Application of the SEMF process to financial time series

Peter Zeeuw van der Laan

08/19/11

Abstract

This thesis demonstrates that the Self-Excited Multifractal process in its original form does not realistically describe financial data and proposes a modification that incorporates a dependency on past absolute returns. The Self-Excited Multifractal (SEMF) process is a multifractal process with an endogenous drive to capture all known stylized facts of financial time series. It is therefore considered a promising candidate for modeling financial data. I propose a method for volatility estimation and model calibration. The method is efficient in a broad range of synthetic SEMF processes ($R^2 > 0.8$). Application of this method to several types of real financial data (currency exchange returns, stock and index returns, on scales of days and 30 minute intervals) has unfeasible results: estimated parameters describe unrealistic SEMF processes and the estimated error terms (innovations) display a significant correlation in their second moment. Dependencies within the estimation imply that financial time series more explicitly depend on past absolute returns than the model. The generalized SEMF process that incorporates this dependency on past absolute returns is introduced. This process also displayed multifractal properties.

Contents

Explanation of symbols

ϑ	the kernel parameter set used for the generation of time series
ϑ'	the kernel parameter set used to calculate the cost function $\mathbf{l}(\vartheta)$ or ϱ
ϑ_{est}	the kernel parameter set estimated with CML or MIC
g	corrected scale according to equation 3.18
ω	term in the exponent that expresses the volatility, see equation 2.27 for a SEMF process
h_0	parameter of the SEMF process that is a measure of non-stationary and kurtosis of the process. See equation 2.28
φ	parameter of the SEMF process that is a measure of the memory strength of the process. See equation 2.28
σ	parameter of the SEMF process that is an indication for time scale. See equation 2.27
$\sigma_{i,est}$	estimated volatility according to equation 3.4
$\sigma_{i,real}$	volatility computed using Monte Carlo simulations according to equation 3.1
d_n	returns, either computed using Monte Carlo simulations according to equation 3.1 or returns as registered for real financial time series
h_n	memory kernel at point n, either exponential or power law. See equation 2.28
ξ_i	innovations of the SEMF process
n_{sp}	starting point of the estimation of the volatility and innovations according to equation 3.4
n_{start}	starting point of the conditional likelihood computation according to equation 3.13
$\mathbf{l}(\vartheta)$	conditional likelihood of a set of innovations estimated using parameter set ϑ according to equation 3.13
$\varrho(d, \vartheta, l)$	clustering according to equation 3.22
R^2	measure of quality of the estimation of the volatility, see equation 3.7
$P_{R2\geq 0.8}$	percentage of the estimations of the volatility of the simulated time series that had a quality of $R^2 = 0.8$ or higher
T_{R2}	total time needed to obtain estimations with quality R^2, equal to $T_{R2} = L_{R2} + n_{R2}$ where L_{R2} is the time used for the computation of R^2 and n_{R2} are all the data points taken into account before the computation of R^2 to estimate the volatility

f_θ	fraction that indicates the deviation of the parameter used for estimation with respect to the parameter used for simulation
α_θ	indication of the accuracy required in parameter θ to obtain estimations of the volatility with accuracy $R^2 = 0.8$, see equation 3.10
CML	Conditional Maximum Likelihood
MIC	Minimum Innovation Volatility Clustering
k_{abs}	parameter in the generalized SEMF process that regulates the dependence of ω with past absolute returns

Chapter 1

Introduction

Financial markets have the responsibility to distribute resources in the most efficient way. Therefore they play an important role in modern society. However in the last decades we have seen many crises in these financial markets. Such crises severely limit the efficiency of market operation and impose costs to our society. Therefore there is constant research to improve our understanding of the dynamics of financial markets. Specifically, models for financial returns can be used for risk management, derivative pricing and forecasting.

The modeling of financial asset prices started with Bachelier in 1900 who proposed the random walk process as a model for asset returns. Since the random walk is frequently used by physicists to model other natural phenomena, the link between physics and finance is as old as the research field itself.

Empirical research however points out that this model does not realistically describe financial time series. Stylized facts [1, 2]were found in all financial time series and many are not accounted for by the random walk process: long range dependence, heavy tails in the probability distribution function of the returns, multifractal properties and time reversal asymmetry.

Classical exogenous models, that are driven by external influences, were not effective in explaining these stylized facts. A new attempt was made by a class of the so called endogenous models, which are driven by the process itself. The self organizing mechanisms from these models are also found in other natural systems, such as turbulence, seismicity and neuron firing. It is largely believed that feedback mechanisms play an important role in financial time series [3, 4, 5, 6, 7, 8].

A prominent example of this approach is the GARCH process, introduced by Bollerslev in 1986 [9]. This process is defined by a historical term that only influences the magnitude (volatility) of the process increments and a stochastic error term (innovation), which represents the random part of the process. The GARCH process incorporates a great deal of stylized facts, such as absence of auto-correlations and a degree of auto-correlation in the absolute returns. However it fails to incorporate all stylized facts. It shows reasonable performance in forecasting absolute and squared future returns.

Several modifications of the GARCH process have been introduced to incorporate the remaining stylized facts, but non of the modifications incorporates all. For example the well-known exponential GARCH (EGARCH) process introduced by Nelson in 1991 [10] displays the leverage effect, but no exact long-range dependence.

In the last decade so called 'Econophysicists' have proposed to make a new attempt using Multifractal processes. The notion of multifractality, based on self similarity and the notion of fractals, introduced by Mandelbrot [11], describes processes which have a non-trivial scaling in contrast to the simple scaling properties of fractal processes like fractional Brownian motion. Since multifractal properties describe how the dynamics of processes are related to dynamics at larger scales, they capture the dynamics that impact the process the most. Since financial time series have multifractal properties, any model that does not take these properties into account is limited in its validity and performance. The notion of multifractality is also frequently used

in other fields of physics where apparent random movements account for much of the dynamics. Examples include fields as turbulence, seismicity and heartbeat dynamics.

The Markov Switching Multifractal process, introduced by Calvet and Fisher in 2001 [12], represents the most prominent Multifractal approach to date. It showed strong performance in forecasting squared and absolute returns. However, the process does not incorporate several stylized facts such as time asymmetry and gain/loss asymmetry, nor does it have an endogenous drive and full long range dependence.

The Multifractal Random Walk (MRW) [13], introduced by Muzy and Bacry in 2002, demonstrated how to create a process with exact Multifractal properties by constructing exponentials of stochastic long memory processes with power law memory. Saichev, Sornette and Filimonov [15, 14] demonstrated that effective or quasi Multifractal properties of this process are robust under small changes of the memory kernel. Furthermore they demonstrated that this would solve the problem of infinite variance the MRW displayed in the continuous time limit.

Recently Sornette and Filimonov extended framework of the MRW and Quasi multifractality by introducing the Self Excited Multifractal (SEMF) process [16]. They replace the stochastic process with long memory in the previous models with a process that has long memory of past returns, effectively creating an endogenous process. In the SEMF process there is again a distinction between the volatility and the stochastic innovations. They demonstrate that this process has effective Multifractal properties, which are robust to changes in the memory structure of the process. Furthermore they demonstrate that the process incorporates all stylized facts of financial time series. Because it is the only known endogenous Multifractal process, it possibly also has applications to areas such as seismicity and biology (human heart beat rhythm).

Showing all known stylized facts and having an endogenous structure, the SEMF process is a promising candidate for modeling financial time series. However it has not yet been verified to realistically describe real financial time series. Such verification would consist of calibration of the model on real financial time series.

1.1 Contribution and structure of this research

In this research, I perform such calibration and demonstrate that the SEMF process needs to be modified to realistically describe real financial time series. Because of an observede relation between the volatility process and the absolute estimated innovations, an additional dependency on past absolute returns is implied. A generalization of the SEMF model that includes this additional dependency is proposed, and it is demonstrated that this model also includes all the stylized facts.

In chapter 3, a two calibration methods are proposed. The Conditional Maximum Likelihood (CML) method is proposed and is effective in estimating the parameters of the model in synthetic time series with a broad parameter range. The parameter estimates can be used to estimate the volatility with a high accuracy. Furthermore I introduce a new method for parameter estimation that Minimizes the Clustering (the auto-correlation of the second moment) of the estimated Innovations (MIC) and demonstrate its effectiveness. Finally, the chapter presents several scaling relations that describe the relation of the SEMF process at different scales.

In chapter 4, the estimation procedure is applied to a broad range of financial data (stock returns, stock index returns, currency exchange returns) on multiple scales (daily returns and returns on 30 minute intervals). Using both parameter estimation methods parameters are estimated that do not correspond to realistic SEMF processes. For stock and index returns (on both scales), a reasonable memory decay parameter is found but the non-stationary parameter is extremely high. It is demonstrated that, in contrast to the model where the innovations are independent and obey a Gaussian distribution, the model estimates of the innovations show strong clustering and are heavy tailed. Therefore I conclude that the SEMF model in its current form does not realistically describe this financial data. Finally, it is found that the absolute estimated innovations have a strong relation with the magnitude of the process that determines the volatility (ω). This implies that the model should have a dependency on past absolute returns to realistically describe financial time series.

In chapter 5, I introduce the generalization of the SEMF process where the volatility has an additional dependency on past absolute returns. It is demonstrated that this process also displays heavy tails and multifractality. Finally, the similarity of the generalized SEMF process to the EGARCH(1,1) process is shown, which suggests that the EGARCH(1,1) process possibly also has multifractal properties.

Chapter 2

Theoretical Background

For the efficiency of the (international) capital markets many financial instruments such as bonds, common stocks and foreign currencies are publicly traded. Therefore effective trading prices of each of these instruments are of great interest.

As stocks give right to future dividends, it is possible to express the absolute price of a stock by calculating the expectation value of its future cash flows:

$$P'_t = \sum_{i=t+1}^{\infty} E[D_i] \tag{2.1}$$

Because we can also buy a bond for this price P_t which would give right to an interest rate k, we have to discount these future dividends in order to determine the price of the stock relative to the bond:

$$P_t = \sum_{i=t+1}^{\infty} \frac{E[D_i]}{1+k} \tag{2.2}$$

It is difficult to estimate the price this way, because future dividends D_i depend on many unknown factors and which are difficult to predict. To facilitate the pricing process, the Gordon growth model [17] was introduced

$$P_0 = \sum_{i=1}^{\infty} \frac{D_0(1+g)}{(1+k)} \tag{2.3}$$

where D_0 represents the last dividend payout and it is assumed that the dividend grows infinitely with rate g. Due to the limited accuracy of this assumption the model can only provide a limited accuracy in asset price estimation.

Because of the current inability to accurately estimate these fundamental prices, investors acting in financial markets are faced with huge complexity and market prices of financial assets change extremely frequently.

These price changes are denoted by returns d_n, where d_n is calculated using

$$d_n = ln(P_n/P_{n-1}) \tag{2.4}$$

The high variability of the asset prices makes capital markets less attractive for investors, which has a cost to society.

2.1 Stylized facts

The first attempt to model these frequent price changes was presented by Bachelier in 1900, who proposed the famous random walk as a model for asset returns. However, this process does not capture all the stylized facts of financial time series that have since been identified: absence of auto-correlation in the returns, clustering of squared returns and long range dependence, heavy tails in the probability distribution of the returns, multifractality, time reversal asymmetry of the process, gain/loss asymmetry in the distribution of the returns and the presence of critical points. They can be seen as an imprint of a complex system, as laws that the asset pricing process obeys, being the result of the interaction of a large number of market participants [2, 1, 18]. They serve as guidelines for formulating models that aim to describe the pricing process ever since Bachelier's null hypothesis. Below these stylized facts will be presented in more detail.

Absence of auto-correlations

Auto-correlations of asset returns are insignificant above a certain time scale, typically around 20 minutes. Therefore the asset prices at these scales are martingale processes, a property which can be formulated as $E[P_{n+1}\,|I_n|] = P_n$ where I_n represents the available information set at time t. This means that the best estimate of the future price of an asset is its current price. This is also known as the Efficient Market hypothesis.

Intermittency, Volatility clustering and long range dependence

Financial time series display a high variability in the returns, where periods of high variability are preceded or followed by periods of lower variability in the returns.. This is called intermittency, and can be quantified by calculating the volatility clustering ϱ_{VC}, which is the auto-correlation of the absolute returns or squared returns, of the time series:

$$\varrho_{VC} = corr(|d_t|\,, |d_{t+T}|),\ T > 0 \tag{2.5}$$

This quantity results in a positive value in the case of financial time series. This volatility clustering decays over time as a power law with an exponent β between $0.2 < \beta < 0.4$. There is also an auto-correlation of absolute returns at orders higher than one:

$$\varrho_{\lambda} = corr(|d_t|^{\lambda}\,, |d_{t+T}|^{\lambda}),\ T > 0 \tag{2.6}$$

where λ indicates the order. This phenomenon is called long range dependence.

Heavy tails in the probability distribution of the returns

The probability distribution function (pdf) of the returns in the random walk is Gaussian. However, empirically it was found that the tails of the pdf of the returns decay as a power law with an exponent γ around $2 < \gamma < 4$, and are therefore called 'heavy tails'. This means that the probability of extreme returns is higher in reality than in the random walk process. An example of this can be seen in Fig. 2.1. The heavy tails are also sometimes referred to as 'excess kurtosis'.

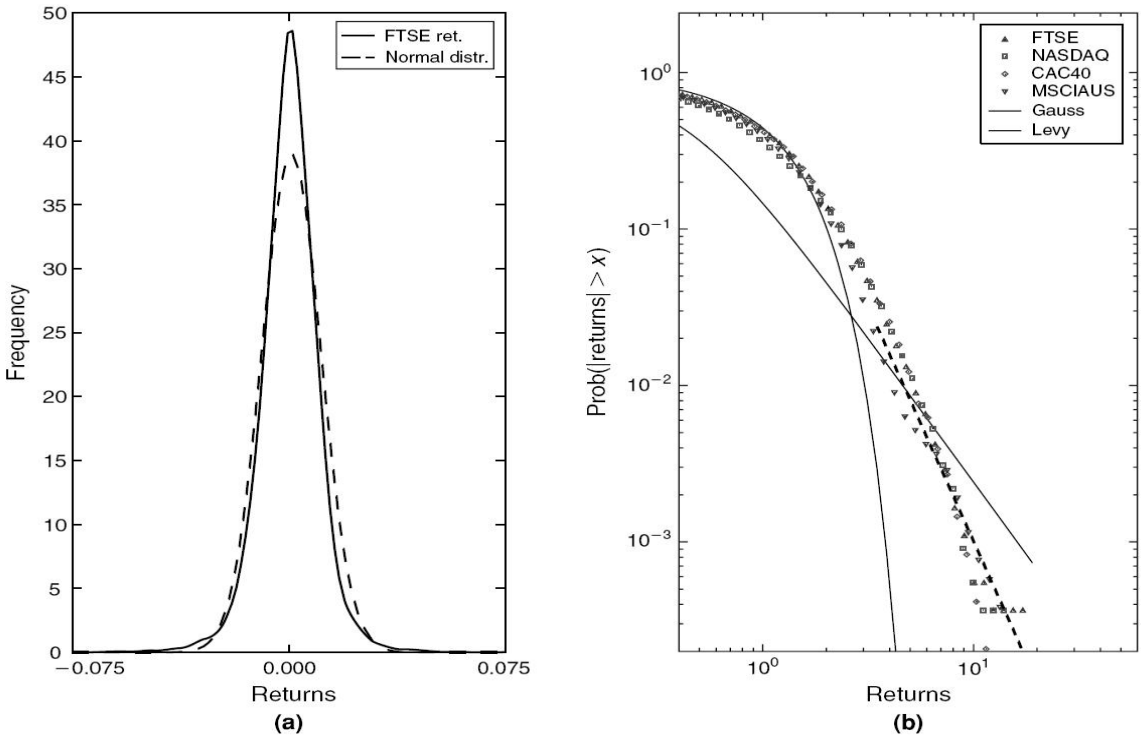


Figure 2.1: The tails of the pdf of stock returns are heavier than Gaussian tails. a). Comparison of the frequency of stock returns of the FTSE index with a process that follows a normal distribution. b) The tails of the empirically obtained pdf of the returns of several stock indices roughly follow a power law distribution. Adapted from [1]

Multifractality

The properties above (heavy tails, long range dependence) are the result of the multifractal properties of the process. A multifractal is the generalization of a mono-fractal, which is an object that has identical properties at different time scales. An example of a mono-fractal and a mono-fractal process can be seen in Figure 2.2.

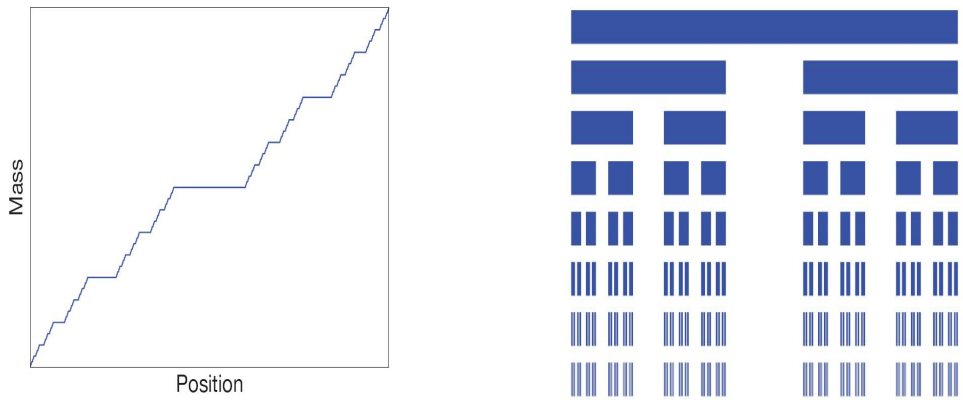


Figure 2.2: Example of a mono-fractal process. A mono-fractal process (Right) is constructed by integrating weights created by using a constant division law (Left)

The properties of a multifractal are related at different time scales, but they do not have to be identical. There are many examples of multifractals in nature such as the texture of leaves, clouds and shorelines. The notion of multifractality was largely developed by Benoît Mandelbrot , and he was also one of the first to apply it to financial time series [22]. A multifractal is defined as the object that is described with nonlinear measure spectrum on the fractal set.

A mono-fractal process is a process $X(t)$ that has trivial scaling properties with regards to its statistical moments $M_q(l)$. These moments are defined as

$$M_q(l) = E(|\delta_l X(t)|^q) = \langle |X(t+l) - X(t)|^q \rangle \tag{2.7}$$

and can be scaled as

$$M_q(l) = K_q l^{\varsigma_q(l)} \tag{2.8}$$

where K_q is a constant. In the case of mono-fractals, the exponents $\varsigma_q(l)$ are independent of l and can be represented as $\varsigma_q = qH$. Since H is a constant, it has a linear scaling spectrum. This proportionality parameter H is denoted as the Hurst exponent, and it is equal to $1/2$ in the case of the earlier mentioned random walk.

If the (generalized) Hurst exponent $H(q)$ is not constant for different orders of the moment q, it describes a non-trivial (non-linear) scaling spectrum:

$$\varsigma_q = qH(q) \tag{2.9}$$

This is called a multifractal singularity spectrum.

Multifractal properties can arise due to power law distributions and long range dependence, which are both present in financial time series. Note that the presented formalism can only be used to detect multifractality in processes $X(t)$ with stationary increments $\delta_l X(t)$.

Time reversal asymmetry

The statistical properties of the process are not identical upon time reversal of the time series. The strongest manifestation of this property is the leverage effect, which can be quantified as

$$L(l) = \frac{E[d_t d_{t+l}^2]}{(E[d_t^2])^{3/2}} \tag{2.10}$$

It describes the negative correlation between past returns and future squared returns that is observed in financial time series. The asset pricing process obtains more stability if positive returns occur and becomes more unstable with negative returns. This correlations decays to zero over time l, where t represents the current time

Gain/loss asymmetry and the presence of critical points

In the stock market, strong price drops or 'crashes' can be observed, but prices tend to rise more gradually. Therefore an asymmetry between positive and negative returns in the pdf is observed. The frequent occurrence of crashes over time (over a hundred year) also implies that these crashes are indeed a part of the dynamics of the process, and not outliers. Because the statistical properties of crashes do not coincide with the statistical properties of the rest of the financial time series, they are sometimes referred to as critical points [19].

2.2 Volatility, Innovations and Self - organization

We have seen that that the returns (with sign) of the financial return process are not auto-correlated. However, the absolute or squared returns do display correlation, even at multiple orders (Long range dependence). Therefore the pricing process is frequently modeled as consisting of a part controlling mainly sign and a part controlling mainly magnitude:

$$d_n = \xi_n \sigma_n \tag{2.11}$$

where ξ_n represents the innovation of the price (controlling sign and partly magnitude of the price return) at time n and σ_n represents the volatility (partly controlling magnitude of the price return) at that time. This separation originates in the random walk process.

In the context of financial time series, the innovations ξ_n are often interpreted as the market innovations. They represent exogenous news that has impact on the expected future dividends of the asset and and therefore on its price (see equation 2.2). Typically a Gaussian distribution is used to model the distribution of the innovations. The use of the Gaussian distribution can be supported by the assuming that the news is a result of many independent events, of which the sum according to the Central Limit theorem would lead to a Gaussian distribution. However in many social-economic environments power law distributions are observed, which implies interacting mechanisms between events (e.g., preferential attachment). In recent popular literature the impact of the news on asset prices is also suggested to have power law properties [20]. Therefore the possibility of different probability distribution functions for the innovations should be kept in mind.

Volatility is a measure of risk at a certain time. It can be measured in historical financial time series ex post (when subsequent returns are known) using the expression for realized volatility:

$$\sigma_{rel}^2(n, L) = \frac{1}{n-1} \sum_{i=n}^{n+L} \left(d_i - \overline{d}\right)^2 \tag{2.12}$$

where

$$\bar{d} = \frac{1}{L_{rel}} \sum_{i=n}^{n+L} d_i \tag{2.13}$$

It is more difficult to determine volatility ex ante (when subsequent returns are unknown). However accurate estimates of this quantity could be used to estimate market risk and also price financial assets more accurately

(e.g., options), making the financial market more efficient. For these purposes many so called volatility models have been formulated.

Two classes of models with regards to financial time series have been introduced: a class that that treats volatility as being stochastic or exogenous, and a class of models that treats volatility as non-stochastic and therefore as an endogenous property.

It is largely conceived that financial time series indeed have strong endogenous properties [4, 6, 7, 8, 3]. A major contribution to this field was made by Shiller [5] in 1981. Shiller used equation 2.2 to calculate the 'fundamental' price of assets in historical time series based the knowledge of future returns multiple years later. He concluded that the high variability in the price could not be justified by changes in the future dividends alone. Therefore this endogenous property can be seen as a criterion on the validity of financial models.

Below a few of the most prominent volatility models and processes are discussed, both endogenous and exogenous.

2.3 GARCH processes

In 1986 Bollerslev introduced the Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) process [9], that treats volatility as an endogenous property. It is one of the most prominent volatility models in this class, it captures several stylized facts and can be applied to financial time series. In the GARCH(p,q) process the current volatility is determined by q terms of past squared returns and p terms of past volatility:

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i d_{n-1}^2 + \sum_{i=1}^{p} \beta_i \sigma_{n-i}^2 \tag{2.14}$$

In this equation the coefficient α_i determines the strength of the memory of past squared returns i time steps ago. Coefficient β_i determines the strength of the memory of past volatility terms. The returns are related to the volatility through equation 2.11, and therefore the process has no auto-correlations of signed asset returns. Furthermore, the GARCH process displays heavy tails and the dependence of current volatility on past squared returns and volatility terms leads to a degree of volatility clustering.

The process has rich applications to financial time series and its relative success in forecasting absolute or squared future returns is well established. The model is often used as a benchmark for other models. The parameters of the process α_i and β_i can be estimated using the conditional maximum likelihood method.

However, the process does not display several stylized facts. Multifractal properties have not been found for this process and it does not incorporate the leverage effect. Furthermore, because of the finite number of past terms in the model, the memory of the process decays exponentially and does not fully capture long range dependence (which requires power law decay).

Because of the relative success of this model but its failure to incorporate several stylized facts, modifications of the GARCH process have been proposed that capture additional stylized facts. The fractionally integrated extension of the GARCH model (FIGARCH) does a better job at capturing long range dependence, as it introduces an infinite number of past lags with hyperbolically decaying coefficients [28, 29]. Several other modifications of the GARCH process such as the Threshold GARCH (TGARCH), the nonlinear GARCH (NGARCH) and the exponential GARCH (EGARCH) incorporate the leverage effect. Because of its relevant structure the EGARCH process will be presented in more detail below.

2.3.1 Exponential GARCH (EGARCH)

In 1991 Nelson introduced the EGARCH model [10]. The main advantage of this process is that it takes into account the leverage effect, as positive returns reduce the amplitude of the returns while negative returns increase it. A major difference with the GARCH process is that the EGARCH process only depends on past innovations where the GARCH process depends on past returns. An EGARCH(p,q) process is given by

$$log\sigma_t^2 = \omega + \sum_{k=1}^{q} \beta_k g(\xi_{t-k}) + \sum_{k=1}^{p} \alpha_k log\sigma_{t-k}^2 \tag{2.15}$$

where ω is a measure of time and β_k and α_k again indicate memory. In the equation $g(\xi_t)$ is defined as

$$g(\xi_t) = \theta\xi_t + \lambda(|\xi_t| - E(|\xi_t|)) \tag{2.16}$$

The parameter θ controls the effect of the absolute returns. The parameter λ controls the effect of the returns with sign, and is primarily related to controlling the strength of the leverage effect in the process. The parameters of the process can be estimated using conditional maximum likelihood estimation. The estimated parameters can also take on negative values, which is not possible in the GARCH process. The success of the EGARCH process in forecasting absolute or squared future returns has been well established since its introduction. However, none of the proposed GARCH modifications captures all stylized facts.

2.4 Multifractal models

Multifractal properties have not been found for the GARCH processes, while possible multifractal properties of the EGARCH(1,1) process have not been studied. The multifractal structure of the asset pricing process most likely accounts for a large part of the dynamics of the process, and any model without a multifractal structure is inherently limited in its validity and performance. Therefore there has been great interest in understanding how multifractal processes can be created, which has resulted in the introduction of a new class of volatility models: Multifractal models.

Below I present the formulation of the Multifractal Random Walk (MRW). The MRW presents a structure which can lead to exact multifractal properties. This structure was used by the Quasi Multifractal model, which showed that effective multifractal properties could still hold for small modifications of the derived structure. In both processes the volatility is an exogenous property. These results were used to formulate the Self-Excited Multifractal process, which describes the volatility as an endogenous process that also incorporates the leverage effect. It therefore incorporates all stylized facts and is a promising process for modeling financial time series. The three models will be discussed in more detail below.

The Markov Switching Multifractal model [12] is a process that has multifractal properties and also displays several of the other stylized facts such as long-range dependence, heavy tails and absence of auto-correlations. It has been successfully applied to financial time series and it embodies of the potential of using multifractal processes for modeling financial time series. Drawbacks of the process are its exogenous volatility, time symmetry and an absence of the leverage effect. Because its formulation has no relation to this research, it is not presented here.

2.4.1 Multifractal random walk

In section 2.1 we have seen that a stationary process displays multifractality if for its moments

$$M_q(l) = E(|\delta_l X(t)|^q) = K_q l^{\zeta_q} \tag{2.17}$$

and ζ_q has a nonlinear dependency on q: $\zeta_q = qH(q)$ with H(q) the generalized Hurst exponent. This scaling property can also be expressed as

$$\delta_{\lambda l} X \sim \lambda^H \delta_l X(t) \tag{2.18}$$

where $\sim$ indicates that the probability distribution functions on either side are similar. This representation was used to derive a cascading rule, that a multifractal process should obey:

$$\delta_{\lambda l} X \sim W_\lambda \delta_l X(t) \tag{2.19}$$

where W_λ is a log-normal random variable that depends only on λ. This cascading rule should be valid when going from coarse to fine scales $\lambda < 1$.

This result was used by Bacry, Delour and Muzy who introduced the Multifractal Random Walk [13]. The log-normal random variable W_λ is written as $e^{\omega_{\Delta t}[k]}$ and the process in continuous time is given by

$$X(t) = \lim_{\Delta t \to 0} \sum_{k=1}^{t/\Delta t} \xi_{\Delta t}[k] e^{\omega_{\Delta t}[k]} \tag{2.20}$$

$\xi_{\Delta t}[k]$ represents the innovation of the process at time step k, and is a white noise process with variance $\sigma^2 \Delta t$. The variables $\omega_{\Delta t}[k]$ are random parameters with a Gaussian distribution and a correlation function:

$$\mathrm{cov}(\omega_{\Delta t}(k_1), \omega_{\Delta t}(k_2)) = \lambda^2 \ln \rho_{\Delta t}\left[|k_1 - k_2|\right],\ \rho_{\Delta t}\left[|k|\right] = \begin{cases} \frac{1}{(|k|+1)\Delta t} \text{ for } |n| \leq L/\Delta t - 1 \\ 1 \text{ otherwise} \end{cases} \tag{2.21}$$

We can derive that such a process displays a multifractal singularity spectrum in the continuous time limit. This scaling spectrum is expressed as

$$\varsigma_q = \left[q - q(q-2)\lambda^2\right]/2 \tag{2.22}$$

and is valid up to the larges scale L. Therefore Bacry, Delour and Muzy demonstrated how to construct a process with exact multifractal properties in the continuous time limit. However, the process has a significant drawback, as the variance goes to infinity for $\Delta t \to 0$.

2.4.2 Quasi Multifractal model

The Quasi Multifractal process, introduced by Saichev and Sornette [15, 14] in 2006, generalized the Multifractal Random Walk and showed that effective multifractal properties were robust under this generalization.

