \mathcal{Y}_t; \theta] \quad \forall\, j = 1, \dots, N, \tag{6.16}$$

which are collected in $\hat{\xi}_{t+1|t}$.
Given a starting value $\hat{\xi}_{1|0}$ and the assumed θ the **optimal inference and forecast** for each time t is found by iterating on the equations

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{\mathbb{1}^T(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{6.17}$$

$$\hat{\xi}_{t+1|t} = P \cdot \hat{\xi}_{t|t}, \tag{6.18}$$

where $\odot$ stands for element-by-element multiplication and $\mathbb{1}$ is an $N \times 1$ vector of 1s. The j-th element of η_t is $f(y_t | s_t = j, x_t, \mathcal{Y}_{t-1}; \theta)$. At the same time one can evaluate the **log likelihood function** $L_T(\theta)$ for the observations $\mathcal{Y}_T$ at θ by

$$L_T(\theta) = \sum_{t=1}^{T} \log f(y_t | x_t, \mathcal{Y}_{t-1}; \theta), \tag{6.19}$$

where

$$f(y_t | x_t, \mathcal{Y}_{t-1}; \theta) = \mathbb{1}^T(\hat{\xi}_{t|t-1} \odot \eta_t). \tag{6.20}$$

A proof for this algorithm can be found in Appendix C. However, the choice of the starting value $\hat{\xi}_{1|0}$ is still an open question. There are several possibilites. One approach is to set $\hat{\xi}_{1|0} = \rho$, where ρ is a constant vector of nonnegative values summing to 1 (e.g. $\rho = N^{-1}\mathbb{1}$). Another possibility is to estimate ρ by maximum likelihood along with θ with constraints $\mathbb{1}^T\rho = 1$ and $\rho_j \geq 0$ for $j = 1, \dots, N$.
Now, let us generalize the probabilities from equations 6.15 and 6.16 by considering $\mathbb{P}[s_t = j | \mathcal{Y}_\tau; \theta]$ for arbitrary t and τ. We collect them in a vector $\hat{\xi}_{t|\tau}$. In case $t > \tau$ this is a forecast about the state at t, whereas in case $t < \tau$ it is the smoothed inference about the state the process was in at t based on information up to time τ.
Recall the formula for the conditional expectation of ξ_{t+m} in equation 6.11. Taking on both sides the expectation conditional on $\mathcal{Y}_t$ yields the optimal ***m*-period-ahead forecast**

$$\hat{\xi}_{t+m|t} = P^m \cdot \hat{\xi}_{t|t}, \tag{6.21}$$

where $\hat{\xi}_{t|t}$ comes from the iteration in equations 6.17-6.18.
Smoothed inferences can be derived by the following algorithm developed by Kim (1993) (a derivation can be found in the Appendix of Hamilton (1994) [H]):

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot (P^T \cdot [\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]), \tag{6.22}$$

where the operator $\oslash$ denotes element-by-element division. Hence, the **smoothed probabilities** $\hat{\xi}_{t|T}$ are computed from 6.22 backwards for $t = T-1, T-2, \ldots, 1$. The starting value for the iteration is $\hat{\xi}_{T|T}$, which is obtained from equations 6.17-6.18.
Recall that in the algorithm from equations 6.17-6.20 θ has been assumed as known. Now, we present how to estimate the value of θ that maximizes the log likelihood function. If $\hat{\xi}_{1|0}$ is a constant vector ρ and the transition probabilities need to satisfy only $p_{ij} \geq 0$ and sum to 1 in each column of P, then the maximum likelihood estimates for p_{ij} fulfill (shown in Hamilton (1990))

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} \mathbb{P}[s_t = j, s_{t-1} = i | \mathcal{Y}_T; \hat{\theta}]}{\sum_{t=2}^{T} \mathbb{P}[s_{t-1} = i | \mathcal{Y}_T; \hat{\theta}]}, \tag{6.23}$$

where $\hat{\theta}$ is the full vector of maximum likelihood estimates. This is essentially the number of regime switches from i to j divided by number of times s was in regime i, which can be estimated based on the smoothed probabilites.
The maximum likelihood estimate of ρ constrained on $\mathbb{1}^T\rho = 1$ and $\rho_j \geq 0$ for $j = 1, \ldots, N$ is found to be the smoothed inference about the initial state:

$$\hat{\rho} = \hat{\xi}_{1|T}. \tag{6.24}$$

Last but not least, the maximum likelihood estimate of the parameter vector α is characterized by the condition

$$\sum_{t=1}^{T} \left(\frac{\partial \log \eta_t}{\partial \alpha^T} \right)^T \hat{\xi}_{t|T} = 0, \tag{6.25}$$

where here η_t is again the vector of densities for y_t from equation 6.12.
Now, we have all the necessary tools to estimate the coefficients of a **Markov-switching regression model** of the form

$$y_t = \beta_{s_t}^T x_t + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2), \tag{6.26}$$

where x_t is a vector of explanatory variables that can include lagged values of y as well. The coefficient vector β_{s_t} takes different values depending on the regime at time t:

$$\beta_{s_t} = \beta_j, \quad \text{if} \quad s_t = j, \quad j = 1, \ldots, N. \tag{6.27}$$

The parameter vector $\alpha = (\beta_1^T, \ldots, \beta_N^T, \sigma^2)^T$ can be estimated by the following procedure:
Since the vector η_t of conditional densities of y_t has the form

$$\eta_t = \begin{pmatrix} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_t - \beta_1^T x_t)^2}{2\sigma^2} \right) \\ \vdots \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y_t - \beta_N^T x_t)^2}{2\sigma^2} \right) \end{pmatrix}, \tag{6.28}$$

condition 6.25 becomes

$$\sum_{t=1}^{T}(y_t - \hat{\beta}_j^T x_t)x_t \cdot \mathbb{P}[s_t = j|\mathcal{Y}_T;\hat{\theta}] = 0 \quad \text{for } j = 1,\ldots,N, \tag{6.29}$$

and

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}(y_t - \hat{\beta}_j^T x_t)^2 \cdot \mathbb{P}[s_t = j|\mathcal{Y}_T;\hat{\theta}]. \tag{6.30}$$

Equation 6.29 shows $\hat{\beta}_j$ to satisfy a weighted OLS orthogonality condition, each observation is weighted by the probability that it came from regime j. Thus, by defining

$$\tilde{x}_t(j) := \sqrt{\mathbb{P}[s_t = j|\mathcal{Y}_T;\hat{\theta}]} \cdot x_t, \tag{6.31}$$

$$\tilde{y}_t(j) := \sqrt{\mathbb{P}[s_t = j|\mathcal{Y}_T;\hat{\theta}]} \cdot y_t, \tag{6.32}$$

$$\tag{6.33}$$

$\hat{\beta}_j$ can be found from an OLS regression of $\tilde{y}_t(j)$ on $\tilde{x}_t(j)$:

$$\hat{\beta}_j = \left(\sum_{t=1}^{T}\tilde{x}_t(j)^T\tilde{x}_t(j)\right)^{-1}\sum_{t=1}^{T}\tilde{x}_t(j)\tilde{y}_t(j) \quad \forall\, j = 1,\ldots,N. \tag{6.34}$$