In the continuous time limit, the autocorrelation property of the $\omega_{\Delta t}[k]$ process can also be expressed as

$$\omega(t) = \int_{-\infty}^{t} dW(t') h(t - t') \tag{2.23}$$

where $W(t)$ denotes a Wiener process and $h(t)$ is defined as

$$h(t) = \frac{h_0}{(1 + t/\Delta t)^{1/2}} \tag{2.24}$$

Saichev and Sornette showed that for an intermediate scale $\tau \ll t \ll L$, the multifractality property also holds if the exponent of memory decay is slightly increased (with $\varphi > 0$):

$$h(t) = \frac{h_0}{(1 + t/\tau)^{\varphi + 1/2}} \tag{2.25}$$

The property of having multifractal properties in this intermediate scale $\tau \ll t \ll L$ was denoted by effective multifractality or quasi-multifractality. Another advantage of the generalized process is that for $\varphi > 0$ there is no infinite variance as $\Delta t \to 0$. Therefore Saichev and Sornette demonstrated how to construct a process with robust effective multifractal properties. However, the volatility in the process only has exogenous components, and it does not incorporate the leverage effect.

2.5 Self Excited Multifractal process

The Self Excited Multifractal process introduced by Filimonov and Sornette in 2010 [16] was the first endogenous process with multifractal properties. In the SEMF model the $\omega(n)$ are dependent on past returns d_n rather than on an exogenous process $W(t)$, such as in the QMF model The SEMF process can therefore be seen as the simplest multifractal generalization of the GARCH process. In discrete time the process is defined by:

$$\omega(n) = \sum_{i=0}^{n-1} d_i h_{n-i-1} \tag{2.26}$$

where $\omega(n)$ is related to the volatility σ_n and the returns d_n by

$$d_n = \sigma_n \xi_n,\ \sigma_n = \sigma \exp\left\{-\frac{\omega(n)}{\sigma}\right\} \tag{2.27}$$

Here h_n represents the memory kernel that allows for multifractal scaling. Three memory kernels are proposed: the power law kernel, exponential kernel and the constant kernel respectively:

$$h_n = h_0 n^{-\varphi - 1/2},\ h_n = h_0 \exp(-\varphi n),\ h_n = h_0 \tag{2.28}$$

where σ, φ and h_0 are the parameters that specify the process for a given kernel type. The process is more robust than the previously proposed multifractal processes (MRW and QMF) in the sense that it exists for a larger domain of kernel types. In Fig. 2.3 we see a sample realization of a SEMF process.

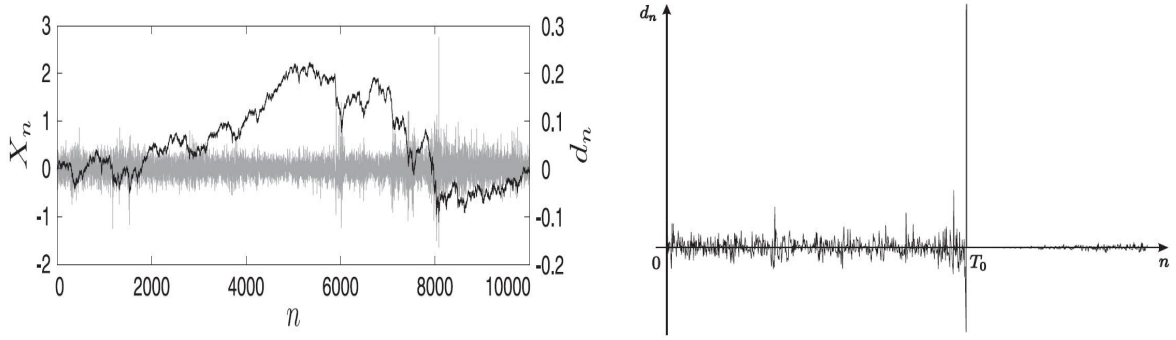
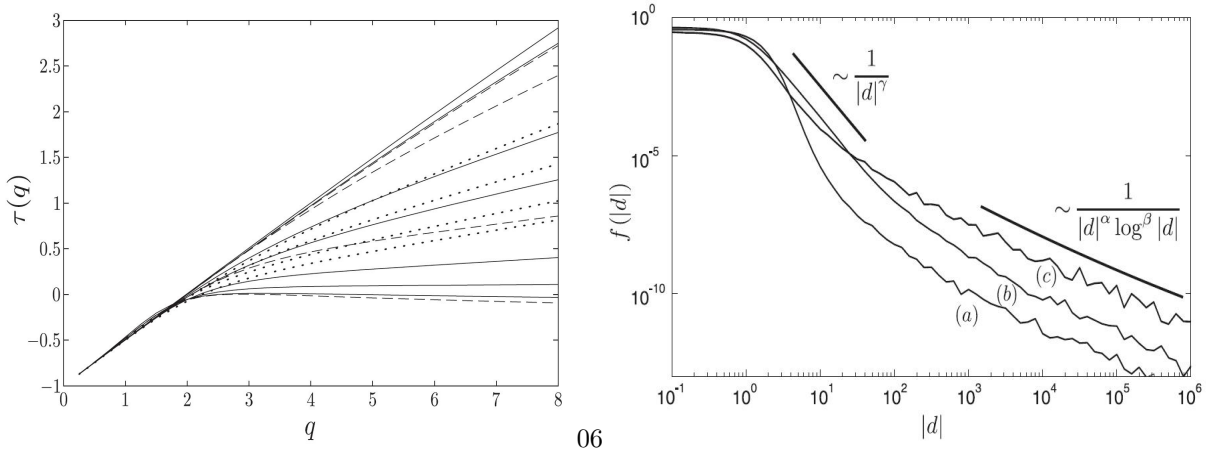


Figure 2.3: ***Left***: Time-series of log price increments d_n (gray) and log price $X_n = P_n = \sum_{i=1}^{n} d_i$ (black) of the discrete SEMF process defined by equation 2.27, for $\sigma = 0.01$ and with the power-law kernel, specified in equation 2.28, with $\varphi = 0.01$ and $h_0 = 0.05$. ***Right****:* Example of an extreme event in a simulated time series where a high return (caused by a burst in volatility) caused subsequent returns to have a very low amplitude. Figures adopted from [16].

It was found that the signed increments of the process are generally not stationary, because 'extreme events' such as illustrated in Fig. 2.3 *(Right)* can dominate the dynamics of the process for large intervals due to the strong memory of the process. These events may be regarded as tipping points indicating change of regime after which the model does not describe reality. The kernel parameter h_0 can be seen as a measure of non-stationary of the process, and it amplifies intermittency in σ_n. φ can be interpreted as the parameter indicating memory strength of the process while σ is a measure of time scale.



06

Figure 2.4: ***Left***: Multifractal scaling exponents (q) of the process d_n for $\sigma = 1$ and (i) power-law kernel (2) with $\varphi = 0.01$ and $h_0 = (0.01; 0.06; 0.08; 0.10; 0.12; 0.14; 0.16)$ (solid lines top to bottom), (ii) exponential kernel (3) with $\varphi = 0.01$ and $h_0 = (0.06; 0.08; 0.10; 0.12)$ (dotted lines top to bottom), (iii) constant kernel h $h_0 = (0.02; 0.04; 0.06; 0.08)$ (dashed lines top to bottom). ***Right:*** Probability density function of the absolute values of the increments d_n for $\sigma = 1$: (a) power-law kernel (2) with $\varphi = 0.1$, $h_0 = 0.1$, (b) constant kernel $h_n = h_0 = 0.02$ and (c) exponential kernel (3) with $\varphi = 0.01$, $h_0 = 0.1$.Figures adopted from [16].

Filimonov and Sornette demonstrate that the process displays effective multifractality (see Fig. 2.4 (*Left*)). In the plot the values on the y-axis $\tau(q)$ are related to the generalized Hurst exponent of equation 2.9 by $\tau(q) = qh(q) - 1$. They also demonstrate that incorporates all other stylized facts: absence of auto-correlations, Long range dependence, Heavy tails (see Fig. 2.4 (*Right*)), time reversal asymmetry (including the Leverage effect), gain/loss asymmetry and the presence of critical points. These properties are robust to small changes of the process parameters σ, φ and h_0 for the different kernel types. The properties were demonstrated for parameters in the range:

$$\varphi_{exp} \sim 0.01,\ \varphi_{pow} \sim 0.1 - 0.01 \tag{2.29}$$

$$h_{0,exp} = 0.05 - 0.10,\ h_{0,pow} = 0.10 - 0.25 \tag{2.30}$$

Furthermore the process is driven endogenously. The process is therefore a promising candidate for volatility forecasting and demonstrating the endogenous drive of financial markets.

Chapter 3

Determination of the volatility and the model parameters in synthetic time series

In this chapter, I propose a procedure to estimate the volatility within a SEMF process and to calibrate the SEMF model. It is demonstrated that the method is effective in a broad parameter range of synthetic SEMF processes.

Section 3.1 demonstrates that we can estimate the volatility and innovations with a high quality ($R^2 = 0.8$) in synthetic time series if the process parameters are known. We see that this is valid for a broad range of parameters for both exponential and power law kernel types. The method used ignores past history and as a result the accuracy increases as a function of time.

Section 3.2 shows that this high estimation quality remains if there is a small uncertainty in the process parameters. This is again valid for a broad parameter range. An equation is presented that approximates how the uncertainty in the individual parameters jointly influence the quality of the estimation (R^2). We see that estimations with the incorrect kernel type do not result in high quality estimations.

Section 3.3 presents two effective parameter estimation methods. The Conditional Maximum Likelihood (CML) method is successful in determining the memory kernel type of a synthetic time series (close to 100% of the time). The parameter estimates of this method are sufficiently accurate to obtain volatility estimations with high quality in synthetic SEMF processes with a broad range of parameters. The volatility can not be obtained with high accuracy using the parameter estimates in synthetic time series with an exponential kernel with a low memory decay and a high non-stationary. For power law kernels, the accuracy in parameter estimation is sufficiently high only for processes with a sufficiently strong memory decay. Furthermore I find scaling relations using this method that relates the process parameters of SEMF time series on different time scales.

Furthermore, I introduce the Minimum Innovation Clustering (MIC) method that estimates the process parameter by minimizing auto-correlations in the second moment of the innovations. It is demonstrated that this method has reasonable accuracy in determining memory kernel type and process parameters. However the method's accuracy is lower than the accuracy of the CML method. Since this method optimizes the independence of the innovations and the CML method optimizes the likelihood, it can serve as a second option in parameter estimation in SEMF processes.

3.1 Diagnostic of the volatility and the innovations

This section proposes an method for estimating the volatility and innovations if the process parameters are known and demonstrate that it results in high quality estimations in a broad parameter range.

We see in section 3.1.2 that the estimations of both quantities converge to their real values over time if we ignore the non-realistic volatility bursts. In section 3.1.3 it is found that both estimations are much more accurate for an exponential kernel than for a power law kernel while increasing h_0 (non-stationary) and decreasing φ (memory strength) have a negative influence on the quality of the estimation. Volatility estimations in time series with a power law memory kernel are of high accuracy ($R^2 = 0.8$) for low non-stationary (low h_0) and limited memory strength (high φ). Finally, in sections 3.1.4 and 3.1.5, we see a positive bias for the volatility estimation in the case of power law memory kernels and see that the initial condition of the synthetic time series ($\omega \neq 0, \sigma_{n_{SP}} \neq 1$) has a strong influence on the accuracy of the estimation for a given parameter set.

The influence of uncertainty on the estimation accuracy and the possibility of parameter estimation will be discussed in further sections of this chapter.

3.1.1 Algorithms for volatility estimation in synthetic time series

I investigate the estimation of volatility in synthetic time series. The synthetic time series d_n were created using Monte Carlo simulation. The synthetic innovations $\xi_{n,real}$ (where real stands for synthetic) were simulated i.i.d. using a Gaussian distribution with $\mu = 0$ and $\sigma = 1$. As an initial condition, I take $\sigma_{1,real} = 1$ in each estimation.

For specified values for the input variables (σ, φ, h_0), the returns d_n and subsequent volatility $\sigma_{n,real}$ using equation 2.27 that defines the SEMF process:

$$d_n = \sigma_{n,real}\xi_{n,real},\ \sigma_{n,real} = \sigma \exp\left\{-\frac{\sum_{i=0}^{n-1} d_i h_{n-i-1}}{\sigma}\right\} \tag{3.1}$$

where h_{n-i-1} refers to one of the memory kernels specified in equation 2.28. I only consider the exponential and power law kernel since a process with a constant kernel is not a realistic model for financial time series. Because of the normalization of the returns d_n by σ in the expression for ω (see equation 2.27), σ is only a measure of scale and has no impact on the dynamics of the process. One can demonstrate this property with the following example:

$$d_n^1 : (\sigma, h_0),\ d_n^2 : (\alpha\sigma, \alpha h_0) \tag{3.2}$$

$$d_n^2 = \alpha d_n^1 \tag{3.3}$$

Therefore I only consider different cases of h_0 and φ and $\sigma = 1$ for all simulations. Because of the infinite memory of the process the computation time increased non-linearly with the length of the time series. I took an optimal range of 10^4 data points for each time series.

In the synthetic time series, the volatility $\sigma_{n,est}$ is estimated using

$$\sigma_{n,est} = \sigma_{est} \exp\left\{-\frac{\sum_{i=n_{sp}}^{n-1} d_i h_{n-i-1}}{\sigma_{est}}\right\} \tag{3.4}$$

I assumed for the initial condition that $\sigma_{np,est} = 1$. In real financial time series it is possible that the initial condition does not have this value. Therefore I took $n_{sp} \gg 1$, and $\sigma_{nsp,real}$ can take on any value. This way a large part of the history of the synthetic time series was ignored. The volatility was then estimated for the

interval $n_{sp} < n \leq n_{end}$ where n_{end} is the last data point of the simulated time series. For now I assume that the kernel parameters σ_{est}, $h_{0,est}$, φ_{est} and kernel type are known. The innovations can then be estimated by

$$\xi_{n,est} = d_n/\sigma_{n,est} \tag{3.5}$$

An example of a synthetic time series where the volatility was estimated can be seen in Fig. 3.1.

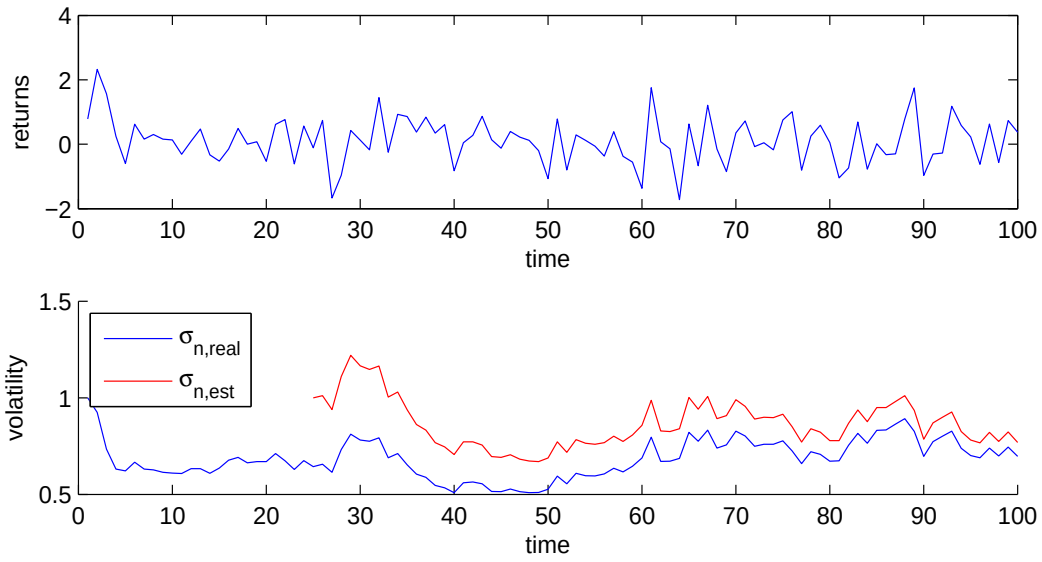


Figure 3.1: Example of volatility estimation in a synthetic time series with $n_{sp} = 25$

3.1.2 Convergence and non-convergence for the estimation of ξ_i

As a first step it is of interest whether the estimation converges over time. I investigate the convergence of the estimated innovations $\xi_{i,est}$ to the real value in the synthetic time series $\xi_{i,real}$ over time. Since the innovations and volatility are proportional, convergence of the estimation of the innovations would also be valid for the volatility. I simulated a large number of time series using equation 3.1 and estimated the volatility and innovations using equations 3.4 and 3.5. I calculated the squared error of the estimation as a function of time i:

$$SE(i) = \{\xi_{i,est} - \xi_{i,real}\}^2 \tag{3.6}$$

Simulations were done for the exponential and power law kernel for parameter ranges where the process displayed the stylized facts (see equations 2.29 and 2.30).

Examples of time dependence of squared errors SE(i) is presented in Fig. 3.2. This Figure illustrates four different behaviors:

1. The squared error decays over time (when a power law kernel was used)
2. The squared error decays exponentially (when an exponential kernel kernel was used)
3. The squared error decays over time, rises suddenly and then remains constant over time (occurred for both kernel types)

4. The size of the squared errors is constant over time (occurred for both kernel types)

For case 1 and 2, we can conclude that the estimation $\xi_{i,est}$ convergences to $\xi_{i,real}$ over time. Simulations corresponding to case 3 and 4 are non-convergent.

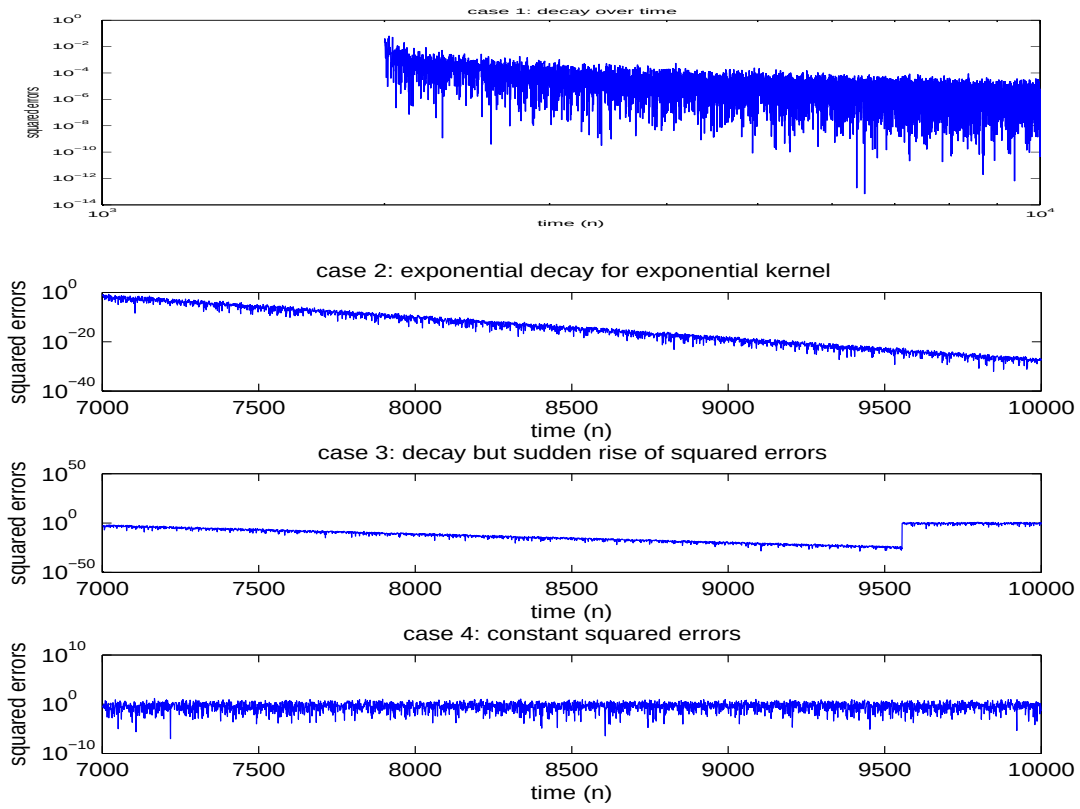


Figure 3.2: Overview of the four different observed behaviors of the squared errors when estimating $\xi_{i,real}$. $n_{sp} = 2001$ in case 1 (power law kernel), $n_{sp} = 7001$ in the other cases (exponential kernels).

Case 1 and 2: Convergence

When the estimated innovations converges to their real value, the behavior is different for the two kernel types. In the case of an exponential kernel the results show that the squared errors decays more strongly with increasing φ while h_0 has no influence on the decay. The squared errors can be fitted with an exponential decay exp(-2φ) quite accurately. We can conclude that a higher φ led to a higher accuracy in the estimations $\xi_{i,est}$.

The returns in the time series before n_{sp} introduce an error in the estimation of the volatility at any point in time afterward because of the infinite memory. Because a higher φ leads to a faster decay of the 'memory' of the volatility, a higher φ reduces the impact of these returns more quickly. This decreases the remaining error in a higher pace, which explains the observed behavior.

For the power law kernel, a log-log plot (see Fig. 3.2) reveals a steady decay of the squared errors and convergence of $\xi_{i,est}$ to its real value $\xi_{i,real}$. The plot is too noisy too verify a possible power law like decay.

The qualitative influence of h_0 and φ is similar to the case of a power law kernel, but no exact relations have been found.

Case 3 and 4: Exclusion

It was found that the behavior corresponding to case 3 and 4 is caused by bursts in the volatility (see Fig. 2.3) of the simulated time series. As the volatility increases due to a series of negative $\xi_{i,real}$ and corresponding d_n, a subsequent positive d_n is likely to have an 'extreme' amplitude. This is followed by a very low volatility $\sigma_{i,real}$ and because of the small amplitude of the subsequent returns d_n the time series takes a long time to recover to its 'normal' dynamics. This behavior was also observed by Sornette and Filimonov [16]. In a numerical environment, this behavior results in returns d_n, volatility $\sigma_{i,real}$ and estimated volatility $\sigma_{i,est}$ that are too small to compute. This either occurs before n_{sp} (case 4) or after (case 3).

Because bursts of this magnitude are not observed in real financial time series, simulations with these bursts were excluded from the results. Therefore one can conclude that the estimation of $\xi_{i,est}$ converges to it's real value $\xi_{i,real}$ since all non-convergent cases are excluded. Because of the relation between ξ_i and σ, the estimation of the volatility also converges over time.

3.1.3 Obtaining high quality estimations

In the previous section it was demonstrated that the estimation $\sigma_{n,est}$ converges to its real value $\sigma_{n,real}$ over time. Therefore one can conclude that the accuracy of the estimation is an increasing function of time. Commonly $R^2 = 0.8$ indicates that an estimation in a stochastic process is of a high quality. As a next step it is investigated how many data points (or the total interval length required T_{R2}) are needed to obtain estimates with this high accuracy for an exponential kernel and a power law kernel.

I use R^2 to quantify the accuracy of the estimation of σ_{est} over an interval $n_{R2} \leq n < n_{R2} + L_{R2}$. Here n_{R2} is the moment in time after n_{sp} where the computation of R^2 starts and L_{R2} is the interval length over which R^2 is calculated (see the example in Fig. 3.3).

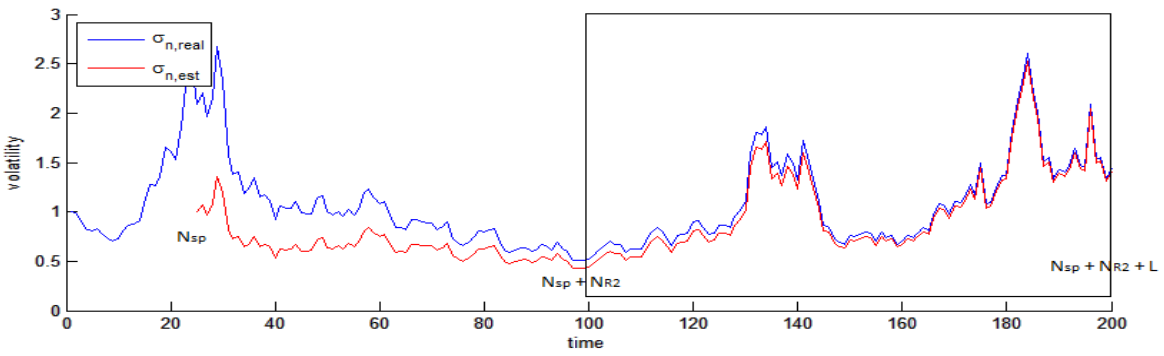


Figure 3.3: Example of a the computation of R^2 (in the black square) in a synthetic time series. In this case $n_{sp} = 25$, $n_{R2} = 75$ and $L = 100$

Because the quality of the estimation increases over time, I calculate R^2 over the last data points of the interval. I found that the computation of R^2 gives stable values if I calculate it 100 data points. Therefore I used $L_{R2} = 100$. R^2 is expressed as

$$R^2(n_{R2}, L_{R2}) = 1 - \frac{\sum_{i=n_{sp}+n_{R2}}^{n_{sp}+n_{R2}+L_{R2}} \left(\sigma_{i,real} - \sigma_{i,est}\right)^2}{\sum_{i=n_{sp}+n_{R2}}^{n_{sp}+n_{R2}+L_{R2}} \left(\sigma_{i,real} - \overline{\sigma}_{real}\right)^2} \tag{3.7}$$

where

$$\overline{\sigma}_{real} = \frac{1}{n} \sum_{i=n_{sp}+n_{R2}}^{n_{sp}+n_{R2}+L_{R2}} \sigma_{i,real} \tag{3.8}$$

By defining $T_{R2} = L_{R2} + n_{R2}$, one can interpret T_{R2} as the total time needed to obtain a series of volatility estimations with length $n = 100$ with corresponding accuracy R^2. An example of a synthetic time series where R^2 is computed for the indicated interval can be seen in Fig. **??** (in comparison with other quality indicators).

I simulated 10^3 time series for different sets of φ and h_0 for both kernel types. Because the process increments have heavy tails, is highly intermittent and is non-stationary, the high number of experiments is needed to ensure the representativeness of the results. In the remainder of this research 10^3 time series are simulated for each time series (unless indicated otherwise).

I estimate the volatility with $n_{sp} = 2001$ and calculate R^2 for the indicated interval length at multiple moments in time T_{R2}. After applying equation 3.4 and estimating σ_{est}, R^2 was calculated for the indicated interval length at multiple moments in time T_{R2}. The percentage of simulations that have reached the threshold $R^2 \geq 0.8$, denoted by $P_{R2\geq 0.8}(T_{R2})$, was evaluated over time T_{R2}. Also look at the influence of φ and h_0 on $P_{R2\geq 0.8}(T_{R2})$ is investigated for both kernel types.

Exponential kernel

Figure 3.4 shows $P_{R2\geq 0.8}(T_{R2} = 200)$ as a function of φ for various h_0 for an exponential kernel. The plot demonstrates that, for exponential kernels, the volatility estimation is of lower quality for simulations with strong non-stationary (high h_0) and strong memory (low φ). It shows that $P_{R2\geq 0.8}$ increases with φ, indicating a higher quality of estimation, while a higher h_0 leads to a lower $P_{R2\geq 0.8}$.

The influence of φ corresponds to observations in section 3.1.2. The increased $P_{R2\geq 0.8}$ for higher φ can be explained by the faster decay of memory at higher φ. This decreases the time needed to 'forget' the returns d_n before the starting point of the estimation n_{sp} that causes the error in estimation.

A higher h_0 leads to an increased non-stationary and kurtosis of the returns d_n, including the returns d_n before n_{sp}. Therefore the expectation value of the deviation from the unit initial condition ($E[|\sigma_{nsp,real} - 1|]$) increases as a function of h_0. Since I assume a unit initial condition for each estimation, this introduces a persistent error in the volatility estimation at subsequent moments in time ($n > n_{sp}$). For a given value of φ, estimations in time series with high h_0 take a longer time to 'forget' their corresponding large errors than in time series with low h_0, leading to a lower $P_{R2\geq 0.8}(T_{R2})$ at any T_{R2}.

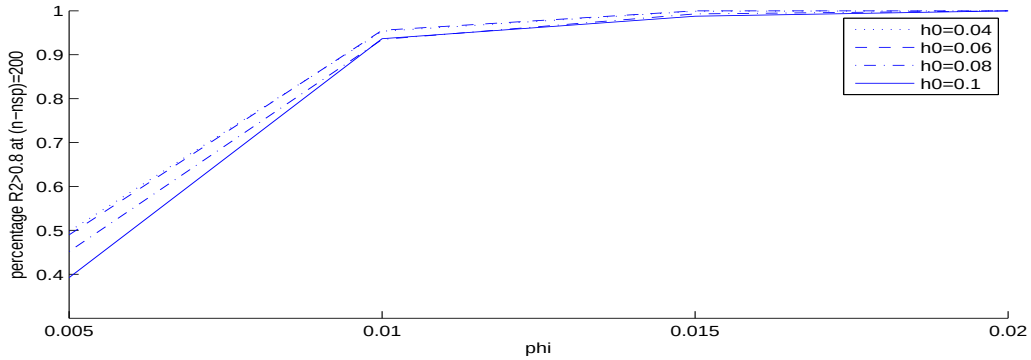


Figure 3.4: Dependence of $P_{R2\geq 0.8}$ on φ for various h_0 with $T_{R2} = 200$ for an exponential kernel

At $T_{R2} = 1500$, 100% of the simulations with an exponential kernel reaches R^2=0.8, even though the exact T_{R2} varies with φ and h_0 (T_{R2} corresponding to 100% is high for low φ and high h_0). Therefore one can conclude that the volatility estimator proposed in equation 3.4 can estimate volatility with high accuracy for exponential kernels for the indicated parameter ranges.

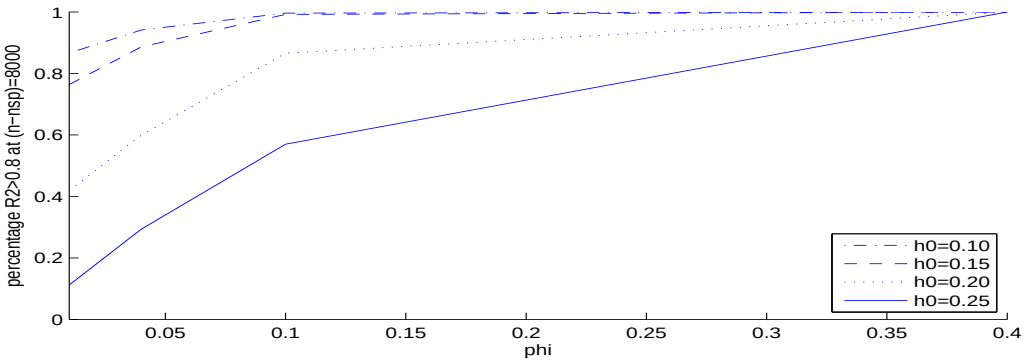


Figure 3.5: Dependence of $P_{R2\geq 0.8}$ on φ for various h_0 with $T_{R2} = 8000$ for a power law kernel

Power law kernel

Fig. 3.5 shows that $P_{R2\geq 0.8}$ is significantly lower for a power law kernel than for an exponential kernel, even if T_{R2} is much higher. An exponential decays much more rapidly over larger time spans than a power law. Therefore an estimation in a time series with a power law memory kernel needs a much longer T_{R2} to 'forget' the returns d_n before n_{sp}.