Hence, $\hat{\sigma}^2$ in equation 6.30 is just $1/T$ times the combined sum of the squared residuals of the above regressions.
The full algorithm to obtain the maximum likelihood estimate for the parameter vector θ works as follows (an application of the expectation-maximization (EM) algorithm that was developed by Dempster, Laird and Rubin (1977)): In case of a fixed starting value $\hat{\xi}_{1|0} = \rho$ and given an initial guess $\theta^{(0)}$ we find first estimates $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$ from the algorithm in 6.17-6.18. Using these estimates we obtain the smoothed probabilities $\hat{\xi}_{t|T}$ by iterating on 6.22. Now, we are able to evaluate 6.23, 6.34 and 6.30 to find estimates $\hat{p}_{ij}$, $\hat{\beta}_j$ and $\hat{\sigma}^2$ based on $\theta^{(0)}$. Hence, we derive a new estimate $\theta^{(1)}$ for the full parameter vector. In the same fashion we compute $\theta^{(2)}, \theta^{(3)}, \ldots$ until the change between two consecutive estimates is smaller than some predefined convergence criterion. It can be shown that this algorithm numerically maximizes the value of the likelihood function.
In case ρ is to be estimated by maximum likelihood as well, 6.24 is added to the equations that are reevaluated in each iteration.

6.2. Synthetic tests

Simulation of a Markov switching model

Let us consider the following form of the Markov switching regression model

$$r_t = \mu + \gamma_{s_t} r_{t-1} + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2). \tag{6.35}$$

The only switching parameter is γ, where we interpret a positive value as an indication for a bubble regime. The idea of the first simulation experiment is to simulate three different regimes for γ (0, 0.3 and 0.7). Then, we fit an MS regression model and compare the estimated parameters with the simulated ones. It is also interesting if the correct time periods can be identified. Figure 6.1 shows the result of this synthetic test, where we use realistic parameters for μ and σ estimated from the NASDAQ Composite time series for simulation. We plot the price of the NASDAQ time series, the simulated price trajectory and the smoothed states probabilities. Moreover, we give the simulated and estimated parameters of each regime. We do not only find similar values for the fixed parameters μ and σ from the simulated trajectory, but also the approximate values for γ (0, 0.26 and 0.74) and their time periods. This synthetic test shows that the MS model where only γ switches is consistent in the sense that it estimates similar parameters than we estimated.

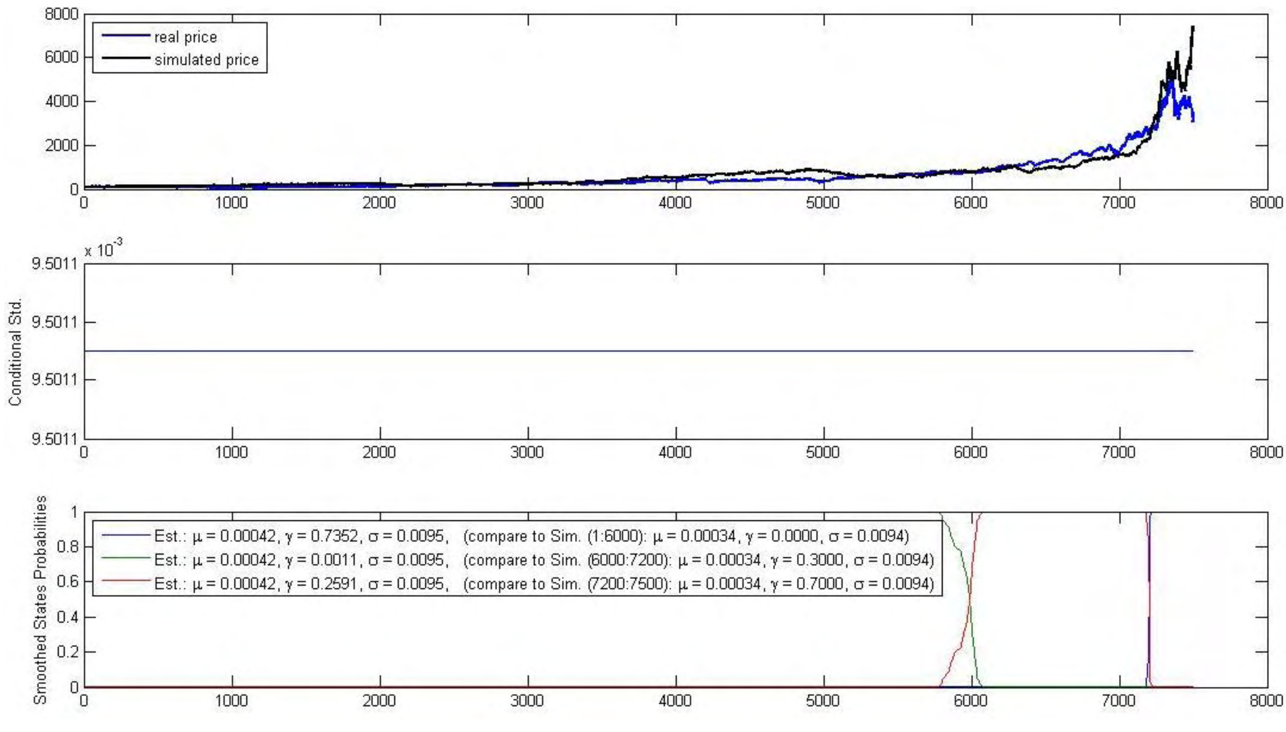


Figure 6.1.: Simulation experiment for MS model (only gamma switches): 6000 days $\gamma = 0$, 1200 days $\gamma = 0.3$ and 300 days $\gamma = 0.7$

Another interesting simulation experiment is to let μ and σ switch as well. In this case the Markov switching regression model has the form

$$r_t = \mu_{s_t} + \gamma_{s_t} r_{t-1} + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma_{s_t}^2). \tag{6.36}$$

The results in Figure 6.2 show that we find again quite similar parameters as we simulated, especially the volatility levels are estimated quite precisely. Obviously, the time period of each regime is identified.

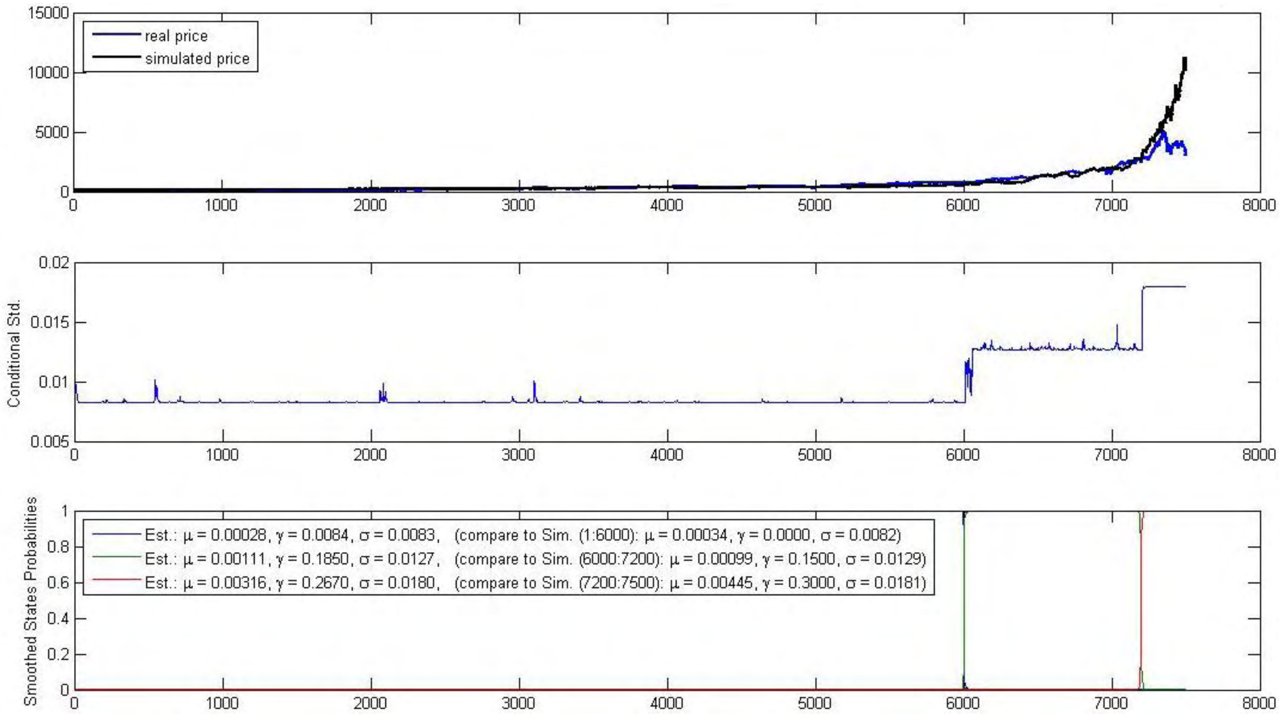


Figure 6.2.: Simulation experiment for MS model (all parameters switch): 6000 days $\gamma = 0$, 1200 days $\gamma = 0.15$ and 300 days $\gamma = 0.3$

Fitting an MS model to a simulated *FTS GARCH* trajectory

The following synthetic test should test if the MS model is able to detect bubbles generated by another class of models. Thus, we simulate 6000 days of a *GARCH* model, then alternate *GARCH* and *FTS GARCH* ($\gamma = 0.1$) regimes for another 1200 days and finally generate an *FTS GARCH* regime with doubled $\gamma = 0.2$ for 300 days. We use returns as regression component for the simulation and choose 100 days as length of the alternating regimes (Note this is the same kind of bubble as we simulated several times in Chapter 5). We fit the following MS regression model to the simulated trajectory:

$$r_t = \mu_{s_t} + \gamma_{s_t} r_{t-1} + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2_{s_t}). \tag{6.37}$$

The results are shown in Figure 6.3. In this test the following problem of the MS model arises: The bubble regime in red color (positive γ) is the regime with the highest volatility and subsequently every peak in the level of volatility leads to a jump to the bubble regime, even if the price is decreasing. This means the MS model depends too much on the volatility level and consequently we are not able to detect the *FTS GARCH* bubble regime appropriately. So, we could try to reduce this dependence by using again one of the 3 methods we have already used for the noise reduction of the regression component in an *FTS GARCH* model (compare Chapter 5.5). Moreover, we have also tried to overcome this problem by simply fixing the parameter σ. The best results for synthetic as well as empirical data are found by using a combination of both approaches: fixing the volatility parameter and using average returns. The resulting **MS regression model using average returns** has the form

$$r_t^\tau = \mu_{s_t} + \gamma_{s_t} r_{t-1}^\tau + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2), \tag{6.38}$$

where the average returns r_t^τ are calculated as usual:

$$r_t^\tau := \log\left(\frac{P_{t\cdot\tau}}{P_{(t-1)\cdot\tau}}\right), \quad \text{for } t = 1, \dots, \left[\frac{n}{\tau}\right] =: T. \tag{6.39}$$

Looking at the results in Figure 6.4 for the detection of a simulated *FTS GARCH* bubble by using the new approach above, we see that in the period of alternating regimes we sometimes jump to a bubble regime for a short time (in red color), and even more importantly, we find quite a strong bubble regime with $\mu = 0.04$ and $\gamma = 0.49$ for the last 200 days of the simulated trajectory (in green color). Hence, this method clearly identifies the simulated bubble regime at the end.

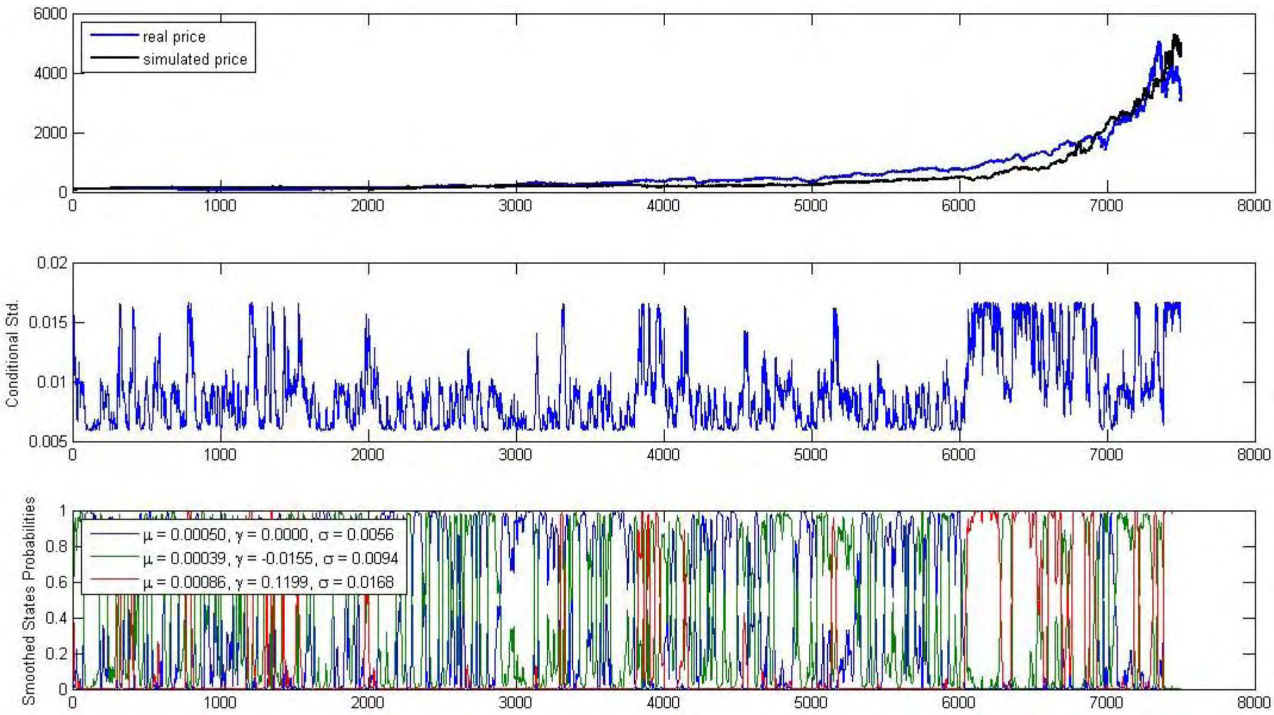


Figure 6.3.: Simulated trajectory generated by 6000 days of *GARCH*, then alternating *GARCH* and *FTS GARCH* ($\gamma = 0.1$) regimes for another 1200 days (regime length = 100 days) and finally *FTS GARCH* with doubled $\gamma = 0.2$ for 300 days ($y_t = r_t$). Estimation of an MS regression model where all parameters switch.