Secondly Fig. 3.5 shows that also for a power law $P_{R2\geq 0.8}$ increases with φ and decreases with h_0. As for an exponential kernel, φ represents the decay of 'memory' in a power law kernel. Therefore the qualitative influence of φ is the same, namely 'forgetting' the returns d_n before n_{sp} at a faster rate. The role of h_0 is the same in a power law kernel as an exponential kernel, being a measure for non-stationary and the kurtosis of the time series. In both kernel types larger h_0 lead to higher variability in the returns d_n before n_{sp}, therefore increasing the error at n_{sp}, leading to lower $P_{R2\geq 0.8}$ at given φ.

To illustrate the big difference in the quality of the volatility estimation for different parameters with a power law kernel, I calculate the average R^2 at $T_{R2} = 8000$ and present the landscape as a function of φ and

h_0 in Fig. 3.6. This allows us to more accurately determine for which parameters sets the volatility estimation is accurate. We see that the R^2 decays rapidly if roughly φ/h_0 exceeds a certain value.

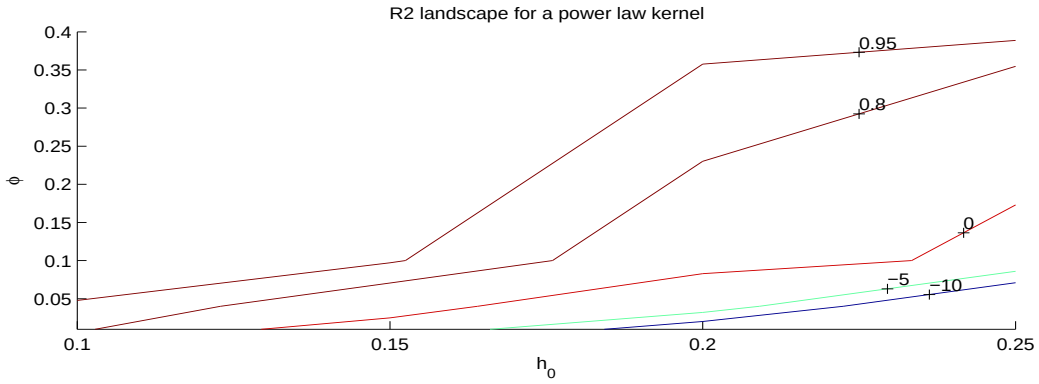


Figure 3.6: The R^2 landscape of a power law kernel as a function of h_0 and φ for $T_{R2} = 8000$.

The quantiles of the estimated volatility relative to the real process volatility σ_{est}/σ with a 95% confidence level were also calculated. Specifically, of a set of 10^3 simulated time series, I excluded the results with the 2.5% lowest and highest σ_{est}/σ. Figure 3.7 shows the maximum and minimum values for different parameters of the resulting collection of results. We see that indeed for high h_0 and low φ the estimation of the volatility is very inaccurate, since the volatility can be estimated four times too high.

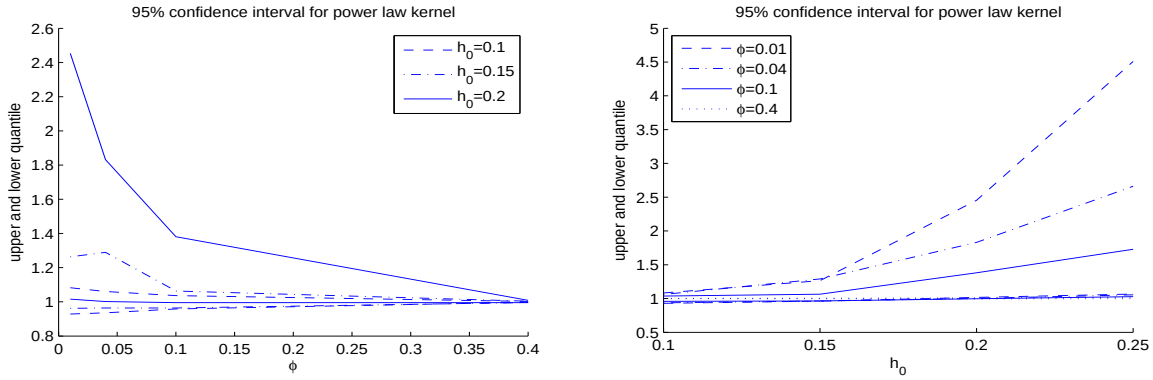


Figure 3.7: Upper and lower quantiles as a function of φ (Left plot) and h_0 (Right plot) for $T_{R2} = 8000$. Each point is based on 4000 simulations.

3.1.4 Influence of the initial volatility

We have seen that for both kernel types, increasing h_0 had a negative influence on the quality of volatility estimation R^2. Since a higher h_0 results in a higher variability in $\sigma_{nsp,real}$, I proposed an explanation that $\sigma_{nsp,real}$ had a strong influence on the quality of the estimation of the volatility (R^2). In this section it is demonstrated that $\sigma_{nsp,real}$ indeed has a strong influence on the estimation quality R^2, supporting the hypothesis.

In order to find the influence of $\sigma_{nsp,real}$ on R^2 at different time steps, I simulated 10^3 time series, recorded $\sigma_{nsp,real}$, estimated the volatility and computed R^2 at different T_{R2}. Also the probability distribution of the

volatility σ_{real} was computed. This was done by simulating 10^4 time series and producing a histogram of $\sigma_{n,real}$ with $n = 5500$ (halfway between n_{sp} and end of time series) for different kernel types and parameter.

The influence of $\sigma_{nsp,real}$ on R^2 at different time steps is displayed in Fig. 3.8 (Left). The Figure shows that large errors (low R^2) occurred if $\sigma_{nsp,real}$ deviated strongly from 1. We can explain this effect because a unit initial condition is assumed in each estimation ($\sigma_{nsp,est} = 1$). Specifically, Fig. 3.8 (Left) shows that for large $|\log(\sigma_{nsp,real})| = |\omega(n_{sp})|$, large intervals T_{R2} were needed to reach the $R^2 = 0.8$ threshold and for some very big $|\log(\sigma_{nsp,real})|$ this threshold was never reached. Fortunately, we also see that the majority of the $|\log(\sigma_{nsp,real})|$ are small.

Figure 3.8 (Right) shows the probability distribution function (pdf) of the volatility in synthetic time series $\sigma_{n,real}$. The pdf indicates the peak around 1 and the relatively fact decay of the tails. Therefore the pdf confirms that the majority of the $|\log(\sigma_{nsp,real})|$ are small, because the plot has a peak around 1. The same properties are observed for all investigated parameter sets.

Therefore, we can conclude that the volatility at the starting point of the estimation has a strong influence on the quality of the estimation.

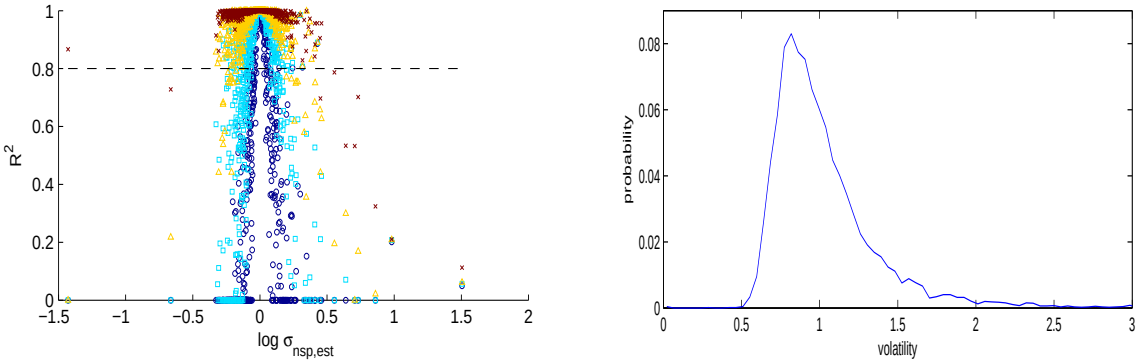


Figure 3.8: **Left**: Dependence of R^2 on $|\log(\sigma_{nsp,real})|$ at different T_{R2}, specifically $T_{R2} = 3000$ (Brown asterisk), $T_{R2} = 1000$ (Yellow triangles), $T_{R2} = 300$ (Light blue squares), $T_{R2} = 100$ (Dark blue rounds). The dotted line represents the threshold $R^2 = 0.8$. Exponential kernel with h_0=0.06 and φ=0.01. **Right:** Probability distribution of $\sigma_{n,real}$ with $n = 5500$. Exponential kernel with h_0=0.04 and φ=0.01. The median of $\sigma_{n,real} = 0.95$ and average $\sigma_{n,real} = 1.04$.

3.1.5 Bias in the estimation of the volatility for a power law kernel

This subsection demonstrates that the estimation of the volatility is positively biased ($\sigma_{n,est} > \sigma_{n,real}$). I define the bias as $\sigma_{n,est}/\sigma_{n,real}$. It was found that this bias is much stronger and persistent for a power law kernel than for an exponential kernel. The bias was investigated and quantified for a power law and an exponential kernel.

Power law kernel

I have simulated 10^3 time series and estimated the volatility $\sigma_{n,est}$ for several power law kernel parameters. Then, $\sigma_{n,est}/\sigma_{n,real}$ was determined at different moments in time. The biases $\sigma_{n,est}/\sigma_{n,real}$ corresponding to individual estimations are displayed in Fig. 3.9 (Left) as blue lines, while the fat red line indicates the average of all biases. It shows that the estimation is positively biased as 98% of the volatility estimations were too high ($\sigma_{n,est}/\sigma_{n,real} > 1$).

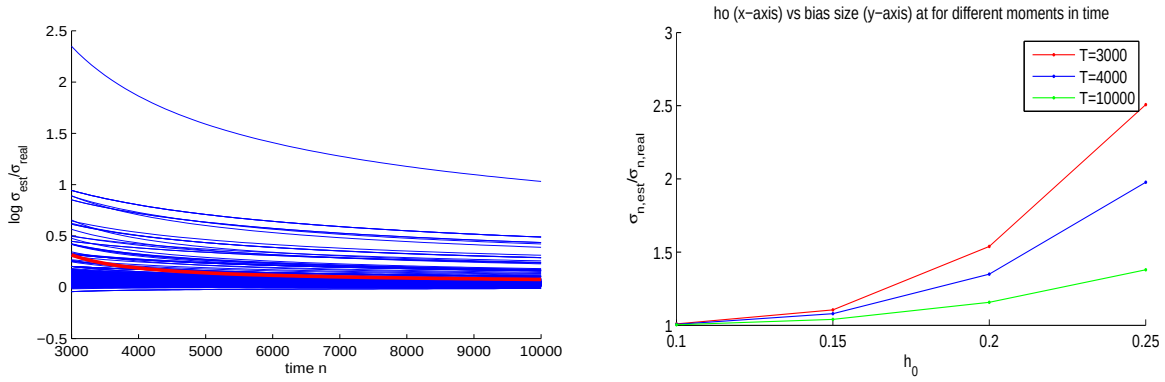


Figure 3.9: **Left**: Bias $\sigma_{n,est}/\sigma_{n,real}$ of individual estimations (blue lines) as a function of time. The fat red line represents the average of all the observed biases. Power law kernel with h_0=0.2 and φ=0.01. **Right**: Dependence of the bias size $\sigma_{n,est}/\sigma_{n,real}$ on h_0, also at different n (indicated in by T).

Similar behavior was observed for the other power law kernel parameter sets. Fig. 3.9 (Right) shows the bias size at different n as a function of h_0. It can be seen that the average bias increased with h_0.

Using equations 3.1 and 3.4 I derived an expression for the bias, which is presented in equation 3.9. We see that this expression supports the observed qualitative behavior of increased bias size as a function of h_0. In the interpretation of equation 3.9 it should be kept in mind that h_0 and φ also influence the statistics of the returns d_n.

$$\frac{\sigma_{est}}{\sigma_{real}} = \left[\exp\left\{-\sum_{i=1}^{n_{sp}} d_i (n-i-1)^{-\varphi-1/2}\right\}\right]^{h_0} \tag{3.9}$$

The occurrence of a positive bias indicates that on average $\sum_{i=1}^{n_{sp}} d_i$ was positive (see equation 3.9). An explanation for this has not been found.

Exponential kernel

Fig. 3.10 shows the bias of individual estimations for an exponential kernel. It demonstrates that the bias of the exponential kernel was very small and decayed very fast. This result was representative for other kernel parameters. Therefore we can conclude that the bias for an exponential kernel was insignificant.

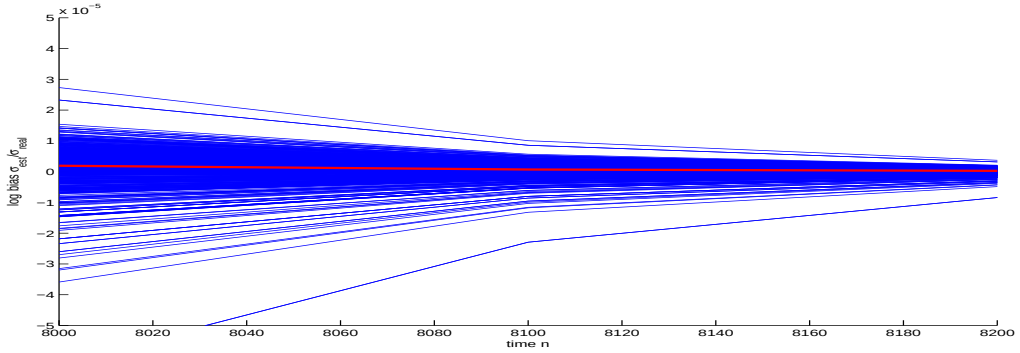


Figure 3.10: Bias $\sigma_{n,est}/\sigma_{n,real}$ of estimations based on single time series (blue lines) as a function of time. is observed when estimating σ_{est}. The red line represents the average of all the observed biases. Exponential kernel with h_0=0.06 and φ=0.01.

3.1.6 Conclusion

I have proposed a method for estimating volatility and innovations in pure SEMF processes that assumes a unit value for the initial condition. We have seen that this method is effective in estimation the volatility with a high quality for a broad range of parameters. For a power-law kernel with a high non-stationary (high h_0) and low memory decay (low φ) the method is not effective.

We have seen that the estimation converges over time if volatility bursts, that are also observed in [16], are excluded. Furthermore, increasing h_0 has a negative effect on the quality of the estimation (R^2). A higher h_0 increases the variability of the returns at the start of the estimation ($\sigma_{nsp,real}$), and therefore the error made by assuming neutral history ($\sigma_{nsp,est} = 1$) at this point. Increasing φ has a positive effect on the estimation quality (R^2), because it increases the decay over time of memory of past events that are not taken into account.

Furthermore it was determined that $\sigma_{nsp,real}$ has a large influence on the quality of the estimation (R^2). Moreover, a positive bias in the estimation procedure is observed for a power law kernel. I found that the size of the bias is largely a function of h_0 and time.

In the appendix B I propose different estimators to improve the volatility estimation quality in the problematic parameter ranges for a power-law kernel. They are based upon the large influence of the initial condition and the observed bias. However I find that the improvements in quality made by these changes is very small. Therefore I do not use them and choose not present them here in detail.

3.2 Influence of uncertainty in the kernel type and parameters on the estimation quality

We have seen that the volatility of a pure SEMF process can be estimated in most parameter ranges with a high quality if the process parameters are known. In non-synthetic time series however the process parameters are not known and have to be estimated. Most likely estimated process parameters have a degree of uncertainty. Therefore it is of interest to known whether the high estimation quality holds if there is uncertainty in the parameters used for estimation.

This section demonstrates that the volatility and therefore also the corresponding innovations of a pure SEMF process can still be estimated with high quality if there is a small uncertainty in the process parameter used for estimation.

First we see that if the different kernel type is used for estimation, the estimation does not converge to the real value and no high quality of estimation can be achieved. Second we see that if there is a small deviation in the parameter used for estimation, the estimation quality R^2 decreases.

Furthermore we see that a joint uncertainty in the three parameters σ_{est}, φ_{est} and $h_{0,est}$ has a combined influence on the quality of estimation R^2. This joint influence can be approximated by addition of the squared uncertainties in the individual parameters, normalized by the sensitivity of the estimation quality to each individual parameter. I also quantify this sensitivity of the quality to the individual parameters with an approximation for different kernel types and parameters. It is found that particularly σ requires a high accuracy (around 5%). Finally, we see that the estimation quality becomes more sensitive to uncertainty for higher h_0 and lower φ.

Knowing that we can estimate the volatility and innovations with accuracy even if there is uncertainty in the kernel parameters, we look at methods for parameter estimation in the next section.

3.2.1 Uncertainty in kernel type

It is not known which of the proposed kernel makes the optimal approximation of real financial data. Therefore we first investigate the consequences of fitting time series, which were generated using one of the two proposed kernels, with the wrong kernel type. The resulting squared errors of the estimated innovations were computed as a function of time. Examples of the evolution of the squared error as a function of time for both cases can be seen in Fig. 3.11.

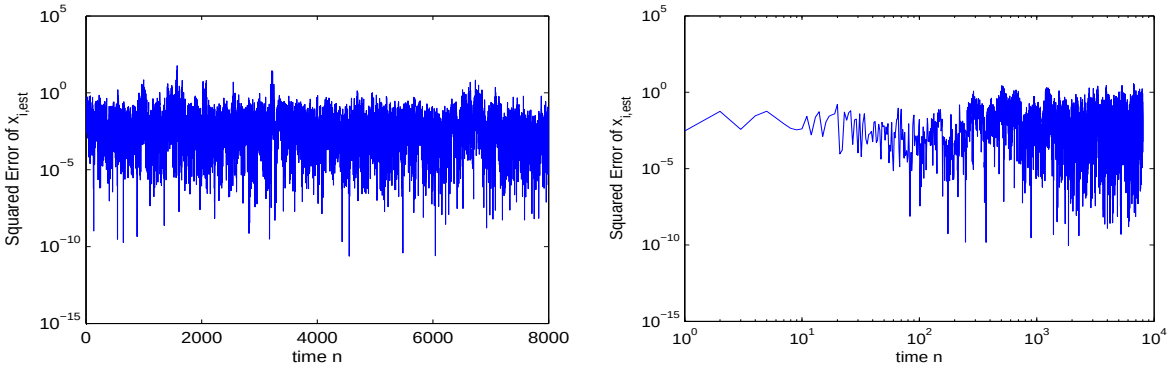


Figure 3.11: Influence of fitting a time series generates by one kernel type with the other kernel type on the squared errors as a function of time. **Left:** Time series with an exponential kernel fitted with a power law kernel. **Right:** Time series with a power law kernel fitted with an exponential kernel

It can be seen that the squared error does not continuously decrease as a function of time and therefore that $\xi_{n,est}$ does not converge to $\xi_{n,real}$ over time. This results in negative values for R^2. These results were representative for broad ranges of process parameters.

Therefore we can conclude that accurate estimations $\xi_{n,est}$ require a significant resemblance between the kernel used for the estimation and the kernel that 'generated' the time series. The resemblance between the exponential kernel and power law kernel is too small for accurate $\xi_{n,est}$.

3.2.2 Qualitative influence of uncertainty in kernel parameters

As a next step e the qualitative influence of uncertainty in kernel parameters is investigated. I generate time series with certain kernel parameters and then estimate the volatility with a small deviation in one of the

kernel parameters and evaluate R^2 (again $L_{R2} = 100$) as a function of time T_{R2}. A representative example of the results is displayed in Fig. 3.12.

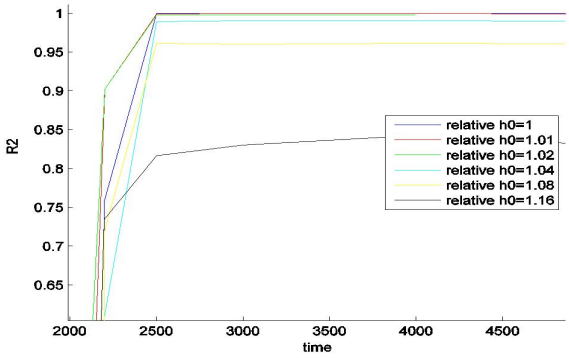


Figure 3.12: Influence of fitting with a different h_0 on the evolution of R^2 over time. Exponential kernel with $h_0 = 0.06$ and $\varphi = 0.01$

We see that the estimation quality R^2 decreases as a function of $h_{0,est}/h_0$. Similar results were observed for different parameter ranges and it was found that the effect was symmetric (the effect of $|h_{0,est}/h_0 - 1|$ on R^2 is the same for $h_{0,est}/h_0 < 1$ as for $h_{0,est}/h_0 > 1$). The effect on the estimation quality was observed to increase in strength for kernels with slower memory decay (smaller φ).

3.2.3 Uncertainty in all parameters

The influence of uncertainty in multiple parameters on the quality of the volatility estimation R^2 is investigated. I simulated time series with both kernel types with given parameters h_0, φ and σ. The $\sigma_{n,est}$ and $\xi_{n,est}$ were estimated with parameters h_0', φ', σ' where $h_0' = f_{h_0} h_0$, $\varphi' = f_\varphi \varphi$ and $\sigma' = f_\sigma \sigma$. f_{h_0}, f_φ and f_σ indicate the relative parameters used for the fit. Then, for several sets of $(f_{h_0}, f_\varphi, f_\sigma)$ the estimation quality R^2 (for $T_{R2} = 8000$ and $L_{R2} = 100$) was calculated for the corresponding estimated volatility $\sigma_{n,est}$. Examples of the three cross sections ($f_\sigma = 1$, $f_\varphi = 1$ and $f_{h_0} = 1$) of the resulting 3D landscape of R^2 are displayed in Fig. 3.13. These cross sections are also representative for cross sections where there was uncertainty in all three parameters (e.g., $f_\sigma = 1.1$ for the cross-section of f_φ and f_{h_0}).

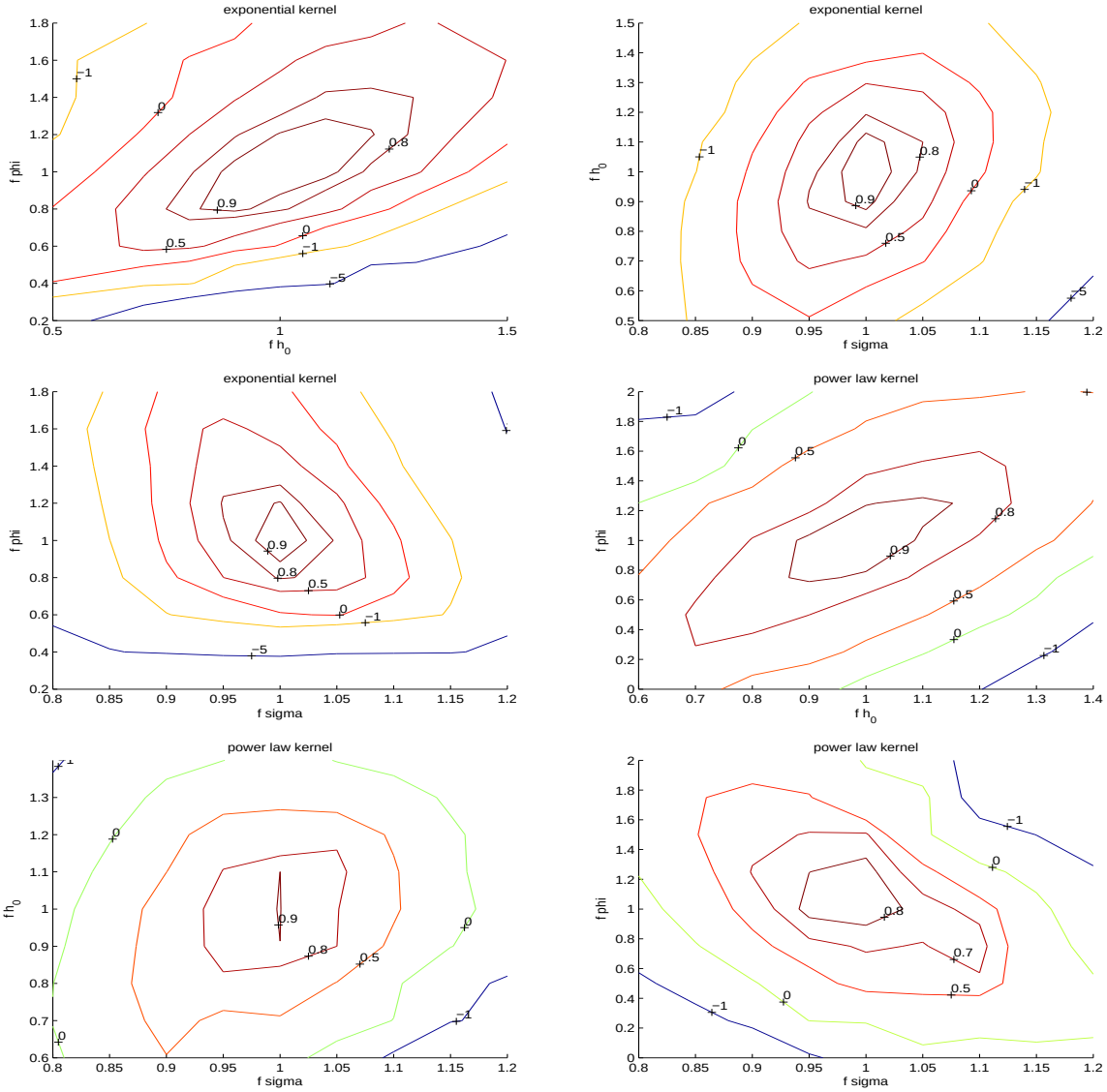


Figure 3.13: Examples of landscapes of R^2 as a function of f_{h_0}, f_φ and f_σ for both an exponential kernel ($\varphi_{exp} = 0.01$, $h_{0,exp} = 0.04$) and a power law kernel ($\varphi_{pow} = 0.1$, $h_{0,pow} = 0.15$) .

It can be seen that the isoplots of R^2 form asymmetric ellipses for the representative kernel ranges. Although the qualitative influence of deviations of the parameters on R^2 is quite similar, it can be seen that that the quantitative influence differs significantly (e.g., the influence of deviations in σ on R^2 is stronger than the influence of deviations in φ). Despite the asymmetry, the joint influence on R^2 can be approximated by an ellipse:

$$I_{R2}(R^2) = \alpha_{h_0}^2 \left|f_{h_0} - 1\right|^2 + \alpha_\varphi^2 \left|f_\varphi - 1\right|^2 + \alpha_\sigma^2 \left|f_\sigma - 1\right|^2 \tag{3.10}$$

where a fixed $I_{R2}(R^2)$ forms an isoplot for a certain R^2. The coefficients are determined by $\alpha_{h_0} = \left|(|f_{h_0+} - 1| + |f_{h_0-} - 1|)/2\right|^{-1}$ where f_{h_0+} is the higher value of f_{h_0} when $f_\varphi, f_\sigma = 1$ and R^2 is still on

$I_{R2}(R^2)$.

In the representative example of the exponential kernel, for $I_{R2}(R^2 = 0.8)$, $\alpha_{h_0}^{-1} = 0.2$, $\alpha_{\varphi}^{-1} = 0.2$, and $\alpha_{\sigma}^{-1} = 0.05$. Within this approximation, this means that for this kernel specification h_0, φ and σ require an accuracy lower than 20%, 20% and 5% respectively in order to have estimations with at least $R^2 = 0.8$.

I repeat the procedure for broader parameter ranges ($\varphi_{exp} = 0.005 - 0.02$, $h_{0,exp} = 0.04 - 0.10$) for both kernel types ($\varphi_{pow} = 0.01 - 0.4$, $h_{0,pow} = 0.10 - 0.20$). It is found that approximately $\alpha_{h_0}^{-1} = 0.20 - 0.40$, $\alpha_{\varphi}^{-1} = 0.15 - 0.35$, and $\alpha_{\sigma}^{-1} = 0.05 - 0.10$ for $I_{R2}(R^2 = 0.8)$ for both kernel types. For the power law kernel $\alpha_{h_0}^{-1}$ tends to be slightly lower (the required accuracy in h_0 tends to be slightly higher).

It was found that, for both kernel types, uncertainty in h_0 or φ has a larger decreasing effect on R^2 if h_0 is high and φ is low. We have seen in section 3.1.3 that low φ (corresponding to a strong memory) and high h_0 (corresponding to strong non-stationary) leads to lower R^2 if the parameters are known. Therefore the uncertainty in both parameters at these ranges aggravate the already existing high inaccuracies, explaining the stronger decrease in R^2.

Uncertainty in σ tends to have a larger effect for higher h_0 and φ for an exponential kernel but the effect is unclear for a power law kernel. An explanation for this has not been found.

Summarizing, uncertainty in the three parameters have a joint negative influence on the quality of the volatility estimation. The joint influence can be approximated by a sphere. Particularly σ requires high accuracy and also a higher accuracy is required for higher h_0 and lower φ.

3.3 Estimation of the kernel type and parameters

This section presents two methods that can be used to estimate the parameters in a pure SEMF process: the Conditional Maximum Likelihood (CML) method and the Minimum Innovation Volatility Clustering (MIC) method.

Both methods rely optimize different properties. The CML method optimizes the likelihood of the innovations according to a Gaussian distribution to estimate the process parameters. The MIC method aims to optimize the independent, identically distributed (i.i.d.) property of the innovations. Therefore it minimizes the volatility clustering in the innovations, to estimate the process parameters.