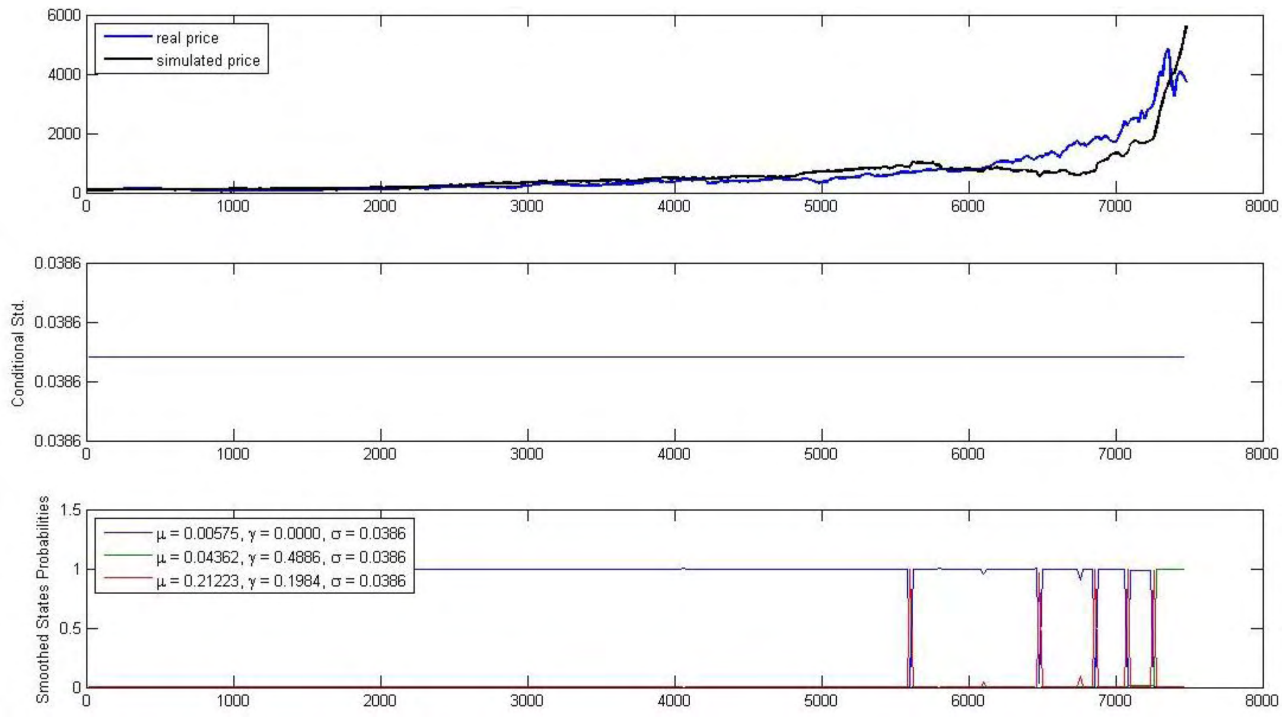


Figure 6.4.: Simulated trajectory generated by 6000 days of *GARCH*, then alternating *GARCH* and *FTS GARCH* ($\gamma = 0.1$) regimes for another 1200 days (regime length = 100 days) and finally *FTS GARCH* with doubled $\gamma = 0.2$ for 300 days ($y_t = r_t$). Estimation of an MS regression model for the averaged simulated returns with $\tau = 20$ (μ and γ switch).

6.3. Detecting bubbles in real data

6.3.1. Usage of average returns

We would like to apply the MS model using average returns from equation 6.38 to real data. Recall it manages to detect the bubble period in the above synthetic test. Figure 6.5 shows the results for the NASDAQ time series. We fit 3 different regimes and use average returns calculated over $\tau = 40$ days. We find the MS model to indicate a regime switch from $\gamma = 0$ to $\gamma = 0.67$ as the peak of NASDAQ in 2000 is approached. This can be interpreted as a signal for a bubble regime (compare the indication for a bubble using *FTS GARCH* with $\tau = 40$ in Figure 5.43).

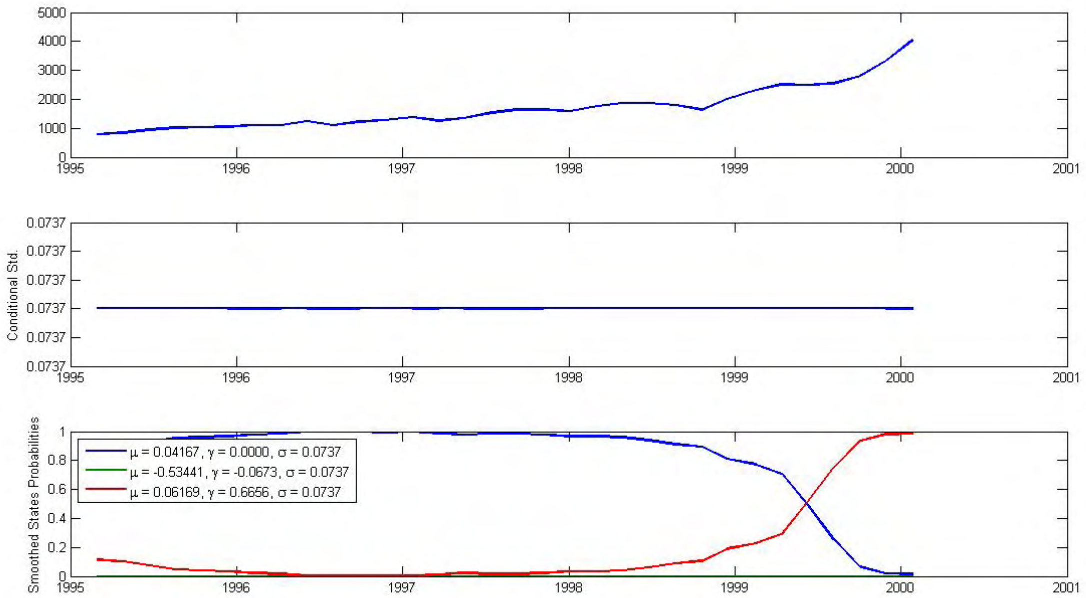


Figure 6.5.: NASDAQ Composite (November 4th, 1994 - March 27th, 2000): MS model for average returns ($\tau = 40$ days) with 3 different regimes (μ and γ switch).

6.3.2. Usage of EWMA returns

In case of the *FTS GARCH* model in Chapter 5, using EWMA returns as regression component yields more plausible results than using uniformly averaged returns. Hence, we would like to apply also a slightly different version of the MS model in equation 6.38 by using exponentially weighted moving average returns. Recall that EWMA returns with horizon τ are calculated from the daily return series $r_1, \dots, r_n$ by

$$r_t^{EWMA} := EWMA(r, \tau, \alpha)_t := \frac{1-\delta}{1-\delta^\tau} \cdot \sum_{k=0}^{\tau-1} \delta^k \cdot r_{t \cdot \tau - k}, \qquad \text{for } t = 1, \dots, \left[\frac{n}{\tau}\right] =: T, \tag{6.40}$$

where $\delta = e^{-\alpha}$ and $\alpha > 0$ constant.
The **Markov switching model using EWMA returns** as regression component has the form

$$r_t^\tau = \mu_{s_t} + \gamma_{s_t} r_{t-1}^{EWMA} + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim N(0, \sigma^2). \tag{6.41}$$

Let us now analyze the results for DAX, Gold and NASDAQ in Figures 6.6-6.8.
For the DAX time series we use a horizon of $\tau = 40$ days. We find a non-bubble regime with $\gamma = 0$ between 1993 and 1996. Around the peaks in 1997, 1998 and 2000 the model switches to a bubble regime with $\gamma = 0.94$, after each peak it switches back to $\gamma = 0$. Hence, the MS model finds super-exponential growth during each of the 3 bubble periods.
Concerning the Gold price we concentrate on the analysis of the two peaks in 2006 and 2008. The smoothed state probabilities show indication for a bubble regime already between 2002 and 2006 with $\gamma = 0.16$. Subsequently, there is an even stronger bubble period with $\gamma = 0.42$ until 2008 where the model switches to a negative feedback regime ($\gamma = -1.33$). Comparing it with the results for

the $FTS\ GARCH$ model using EWMA returns with $\tau = 60$ days as regression component, there is quite a strong signal already in the initial phase of the bubble. Using the $FTS\ GARCH$ model we find strong evidence for a bubble around 2008, but between 2002 and 2005 there is not such a strong indication.

Looking at the results for the NASDAQ using EWMA returns with $\tau = 40$ days as regression component, we see a quite similar picture as for average returns with $\tau = 40$. Between 1991 and 1997 there is a non-bubble regime ($\gamma = 0$) and as the peak in 2000 is approached the state probability of a bubble regime ($\gamma = 0.3$) increases continuously.

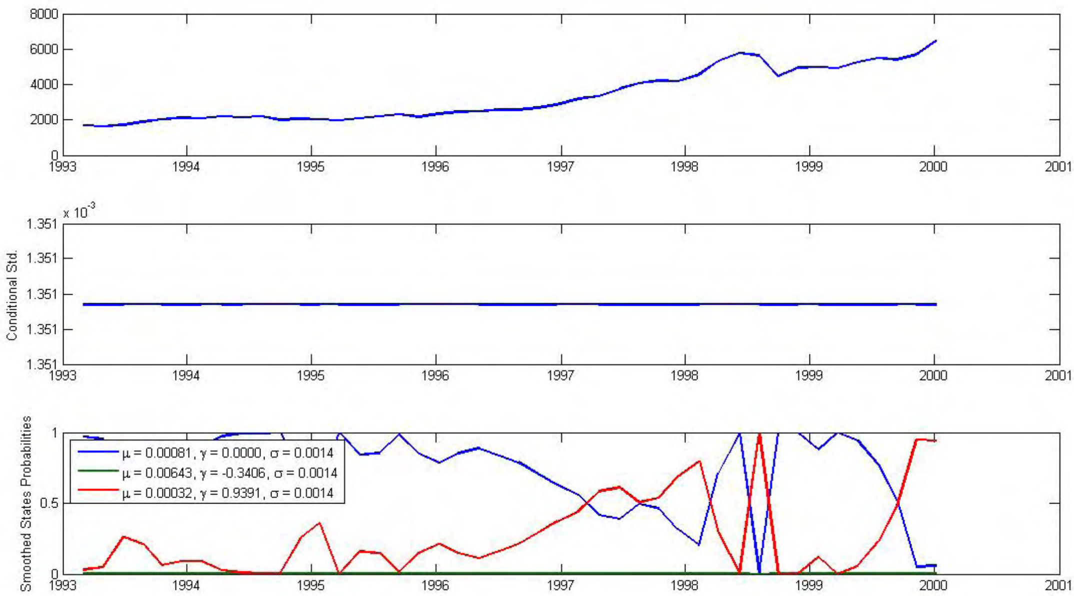


Figure 6.6.: DAX (January 6th, 1993 - March 7th, 2000): MS model for EWMA returns ($\tau = 40$ days) with 3 different regimes (μ and γ switch).

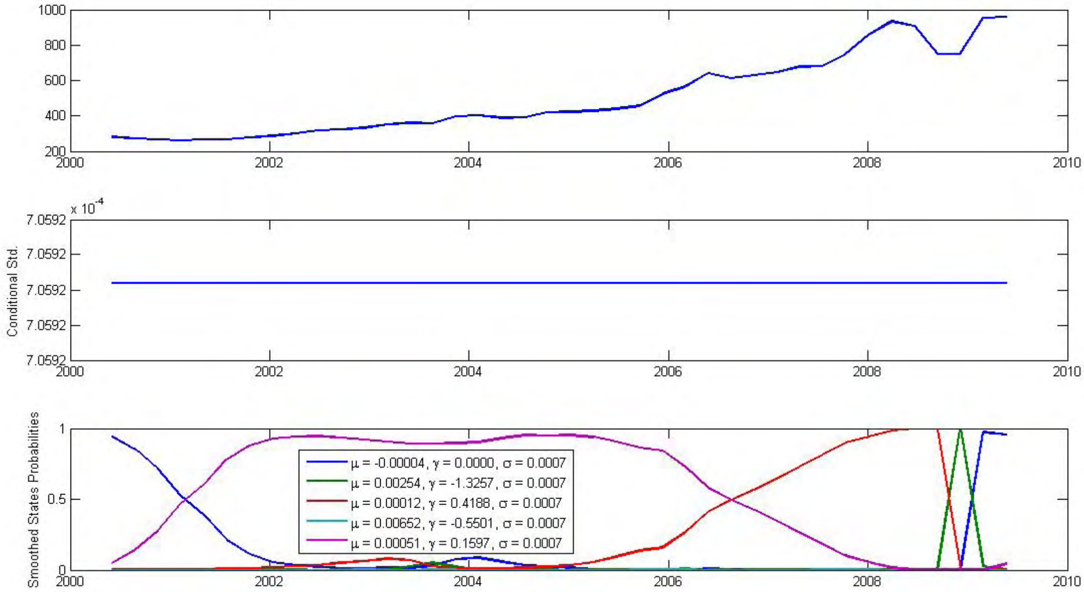


Figure 6.7.: Gold (March 13th, 2000 - August 28th, 2009): MS model for EWMA returns ($\tau = 60$ days) with 5 different regimes (μ and γ switch).