Section 3.3.1 demonstrates that the conditional likelihood landscape has a unique global maximum when the correct kernel type is used. We see that the method is successful in determining the correct kernel type (the method estimates the correct kernel type in close to 100% synthetic time series) . I compute the accuracy of the method for different kernel types and parameters, and determine that the method is sufficiently accurate to obtain volatility estimations with high quality. Furthermore, using the CML method, I present several scaling relations that describe how the parameters of a process at different scales are related.

Section 3.3.2 introduces the Minimum Innovation Clustering (MIC) method and demonstrates that this MIC method can also be used for SEMF process parameter estimation. It is found that the accuracy of the CML method is higher. However, the method is still of interest as it optimizes a different property than the CML method and can provide a 'second opinion' in the case of real financial time series.

Because the process parameters influence the distribution of the process returns, it is also investigated whether the distribution of the returns could be used for parameter estimation. However this was not successful. An overview of this method can be found in the appendix section C.1. It was also found that minimizing the Kendall-tau coefficient, which is the correlation between ranks of the innovations and the returns and therefore a measure of endogeneity of the process, can not be used for parameter estimation (see appendix section C.2).

Below the details of the two (relatively) successful methods are discussed.

3.3.1 Conditional Maximum Likelihood (CML)

Methods that optimize the likelihood of the innovations or a different input variable (Maximum Likelihood) are frequently used for GARCH and MSM processes. Parameter of the process are swiped until the input values have the highest joint probability. When past returns are taken into account when optimizing the likelihood, the method is called Conditional Maximum Likelihood (CML).

The likelihood of a process depends on the distribution by which the process is generated. However, maximizing the Gaussian likelihood for distributions with heavy tails but a finite fourth moment [21] is still effective for parameter estimation. The method is therefore sometimes denoted as Quasi-Maximum Likelihood, as it is not the 'real' likelihood that is maximized. Therefore the method should also be applicable to processes with a non-Gaussian distribution for the innovations (see section 2.2).

It is investigated whether we can estimate the memory kernel type and parameters using the Conditional Maximum Likelihood (CML) method. In the proposed model, the innovations ξ_i follow a Gaussian distribution with zero mean $\mu = 0$ and unit standard deviation $\sigma = 1$. A Gaussian distribution is defined by

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{3.11}$$

One can estimate the innovations $\xi_{i,est}(\theta')$ with a certain parameter set $\theta' = (\sigma', h_0', \varphi')$ and calculate how likely the innovations are according to this distribution. The innovations $\xi_{i,est}(\theta')$ can be estimated with parameter set $\theta' = (\sigma', h_0', \varphi')$ using $\xi_{i,est}(\theta') = d_n/\sigma_{n,est}(\theta')$, where $\sigma_{n,est}(\theta')$ is given by

$$\sigma_{n,est}(\sigma', h_0', \varphi') = \sigma' \exp\left\{ -\frac{\sum_{i=n_{sp}}^{n-1} d_i h'_{n-i-1})}{\sigma'} \right\} \tag{3.12}$$

where $h'_{n-i-1} = h_0(n-i-1)^{-\varphi-1/2}$ for a power law kernel and $h'_{n-i-1} = h_0 exp(-\varphi n)$ for an exponential kernel.

Using the relation $\xi_{i,est} = d_i/\sigma_{i,est}$, the likelihood of the estimated innovations conditional of the parameters $-\mathbf{l}_n(\theta)$ or the cost function $\mathbf{l}_n(\theta)$ can be expressed as:

$$\mathbf{l}_n(\vartheta) = \frac{1}{n} \sum_{i=n_{start}}^{n} l_i(\vartheta) \text{ where } l_i(\vartheta) = \frac{d_i^2}{\sigma_{i,est}^2} + log(\sigma_{i,est}^2) \tag{3.13}$$

where $n_{start} = n_{sp} = 2001$ for Monte Carlo simulations and $n_{start} = 1$ for real financial time series. The optimal parameter set is defined as:

$$\hat{\vartheta}_n = \arg\min_{\vartheta} \mathbf{l}_n(\vartheta) \tag{3.14}$$

where $\hat{\vartheta}_n$ are the parameter estimates. This parameter estimation process is identical in the case of GARCH processes [1]. The minimization is done using a standard sweeping algorithm (Matlab: fminsearch, which uses the simplex search method [30]).

Since the procedure is tested on Monte Carlo simulated time series with known parameters, the cost function is calculated relative to the cost function of the real time series $\mathbf{l}_n^{rel}(\vartheta)$

$$\mathbf{l}_n^{rel}(\vartheta) = \frac{1}{n} \sum_{i=1}^{n} \frac{d_i^2}{\sigma_{i,est}^2} + log(\sigma_{i,est}^2) - \frac{1}{n} \sum_{i=1}^{n} \frac{d_i^2}{\sigma_{i,real}^2} + log(\sigma_{i,real}^2) \tag{3.15}$$

in the case of likelihood landscapes.

3.3.1.1 Synthetic conditional likelihood landscapes

It is important to determine whether the proposed maximum likelihood method can distinguish time series generated with different kernel types and whether the resulting estimated parameters are robust for a single time series. For these purposes I investigate the likelihood landscape of time series fitted by both kernel types and also investigate in which cases there is a local minimum.

I simulate time series with an exponential and power law kernel and calculate the relative cost function (see equation. 3.15) for different parameters $\theta' = (\sigma', h_0', \varphi')$ on a large scale. For each time series this was done for both kernel types.

Examples of how the relative cost function $\mathbf{l}_n(\theta')$ of the innovations depends on h_0' and φ' when $\sigma' = \sigma$ in synthetic time series can be seen in Fig. 3.14 for different combinations of kernel types. The correct kernel type indicates that the memory kernel used for simulation was also used for estimation of the innovations.

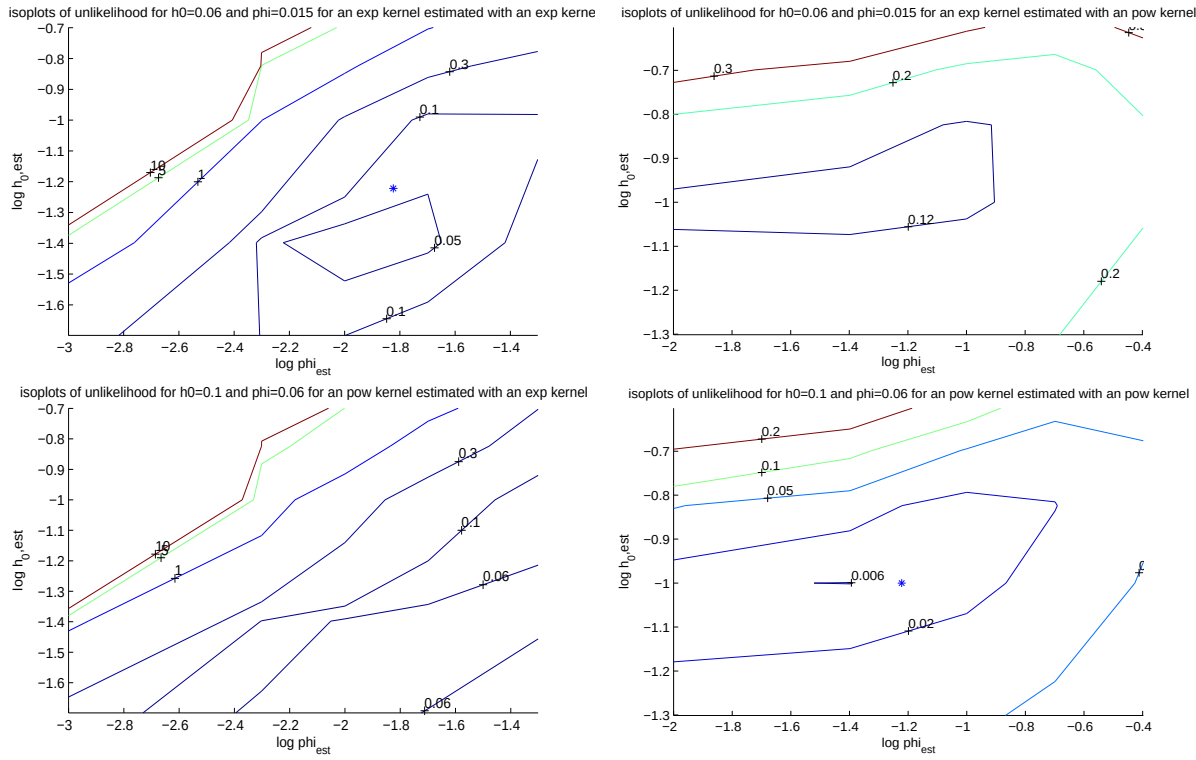


Figure 3.14: Relative negative likelihood (cost function) $-\mathbf{l}_n(\theta)$ as a function of log h_0' and φ' for both kernel types. The blue star indicates the parameter set (h_0, φ) that was used to simulated the time series (for the same kernel type) and $\sigma = \sigma' = 1$. Minimum cost function from top left to bottom right respectively: 0.0170; 0.0775; 0.0367; 0.0132

It can be seen that for both kernel types there is a local minimum if the kernel type is correct. We then also see a global minimum near to the parameter set that was used for simulation. When a different kernel is used for estimation, the minimum is not as clearly defined as it is on the boundary of the used scale (even though the scale is representative of the parameter range in which the process displays multifractality). Furthermore for these particular experiments the estimation using the correct kernel type also has a higher likelihood.

The landscape does not change qualitatively if also a different σ' was used for the fit. The landscape was also computed and investigated as a function of (σ', h_0') and (σ', φ'). It was found that the likelihood has

a strong minimum around σ' regardless the kernel type used. This implies that an accurate estimation of σ is also possible if a different memory kernel type is used for the fit. Figure 3.15 shows an example of this property for a time series generated with an exponential kernel.

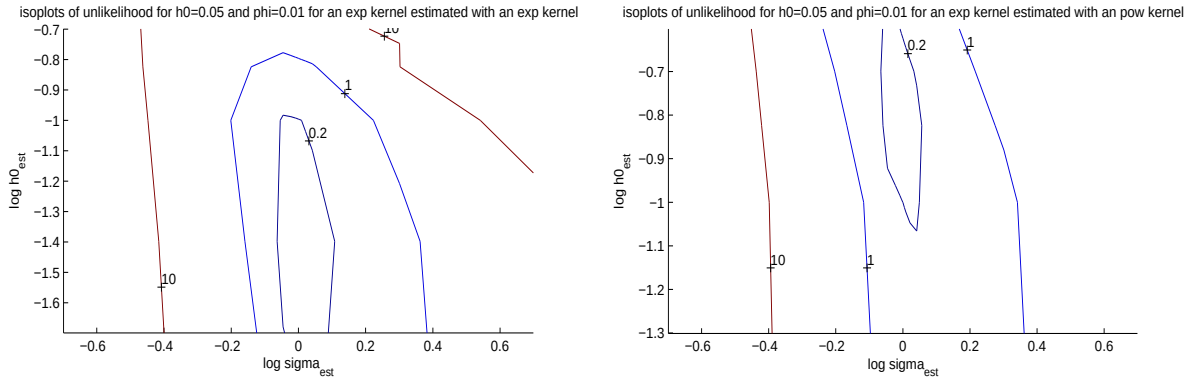


Figure 3.15: Relative negative likelihood (cost function) $-\mathbf{l}_n(\theta)$ as a function of $\log h_0'$ and $\log\varphi'$for an exponential kernel fitted by both kernel types. For the generation of the time series I used $\sigma = 1$, $h_0 = 0.05$ and $\varphi = \varphi' = 0.01$.

The computation of likelihood landscapes on larger scales supported all these findings.

Summarizing, we have seen that the likelihood landscape of an estimation made with the correct memory kernel has a unique minimum. Furthermore, there is a strong minimum around σ' even if a different kernel type is used for estimation than for generation of the time series.

3.3.1.2 Synthetic results of conditional max likelihood

As a next step, it is investigated the accuracy of the proposed method in parameter estimation. I simulated 100 time series for four kernel parameter sets for both kernel types ($\varphi_{exp} = 0.005, 0.02$, $h_{0,exp} = 0.04 - 0.10$) and ($\varphi_{pow} = (0.01, 0.4)$, $h_{0,pow} = 0.10, 0.20$). The conditional likelihood landscape was swiped using the correct kernel type to obtain the parameter set (φ_{est}, $h_{_{0.est}}$, σ_{est}). Figure 3.16 (time series with an exponential memory kernel) and Figure 3.17 (power law memory kernel) show the resulting parameters (φ_{est}, $h_{_{0.est}}$, σ_{est}) for the four different parameter sets used for generation of time series (φ, $h_{_0}$, σ). The landscape was also swiped using the different kernel type and the corresponding highest likelihood was compared with the highest likelihood of the correct kernel type.

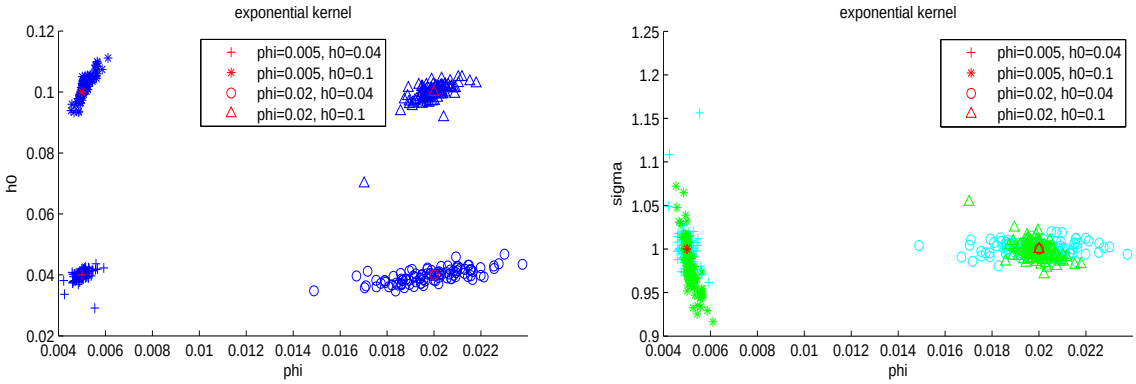


Figure 3.16: Results of fitting the parameters using conditional maximum likelihood of 100 Monte Carlo simulations with an exponential kernel. The red symbols indicate the parameters that were used to simulate the time series. $\sigma_{real} = 1$ for all simulations. *Left*: φ_{est} and $h_{_{0.est}}$. *Right*: φ_{est} and σ_{est}.

I found that 99-100% of the experiments has the lowest conditional likelihood for the correct kernel type. The CML method is therefore effective in distinguishing kernel type for time series simulated using an exponential memory kernel.

Figure 3.16 demonstrates that for an exponential kernel the CML method is effective in distinguishing the four parameters sets. The uncertainty in the three parameters φ_{est}, h_{0est} and σ_{est} is different for each of the four parameter set used for generation of the time series. The accuracy in φ_{est} when φ is high increases with h_0. This is possibly because a higher non-stationary (higher h_0) leads to a stronger variability in the returns. The memory (φ) the time series has of these past returns is then probably easier to calibrate (high accuracy in φ_{est}).

The accuracy in σ increases with φ. This can be explained by the fact that a lower φ corresponds to a slower decay of memory. This increases the length of the intervals with lower or higher volatility that result from the non-stationary of the system. For higher φ these intervals tend to be shorter. This increases the accuracy in σ_{est}.

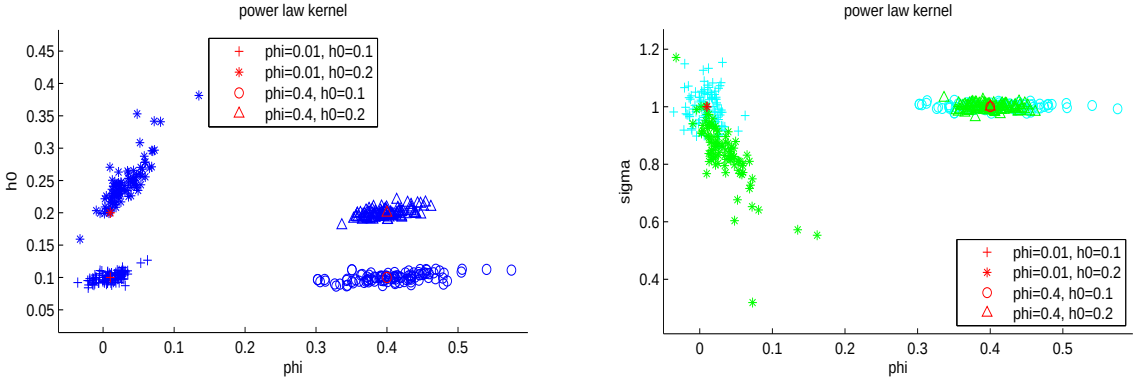


Figure 3.17: Results of fitting the parameter using conditional maximum likelihood of 100 Monte Carlo simulations with a power law kernel. Left: φ_{est} and $h_{_{0.est}}$ where four extreme values with high h_0 are not displayed in the plot. Right: φ_{est} and σ_{est} where five extreme values with h_0 are not displayed in the plot. $\sigma_{real} = 1$ for all simulations.

Figure 3.17 shows the results for a power law kernel. Overall, the uncertainty in the parameters estimated with this method is bigger for time series with a power law memory kernel than for an exponential memory kernel. Particularly the estimation of h_0 and σ seems less accurate. This is because a system with a power law memory kernel is more non-stationary than one with an exponential memory. Also, estimations of systems with a power law kernel with low φ ($\varphi = 0.01$) sometimes result in negative φ.

As for the exponential kernel, the accuracy in φ increases with h_0 (when φ is high) and the accuracy in σ increases with φ. This is probably due to the same mechanisms mentioned for exponential kernels.

The non-stationary parameter $h_{0,est}$ seems positively biased for large h_0 and low φ while σ seems negatively biased for low φ and high h_0. This is probably related to the positive bias observed earlier (see 3.1.5), which was particularly strong in this parameter range. Also, the accuracy in $h_{0,est}$ increases with φ if h_0 is high. This corresponds to our understanding of the system, as a stronger decaying memory (higher φ) is supposed to lead to a better calibration of the influence of (known) short term events (h_0).

As a next step, the relative quantiles for each parameter set with a 95% confidence level were determined. As the quality of the estimation R^2 is more sensitive to some parameters than others, I used the method introduced in 3.2.3 to estimate $\alpha_{h_0}^{-1}$, α_{φ}^{-1}, and α_{σ}^{-1} for $I_{R2}(R^2 = 0.8)$ for each parameter set. For each parameter estimation, $I_{R2}(R^2)$ was calculated and the 5 (5% of 100 experiments) with the highest values (corresponding to the lowest R^2) were excluded. From the remaining collection the lowest and highest relative (e.g., $\sigma_{est}/\sigma_{real}$) parameter estimation were taken. These values correspond to the quantiles with 95% confidence level. The results are displayed in Fig. 3.18 for both kernel types.

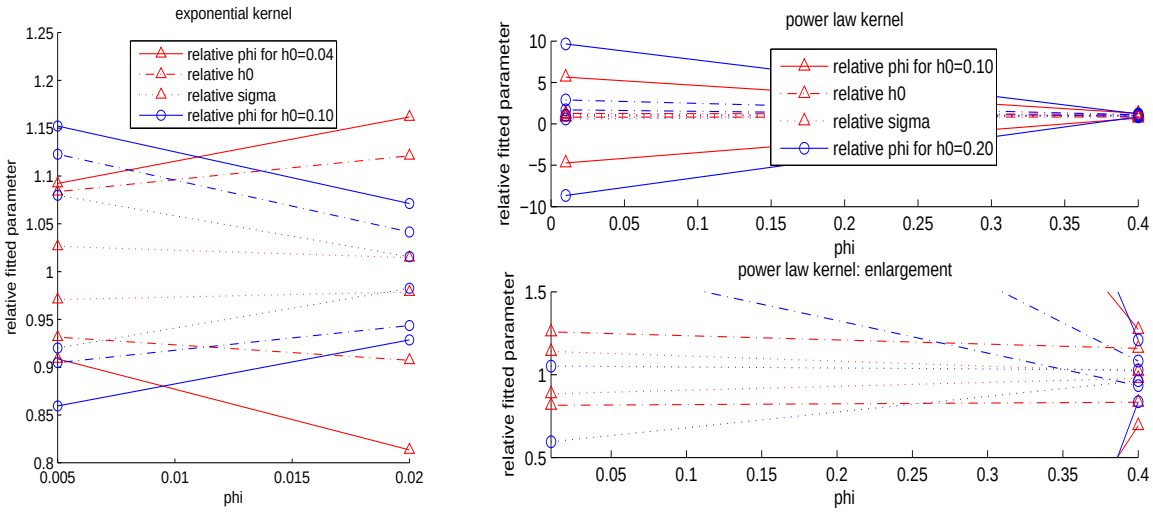


Figure 3.18: Quantiles with 95% confidence in terms of the relative estimated parameters for the conditional maximum likelihood methodology. **Left**: Quantiles for exponential kernels. **Right**: Quantiles for power law kernels with an enlargement below.

We compare the results with the accuracy required for high quality estimation of the innovations ($R^2 = 0.8$) in section 3.2.3. For an exponential kernel, the maximum uncertainty is higher than the values of these quantiles for almost all parameter sets (except for high h_0 and low φ). Therefore at least 95% of the CML parameter estimations result in volatility estimations with an accuracy of $R^2 = 0.8$ in a pure SEMF process. However for low φ and high h_0 this requirement is not met as the uncertainty in σ is too big.

For power law kernels, we see that the uncertainty in φ is very high, especially if the memory decay φ is low. Therefore for this parameter range the tested methodology is not accurate enough to allow for

accurate volatility estimations. For higher φ, there is still a high uncertainty in this parameter but most of the estimations result in high quality innovations.

It was found that there was no correlation between the resulting conditional likelihood and the accuracy of the parameters.

Exclusion of early likelihoods

In section 3.1.2 we saw that the accuracy of the estimated innovations increases over time. This implies that the conditional likelihood of the estimated innovations right after n_{sp} are the least accurate. Therefore it is investigated whether ignoring these 'early' likelihoods improves the accuracy of the estimated parameters. I substitute $n_{start} = n_{sp}$ in equation 3.13 by $n_{start} = n_{sp} + n_{excl}$ to create a new estimator:

$$\mathbf{l}_n(\vartheta) = \frac{1}{n} \sum_{i=n_{sp}+n_{excl}}^{n} l_i(\vartheta) \text{ where } l_i(\vartheta) = \frac{d_i^2}{\sigma_{i,est}^2} + log(\sigma_{i,est}^2) \tag{3.16}$$

The estimation of $\sigma_{i,est}$ still starts at n_{sp}. I simulated 100 time series for different parameter sets and estimated the parameters in three ways:

1. with the 'normal' conditional maximum likelihood of equation 3.13
2. using equation 3.16, with $n_{excl} = 1000$
3. using equation 3.16 with $n_{excl} = 200$

The total length of the used time series remained 10^4 data points. It was found that $n_{excl} = 200$ provides the most accurate results for the estimation of $h_{0,est}$ and σ_{est}. Furthermore it was established that the n_{excl} optimal for parameter estimation varied for different kernel types and parameters h_0 and φ. This implies that an optimal parameter estimation procedure requires an iteration: The parameters are estimated with some general n_{excl}, and then the procedure is repeated with an n_{excl} optimal for the parameter range found in the first estimation.

Since the conditional maximum likelihood method was already quite accurate and the improvements are relatively small, I decided to ignore the iteration step and fix $n_{excl} = 200$ for all of the estimations.

3.3.1.3 Influence of scale on fitted parameters

It has been demonstrated that the SEMF process exhibits multifractal properties [9]. However, it is not known if and how the process parameters h_0, φ, σ are related at different time scales. This knowledge is needed to interpret the fitted process parameters of real financial time series. Therefore I apply the CML method at different time scales of a synthetic SEMF process in an attempt to find these relations.

I simulated 100 time series of length $2 * 10^4$ and calculate the log price $p(n)$ (which is the sum of all previous returns). These time series are generated for both kernel types with parameter sets optimal for CML ($\varphi_{exp} = 0.02$, $h_{0,exp} = 0.10$ and $\varphi_{pow} = 0.4$, $h_{0,pow} = 0.20$). Then 3 additional returns time series for each kernel on different grid scales $g = 2, 3, 4$ are computed :

$$d_g(n) = p(ng) - p((n-1)g) \tag{3.17}$$

where $n = 1, 2, 3, .., (2 * 10^4)/g$ .The resulting $d_g(n)$ have different lengths (10^4, $0.67 * 10^4$ and $0.5 * 10^4$ respectively). As a next step, the CML method is used to estimate the parameters at every grid scale. The average relative fitted parameter (e.g., $\varphi_{g,exp}/\varphi_{exp}$) is then computed as a function of grid scale g. The results are displayed in Fig. 3.19.

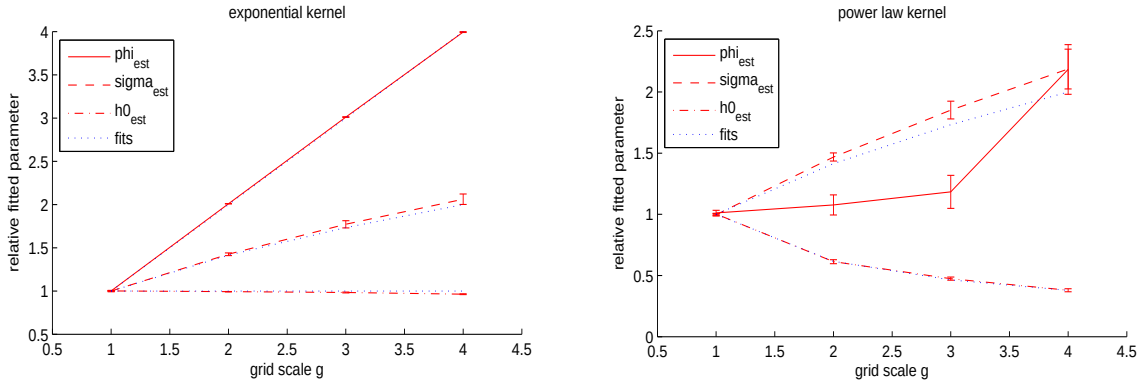


Figure 3.19: The average relative estimated (with CML) parameter at different time scales with attempted fits (dotted blue lines). The 80% confidence level quantiles are indicated by the error bars as a measure of uncertainty. **Left**: Exponential kernel. **Right**: Power law kernel

We see that for the exponential kernel the parameters are scaled in an orderly fashion. Accurate scaling relations are found:

$$\sigma_{g,exp} = \sigma_{1,exp}\sqrt{g} \tag{3.18}$$

$$\varphi_{g,exp} = g\varphi_{1,exp} \tag{3.19}$$

$$h_{0,g,exp} = h_{0,1,exp} \tag{3.20}$$

where $\vartheta_{g,exp}$ indicates the parameter value for an exponential kernel at scale g . As σ is a measure of variability at a certain time scales, it is not surprising to see that it scales the same way as standard deviation is added: $\sigma^2_{total} = \sigma^2_1 + \sigma^2_2$.

Computing ω on a rougher scale g roughly consists of ignoring a set g of intermediate time multiplications with $exp(-\varphi)$, so that $exp(-\varphi)^g$. Therefore φ is scaled linearly with g, which explains the scaling relation for φ.

The intermittency parameter h_0 does not change as a function of scale. This makes sense because the scaling relations for σ andφ already adjust the parameters of the process to account for all changes. Therefore for an exponential kernel the intermittency and non-stationary parameter h_0 is equal at different time scales.

For a power law kernel the same scaling relation for σ is valid:

$$\sigma_{g,pow} = \sigma_{1,pow}\sqrt{g} \tag{3.21}$$

The non-stationary parameter h_0 is found to roughly scale by $h_{0,g} = g^{-0.7}h_{0,1}$. For φ we can at most conclude that the relation is convex. The absence of a smooth scaling relation (that was found for an exponential kernel) for a power law kernel can be explained by the non-linear influence of a power law memory kernel over time (in contrast to the exponential kernel).

It was found that the scaling relation for h_0 does not realistically describe the system dynamics. The scaling relation implies that for small σ (very small g) the h_0 corresponding to the same level of non-stationary is much bigger. However, when I compute a time series with smaller σ and corresponding higher h_0, I find that such a time series has much higher degree of non-stationary. Therefore the empirically observed scaling relation for h_0 for a power law kernel is not valid. Therefore it is assumed that equation 3.20 is also valid for a power law kernel.

3.3.1.4 Summary

The parameters of a pure SEMF process can be estimated with a high accuracy in a broad parameter range using Conditional Maximum Likelihood. The CML method can distinguish the two kernel types with a very high accuracy (around 99-100%).

For an exponential kernel the parameters can be estimated with sufficient accuracy to obtain volatility estimations with a high quality ($R^2 = 0.8$). This high quality can be obtained for all tested parameters except for systems with a high non-stationary ($h_0 = 0.10$) and weak memory decay ($\varphi = 0.005$). For a power law kernel the method was able to distinguish between parameter sets. The method is accurate enough to obtain high quality volatility estimations ($R^2 = 0.8$), unless if the memory decay was very weak ($\varphi = 0.01$).

I demonstrated that the process parameters have smooth scaling relations for an exponential kernel. For a power law kernel we only saw a smooth scaling relation for σ.

3.3.2 Estimating the parameters by Minimizing Innovation Clustering

The possibility of estimating the kernel type, $h_{0,est}$ and φ_{est} by minimizing volatility clustering $\varrho(d, \theta, l)$ in the estimated innovations is investigated. I determine the accuracy of the naive method with one delay for different kernel types and parameters. The method is then optimized with regards to the number of delays n_l and the values of the delays l used for the computation of the volatility clustering.

3.3.2.1 Minimum Innovation Clustering (MIC)

I propose to estimate the process parameters based on the supposed independence of innovations ξ_i. Therefore the clustering of the estimated innovations is minimized, which is expressed as:

$$\varrho(d, \theta', l) = \sum_{n=1}^{N} corr(\xi^2_{n+l,est}, \xi^2_{n,est}) \tag{3.22}$$

where d is a times series of log returns, θ' the proposed parameter set and l the delay. The correlation function is defined as

$$corr(x, y) = \frac{E\left[(y - \mu_y)(x - \mu_x)\right]}{\sigma_x \sigma_y} \tag{3.23}$$

where μ_x is the average value of the process x and σ_x is the standard deviation of the process. The $\xi^2_{n,est}$ depend on the estimation parameters θ' and the input time series d_n according to the equations **??** and 3.12 described before in section 3.3.1.