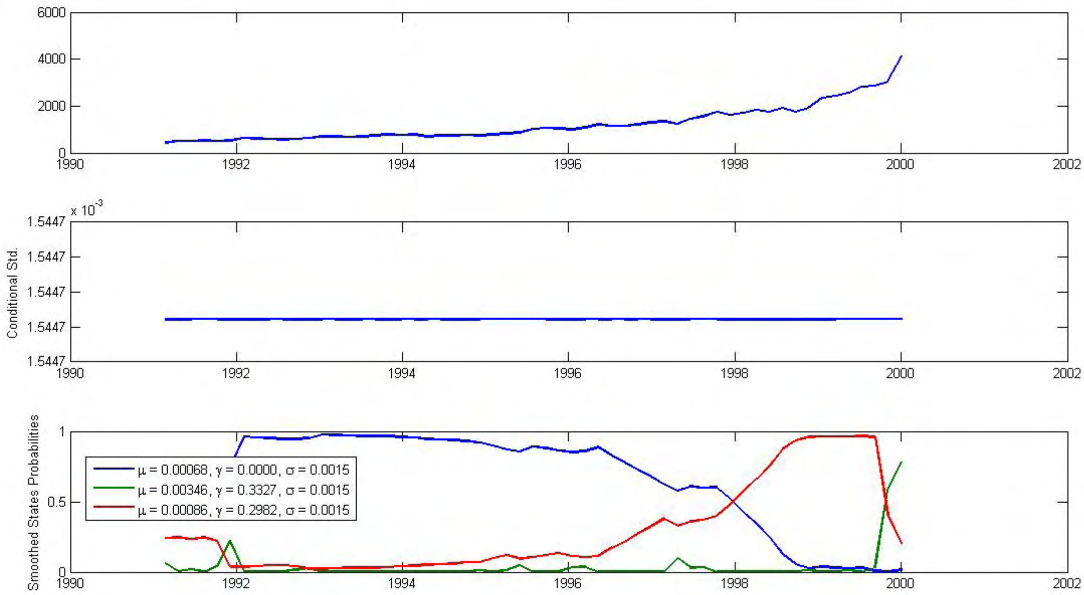


Figure 6.8.: NASDAQ Composite (December 26th, 1990 - March 27th, 2000): MS model for EWMA returns ($\tau = 40$ days) with 3 different regimes (μ and γ switch).

7 Conclusion

Let us give a summary of the main results of this thesis. Based on the evidence for super-exponential growth during bubble periods, we propose to enhance the standard $GARCH(1,1)$ model by adding a regression component that incorporates the possiblity for positive feedback effects (in accordance with the ideas of Corsi and Sornette (2012)). The focus lies on the analysis of different versions of this so called $FTS\ GARCH$ model, where we use various regression components.
Using log price as regression component, the method works quite well. This method is also in accordance with Huesler, Sornette and Hommes who find that traders anchor their expectations more on price than on return. We apply it to different time windows of the DAX, Gold, NASDAQ and NIKKEI time series and find high significance for a positive coefficient γ of the regression component around peaks in the price series. The problem of using log price is its non-stationarity. Even though it is an improvement to use log price instead of price, γ is still dominated by the higher price levels at the end of the chosen time windows. As we have seen in one of the synthetic tests, there is also high significance for a positive γ if we simulate a single regime shift to a period with higher average return but let $\gamma = 0$ (only exponential growth).
Alternatively, one can use return as regression component. Using daily returns the results for simulated as well as empirical data are not that clear. We find either quite a low signal or an even too strong signal that can no longer be used as an early warning signal at the right time. Thus, we propose three different methods to reduce the noise of the return series. The first one is the Kalman filter that helps us to find an almost perfect bubble signal for the NIKKEI time series around 1990. In case of the NASDAQ the result is improved but still not very satisfying. Then, we form either uniformly or exponentially weighted averages over a certain horizon of the daily returns and use them as regression component in the $FTS\ GARCH$ model. We use 20, 40 and 60 day horizons. For an appropriately chosen time horizon and an optimal starting date of the price series, this methods succeed to identify the peaks in the DAX, Gold and NASDAQ price series, although the empirical results using the EWMA returns are even more meaningful. Finally, we include a second regression component (the absolute return) in the $FTS\ GARCH$ modeling positive feedback of volatility on return. This approach enables us to test for a "fearful singular bubble", meaning super-exponential price growth accompanied by volatility growth. We find evidence for a fearful bubble in the DAX and the NASDAQ around 2000 and for the Gold price in 2006 and from 2009-2013. As an alternative we simplify this form of the $FTS\ GARCH$ model a bit by using the sum of exponentially weighted return and exponentially weighted absolute return as only one regression component. This sum reflects an average of only the positive returns. The idea is there could be stronger feedback on this

combined form during bubbles, which indeed can be observed in the empirical results (all bubble peaks are clearly identified).
In the last part of the thesis we use another class of models for the detection of financial bubbles, namely the Markov switching model. We find meaningful results only if we work with averages of returns (either uniformly or exponentially weighted). Also, we need to fix the volatility level, otherwise there is too high dependence of the regime on this parameter. This is a big disadvantage compared to the *FTS GARCH* model which considers time-varying volatility. Estimating a Markov switching model for DAX and NASDAQ, we manage to identify the bubble peaks in these two time series. There is a high smoothed state probability for a bubble regime right before each of the peaks. For the Gold price we find indication for a positive γ even for quite a long period (2002-2008).
Finally, we should keep in mind the high dependence on the date chosen as the beginning of the time series (the parameter s). This seems to be a disadvantage of both the *FTS GARCH* and the Markov switching model as soon as we work with averages of returns.

Appendix - The Johansen-Ledoit-Sornette (JLS) model

The idea of Johansen, Ledoit and Sornette (1999) [JSL] is to model the bubble behavior of a price P_t as a jump diffusion process:

$$\frac{dP_t}{P_t} = \mu(t)dt + \sigma(t)dW_t - \kappa dJ_t, \tag{A.1}$$

where $\mu(t)$ is the drift, $\sigma(t)$ the volatility and W_t a standard Brownian motion. The jump process J_t evolves with the dynamics

$$dJ_t = \begin{cases} 1, & \text{with prob. } h(t)dt, \\ 0, & \text{with prob. } (1-h(t))dt, \end{cases} \tag{A.2}$$

which is governed by the hazard rate $h(t)$ reflecting the instantaneous probability of a crash at time t conditioned that it has not happened yet. So the main idea of this dynamics is that the price evolves as a geometric Brownian motion with the possibility of a jump of size $-\kappa P_t$ modelling a crash (κ is a constant percentage). Recall that according to the ideas of Blanchard and Watson (1982) [BW] in a rational expectations bubble the no-arbitrage condition needs to hold. This in turn means the price process should be a martingale:

$$\mathbb{E}_t[P_s] = P_t \quad \forall\, s > t \geq 0 \quad \Longleftrightarrow \quad \mathbb{E}_t[dP_t] = 0. \tag{A.3}$$

Taking the expectation of the dynamics under the condition that no crash has occured yet

$$\mathbb{E}_t\left[\frac{dP_t}{P_t}\right] = \mu(t)dt - \kappa\mathbb{E}_t[dJ_t] = (\mu(t) - \kappa h(t))dt \tag{A.4}$$

shows that P_t is a martingale iff

$$\mu(t) = \kappa h(t). \tag{A.5}$$

Therefore, the unconditioned return of the asset is 0, whereas conditioned on staying in the bubble (no crash occurs) it is $\mu(t) = \kappa h(t)$. From this relation we infer the following: the higher the probability of a crash the faster the price increases (the higher the return) during a time of no crash. Intuitively, investors are compensated for the increasing risk of a crash.
Before the crash, the price P_t has the dynamics

$$\frac{dP_t}{P_t} = \kappa h(t)dt + \sigma(t)dW_t. \tag{A.6}$$

By integration one finds the expected price to be equal to

$$\mathbb{E}[P_t] = P_0 \mathbb{E}\left[\exp\left(\kappa \int_0^t h(s)ds + \int_0^t \sigma(s)dW_s\right)\right] = P_0 \exp\left(\kappa \int_0^t h(s)ds\right). \tag{A.7}$$

Now, by choosing an explicit form for the hazard rate $h(t)$ we determine the evolution of the bubble. The first reasonable choice (leading to price growth according to a power law) we consider is

$$h(t) = B(t_c - t)^{\beta-1}, \tag{A.8}$$

with a constant $B > 0$ and the critical point $t_c > 0$ (theoretical date of bubble burst). Please notice, for an economically meaningful model β should be between 0 and 1. This guarantees there is a finite upper bound P_c for the price as time t_c is approached and no crash has happened (however, the slope of P_t is unbounded as $t \longrightarrow t_c$). Also, t_c is only the most probable time for the crash, but it can occur at any time before t_c as well (although not very likely). The probability that t_c is reached without a crash in price is positive (otherwise rational agents could anticipate the crash):

$$1 - \int_0^{t_c} h(s)ds > 0 \tag{A.9}$$

Plugging the explicit form for $h(t)$ from (A.8) into the expected price in (A.7) yields:

$$\log(\mathbb{E}[P_t]) = \log(P_0) + \kappa B \int_0^t (t_c - s)^{\beta-1} ds = \tag{A.10}$$

$$= \log(P_0) + \frac{\kappa B}{\beta} t_c^{\beta} - \frac{\kappa B}{\beta}(t_c - t)^{\beta} =: \log(P_c) - \alpha(t_c - t)^{\beta}, \tag{A.11}$$

where $\alpha = \frac{\kappa B}{\beta}$ and $P_c = P_0 \cdot \exp(\alpha t_c^{\beta})$.
This can be generalized further by using the following hazard rate that includes log-periodic oscillations:

$$h(t) = B(t_c - t)^{\beta-1} + C(t_c - t)^{\beta-1}\cos(\omega \log(t_c - t) + \phi), \tag{A.12}$$

Substituting the above form of $h(t)$ into (A.7) and integrating yields the log-periodic power law (LPPL)

$$\log(\mathbb{E}[P_t]) = \log(P_c) - \alpha(t_c - t)^{\beta} - \gamma(t_c - t)^{\beta}\cos(\omega \log(t_c - t) + \phi), \tag{A.13}$$

with $\gamma = \frac{\kappa C}{\sqrt{\beta^2 + \omega^2}}$.

Note that the JLS model does not specify what happens beyond the critical time t_c. At t_c the bubble regime ends and a new regime (e.g. a crash or growth at a new rate) starts. Nevertheless, it can be used to predict crashes (or rebounds in case of an anti-bubble). For a detailed description of the prediction method and the development of trading strategies using the JLS model in equity markets we refer to Yan, Rebib, Woodard and Sornette (2011) [YRWS].

B Appendix - Kalman filter: Derivation of the optimal Kalman gain

We would like to show the derivations of the optimal Kalman gain K_k from (5.39) as well as the optimal error covariance matrix P_k from (5.41). Since the Kalman filter seeks for an estimate that minimizes the mean-square error (which is equivalent to minimize the trace of the error covariance matrix), we start with the computation of the error covariance matrix by using the measurement update for $\hat{x}_k$ from (5.36):

$$
\begin{aligned}
P_k &= \mathrm{Cov}(e_k) = \mathrm{Cov}(x_k - \hat{x}_k) = \mathrm{Cov}(x_k - \hat{x}_k^- - K_k(z_k - H\hat{x}_k^-)) = && \text{(B.1)}\\
&= \mathrm{Cov}(x_k - \hat{x}_k^- - K_k(Hx_k + v_k - H\hat{x}_k^-)) = && \text{(B.2)}\\
&= \mathrm{Cov}((I - K_kH)(x_k - \hat{x}_k^-) - K_kv_k) = && \text{(B.3)}\\
&= (I - K_kH)P_k^-(I - K_kH)^T + K_kR_kK_k^T = && \text{(B.4)}\\
&= P_k^- - K_kHP_k^- - P_k^-H^TK_k^T + K_kHP_k^-H^TK_k^T + K_kR_kK_k^T && \text{(B.5)}
\end{aligned}
$$

Next, we set the derivative of the trace of the error covariance matrix by the Kalman gain equal to 0 and solve for the optimizing K_k:

$$
\frac{\partial \mathrm{tr}(P_k)}{\partial K_k} = -2P_k^-H^T + 2K_k(HP_k^-H^T + R_k) \stackrel{!}{=} 0 \qquad \text{(B.6)}
$$

$$
\iff \quad K_k = P_k^-H^T(HP_k^-H^T + R_k)^{-1} \qquad \text{(B.7)}
$$

Using the optimal Kalman gain from above, one can easily find the optimal error covariance matrix:

$$
P_k = (I - K_kH)P_k^- + \underbrace{(K_k(HP_k^-H^T + R_k) - P_k^-H^T)}_{=0 \text{ for the optimizing } K_k} K_k^T = (I - K_kH)P_k^- \qquad \text{(B.8)}
$$

□

C

Appendix - MS model: Derivation of the inference algorithm

We verify the algorithm for the optimal state inference and forecast as well as the evaluation of the likelihood function from equations 6.17-6.20 (Hamilton (1994) [H]).
Note that the j-th element of the forecast $\hat{\xi}_{t|t-1}$ can also be written as $\mathbb{P}[s_t = j|x_t, \mathcal{Y}_{t-1}; \theta]$ since x_t is exogenous and hence does not contain information about s_t beyond $\mathcal{Y}_{t-1}$. Using this form we find the j-th element of the vector $\hat{\xi}_{t|t-1} \odot \eta_t$ to be the conditional joint density-distribution of y_t and s_t:

$$\mathbb{P}[s_t = j|x_t, \mathcal{Y}_{t-1}; \theta] \cdot f(y_t|s_t = j, x_t, \mathcal{Y}_{t-1}; \theta) = p(y_t, s_t = j|x_t, \mathcal{Y}_{t-1}; \theta). \tag{C.1}$$

By summing these joint-density distributions we get the density of y_t conditioned on past observations and θ:

$$f(y_t|x_t, \mathcal{Y}_{t-1}; \theta) = \mathbb{1}^T(\hat{\xi}_{t|t-1} \odot \eta_t), \tag{C.2}$$

as claimed in 6.20.
Dividing the joint density-distribution in C.1 by the density of y_t from C.2 we get the conditional distribution of s_t:

$$\frac{p(y_t, s_t = j|x_t, \mathcal{Y}_{t-1}; \theta)}{\mathbb{1}^T(\hat{\xi}_{t|t-1} \odot \eta_t)} = \frac{p(y_t, s_t = j|x_t, \mathcal{Y}_{t-1}; \theta)}{f(y_t|x_t, \mathcal{Y}_{t-1}; \theta)} = \mathbb{P}[s_t = j|y_t, x_t, \mathcal{Y}_{t-1}; \theta] = \mathbb{P}[s_t = j|\mathcal{Y}_t; \theta] \tag{C.3}$$