The the optimal parameter set is defined as

$$\hat{\vartheta}_n = \arg\min_{\vartheta} \varrho(d, \theta, l) \tag{3.24}$$

for a given delay l and time series of log returns d. The minimization is again done using the standard sweeping algorithm. The time scale σ can not be estimated using this method because the covariance is normalized in equation 3.22. Without this normalization the function $\varrho(d, \theta, l)$ would have ever decreasing values for increasing σ_{est}. Therefore, I fixed this parameter to either unit value ($\sigma' = 1$) for theoretical estimations or to a value obtained by conditional maximum likelihood ($\sigma' = \sigma_{est}$) for applications on real financial time series.

It is also possible to calculate the value for multiple delays l:

$$\varrho(d, \theta, l_1, l_2, l_3, ..) = \varrho(d, \theta, l_1) + \varrho(d, \theta, l_2) + \varrho(d, \theta, l_3) + .. \tag{3.25}$$

We will optimize the number of delays n_l and their values $(l_1, l_2, .., l_{n_l})$ in order to get the highest accuracy of estimation.

3.3.2.2 Results on kernel type and parameters estimation

To investigate the accuracy of the proposed method, I simulated 100 time series with different kernel parameters and types ($\varphi_{exp} = 0.01, 0.05$, $h_{0,exp} = 0.04 - 0.10$) and ($\varphi_{pow} = (0.01, 0.4)$, $h_{0,pow} = 0.10, 0.20$). The volatility clustering $\varrho(d, \theta, l)$ was minimized using both kernel types for each time series. Table 3.1 shows the percentage of time series for which $\varrho(d, \theta, l)$ was lower for the kernel type with which the time series was generated. $\varrho(d, \theta, l)$ was computed for $n_l = 1$ and $l = 1$ in this case.

Exponential kernel	$h_0 = 0.04$	$h_0 = 0.10$
$\varphi = 0.01$	82%	100%
$\varphi = 0.05$	67%	97%

Power law kernel	$h_0 = 0.10$	$h_0 = 0.20$
$\varphi = 0.01$	85%	78%
$\varphi = 0.4$	92%	100%

Table 3.1: The percentage simulations for which the correct kernel type was determined. This was done by comparing the minimum $\varrho(d, \theta, l)$ for both kernel types. The correct kernel type was the kernel type used for Monte Carlo simulation of the time series.

We see that the method is reasonably effective in determining the kernel type. For lower h_0 ranges, the method loses accuracy.

The conditional maximum likelihood procedure is more effective in determining kernel type (see section 3.3.1.1, the success rate was around 99%).

For the same time series, Fig. 3.20 displays the optimal parameters for which $\varrho(d, \theta, l)$ with $l = 1$ in the case the right kernel type was determined for an exponential kernel and a power law kernel.

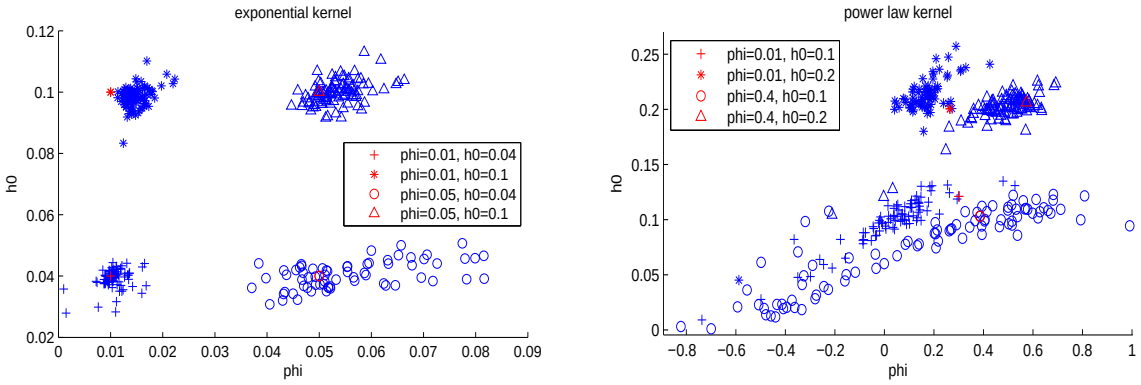


Figure 3.20: Results of estimating the parameters $h_{0,est}$ and φ_{est} by minimizing the volatility clustering of the innovations of 100 Monte Carlo simulations for both kernel types kernel. **Left**: The results for an exponential kernel **Right**: The results for a power law kernel

For both kernel types we see that the estimation of φ_{est} is very inaccurate if h_0 is low, particularly if φ is high. This parameter range corresponds to low non-stationary and strong decay of memory. Strong memory decay (high φ) results in a relatively low degree of volatility clustering in the returns. Therefore minimizing the volatility clustering in the innovations corresponds to calibrating φ' to remove a relatively small effect, resulting in a larger uncertainty in φ.

For the CML method we have also seen that a lower h_0 could have a negative effect on the estimation accuracy of φ_{est}. The explanation for this negative effect is the same for the MIC. Because of the low variability in the returns because of the low h_0, the estimation procedure does not have enough 'calibration points'.

For an exponential kernel, we can observe that for high h_0 the estimated φ_{est} is positively biased, while for a power law kernel φ_{est} is negatively biased. An explanation for this has not been found.

3.3.2.3 Optimal number of delays

Here it is demonstrated that the accuracy of the method can not be significantly improved using n_l greater than one, and that $l = 1$ produces the most accurate results.

I simulated 10 time series and minimized the volatility clustering $\varrho(d, \theta', l)$ as a function of θ' for $n_l = 1, 10, 100$. The delays l were chosen in a way such that $(l_1, .., l_{n_l}) = (1, ..., n_l)$. It was found that the accuracy does not significantly increase as a function of n_l. This is because minimizing $\varrho(d, \theta', l)$ as a function of θ' for small l yields very similar results as for large l. We see an example of this similarity in Fig. 3.21, where the parameters resulting from the methodology for $l = 1$ and $l = 50$ are displayed ($n_l = 1$). We see that the values and structure of the results are very similar. There is only a small difference observed for high h_0 and high φ. Therefore using $n_l > 1$, which means taking the average result of minimizing several $\varrho(d, \theta, l)$ with different l, does not significantly improve the accuracy of the MIC methodology.

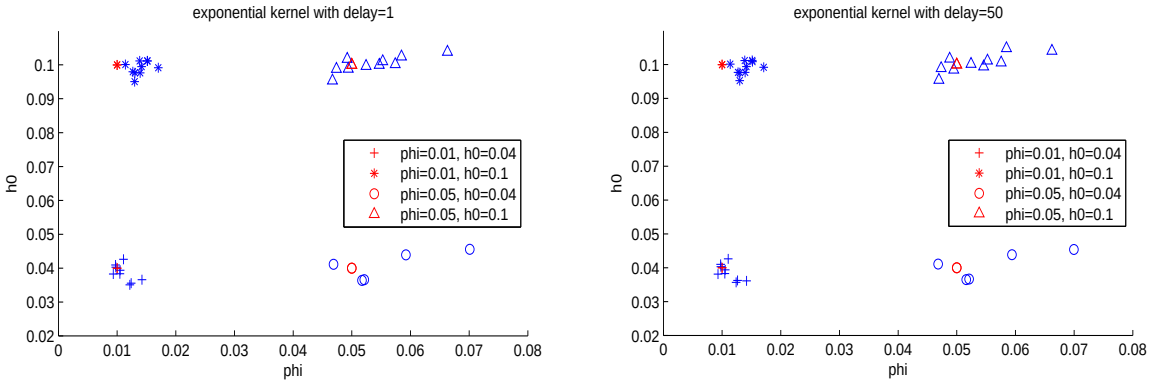


Figure 3.21: The results of minimizing volatility clustering of the innovations $\varrho(d, \theta, l)$ with different delays l. Because the results are very similar, the accuracy of the method is not increased significantly by averaging over multiple delays ($n_l > 1$).

Therefore Fig. 3.20 present the most accurate results that are obtained using this method. The results are not more accurate than those obtained by the CML method, and therefore no parameter quantiles are determined. However, because the method relies on a different principle for estimation, it can provide a second opinion in the case of parameter estimation for real financial time series.

3.3.3 Conclusion parameter estimation

We have seen two methods that can estimate the parameters of a pure SEMF process: the Conditional Maximum Likelihood (CML) method and the Minimum Innovation Volatility Clustering (MIC) method.

It was demonstrated that, using CML, the process parameters and memory kernel can be estimated with sufficient accuracy to obtain volatility estimations with high quality ($R^2 = 0.8$). The parameter estimation is more accurate for time series with an exponential memory kernel than for a power law memory kernel. We have seen that the CML method was less accurate for combinations of low h_0 and high φ and vice versa (high h_0 and low φ).

Furthermore I demonstrated that the parameters of a pure SEMF process can also be estimated by minimizing the clustering of the estimated innovations (MIC). The method proved less accurate than the CML method but is of interest to this research because it uses a different principle for parameter estimation. Therefore it can provide a second in parameter estimation in real financial time series.

Finally, using the CML method, I have also presented scaling relations for the process parameters in the case of an exponential memory kernel. For power law memory kernels, a scaling relation was only found for σ.

3.4 Conclusion

I have proposed a procedure for estimating the volatility and the innovations in SEMF processes. Furthermore the effectiveness of this procedure has been demonstrated in a broad range of process parameters for both kernel types.

If the parameters are known, the volatility and innovations can be estimated by assuming a neutral history at a certain point and computing the volatility using the SEMF process definition at subsequent points. We have seen that this estimation is more accurate for SEMF processes with an exponential memory kernel than for processes with a power law memory kernel. The estimation quality of the volatility (in terms of R^2) decreases with increasing h_0 and decreasing φ. For most parameter ranges a high estimation quality ($R^2 = 0.8$) can be obtained. This method remains successful (high R^2) if there is a limited degree of uncertainty in the parameters used for volatility estimation. Using Conditional Maximum Likelihood, it is shown that it is possible to obtain the parameters of a pure SEMF process with in this limited degree of uncertainty. I also demonstrate that the process parameters at different scales are related. Finally I present the Minimum Innovation Clustering method that can serve as a second opinion for parameter estimation in real financial time series.

Chapter 4

Application of the process to real financial time series

We have seen that in synthetic SEMF processes (within a broad range of parameters) we can estimate the volatility and parameters with a high accuracy. In this chapter, the volatility and parameter estimation procedure from chapter 3 is applied to real financial data. I find parameters that describe unrealistic processes and innovations that are very similar to the returns. Since the estimation procedure is successful for synthetic SEMF processes, it is concluded that the SEMF model in its current form does not realistically describe real financial time series. We find indications that in real financial time series the volatility process (ω) also depends on past absolute returns.

Section 4.1 presents estimates of the process parameters using CML and MIC in real financial time series. The financial time series include daily returns of stocks and indices, and returns on 30 minute intervals from stocks and currency exchanges. Stable, non-negative parameters are obtained for most of the tested financial time series. However all but one of the estimated parameters describe SEMF processes with no physical meaning as they are dominated by volatility bursts. We see a big sloppiness of the cost function used for CML parameter estimation as a function of different parameters. Therefore the parameter estimates are inaccurate.

In section 4.2 the volatility and innovations are estimated using these parameter estimates. We see that for all time series, including the time series with realistic parameters, the dynamics of the innovations is similar to the returns. There is significant clustering in the innovations. The probability distribution function of the innovation has heavy tails. We also see that the structure of the innovations is very similar to the returns.

Section 4.3 explores the possibility of forecasting but conclude that high performance in forecasting of the dynamics is most unlikely.

Section 4.4 more deeply investigates the estimations and find a positive relation between the absolute innovations and the absolute value of the process that determines the volatility (ω). This relation implies that in real financial time series this process (ω) has also a relation with past absolute returns.

4.1 Parameter estimation

We have seen in section 3.3 that two methods, CML and MIC, are effective in estimating the parameters in synthetic SEMF processes. Even though the CML method is more accurate, both methods are useful for parameter estimation in real financial time series because they rely on different properties of the innovations. Therefore I apply both methods to real financial time series in this chapter.

4.1.1 Description of data

The procedure is applied to a representative set of financial data on different time scales: returns on days and on intraday intervals of 30 minutes.

First, the procedure is applied to the daily returns of common stocks and stock indices over intervals of several decades. These large intervals are chosen because the quality of the volatility estimation increases with the number of data points used (see section 3.1.3).

The data is obtained from two different sources (online stock database [24] and Bloomberg [25]) to ensure the quality of the results (errors or anomalies in data from either source can be detected). The original data from [24] incorporates dividends (adjusted close), but I manually compensate for stock dilutions, which have occurred at well recorded dates. It is found that time series of the Dow Jones daily returns from two different sources are identical. Because of the possible change in dynamics over time, I do not use data that covers time spans larger than 60 years. Therefore the Dow Jones index returns is only used from 1950 onwards.

However, time series that cover several decades may include different regimes of the process (e.g., bubbles). The different regimes could disturb the estimation procedure. Furthermore, the dynamics of the process may have changed over time, which would also disturb the estimation. Therefore, I also apply the procedure to returns on intervals of 30 minutes over a period of several months. I also use currency exchange data on the same time scale, as it represents a different form of financial data than stock and index returns.

The full specification of the used data can be found in Table 4.1.

Asset name	Source	Start (dd-mm-yyyy)	Finish (dd-mm-yyyy)	Time Scale
IBM [A]	[24]	02-01-1962	20-05-2011	Daily
IBM [B]	[25]	06-02-2002	13-06-2011	Daily
Shell	[25]	15-07-2005	10-06-2011	Daily
General Electric	[24]	02-01-1962	20-05-2011	Daily
Coca-cola	[24]	02-01-1962	23-05-2011	Daily
Dow Jones [A]	[24]	01-10-1928	13-06-2011	Daily
Dow Jones [B]	[25]	01-10-1928	13-06-2011	Daily
S&P 500	[24]	03-01-1950	23-05-2011	Daily
IBM	[25]	18-01-2011	02-08-2011	30 min
GE	[25]	18-01-2011	02-08-2011	30 min
Dow Jones	[25]	18-01-2011	02-08-2011	30 min
USD/EUR	[25]	18-01-2011	02-08-2011	30 min
Yen/USD	[25]	18-01-2011	02-08-2011	30 min
CHF/EUR	[25]	18-01-2011	02-08-2011	30 min

Table 4.1: Specification of the used data sets that were used for the application of the SEMF process. USD stands for US dollar, and CHF denotes the Swiss franc.

4.1.2 Parameter estimates using CML

Recall that the CML method estimates the parameters by minimizing the cost function $\mathbf{l}_n(\vartheta')$ (i.e. maximizing the likelihood)

$$\theta_{n,est} = \arg\min_{\vartheta'} \mathbf{l}_n(\vartheta') \tag{4.1}$$

where the cost function $\mathbf{l}_n(\vartheta')$ is determined using the Gaussian distribution of the innovations. The accent ' indicates that θ' is the input parameter for the cost function. It is defined as

$$\mathbf{l}_n(\vartheta') = \frac{1}{n} \sum_{i=n_{sp}+n_{excl}}^{n} l_i(\vartheta') \text{ where } l_i(\vartheta') = \frac{d_i^2}{\sigma_{i,est}^2} + log(\sigma_{i,est}^2) \tag{4.2}$$

I used $n_{exl} = 250$ and $n_{sp} = 1$. The volatility is estimated using parameter set ϑ' ignoring the returns before n_{sp}:

$$\sigma_{n,est}(\sigma', h_0', \varphi') = \sigma' \exp\left\{ -\frac{\sum_{i=n_{sp}}^{n-1} d_i h'_{n-i-1})}{\sigma'} \right\} \tag{4.3}$$

where $h'_{n-i-1} = h_0(n-i-1)^{-\varphi^{-1/2}}$ for a power law kernel and $h'_{n-i-1} = h_0 exp(-\varphi n)$ for an exponential kernel.

4.1.2.1 Application to daily returns

The CML method is applied to the time series of daily stock returns. The estimated parameters for both kernel types are presented in Table 4.2.

Asset name	Exp - φ_{est}	Exp - $h_{0,est}$	Exp - σ_{est}	Pow - φ_{est}	Pow - $h_{0,est}$	Pow - σ_{est}	Optimal kernel
IBM [A]	0.031	1.93	0.016	0.29	3.49	0.016	Exponential
IBM [B]	0.008	2.73	0.015	-0.10	5.71	0.015	Exponential
Shell	-0.003	0.17	0.025	-0.15	5.68	0.023	Power law
GE	0.012	1.78	0.016	0.18	4.72	0.017	Exponential
Coca-cola	0.021	1.83	0.016	0.21	4.09	0.016	Exponential
Dow Jones [A]	0.029	4.74	0.009	0.25	10.70	0.010	Exponential
Dow Jones [B]	0.028	4.72	0.009	0.24	10.70	0.010	Exponential
S&P 500	0.023	4.45	0.009	0.23	11.14	0.010	Exponential

Table 4.2: Resulting parameters of the application of the conditional maximum likelihood methodology for both kernel types on daily returns of stocks and stock indices. I present the parameter estimates with an exponential kernel (columns with Exp) and a power-law kernel (columns with Pow). The optimal kernel is the kernel with the highest likelihood.

For three time series (GE, IBM [B] and Shell) the estimates are invalid. For the GE and IBM [B] daily stock returns, the parameter estimation is unstable because starting the procedure with different initial conditions results in a different parameter estimation. The Shell parameter estimate with a power-law kernel for φ_{est} is negative. In a SEMF process with negative φ the memory of a past event increases over time. This is an unrealistic condition for financial time series. Therefore results with negative φ_{est} are not usable.

We see that for all other time series the exponential kernel has the highest likelihood. The range of estimated φ_{est} corresponds to a realistic memory decay. However, for each of these time series the resulting non-stationary parameter h_0 takes on an extremely high value ($h_0 > 1$). A process defined by such a h_0 is completely dominated by volatility bursts. A process with these volatility bursts does not describe real financial time series and has is no physical meaning. The parameter estimations with a power law memory kernel also result in extremely high values for h_0.

We can conclude that the estimated parameter sets do not correspond to a realistic pure SEMF process with either one of the proposed memory kernels. To get more insight into the accuracy of these estimates, the likelihood landscape of the time series is investigated. We will see that the likelihood landscapes are very flat and therefore that the parameter estimates are most likely inaccurate.

Conditional likelihood landscape of daily financial returns

I computed the conditional likelihood landscape for as a function of the parameters φ', σ' and h'_0 for both kernel types. The landscape is expressed in terms of the cost function $\mathbf{l}_n(\vartheta)$. Cross-sections of the landscapes for the IBM [A] daily stock returns set can be seen in Fig. 4.1. These landscapes are representative for the tested data sets (apart from the invalid results for IBM [B], Shell and GE).

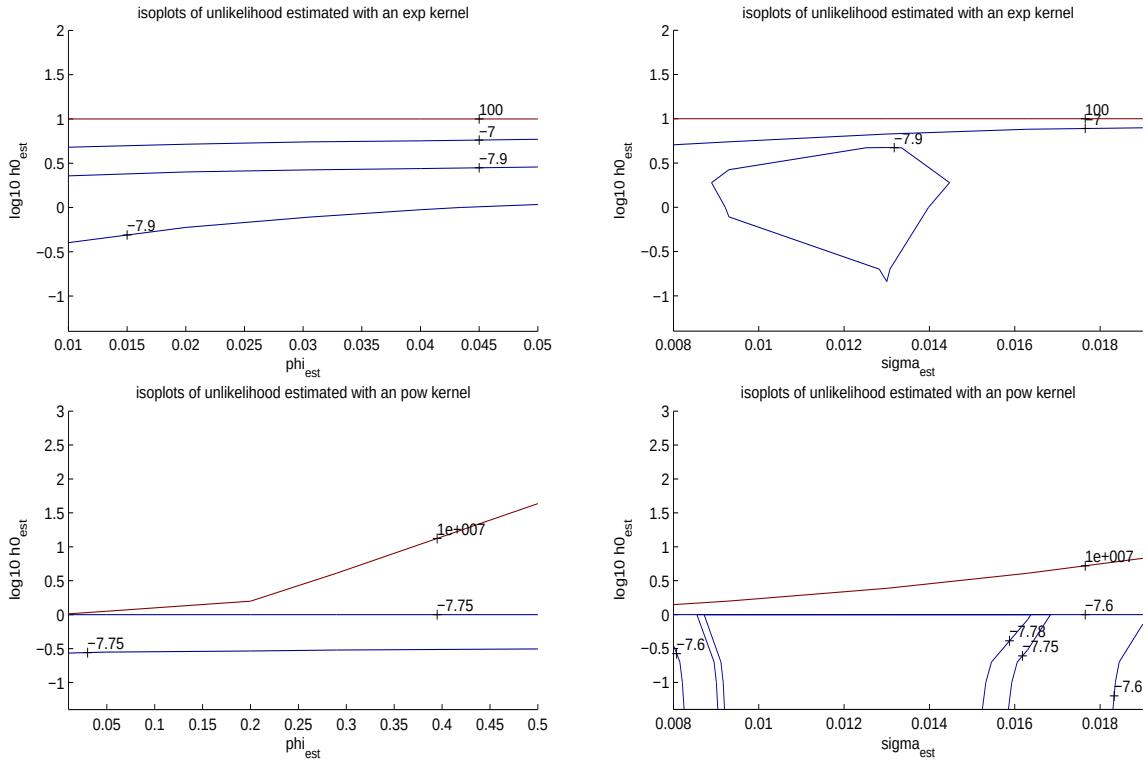


Figure 4.1: Conditional likelihood landscape (expressed by the cost function $\mathbf{l}_n(\vartheta)$) for the IBM [A] sample for both kernel types. **Left:** The cost function for different h'_0 and φ' while σ_{est} takes on the maximum likelihood value from Table 4.2. **Right:** The cost function for different h'_0 and σ' while φ; takes on the maximum likelihood value from Table 4.2.

We see that the cost function as a function of φ' is very flat for both kernel types. The cost function as a function of $log[h'_0]$ has a sharp decay above a certain threshold for both kernel types ($h'_0 \approx 10$ for an exponential kernel and $h'_0 \approx 1-10$ for a power law kernel). However for decreasing $log[h'_0]$ the cost function is also very flat. Since a flat landscape corresponds to an inaccurate estimate of the parameter, we can conclude that the estimates of these two parameters are not accurate. Since the landscape is flat as a function of h_0 on a log scale, we see that also the order of size of the h_0 can be inaccurate. The landscape as a function of σ has a strong minimum around a certain value in each of the plots. Therefore the estimate of σ_{est} is most likely accurate.

The exact uncertainty in the parameter estimation is unknown. However most likely a joint uncertainty in the parameters of this magnitude does not result in accurate estimations of the volatility (see section 3.2.3).

In the plots we see that the likelihood landscape is very different from the likelihood landscapes of synthetic time series (see section 3.3.1.1). Therefore we can conclude that the financial time series behaves differently than the pure SEMF process with either one of these two kernels (exponential and power law).

4.1.2.2 Application to returns of 30 minute intervals with CML

I also estimate the parameters in time series of returns on 30 minute intervals. The results of the parameter estimations can be seen in Table 4.3.

Asset name	Exp - φ_{est}	Exp - $h_{0,est}$	Exp - σ_{est}	Pow - φ_{est}	Pow - $h_{0,est}$	Pow - σ_{est}	Optimal kernel
IBM	0.0073	8.11	0.0030	-0.22	11.20	0.0034	Exponential
GE	0.014	6.41	0.0041	-0.068	9.31	0.0042	Exponential
Dow Jones	0.012	16.72	0.0021	0.10	0.17	0.0021	Exponential
USD/EUR	0.011	0.084	0.0010	0.11	0.21	0.0010	Exponential
Yen/USD	0.014	0.068	0.0010	-0.21	32.43	0.0009	Power Law
CHF/EUR	0.013	0.071	0.0011	0.14	0.17	0.0011	None

Table 4.3: Resulting parameters of the application of the conditional maximum likelihood methodology for both kernel types for time series of returns of 30 minute intervals. The preferred kernel is the kernel with the highest likelihood.

We see one result that is clearly invalid. The CML estimation in the Yen/USD time series results in a power law kernel with a negative φ_{est}. Results with a negative φ_{est} do not describe realistic processes. In the other time series the exponential kernel type is estimated, except for the CHF/EUR which does not estimate any kernel type.

For the stock returns, again extremely high values for $h_{0,est}$ are estimated. I was not able to simulate a process with such parameters without the dominating presence of volatility bursts. For the currencies, lower values of h_0 are estimated that describe SEMF processes with a physical meaning. For all results, φ_{est} and σ_{est} are around 0.01 and 0.001 respectively. These values are close to the initial parameters used in the optimization algorithm.

The parameter estimates of the stock returns on the 30 minute scale (IBM and Dow Jones) are compared with the estimates on the daily scale (IBM [A] and Dow Jones [A]). We have seen scaling relations using CML parameter estimation in section 3.3.1.3. These scaling relations describe how the parameters of a SEMF process with an exponential kernel are related on different time scales g. Here I investigate whether these scaling relations are valid for these parameter estimates.

Because a trading day consists of 8 hours, the scale difference g (see equation 3.18) in this case is equal to 16. We see that σ_{est} is about a factor 4-5 higher in the daily returns than in the returns in 30 minute intervals. Since $\sigma_{g,exp} = \sigma_{1,exp}\sqrt{g}$, this factor corresponds to a scale g around 16. The accuracy of this scaling relation supports the hypothesis that the estimation of σ is accurate. However for φ_{est} the scaling relation does not hold. We see that the values for φ_{est} differ about a factor 2-4, which according to $\varphi_{g,exp} = g\varphi_{1,exp}$ corresponds to a g=2-4. This is incorrect, because we known that g should be 16. This supports the notion that, based on the likelihood landscapes, the estimation of φ_{est} is inaccurate (at least in one of the two time scales).

For the power law kernel estimates (excluding the results with negative φ), we see comparable values for φ_{est} and σ_{est}. The estimated value of h_0 is within the multifractality range (see equation 2.30) where processes without volatility bursts can be obtained.

We have seen that, in contrast to the parameter estimates of daily returns, some of the estimated values of h_0 are not extremely high. Therefore the likelihood landscapes of these time series are investigated. We see that also these estimates have a high uncertainty.

Conditional likelihood landscape of financial returns on 30 minutes

Figure shows likelihood landscapes for the Yen/USD returns on intervals of 30 minutes. These landscapes are representative for the landscapes of other time series for which the estimated value of $h_{0,est}$ is not extremely high (also for power law memory kernels).

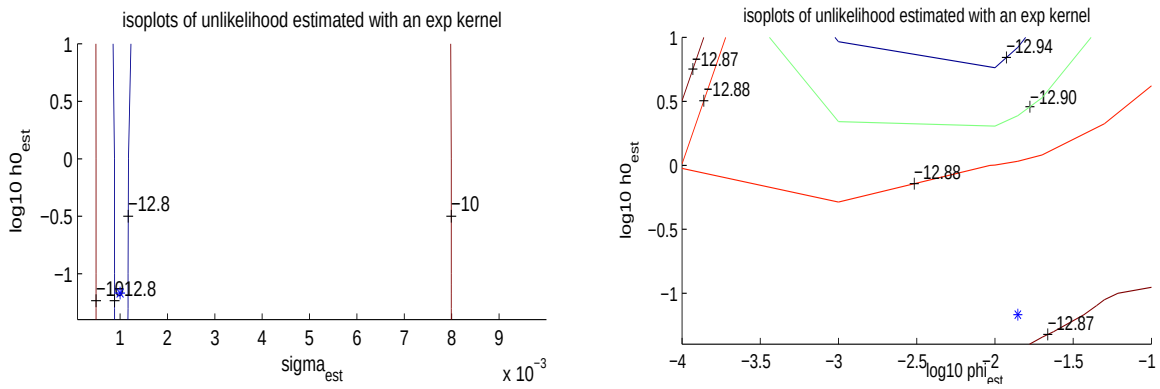


Figure 4.2: likelihood landscape USD/EUR 30 min scale by CML meaning of star

We see that the landscape as a function of log h_0' and log φ' again is very flat. Since the isoplots do not circle with decreasing values around the parameter found with CML (blue star), it is not sure whether the estimated parameters correspond to a local or global maximum likelihood. However the landscape as a function of σ' has a strong minimum cost function around the estimated value. Therefore σ_{est} is most likely relatively accurate.

There is an absence of a broad likelihood maximum around the estimated parameters and also the likelihood landscape as a function of φ and h_0 is very flat. We can conclude that when the estimated $h_{0,est}$ was not extremely high the estimation was also inaccurate.

The likelihood landscapes of the stock returns on 30 minute intervals (with high $h_{0,est}$) were similarly flat to the landscapes of daily returns (see Fig. 4.1).

4.1.2.3 Conclusion

We have seen that, using the CML method, we can estimate stable, non-negative parameters in daily stock and index returns (IBM [A], Coca Cola, Dow Jones [A,B] and S&P500) and for returns on 30 minute intervals (IBM, GE and the Dow Jones). The exponential memory kernel has the highest likelihood in these time series. However, the method estimates an extremely high h_0 for which a pure SEMF process (where the innovations are uncorrelated) is not realistic (because it is dominated by volatility bursts).

Likelihood landscapes as a function of h_0' and φ' are very flat. This flatness indicates that the parameter estimations are not very accurate. The likelihood landscape as a function of σ' has a deep minimum around the estimated value. Furthermore the estimated values σ_{est} on different time scales obey scaling relations (see section 3.3.1.3). Therefore the estimation of σ_{est} is most likely accurate. The method does not estimate a power law kernel with usable parameters for any of the financial time series.