Collecting equation C.3 for $j = 1, \ldots, N$ in a vector yields

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{\mathbb{1}^T(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{C.4}$$

as claimed in 6.17.
Recall equation 6.8, where we expressed the Markov chain s_t in form of a first-order vector autoregression for ξ_t:

$$\xi_{t+1} = P \cdot \xi_t + v_{t+1}. \tag{C.5}$$

Taking the expectation conditional on $\mathcal{Y}_t$ yields

$$\mathbb{E}[\xi_{t+1}|\mathcal{Y}_t] = P \cdot \mathbb{E}[\xi_t|\mathcal{Y}_t] + \mathbb{E}[v_{t+1}|\mathcal{Y}_t]. \tag{C.6}$$

Since v_{t+1} is a martingale difference sequence with respect to $\mathcal{Y}_t$ its conditional expectation is 0. Hence,

$$\hat{\xi}_{t+1|t} = P \cdot \hat{\xi}_{t|t}, \tag{C.7}$$

as claimed in 6.18. □

Bibliography

[ADKM] T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. V. Mikosch (2009):
Handbook of Financial Time Series
Springer-Verlag Berlin Heidelberg.

[AFL] Y. Ait-Sahalia, J. Fan and Y. Li (2013):
The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency
Available at http://www.princeton.edu/~yacine/leverage.pdf

[AH] S. V. Aradhyula and M. T. Holt (1988):
GARCH Time Series Models: An Application to Retail Livestock Prices
Available at http://ideas.repec.org/p/ias/cpaper/88-wp29.html

[AS] J. V. Andersen and D. Sornette (2004):
Fearless versus Fearful Speculative Financial Bubbles
Physica A 337 (3-4), 565-585 (2004)
Available at http://arXiv.org/abs/cond-mat/0311089.

[B] T. Bollerslev (1986):
Generalized Autoregressive Conditional Heteroskedasticity
Journal of Econometrics 31 (1986) 307-327.

[BHHH] E.K. Berndt, B.H. Hall, R.E. Hall and J.A. Hausman (1974):
Estimation and Inference in non-linear structural models
Annals of economic and social measurement. 3/4. 1974.

[BT] H. Buening and T. Thadewald (2004):
Jarque-Bera test and its competitors for testing normality: A power comparison
Discussion Paper: Economics, No. 2004/9.
Available at http://hdl.handle.net/10419/49919

[BW] O. J. Blanchard and M. W. Watson (1982):
Bubbles, Rational Expectations and Financial Markets
NBER Working Paper No. 945.

[C] R. Cont (2001):
Empirical properties of asset returns: stylized facts and statistical issues
Quantitative Finance Volume 1 (2001) 223-236.

[CS] F. Corsi and D. Sornette (2012):
Follow the money: The monetary roots of bubbles and crashes
Swiss Finance Institute Research Paper No. 11-60
Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1966038

[E] R. F. Engle (1982):
Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation
Econometrica, Vol. 50, No. 4. (Jul., 1982), 987-1007.

[H] J. D. Hamilton (1994):
Time Series Analysis
Princeton University Press, Princeton, New Jersey.

[HSH] A. Huesler, D. Sornette and C. H. Hommes (2012):
Super-exponential bubbles in lab experiments: evidence for anchoring over-optimistic expectations on price
Journal Economic Behavior and Organization (submitted 2 May 2012).
Available at http://arxiv.org/abs/1205.0635

[HSTV] C. Hommes, J. Sonnemans, J. Tuinstra and H. van de Velden (2007):
Expectations and Bubbles in Asset Pricing Experiments
Journal of Economic Behavior and Organization.
Available at http://dare.uva.nl/document/127769

[JS] A. Johansen and D. Sornette (2008):
Finite-time singularity in the dynamics of the world population, economic and financial indices.
Physica A 294 (3-4), 465-502 (15 May 2001).

[JSJ] A. Johansen, I. Simonsen and M. H. Jensen (2006):
Optimal Investment Horizons for Stocks and Markets
Available at http://arxiv.org/abs/physics/0601002

[JSL] A. Johansen, D. Sornette and O. Ledoit (1999):
Predicting Financial Crashes Using Discrete Scale Invariance
Available at http://arxiv.org/abs/cond-mat/9903321

[KS] T. Kaizoji and D. Sornette (2009):
Market Bubbles and Crashes
Encyclopedia of Quantitative Finance (Wiley).
Available at http://arxiv.org/abs/0812.2449

[M1] B. Mandelbrot (1963):
The Variation of Certain Speculative Prices
The Journal of Business, Vol. 36, No. 4 (Oct. 1963), 394-419.

[M2] The MathWorks (2000):
GARCH Toolbox User's Guide Version 1 for MATLAB
Available at http://www.weizmann.ac.il/matlab/pdf_doc/garch/garch.pdf

[M3] P. S. Maybeck (1979):
Stochastic models, estimation and control
Academic Press, Mathematics in Science and Engineering Volume 141.

[MPR] A. Mora-Galan, A. Perez and E. Ruiz (2004):
Stochastic Volatility Models and the Taylor Effect
Available at http://ideas.repec.org/p/cte/wsrepe/ws046315.html

[PS] B. Perner and M. Schlener (2011):
Empirical Methods in Finance and Risk Management
Financial and Actuarial Mathematics, TU Vienna (not published lecture notes, only hardcopy available).

[RAB] P.-A. Reigneron, R. Allez and J.-P. Bouchaud (2011):
Principal Regression Analysis and the Index Leverage Effect
Physica A: Statistical Mechanics and its Applications, 390.

[S1] M. Sewell (2011):
Characterization of Financial Time Series
Available at http://www.cs.ucl.ac.uk/fileadmin/UCL-CS/images/Research_Student_Information/RN_11_01.pdf

[S2] D. Sornette (2003):
Why Stock Markets Crash - Critical Events in Complex Financial Systems
Princeton University Press, Princeton and Oxford.

[WB] G. Welch and G. Bishop (1995):
An Introduction to the Kalman Filter
Available at http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf

[WJ] P. Winker and V. Jeleskovic (2006):
The Unconditional Distribution of Exchange Rate Returns: Statistics, Robustness, Time Aggregation
Available at http://www.essex.ac.uk/ccfea/research/WorkingPapers/2006/008_Winker_Jeleskovic__Uncond_Dist_XRateReturns.pdf

[YRWS] W. Yan, R. Rebib, R. Woodard and D. Sornette (2011):
Detection of Crashes and Rebounds in Major Equity Markets
Available at http://arxiv.org/abs/1108.0077

Software

- The MathWorks (2012):
  MATLAB R2012b Win32
- B. Moore (2012):
  Kalman Filter Package for MATLAB
  https://www.mathworks.com/matlabcentral/fileexchange/38302-kalman-filter-package
- M. Perlin (2010):
  MS Regress - The MATLAB Package for Markov Regime Switching Models
  Available at SSRN: http://ssrn.com/abstract=1714016
- Z. Ding (2012):
  An Implementation of Markov Regime Switching Model with Time Varying Transition Probabilities in MATLAB
  Available at SSRN: http://ssrn.com/abstract=2083332