The CML method estimates parameters are in the multifractality range (low $h_{0,est}$) for the USD/EUR currency exchange. However, the flat likelihood landscape as a function of the parameters h_0' and φ' indicates that these estimates are also inaccurate. Furthermore, the likelihood landscape shows that the estimated parameter (with low $h_{0,est}$) set is only a local (due to the noise in the cost function) and not a global minimum (which would correspond to the conditional maximum likelihood estimate of all possible parameters).

We have seen that the likelihood landscape is significantly different for the financial time series than for a pure SEMF processes. I conclude that financial time series behave significantly different from pure SEMF processes. The results show that the CML method is ineffective, because of the large uncertainty in the parameters. As a next step I also estimate the parameters using MIC and investigate whether these parameter estimates are more realistic and accurate.

4.1.3 MIC parameter estimation

For comparison the parameters in the financial time series are also estimated using the MIC method. The method is fully described in section 3.3.2, but the algorithm is repeated briefly below.

The Minimum Innovation Clustering method minimizes the correlation between subsequent squares of innovations:

$$\varrho(d, \theta', 1) = \sum_{n=1}^{N} corr(\xi^2_{n+1,est}, \xi^2_{n,est}) = \sum_{n=1}^{N} \frac{E\left[(\xi^2_{n+1,est} - \mu_{\xi^2})(\xi^2_{n,est} - \mu_{\xi^2})\right]}{\sigma^2_{\xi^2}} \tag{4.4}$$

Here d is a times series of log returns, θ' the proposed parameter set and l the delay. μ_{ξ^2} is the average value of the process $\xi^2_{n,est}$ and σ_{ξ^2} is the standard deviation of this process. The $\xi^2_{n,est}$ depend on the used estimation parameters θ' and the input time series d_n (see equations 3.4 and 3.12).

The the optimal parameter set is defined as

$$\hat{\vartheta}_n = \arg\min_{\vartheta} \varrho(d, \theta, l) \tag{4.5}$$

Effectively this method optimizes the independence of the estimated innovations.

We can only estimate kernel type, φ_{est} and $h_{0,est}$ using this method, and not σ_{est}. I take $\sigma_{est} = \sigma'$ where σ_{est} is the value estimated with CML, for all MIC estimations. However this is should not affect the estimations of φ_{est} and $h_{0,est}$ because the function for the volatility clustering 4.4 is normalized for the standard deviation.

4.1.3.1 Daily stock returns

The MIC method is applied to the real financial time series. Figure 4.4 presents the results of the estimation.

Last experiment-Asset name	Exp - φ_{est}	Exp - $h_{0,est}$	Pow - φ_{est}	Pow - $h_{0,est}$	Optimal kernel
IBM [A]	0.0152	1.74	0.428	5.09	Exponential
IBM [B]	-0.001	3.20	-0.117	5.35	Exponential
Shell	0.396	3.63	0.086	4.15	Power law
GE	-0.004	8.00	0.201	4.97	Exponential
Coca-cola	0.086	3.59	0.357	5.57	Exponential
Dow Jones [A]	-0.005	7.53	0.458	3.54	Exponential
Dow Jones [B]	-0.046	2900	0.367	12.31	Exponential
S&P 500	-0.012	14.61	0.310	11.01	Exponential

Table 4.4: Resulting parameters of the application of the MIC methodology for both kernel types. The preferred kernel is the kernel with the lowest value for $\varrho(d, \vartheta, l)$. daily returns

We see that in seven out of eight time series the exponential kernel is estimated. However, five out of seven of these results estimate negative values for φ_{est}, which does not correspond to a process with multifractal properties or even a realistic process at all (e.g., $h_0 = 2900$) Therefore those results are invalid and not discussed further.

For the other two time series (IBM [A] and Coca Cola), the estimated values of $h_{0,est}$ are also extremely high. The estimated values of $h_{0,est}$ are even higher than the values estimated using CML. The estimated values of φ_{est} indicates a relatively strong degree of memory decay, and these estimates are lower for the CML method. The CML method also estimated the exponential memory kernel in these time series.

The MIC method estimates the power-law kernel in the Shell daily returns. This was also the case for CML. However, in contrast to the CML estimate, here φ_{est} has a non-negative value. Again, $h_{0,est}$ has an extremely high value.

4.1.3.2 Returns on intervals of 30 minutes

The MIC method is also applied to the returns on intervals of 30 minutes. The estimated parameters are presented in Table 4.5.

Asset name	Exp - φ_{est}	Exp - $h_{0,est}$	Pow - φ_{est}	Pow - $h_{0,est}$	Optimal kernel
IBM	0.0213	8.71	45.37	-16.36	Power Law
GE	0.0085	10.85	-0.0080	29.09	Exponential
Dow Jones	0.0098	17.45	0.74	-27.65	Exponential
USD/EUR	30.09	60.5774	122.31	60.58	None
Yen/USD	-0.0063	0.0014	-2.90	553990	Power Law
CHF/EUR	0.0001	7.73	-0.44	10.37	Exponential

Table 4.5: Resulting parameters of the application of the MIC methodology on time series with a 30 minute time scale for both kernel types. The preferred kernel is the kernel with the lowest value for $\varrho(d, \vartheta, 1)$ (see equation 4.4)

We see that none of the estimations for power law kernels have realistic values. Therefore the power law results are ignored.

For the estimations with an exponential memory kernel, we see that the Yen/USD time series has a negative estimate for φ_{est}. Furthermore the USD/EUR time series has $\varphi_{est} = 30.1$, which corresponds to process without a memory. Therefore these results are invalid.

For the remaining four parameter estimations with an exponential kernel, the φ_{est} is realistic but the estimated $h_{0,est}$ is again extremely high. The estimated $h_{0,est}$ are even higher than the CML estimates.

Because of the overall unrealistic results, I conclude that the MIC estimation in these time series failed and that these results are not reliable.

Summary

For IBM [A] and Coca Cola daily returns we have seen stable and non-negative parameter estimates using the MIC method. Like the CML method, the MIC estimates and exponential kernel with an extremely high $h_{0,est}$. However there is a large difference between the estimated parameters. For the Shell daily stock returns, the method estimates a power law kernel with stable and non-negative parameters. The parameter estimations on shorter time scales (returns on 30 minute intervals) are inaccurate.

4.1.4 Summary

Table 4.6 provides the selection of parameter estimates that are the most relevant. All parameter estimates with negative parameters and extremely fast memory decay ($\varphi > 1$) have been disregarded. I also disregard results where no kernel type is estimated.

Based on estimation in synthetic time series, the CML method has a higher accuracy in parameter estimation. The CML estimations are also more frequently positive than the MIC, which supports the notion of superior accuracy using the CML method. Therefore I choose to ignore the MIC parameter estimates if the estimated parameters were similar. (e.g., Dow Jones). However the estimates by MIC also indicated an extremely high $h_{0,est}$.

Name	Time scale	Method	Estimated kernel type	φ_{est}	$h_{0,est}$	σ_{est}
IBM [A]	Daily	CML	Exponential	0.031	1.93	0.016
IBM [A]	Daily	MIC	Exponential	0.049	3.26	0.016*
Shell	Daily	MIC	Power law	0.086	4.15	0.023*
Coca Cola	Daily	CML	Exponential	0.021	1.83	0.016
DJ	Daily	CML	Exponential	0.029	4.74	0.009
S&P500	Daily	CML	Exponential	0.023	4.45	0.009
IBM	30 minutes	CML	Exponential	0.0073	8.11	0.0030
GE	30 minutes	CML	Exponential	0.014	6.41	0.0041
Dow Jones	30 minutes	CML	Exponential	0.012	16.72	0.0021
US/EUR	30 minutes	CML	Exponential	0.011	0.084	0.0010

Table 4.6: Selection of the best parameter estimates using both methods. * The MIC method can not be used to estimated σ, therefore I take σ from the CML method (which has high accuracy)

Both methods indicate a preference for the exponential kernel. Only the stock return time series of Shell is optimally fitted by a power law kernel for both methods. However, the length of this time series was relatively short (2500 data points) and in the case of CML the resulting φ_{est} was negative. Therefore the accuracy of this estimation is controversial.

Furthermore both methods indicate that an optimal fit is made with an extreme value of h_0. Since this parameter does not describe realistic SEMF processes, we can conclude that the SEMF model does not accurately describe the data. However, the estimated parameters do most likely correspond to the most accurate volatility estimations that can be made with this model. Therefore I take deeper look at the resulting innovations and volatility estimations using the estimated parameter.

4.2 Estimated volatility and innovations

The feasibility of the SEMF model is investigated. The stylized facts of the returns should be accounted for by the volatility, since in the model the innovations are a Gaussian white noise. Therefore one can test this feasibility by determining the presence of stylized facts of the returns in the estimated innovations.

4.2.1 Clustering in the innovations

I investigate the degree of clustering in the innovations. The degree of clustering in the innovations can be computed using

$$\varrho(x,1) = \sum_{n=1}^{N} corr(x^2_{n+1,est}, x^2_{n,est}) \tag{4.6}$$

where the $x = \xi$ in this case. I computed this value for all of the selected results and present the results in Table 4.6. For comparison, the clustering of the returns ($x = d$ in equation 4.6) was also calculated.

In the SEMF process, a white noise process following a normal N(0,1) distribution is used to model the innovations. This random process can also display a degree of clustering. I simulated 1000 time series d_n and estimated the innovations using equations 3.4 and 3.5. Then the clustering was determined. It was found that 95% of the time series have a clustering of less than $\varrho = 0.0167$.

Name	Time scale	Method	Estimated kernel type	ϱ returns	ϱ innovations
IBM [A]	Daily	CML	Exponential	0.1509	0.1110
IBM [A]	Daily	MIC	Exponential	0.1509	0.0985
Shell	Daily	MIC	Power law	0.2995	0.2680
Coca Cola	Daily	CML	Exponential	0.1509	0.1100
Dow Jones	Daily	CML	Exponential	0.2044	0.1084
S&P500	Daily	CML	Exponential	0.1418	0.0985
IBM	30 minutes	CML	Exponential	0.0428	0.0421
GE	30 minutes	CML	Exponential	0.0200	0.0073
Dow Jones	30 minutes	CML	Exponential	0.0113	0.0003
US/EUR	30 minutes	CML	Exponential	0.0963	0.0963

Table 4.7: (Volatility) clustering in the returns and the estimated innovations for a selection of time series.

We see that all but one (the returns on 30 minute intervals for the Dow Jones index) of the time series have estimated innovations with a significant clustering (more than the 'random' clustering $\varrho = 0.0167$).

However, in each of the time series there is a significantly lower ϱ in the innovations than in the returns. There is a particularly strong reduction of clustering in the Dow Jones index, IBM (both MIC and CML) and S&P500 daily returns.

We can conclude that there is a strong degree of clustering of the estimated innovations. The clustering was significantly lower than the volatility clustering in the returns.

4.2.2 Examples of estimations in financial time series

The returns, estimated volatility and innovations of a representative set of time series is investigated. I present the estimation using CML in the daily returns of the Dow Jones (Fig. 4.3), because of its representative parameter set and strongly reduced clustering. Furthermore, I present the estimation for IBM [A] (Fig. 4.4) using both parameter estimation methods (CML and MIC) for comparison. Since the MIC parameter estimation of the Shell daily returns is the only estimation of a power law memory kernel with non-negative parameters, the estimation in this time series is also presented (Fig. 4.5).

Furthermore I present the returns on 30 minute intervals of the Dow Jones index using CML (Fig. 4.6) because it is representative for stock returns on this time scale. The CML estimation in US/EUR currency returns (Fig. 4.7) is also presented since it has an estimated $h_{0,est}$ that is not extremely high.

The volatility is normalized by σ_{est} to allow for comparison of $\exp\left\{-\frac{\omega(n)}{\sigma_{est}}\right\}$ between data sets.

Daily returns - CML estimation in the Dow Jones index

Figure 4.3 presents the results for the Dow Jones [A] time series. We see that the structure of the innovations is very similar to the structure of the returns. We also see that the strongly negative return around the data point $1.5 * 10^4$ (which represents the crash of 1987) corresponds to a strongly negative innovation. Because positive returns lead to a reduced volatility in the SEMF model, it does not anticipate crashes related to bubbles. Therefore the model does not anticipate such events (possibly they are associated to a different regime).

We can see a degree of reduction of clustering and the leverage effect in the innovations, particularly in the beginning and end of the sample. This corresponds to the previous calculations of the clustering (see Table 4.7).

However, there is still a strong structure observed in the innovations. This structure is also observed in other time series of daily returns with similarly high $h_{0,est}$ (Coca Cola, S&P 500 and IBM [A]). Furthermore we see a strong presence of outliers in the innovations, indicating persistence of the heavy tails.

With respect to the other estimations, the peaks of the normalized volatility of this data set are quite high.

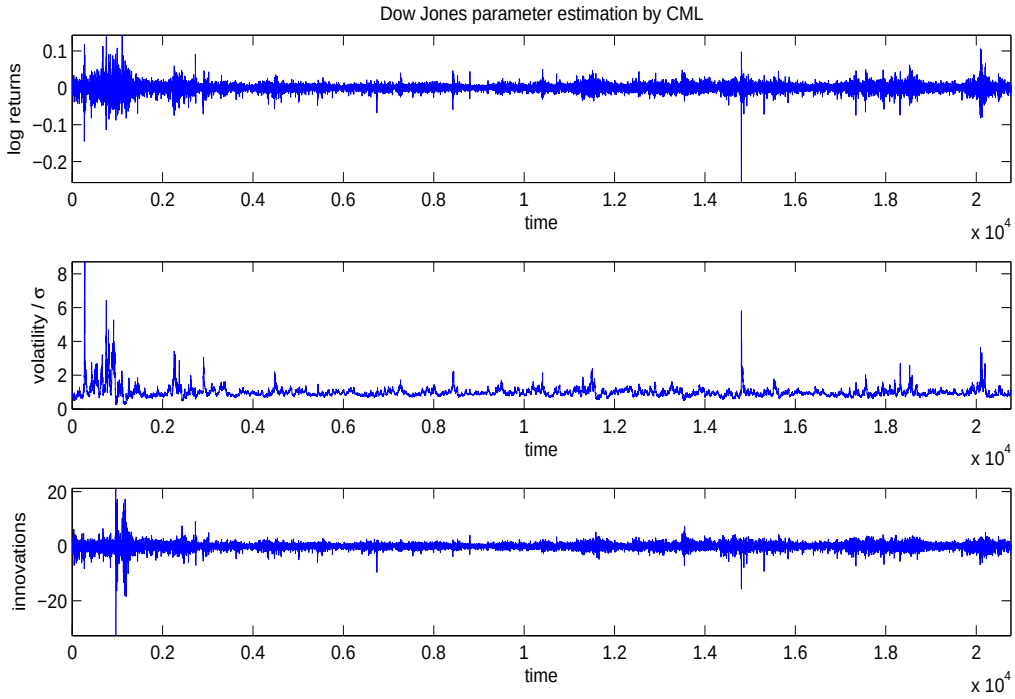


Figure 4.3: Estimation of the innovations and volatility for the Dow Jones [A] index price using the parameters estimated by CML: Exponential kernel with $\varphi_{est} = 0.029$, $\sigma_{est} = 0.009$ and $h_{0,est} = 4.74$

Daily returns - CML and MIC estimation for IBM

Figure 4.4 shows the results for parameter estimation by both CML and MIC. For both estimations the innovations have a strong structure that is very similar to the structure of the returns. The innovations also have frequent outliers. The biggest difference between the time series of the returns and the innovations is the reduction in amplitude of the extreme movements.

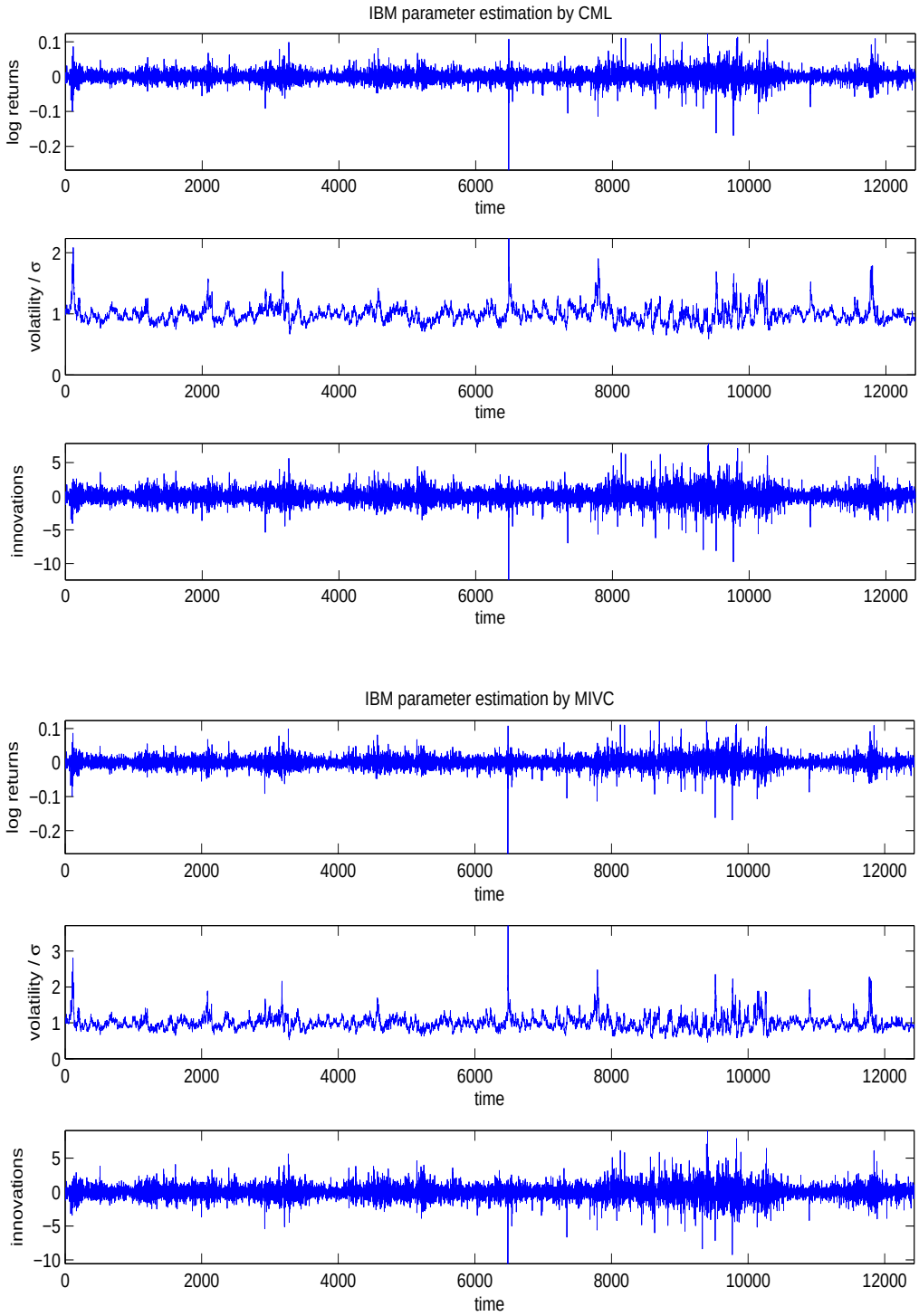


Figure 4.4: Estimation of the innovations and volatility for IBM stock returns using the parameters estimated by CML (Above) and MIC (Below). CML: Exponential kernel with $\varphi_{est} = 0.031$, $\sigma_{est} = 0.016$ and $h_{0,est} = 1.93$. MIC: Exponential kernel with $\varphi_{est} = 0.049$, $\sigma_{est} = 0.016$ and $h_{0,est} = 3.26$

Daily returns - MIC estimation for Shell (Power law)

The results for Shell can be seen in Fig. 4.5. Again we see that the time series of the returns is very similar to that of the innovations; a strong structure in the innovations, similar to the structure of the returns, is again observed.

The similarity between the time series of the returns and innovations is stronger for this time series than for the other, longer, time series. This could be explained by the much shorter length of the time series, which can make the volatility and parameter estimation more inaccurate. However a smaller time series length could also provide more freedom for the fit to minimize clustering.

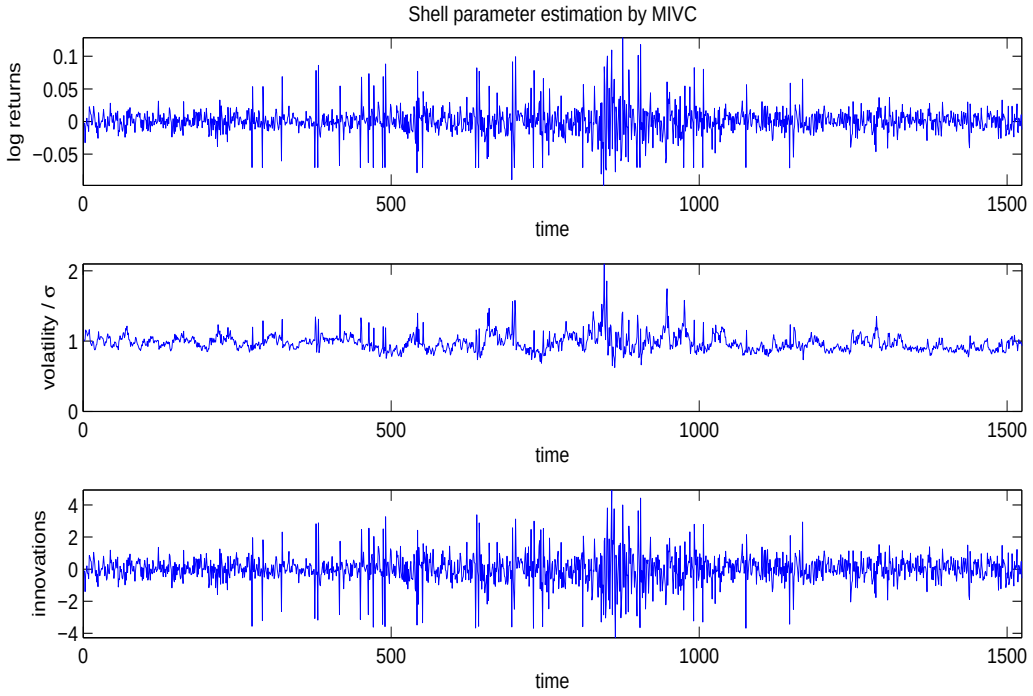


Figure 4.5: Estimation of the innovations and volatility for the Dow Jones index price using the parameters estimated by MIC: Power law kernel with $\varphi_{est} = 0.086$, $\sigma_{est} = 0.023$ and $h_{0,est} = 4.15$

Returns on 30 minutes - CML estimation in the Dow Jones index

In Fig. 4.6 we see the results for the Dow Jones index returns on a time scale of 30 minutes. We see a strong resemblance between the structure of the two time series and on first glance not much has changed. The innovations also have frequent outliers, indicating heavy tails.

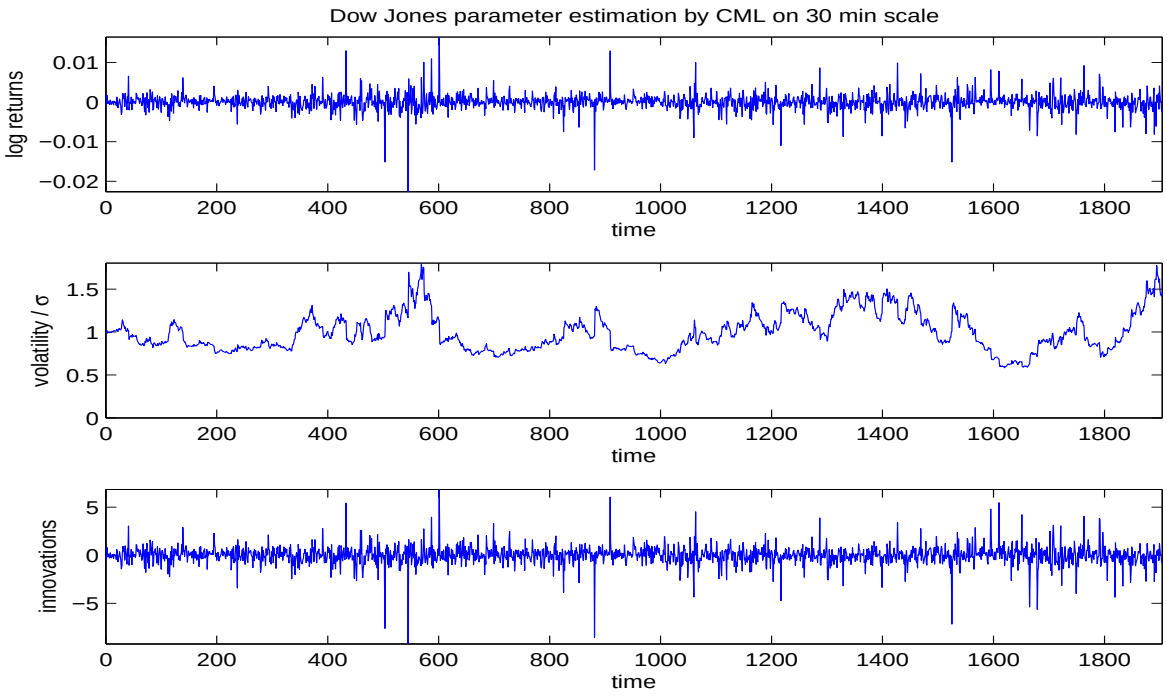


Figure 4.6: CML fit of 30 min scale DJ

Returns on 30 minutes - CML estimation in the USD/EUR currency exchange

The parameter estimates in the USD/EUR currency exchange are the only estimates within the multifractal range (see equations 2.29 and 2.30). The innovations estimated with these parameters are presented in Fig. 4.7.

We see that the estimated volatility is approximately one for all data points and that the innovations only differ from the returns by a factor of scale. We can therefore conclude that the SEMF model with these parameter estimates does not realistically describe this financial time series.

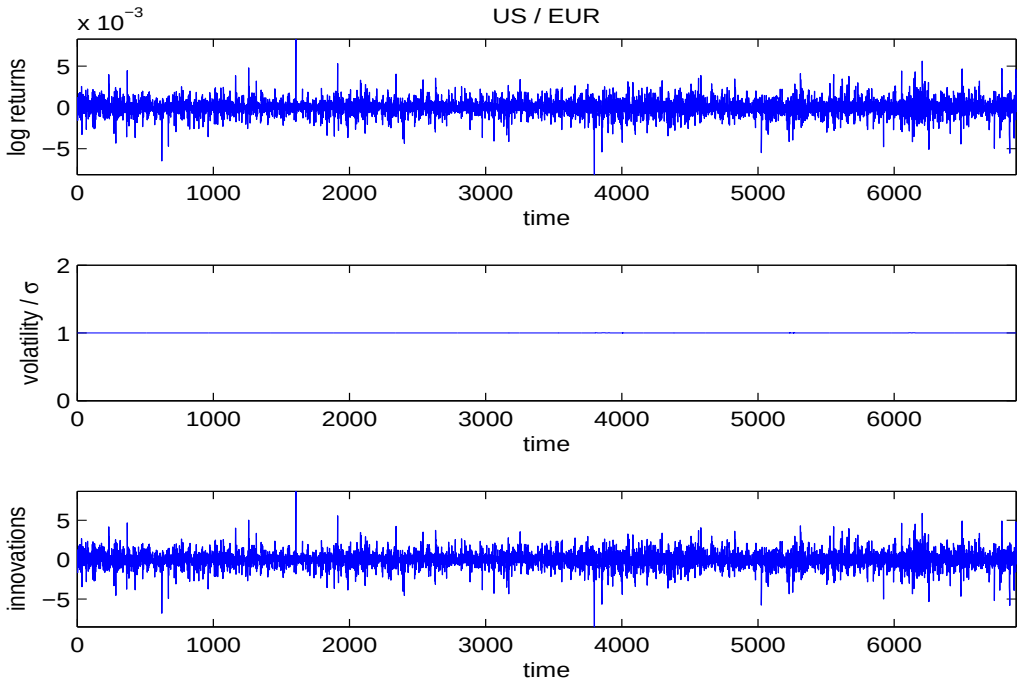


Figure 4.7: CML fit of 30 min scale USD/EUR

4.2.3 Distribution of the innovations

In the SEMF process the innovations are normally (Gaussian) distributed. However we have seen frequent outliers in the estimated innovations of the time series. This indicates heavy tails in the probability distribution function of the innovations.

I compute the histograms of the returns and the innovations, and compare them with a Gaussian distribution. The Gaussian is created using the mean and standard deviations from the returns and the estimated innovations. The results for the Dow Jones index daily returns can be seen in Fig. 4.8. This plot is representative for all other observations.

We see heavy tails for both the returns and the innovations. The weight of the tails of the innovations is not significantly smaller this weight of the case of returns. It is difficult to accurately quantify these tails, but a comparison of kurtosis supports these observations.

This indicates that the application of the SEMF process to these financial time series is not successful in reducing the uncertainty of the dynamics of the stock returns.

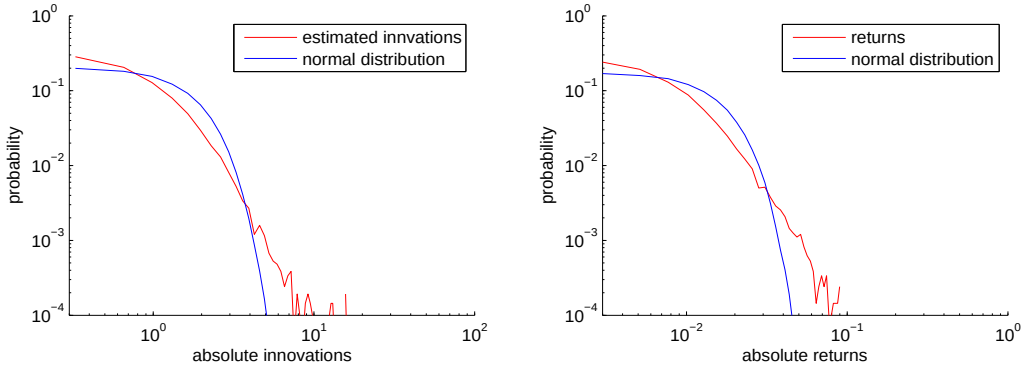


Figure 4.8: The probability distribution function of the innovations and the returns for the Dow Jones. The estimations are fitted for the parameters obtained with CML: Exponential kernel with $\varphi_{est} = 0.029$, $\sigma_{est} = 0.009$ and $h_{0,est} = 4.74$

4.2.4 Conclusion

We have seen a strong similarity of the structure between the innovations and the returns (persistence of clustering) and the presence of frequent outliers in the innovations (heavy tails). These observations demonstrate that either the estimation procedure is not effective or that the current structure of the process is insufficiently accurate. Because both (CML and MIC) estimation procedures are effective in synthetic SEMF time series, we conclude that in its current form the SEMF model can not be applied to these data sets.

4.3 Forecasting

I explore the possibility of forecasting the dynamics of the process and find that when using the estimates in these time series high performance in forecasting is unlikely.

Forecasting future dynamics of the process can consist of giving expectation values of future absolute or squared returns. It can also consist of computing the probability that the sum of the returns, positive or negative, exceeds a certain value in a certain number of time steps. This can be done using Monte Carlo simulation.

For these estimates in the SEMF process, Monte Carlo simulations are not possible. SEMF processes with these high values for h_0 are dominated by volatility bursts and have no physical meaning.

Therefore, forecasting with the current estimates is limited to determining probabilities or expectation values in only one time step. We can calculate these expressions using the estimated volatility and the empirical probability distribution function found for the innovations.

The volatility estimated in these time series are most likely inaccurate. We have seen in section 3.1.3 that increasing h_0 has a decreasing effect on the quality (R^2) of the estimated volatility. Because Monte Carlo simulations are not possible for this parameter range, it is not possible to compute this quality exactly. However I expect that, given the dominating presence of volatility bursts, the quality (R^2) will be very low and the volatility estimation can be off by large factors.

Because the estimated innovations have a heavy tailed distribution, I assume that out of sample innovations follow the same distribution. However this distribution is similar to the distribution of the returns. Therefore the uncertainty of the process is not significantly reduced.

It is therefore unlikely that the current estimations will have a high performance in forecasting and the possibility of forecasting is not explored further.

4.4 Empirical analysis of the model structure

In this section I try to understand why the current form of the SEMF model is not successful in its application to financial time series. A relation between $|\omega_{n,est}|$ and the absolute estimated innovations $|\xi_{n,est}|$ in real financial time series is found. This relation implies that in the SEMF model ω should also depend on past absolute returns.

4.4.1 The relation between the volatility and the innovations in empirical estimates

We recall that in the SEMF model $\omega_{est}(n) = \sum_{i=0}^{n-1} d_i h_{n-i-1}$ is related to the estimated volatility by $\sigma_{est} \exp\left\{-\frac{\omega_{est}(n)}{\sigma_{est}}\right\}$. In the proposed structure of the SEMF process, the innovations ξ_i are independent on the process ω_i that determines the volatility. It is of interest whether this property holds in the estimations in financial time series. I therefore compute the average absolute estimated innovations $|\xi_{i,est}|$ conditional on $\omega_{est}(i)$

$$\left\langle |\xi_{i,est}| \middle| \omega_{est}(i) \right\rangle \tag{4.7}$$

Effectively, these averages are obtained for short intervals of ω_{est} by averaging the corresponding $|\xi_{i,est}|$. This expression is plotted as a function of $\omega_{est}(i)$.

In the SEMF process there is no relation between these two expressions. To demonstrate this independence, a synthetic time series is created and the parameters and volatility are estimated. Figure 4.9 shows the plot for this synthetic time series.

We see a straight line in this plot, which also holds in the magnification around $|\omega_{est}| = 0$. This straight line implies that there is no relation between the two quantities, as the expectation value of $|\xi_{i,est}|$ is not a function of $\omega_{est}(i)$. There is a high degree of uncertainty in the plot for larger values of $|\omega_{est}|$. Because the large deviations from $\omega_{est} = 0$ are not common, there are less data points for averaging the $|\xi_{i,est}|$ for larger values of $|\omega_{est}|$; this introduces uncertainty in the average.

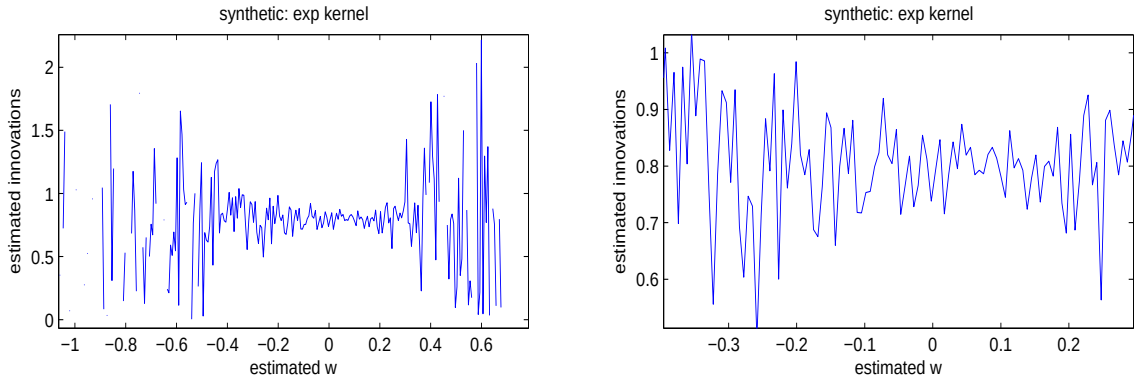


Figure 4.9: The average $|\xi_{i,est}|$ as a function of $\omega_{est}(i)$ for a synthetic time series with an exponential memory kernel with $h_0 = 0.06$, $\varphi = 0.01$, and $\sigma = 1$. The right plot shows a magnification of the left plot in the area around $\omega_{est}(n) = 0$. The volatility is estimated using CML estimates for the process parameters.

I also computed these plots for the financial time series and find that in these estimates there is a relation between the absolute innovations $|\xi_{i,est}|$ and $\omega_{est}(i)$. Figure 4.10 shows the plots for the Dow Jones [A] and

the S&P500 sample using the parameter estimates from table 4.6 (CML, exponential kernel). These plots are representative for all the selected estimates.

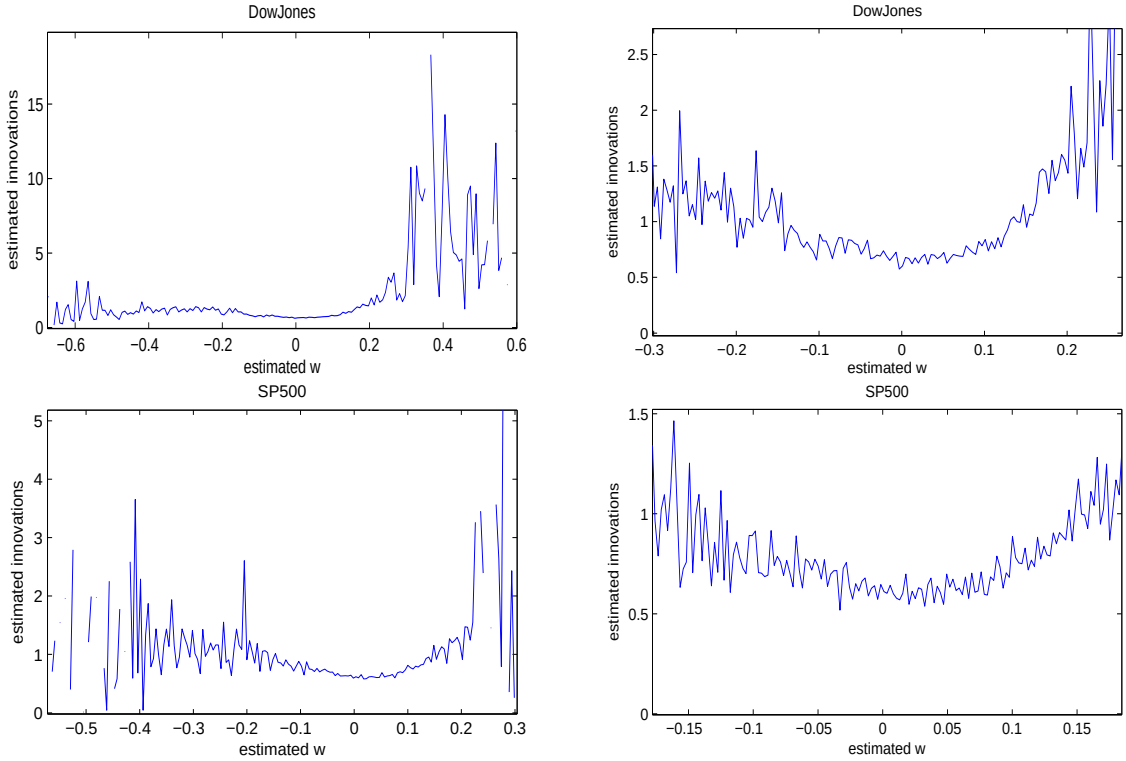


Figure 4.10: The average $|\xi_{i,est}|$ as a function of $\omega_{est}(i)$ for the Dow Jones [A] (above) and the S&P500 (below). The right plot shows a magnification of the left plot in the area around $\omega_{est}(n) = 0$. The estimations are fitted for the parameters obtained with CML: Exponential kernels with $\varphi_{est} = 0.029$, $\sigma_{est} = 0.009$ and $h_{0,est} = 4.74$ for the Dow Jones sample and $\varphi_{est} = 0.023$, $\sigma_{est} = 0.009$ and $h_{0,est} = 4.45$ for the S&P500.

In both plots we see that $|\xi_{i,est}|$ increases for higher $\omega_{est}(i)$. This is particularly clear in the magnifications. The increased average $|\xi_{i,est}|$ for higher $\omega_{est}(i)$ implies that the volatility is estimated too small for higher $\omega_{est}(i)$. A series of positive returns results in a higher $\omega_{est}(i)$. The decreasing effect of positive returns on volatility is therefore stronger in the SEMF process than in real financial time series.

We see that $|\xi_{i,est}|$ also increases for lower $\omega_{est}(i)$, which corresponds to a series of negative returns. In the SEMF process the positive effect of negative returns on volatility is therefore also not strong enough.

Because of the clear relation between $|\xi_{i,est}|$ and $\omega_{est}(i)$, we can conclude that the structure of ω in the SEMF process is not correct. Specifically, the volatility estimate should be higher for larger $|\omega_{est}(i)|$. Because $|\omega_{est}(n)| = \left|\sum_{i=0}^{n-1} d_i h_{n-i-1}\right|$, we see that $|\omega_{est}(n)|$ is large if

$$\left| \sum_{i=0, d_n>0}^{n-1} d_i h_{n-i-1} - \sum_{i=0, d_n<0}^{n-1} d_i h_{n-i-1} \right| \tag{4.8}$$

is large. Because the returns d_n are not auto-correlated, it is likely that in that case the amplitudes of the returns $|d_n|$ are also large. It is therefore likely that in real time series, ω has a dependency on absolute returns $|d_n|$.

Conclusion

We have seen that the absolute innovations $|\xi_{i,est}|$ have a U-shaped relation with $\omega_{est}(i)$. Therefore the estimated volatility in these time series is too low for higher $|\omega_{est}(i)|$. We can possibly estimate the volatility higher for higher $|\omega_{est}(i)|$ if ω has a dependence on absolute returns $|d_n|$.

4.5 Conclusion

I have applied the volatility and parameter estimation procedure of chapter 3 to a representative set of real financial data. We have seen that the estimated parameters correspond to unrealistic SEMF processes and that the estimated innovations are very similar to the returns. We can therefore conclude that the SEMF model does not realistically describe financial data.

The CML parameter estimates of daily returns of stocks and indices are either negative or correspond to unrealistically high non-stationary (extremely high h_0). The CML parameter estimates on stock and currency exchange returns on 30 minute intervals are also unrealistic. For one time series however (EUR/CHF currency exchange), the parameters are in the multifractal range (h_0 0.10). The cost function landscape of the CML method as a function of the parameters was very flat for all time series. Therefore all parameter estimations are likely to be inaccurate.

The MIC parameter estimates are less reliable (due to a higher fraction of negative parameters) but were similar to the unrealistic parameters obtained by CML.

We have also seen that the estimated innovations (using the parameter estimates using the CML and MIC methods) are not Gaussian distributed white noise, but in fact behave similar to the returns. We saw a high degree of volatility clustering in the innovations, a structure of the innovations that was very similar to the returns and a probability distribution function of the innovations that had heavy tails. Specifically, for the time series with parameter estimates within the multifractal range, the returns and innovations differed only by a scale σ.

It was concluded that the model does not accurately describe the data and also high performance in forecasting was ruled out. Finally, we have seen a strong relation between absolute innovations and absolute ω. This implies that in real financial time series ω also has a strong dependency on past absolute returns.

Chapter 5

Introduction of the generalized SEMF process

In previous chapters we have seen that though SEMF processes incorporate all known stylized facts of financial time series, it could not be used for applications in its present form. Section 4.4 suggests that the volatility estimation could be improved if the process that determines the volatility (ω) depends on past absolute returns. In this chapter I introduce the generalized SEMF process where the process that determines the volatility (ω) has a dependency on past returns. It is shown that this process also has multifractal properties for both kernel types. It is also shown that the structure of this process is similar to the EGARCH(1,1) process.

Section 5.1 presents the formulation of the new process. Furthermore, the section presents a derivation that demonstrates the similarity of the generalized SEMF process with an exponential kernel and the EGARCH(1,1) process.

Section 5.2 demonstrates that the process also has multifractal properties if there is an additional dependency of ω on the past absolute returns. Furthermore it demonstrates that the new process also has heavy tails in the probability distribution function of the returns.

Section 5.3 investigates whether the estimation procedure presented in chapter 3 is also effective for the new process. We see that the process parameters and the volatility can be estimated with high accuracy for an exponential kernel. For a power-law kernel, both the parameter estimation accuracy as the quality of the volatility estimation are problematic.

5.1 Formulation of the generalized SEMF process

I propose a process where ω depends on past absolute returns. As a null hypothesis, it is proposed that ω depends on the past absolute returns with the same memory kernel as the past returns (with sign):

$$\omega = \sum_{i=0}^{n-1} \left[d_i - k_{abs} \left| d_i \right| \right] h_{n-i-1} \tag{5.1}$$

where k_{abs} is the parameter that regulates the strength of the influence of absolute past returns on ω. For too high values of k_{abs} such a process diverges to zero or infinity. Therefore, as a null hypothesis, it is assumed that k_{abs} has a small value ($|k_{abs}| < 1$). We have seen that in real financial time series,$|\xi_{i,est}|$ has a positive relation with $|\omega_{est}(i)|$. The latter implies that $|\omega(t)|$ should have a positive relation with d_{t-i}, and that $k_{abs} > 0$.

If a process has the exact form as described above (equation 5.1), then the expectation value of ω is a function of time. For a process with no previous returns, $\omega_1 = 0$. Then, $\omega_2 = \left[d_i - k_{abs} \left| d_i \right| \right] h_0$ and

$\omega_3 = \sum_{i=1}^{2} [d_i - k_{abs} |d_i|] h_{3-i-1}$. Because $E[|d_i|] \neq 0$, the expectation values of ω_t are also a function of time: $E[\omega_1] = 0$, $E[\omega_2] = -k_{abs} h_0 E[|d_i|]$ and $E[\omega_3] = \sum_{i=1}^{2} k_{abs} E[|d_i|] h_{3-i-1}$. This is not a property that has been observed in financial time series and is therefore undesired.

Therefore the expectation value of the returns $E[|d|]$ is subtracted to remove this artificial evolution of the volatility as a function of time at the beginning of a new time series. I define the generalized SEMF process as

$$\omega = \sum_{i=0}^{n-1} [d_i - k_{abs}(|d_i| - E[|d|])] h_{n-i-1} \tag{5.2}$$

Now we compute $E[|d|]$, which in such a process is equal to $E[\sigma |\xi| exp(-\omega)]$. Because $E[d] = 0$ and therefore $E[\omega] = 0$, we see that $E[|d|] = E[\sigma |\xi|]$. The innovations ξ follow an i.i.d. $N(0, 1)$ distribution. When we recall the Gaussian probability distribution function (see equation 3.11), we can compute the expectation value of the innovations $E[|\xi|]$:

$$E[|\xi|] = \int_{-\infty}^{\infty} |\xi| f(\xi) d\xi = 2 \int_{0}^{\infty} \xi \frac{1}{\sqrt{2\pi}} exp[-\frac{\xi^2}{2}] d\xi = \sqrt{\frac{2}{\pi}} \tag{5.3}$$

Therefore $E(|Z_t|) = \sigma\sqrt{2/\pi}$ and the generalized SEMF process reads

$$\omega = \sum_{i=0}^{n-1} \left[d_i - k_{abs}(|d_i| - \sigma\sqrt{2/\pi}) \right] h_{n-i-1} \tag{5.4}$$

Monte Carlo simulations are possible for this process (with both memory kernels) for values of k_{abs} 0.1, since otherwise the process becomes highly non-stationary or divergent. In Fig. 5.1 we can see a sample realization of a generalized SEMF process with power law memory kernel.

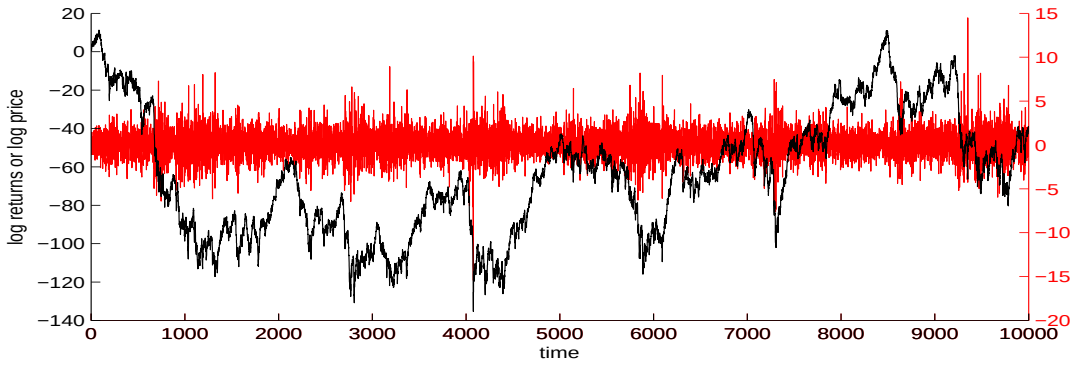


Figure 5.1: Time-series of log price increments d_n (gray) and log price $X_n = P_n = \sum_{i=1}^{n} d_i$ (black) of the generalized SEMF process defined by equation 5.4, for $\sigma = 1$ and with the power-law kernel, specified in equation 2.28, with $\varphi = 0.01$, $h_0 = 0.10$ and $k_{abs} = 0.1$.

5.1.1 Relation with EGARCH(1,1):

The definition of this generalized SEMF process with an exponential kernel is similar to the definition of an EGARCH(1,1) process, where the only difference is that the EGARCH(1,1) has a dependency on past innovations where the generalized SEMF process depends on past returns. We demonstrate this equality below.

An EGARCH(1,1) process is defined as

$$log[\sigma_t^2] = \Omega + \beta_1 g(\xi_{t-1}) + \alpha_1 log[\sigma_{t-1}^2] \tag{5.5}$$

where $g(\xi_t)$ equals

$$g(\xi_t) = \theta\xi_t + \lambda(|\xi_t| - E(|\xi_t|)) \tag{5.6}$$

In such a process the expression for $log[\sigma_{t-1}^2]$ equals

$$log[\sigma_{t-1}^2] = \Omega + \beta_1 g(\xi_{t-2}) + \alpha_1 log[\sigma_{t-2}^2] \tag{5.7}$$

We can substitute this expression into equation 5.5:

$$log[\sigma_t^2] = (1 + a_1)\Omega + \beta_1 g(\xi_{t-1}) + \alpha_1\beta_1 g(\xi_{t-2}) + \alpha_1^2 log[\sigma_{t-2}^2] \tag{5.8}$$

With the following expressions for the process parameters

$$\Omega = 0 \tag{5.9}$$

$$\beta_1 = 2h_0 exp(-\varphi) \tag{5.10}$$

$$\alpha_1 = exp(-\varphi) \tag{5.11}$$

$$\theta = -1 \tag{5.12}$$

$$\lambda = k_{abs} \tag{5.13}$$

the expression reduces to

$$2log\sigma_t = -2h_0 exp(-\varphi)g(\xi_{t-1}) - 2h_0 exp(-2\varphi)g(\xi_{t-2}) - exp(-2\varphi)log[\sigma_{t-2}^2] \tag{5.14}$$

If we iterate the step of substituting $log[\sigma_i^2]$ by equation 5.7, we obtain a structure similar to the generalized SEMF process:

$$\sigma_t = exp(-\omega') \tag{5.15}$$

where the process ω' is defined as:

$$\omega' = -h_0 \sum_{k=1}^{\infty} exp(-k\varphi)g(\xi_k) = -h_0 \sum_{k=1}^{\infty} exp(-k\varphi)\left[\xi_i - k_{abs}(|\xi_i| - E[|\xi_i|])\right] \tag{5.16}$$

Note that this process is identical to the generalized SEMF process if the innovations ξ are replaced by the returns d. Taking $\Omega = 0$ is equivalent to taking $\sigma = 1$ in the SEMF process, but the similarity is also valid for different values of σ.

5.1.2 Conclusion

I have introduced the generalized Self-Excited Multifractal process which has an additional dependency on past absolute returns. We have seen that this generalization of the SEMF process is very similar to the EGARCH(1,1) process.

5.2 Multifractal properties and heavy tails of generalized SEMF process

In this section I demonstrate that the heavy tails and the effective multifractal properties of the original SEMF process are also present in the generalized SEMF process. Therefore they are also possibly present in the EGARCH(1,1) process.

5.2.1 Multifractal properties

The singularity spectrum for the generalized SEMF process is computed. Equations 2.7 and 2.8 can not be used to demonstrate multifractality for non-stationary processes [27]. Multifractal Detrended Fluctuation analysis (MF-DFA) [27] can be used to demonstrate multifractality for non-stationary processes. This method was also used to demonstrate multifractality for the original SEMF process ([16], the singularity spectrum is displayed in Fig. 2.4). The identical procedure is used in this research.

The scaling relation $F_q(s) \sim s^{h(q)}$ is used to find the generalized Hurst exponent $h(q)$ from equation 2.9, where $F_q(s)$ is defined as

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} \left[F^2(v,s)\right]^{q/2} \right\}^{1/q} \tag{5.17}$$

N_s is a number of segments of length s within the time series d_n of length $N = 5e4$. $F^2(v,s)$ is the average of the squared residuals of the linear fit of the time series d_n within the time segment v. This technique is also used by Sornette and Filimonov [16], who found that for $10 < s < N/10$ an excellent scaling regime where $F_q(s) \sim s^{h(q)}$ is observed. Here the exponents $h(q)$ are the slopes of the $F_q(s)$ in the log-log plot in Fig. **??** and represent the generalized Hurst exponent $h(q)$.

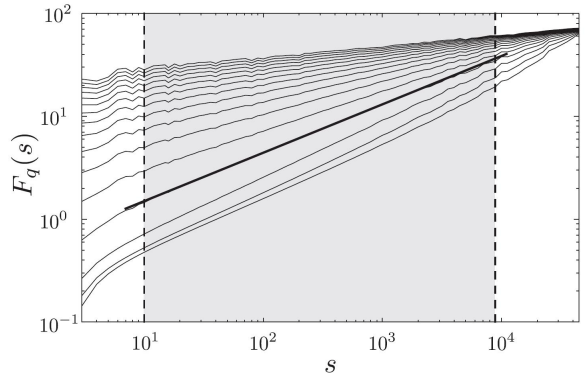


Figure 5.2: Log-log plot of the averaged fluctuations function given by 5.17 as a function of scale s, for q=0.5n, n=1 to 16 (bottom to top), for $\sigma = 1$ and exponential kernel with $\varphi = 0.1$, $k_{abs} = 0.05$ and $h_0 = 0.06$. The grey rectangle delineates the range of scales where the fluctuation function was approximated with a strict power-law $F_q(s) = K_q s^{h(q)}$. The bold line shows the power-law approximation for the case q=2.

.

I simulated M=1000 time series for different parameter sets for both kernel types ($\varphi = 0.01, h_{0,exp} = (0.02, 0.06)$ and $h_{0,pow} = (0.06, 0.10)$). The first half of the time series ($0 < n < N/2$) is not considered, to ensure stability of the processes. I average $F^2(v,s)$ over M time series, before we obtain $F_q(s)$. Figure 5.3 displays the scaling spectra for different values of k_{abs} ($k_{abs} = (0, 0.05, 0.10, 0.15, 0.20)$) for each of these four

sets. The scaling exponent $\xi(q)$ of the standard multifractal structure function shown on the y-axis of Fig. 5.3 is obtained from the generalized Hurst exponent $h(q)$ using the relationship $\xi(q) = qh(q)$.

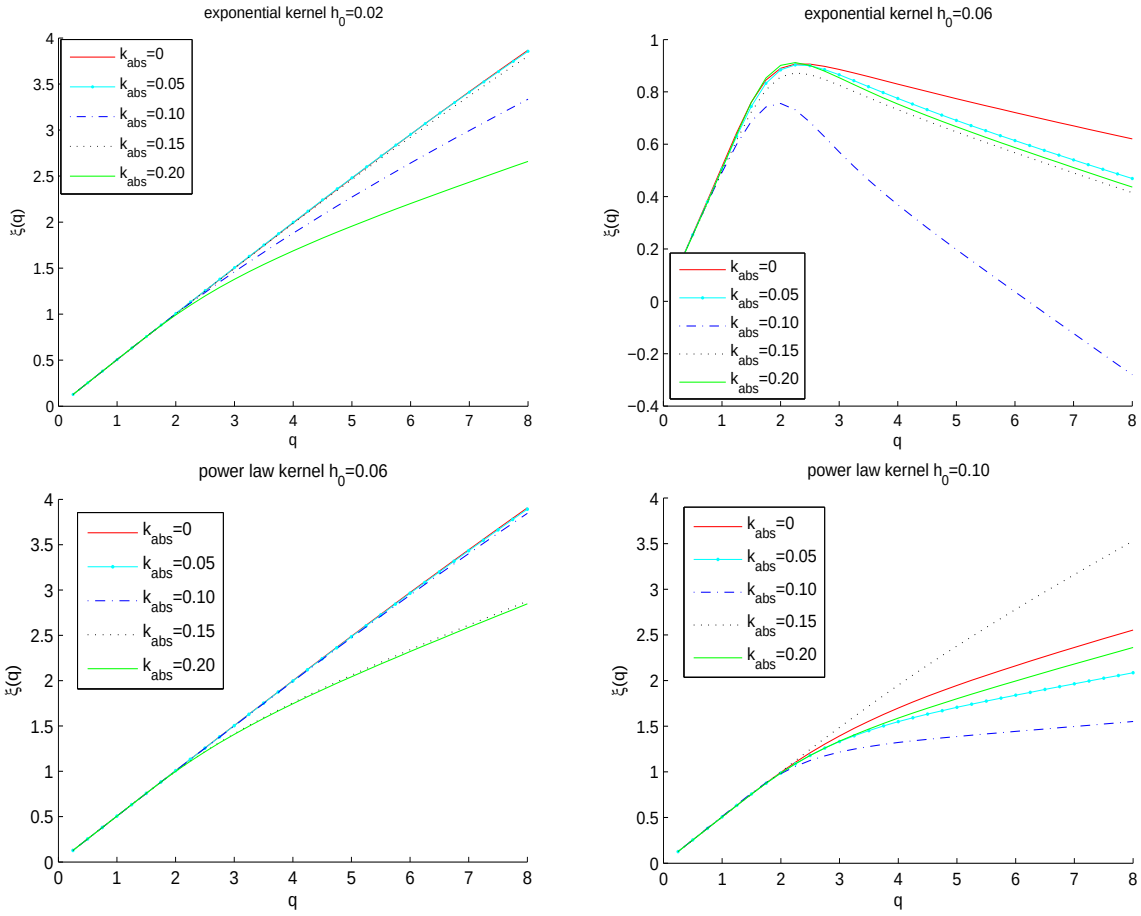


Figure 5.3: Multifractal scaling exponents $\xi(q)$ of the generalized SEMF process d_n for $\sigma = 1$ and $\varphi = 0.01$ with an exponential kernel (Upper two plots) and a power law plot (Lower two plots).

We see that in all plots, for several values of $k_{abs} > 0$, the exponent $\xi(q)$ has a nonlinear dependence on the order q. We can conclude that multifractal properties are observed for $k_{abs} > 0$ for both memory kernel types.

For the plots of both memory kernels with higher h_0, the influence of k_{abs} on the singularity spectrum is ambiguous. In both plots, the non-linearity of the spectrum is increased if we increase k_{abs} from zero to $k_{abs} = 0.05$ and $k_{abs} = 0.10$. However, the non-linearity decreases for $k_{abs} > 0.15$. Since the spectra have wide quantiles, this behavior can be related to the uncertainty of the computation. However, since we observe it for both memory kernels, it is unlikely that such a systemic effect is related to uncertainty. Moreover k_{abs} is a measure of the influence of past absolute returns on the process. The notion of quasi-multifractality was derived for processes with a dependence on past returns (or another stochastic process)). It could be that at these values of k_{abs} the influence of the past absolute returns on the process becomes to great, and 'destroys' the quasi-multifractal properties.

A higher h_0 seems to increase the non-linearity of the spectrum. This was also observed in the original

SEMF process with $k_{abs} = 0$. However for values of h_0 that are too high the process has no physical meaning because volatility bursts dominate the process.

The multifractal properties therefore only hold in certain parameter ranges for the generalized SEMF process.

5.2.2 Heavy tails

It is found that the generalized SEMF process also has heavy tails. I simulated 250 time series with length 10^4 for different values of k_{abs} for the different kernel types. For the resulting collections of returns, histograms were created that can be seen in Fig. 5.4.

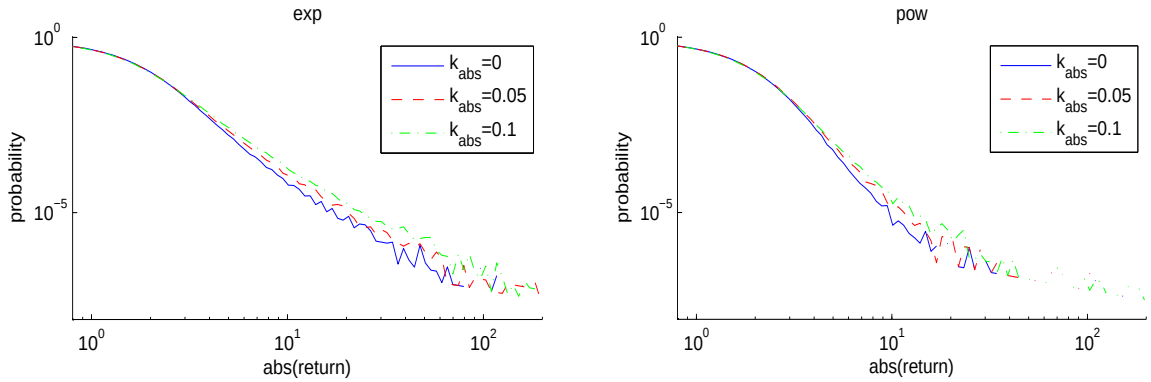


Figure 5.4: The probability distribution function of the generalized SEMF process. The histogram for the exponential memory kernel uses $\varphi = 0.01$, $h_0 = 0.04$ and $\sigma = 1$. For the power-law kernel plot I used $\varphi = 0.1$, $h_0 = 0.10$ and $\sigma = 1$.

We can conclude that the heavy tails in the SEMF process are also present for $k_{abs} > 0$. Therefore the generalized SEMF process also displays this stylized fact. Increasing k_{abs} increases the weight of the tails.

5.3 Estimation of the volatility and process parameters in the generalized SEMF process

This section explores whether the estimation properties found for the pure SEMF process are also valid for the generalized SEMF process with $k_{abs} > 0$. We see that the estimation procedure proposed in section 3.1 that estimates the volatility with known process parameters is also effective if $k_{abs} > 0$. However, we find that the CML estimation method is not effective in accurately determining k_{abs} in the case of a power-law kernel.

5.3.1 Volatility estimation

In this section equation 3.4 (we use the modified structure for ω with $k_{abs} > 0$ from equation 5.4) from section 3.1 is used to estimate the volatility in the generalized SEMF process for known values of the process parameters.

I simulated 25 time series for five different sets of φ and h_0 with four values of k_{abs} for both kernel types. Then, using the same parameters, I estimated the innovations and computed the estimation quality in terms of R^2. The results were averaged for each data set.

It was found that for an exponential kernel the estimation quality R^2 was approximately equal to 1 for all values tested ($\varphi \geq 0.01$,$h_0 \leq 0.10$,$k_{abs} \leq 0.15$). For a power law kernel the effect of k_{abs} is more visible. The resulting R^2 for the different parameter sets used can be seen in Fig. 5.5.

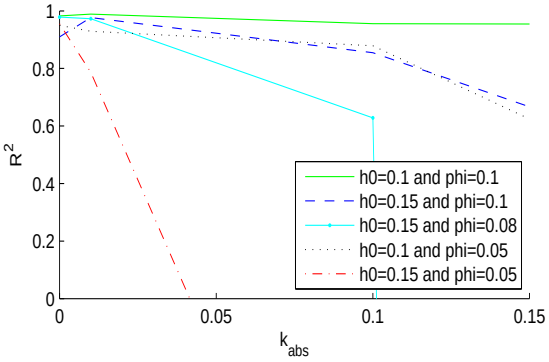


Figure 5.5: The average R^2 decreases as a function of k_{abs} in the case of known parameters for a power law kernel

In the plot, we see that increasing k_{abs} has a negative effect on R^2. The strength of this effect increases with higher h_0 and lower φ. We have seen in section 3.1 that a higher h_0 and lower φ have a negative effect on R^2 when $k_{abs} = 0$. We see that this effect is enforced if $k_{abs} > 0$.

For relatively low values of h_0 and relatively high φ the estimation of the volatility is of high quality ($R^2 > 0.8$). However for combinations of large h_0 and k_{abs}, this quality decreases rapidly.

5.3.2 Parameter estimation in the generalized SEMF process

We also want to know whether we can estimate the parameter of a process with $k_{abs} > 0$ using CML and whether these parameters are sufficiently accurate to obtain high quality volatility estimations. It is found that this is possible for an exponential kernel but for a power law kernel the estimated k_{abs} are too inaccurate.

I simulated 10 time series for both kernel types for three different values of k_{abs} ($k_{abs,exp} = (0; 0, 10; 0, 20)$ and $k_{abs,pow} = (0; 0, 05; 0, 10)$). I used $\sigma = 1$ and $\varphi_{exp} = 0.01$, $h_{0,exp} = 0.06$ and $\varphi_{pow} = 0.1$, $h_{0,pow} = 0.10$. For each time series, I then performed a CML estimation of the kernel type and all four parameters using equation 3.13 (I use the modified structure for ω with $k_{abs} > 0$ from equation 5.4)). As a next step, the volatility was estimated using the parameter estimates. The quality of these estimations of volatility in terms of R^2 was then calculated. We see the resulting R^2 as a function of the estimated $k_{abs,est}$ for the different values of k_{abs} used for simulations for both kernel types in Fig. 5.6.

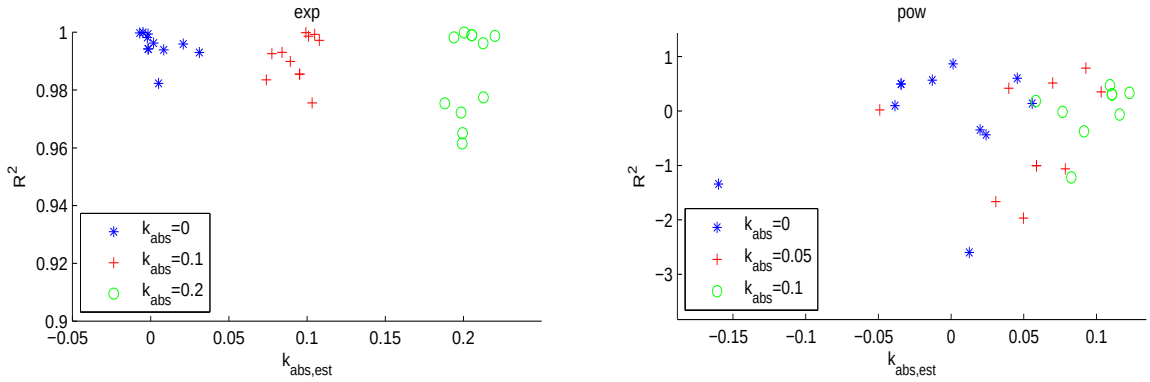


Figure 5.6: CML parameter estimations of a.o. $k_{abs,est}$ in time series with different k_{abs} for both kernel types. The R^2 indicates the accuracy of the estimation of the volatility with the parameter estimates. For the power-law kernel, two results with very negative R^2 for $k_{abs} = 0.05$ and $k_{abs} = 0.10$ are excluded from the plot.

We see that for an exponential kernel $k_{abs,est}$ is estimated accurately and high quality ($R^2 > 0.8$) estimations of the volatility are obtained. However, for a power-law kernel the estimations of the parameters and volatility are less accurate. We see that the CML method determines k_{abs} with an uncertainty of roughly $\Delta k_{abs} \sim 0.10$. Moreover most volatility estimations are of low quality ($R^2 < 0.8$). The parameter type is estimated correctly for all time series with either memory kernel.

The volatility estimation quality R^2 is also low for a few realizations where $k_{abs,est}/k_{abs} \approx 1$. This implies that the lower quality is not only a result of uncertainty in $k_{abs,est}$ but also in the other parameters. Since for $k_{abs} = 0$ accurate volatility estimations are possible for the power-law parameter set, we can conclude that $k_{abs} > 0$ introduces additional uncertainty in the estimations of the other process parameters.

We can conclude that for an exponential kernel we can obtain accurate parameter estimations using CML that lead to an accurate volatility estimation ($R^2 > 0.8$). However, for a power-law kernel, we can estimate k_{abs} but only with limited accuracy. A process with $k_{abs} > 0$ also introduces uncertainty in the estimation of the other parameters. The volatility estimation with the parameter estimates was of a low quality for time series with a power-law kernel. This volatility estimation quality is lower than the quality obtained for pure SEMF processes.

5.4 Conclusion

We have introduced the generalized SEMF process that has an additional dependency on past absolute returns. We have seen that for an exponential kernel this process is similar to the EGARCH(1,1) process.

The SEMF process with $k_{abs} > 0$ has multifractal properties for both kernel types. Therefore we have demonstrated that the generalized SEMF process also has multifractal properties for a certain parameter range (k_{abs} 0.1, φ 0.01, h_0 0.06). The non-linearity of the spectrum increases as a function of k_{abs}. However, for combinations of high h_0 and high k_{abs}, increasing k_{abs} seems to destroy the quasi-multifractal properties and the spectrum returns to linearity. The generalized SEMF process also displays heavy tails for both kernel types. The weight of the tails is increased by a higher k_{abs}. This parameter k_{abs} arranges the dependency of ω on past absolute returns. It also increases the non-linearity of the scaling spectrum, increases the weight of the tails and increases the non-stationary of a process. Therefore its influence on the statistical properties of a generalized SEMF process is very similar to the influence of h_0.

We proposed to use the procedure developed for the original SEMF process for estimation in the generalized SEMF process. Using this procedure for time series with exponential memory kernels, we can estimate the parameters of a synthetic SEMF process with $k_{abs} > 0$ with sufficient accuracy to obtain high quality ($R^2 > 0.8$) estimations of the volatility. However, for a power-law kernel, accurate estimation of the volatility was not possible. We saw that if the parameters are known, k_{abs} also has a negative effect on the estimation quality. Moreover, the accuracy of parameter estimation is low in synthetic time series with $k_{abs} > 0$.

Conclusion

We have seen that the SEMF process needs to be modified to realistically describe financial data. I have proposed a generalization of the SEMF process that includes a dependency on past absolute returns and show that this process also has multifractal properties and heavy tails.

We have demonstrated that the volatility in synthetic SEMF processes can be estimated effectively ($R^2 > 0.8$ for a broad range of parameters and kernel types). If the parameters are known, the volatility can be estimated in a sufficiently large interval (10^4 data points) if we ignore the history of the process. Volatility estimations have a lower quality in synthetic processes with a higher non-stationary (higher h_0) and memory strength (lower φ). The volatility estimation remains of a high quality if there is a limited (~0.10% in each parameter) uncertainty in the kernel parameters. The parameters of a pure SEMF process can be estimated with this required uncertainty for a broad parameter and kernel type range using Conditional Maximum Likelihood (CML). I introduce a new method, based on Minimum Innovation Clustering (MIC), that can also determine the parameters in a pure SEMF process

Application of the estimation procedure in real financial time series has unfeasible results. We have applied the method to stock and stock index returns on days and 30 minute intervals, and to currency exchange returns on 30 minute intervals. Both methods estimate parameters that do not correspond to a SEMF process with a physical meaning (negative parameters of $h_0 > 1$). Furthermore the estimated innovations have a significant clustering and their probability distribution function has heavy tails. Also, the structure of the estimated innovations is very similar to the structure of the returns. Since the pdf of the innovations are heavily tailed and parameters are unrealistic, forecasting and accurate volatility estimations are not possible. Finally, we find a strong correlation between the absolute ω (process that defines the volatility) and the magnitude innovations, which implies that ω in real financial time series also depends on past absolute returns.

A generalized SEMF process that has this dependency on absolute past returns is a promising candidate for application to financial time series. I have demonstrated that this process also has heavy tails and multifractal properties. Because of the small difference with the original SEMF process the generalized process most likely also incorporates long range dependence and the leverage effect. The volatility and parameter estimation method developed for the original SEMF process is used for the generalized SEMF model. The method is effective ($R^2 > 0.8$ for a broad range of parameters) for an exponential kernel in the generalized SEMF process but not for a power-law kernel.

The generalized SEMF process with a power-law kernel most likely fully captures long-range dependence. Since modifications of the estimation procedure (removing an observed bias and using multiple starting points for the volatility estimation, see appendix B) only slightly improved the accuracy of the method, it is unlikely that these estimators improve the estimation for the generalized process. Moreover it is unlikely that parameter estimation using the pdf of the returns for this process is accurate, since the method has proved unsuccessful for the original SEMF process (see appendix C).

In the appendix A I present a proxy for ω. I demonstrate that it accurately estimates ω as a function of average (absolute) past returns in synthetic SEMF processes. When applied to real financial time series the

proxy provides results that are very similar to those of synthetic time series.

Further work on the topic of Self-Excited Multifractal processes may include:

- The possibility of multifractality in the EGARCH(1,1) process in parameters ranges where multifractality is observed in the generalized SEMF process.
- Methods to improve the parameter estimation and possibly the volatility estimation accuracy in generalized SEMF processes with a power-law memory kernel
- The use of the proxy for ω to better understand the relation of ω with past returns in reality, to obtain proxy's of h_0 and k_{abs} in financial time series and to determine the memory kernel of financial data

Acknowledgments

I would like to thank Vladimir Filimonov, Didier Sornette, Jos Thijssen and Peter Denteneer for their contributions to my MSc Applied Physics thesis.

First of all I would like to thank Vladimir Filimonov for many interesting discussions. His analytical abilities and knowledge of volatility (models) have greatly helped me to structure and define my work and ideas. Also his extensive feedback on my report is greatly appreciated. Naturally I also appreciate that I was given such an interesting framework, the Self-Excited Multifractal process that was developed by him and Didier Sornette, for my master thesis.

Furthermore I thank Didier Sornette for the opportunity to do my master thesis at his chair. I have always had a great passion for Physics and Finance and to combine them has been a great experience. Also thank you for your feedback on my work.

I thank Jos Thijssen and Peter Denteneer who have jointly made my thesis on Econophysics a possibility. For me studying Econophysics has been motivating and inspiring.

I express the hope that more Physicist (e.g., TU Delft master students) move into this field and jointly solve the problems we have with risk management in financial markets.

Appendix A

Proxy for volatility process ω

We compute $\langle log[|d_t|]\rangle$ as a proxy for ω in financial time series. We study the relation ω has with past (absolute returns). Therefore we study the relation of the proxy for ω conditional on past (absolute returns)

$$\left\langle \langle log[|d_t|]\rangle \middle| d_{t-i} \right\rangle \tag{A.1}$$

and

$$\left\langle \langle log[|d_t|]\rangle \middle| |d_{t-i}| \right\rangle \tag{A.2}$$

as a function of past (absolute) returns. The results for synthetic time series and real financial time series can be seen in the Figures below.

We see that the proxy for ω is very effective in synthetic time series. Moreover we see that the plots for synthetic time series and real financial time series are very similar. One could use this proxy to better understand the relation of ω with past returns and the memory decay in financial time series.

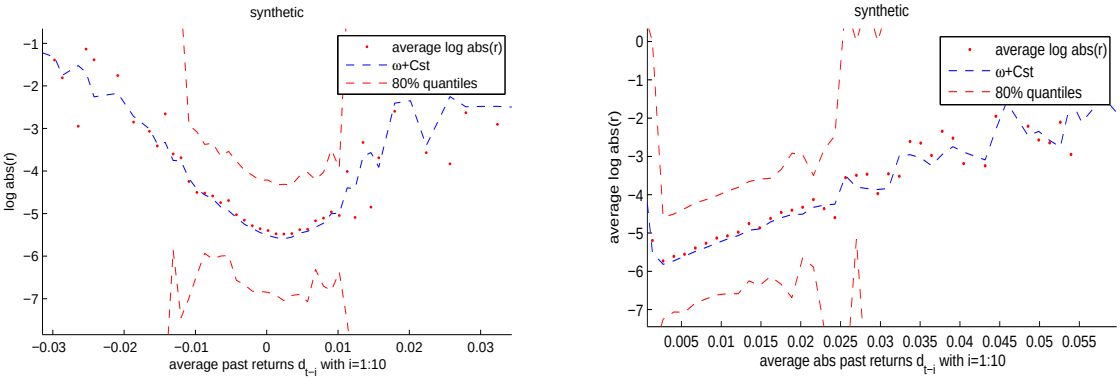


Figure A.1: The proxy of ω compared with the 'real' ω in a synthetic time series. Exponential memory kernel with parameters $h_0 = 0.06$, $\varphi = 0.01$ and $\sigma = 0.01$

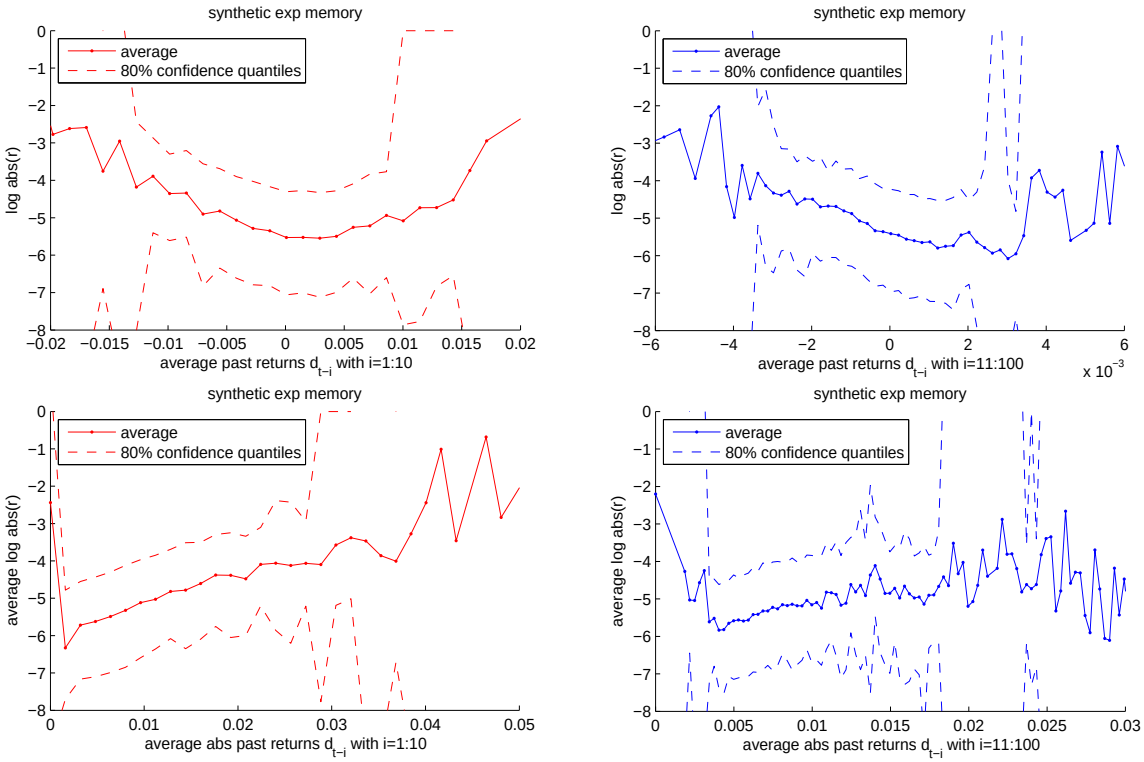


Figure A.2: Synthetic time series with exponential kernel. $h_0 = 0.06$, $\varphi = 0.01$, $\sigma = 0.01$

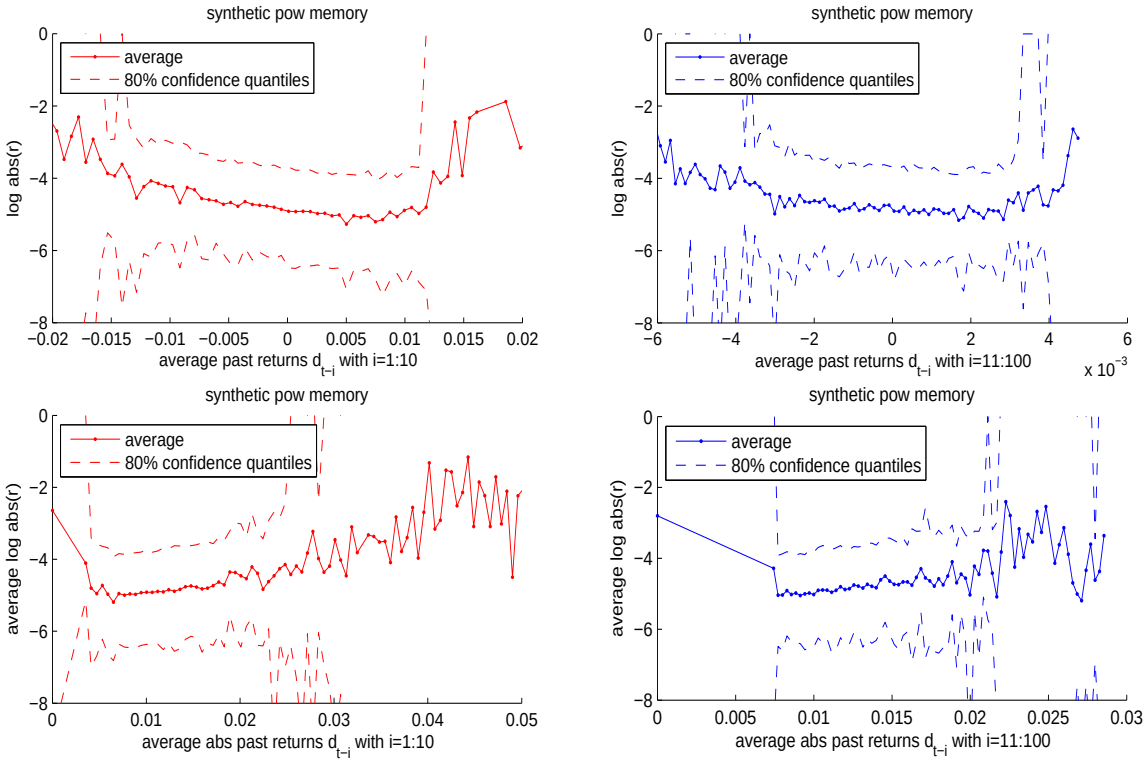


Figure A.3: Synthetic time series with power-law kernel. $h_0 = 0.15$, $\varphi = 0.1$, $\sigma = 0.15$

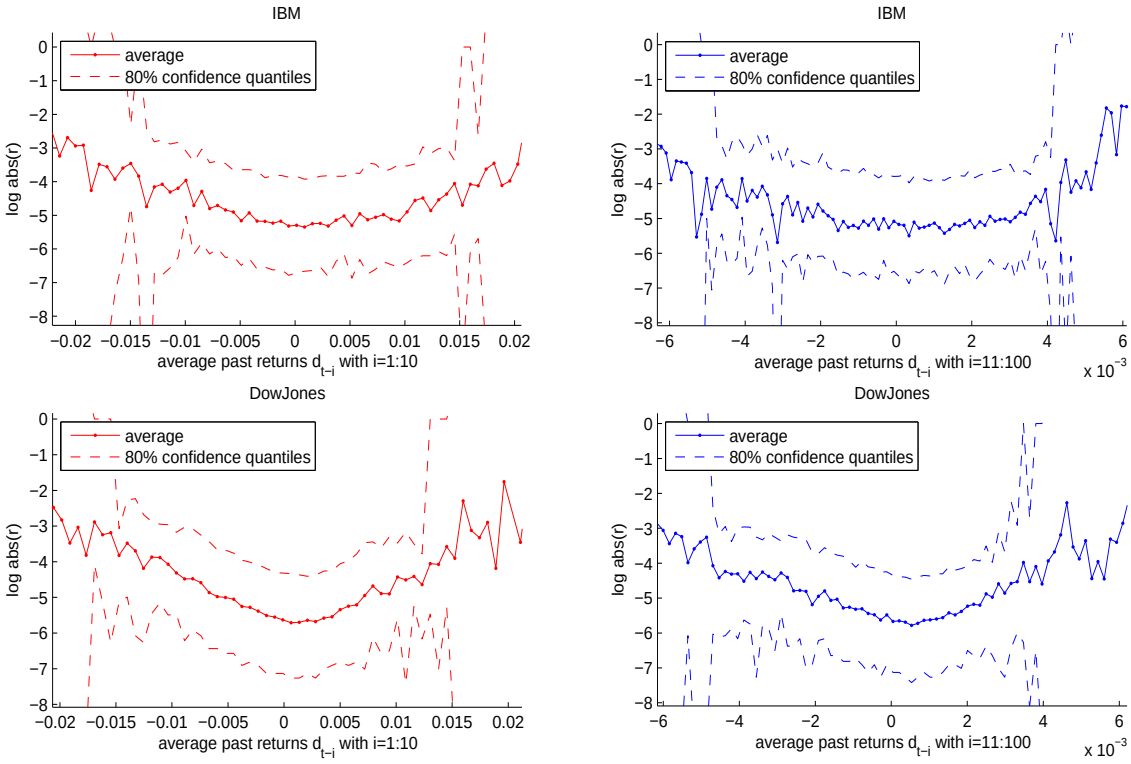


Figure A.4: Dow Jones and IBM

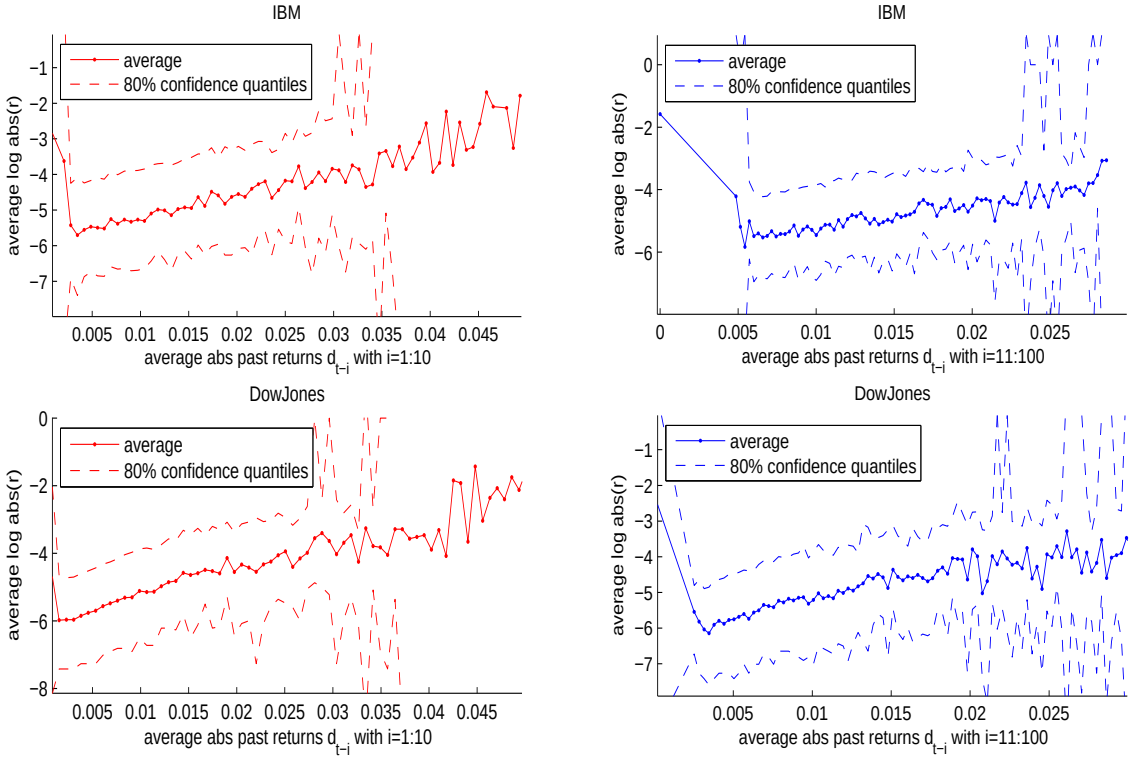


Figure A.5: Dow Jones and IBM

Appendix B

Additional volatility estimators

B.1 Properties of the estimation procedure: initial volatility and bias

As we have found in section 3.1.3, the estimation of volatility is not of sufficiently high accuracy for certain parameter sets for power law kernels. This section shows that these properties can be used to improve the estimation procedure. However, because the CML method resulted in a different domain of kernel parameters, these results were not applied to real data.

B.1.1 Influence of the initial volatility

In section 3.1.4 we have seen that $\sigma_{nsp,real}$ has a strong influence on the estimation of the process, particularly in the case of a power law kernel with a strong memory. In this section we will discuss how this property can be used to improve the estimation procedure.

Therefore a majority of reversals should provide estimations with high quality. However, deviations from 1 and low quality estimations are relatively probable as the probability distribution is quite wide.

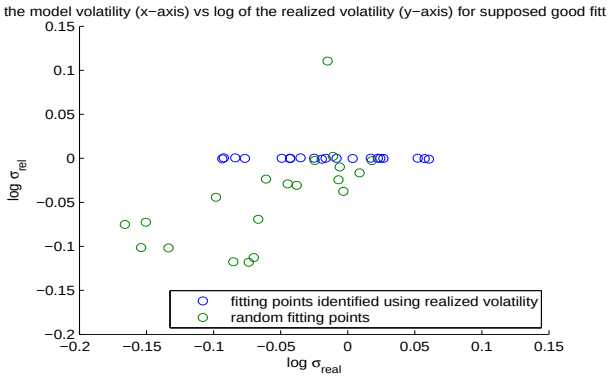


Figure B.1: Example of how n_{sp} with high $|\log(\sigma_{nsp,real})|$ can be excluded using realized volatility.

This problem could possibly be solved by reversing the time series from multiple starting points, expecting the majority of to be close to 1.

An additional tool to exclude starting points with large $|\log(\sigma_{nsp,real})|$ was to calculate the realized volatility σ_{rel}:

$$\sigma_{rel}(n_{rel}, L_{rel}) = \sqrt{\frac{1}{n-1} \sum_{i=n_{rel}}^{n_{rel}+L_{rel}} \left(d_i - \bar{d}\right)^2} \tag{B.1}$$

where

$$\bar{d} = \frac{1}{L_{rel}} \sum_{i=n_{rel}}^{n_{rel}+L_{rel}} d_n \tag{B.2}$$

and where L_{rel} is the interval after the point in time n_{rel} over which the realized volatility σ_{rel} was calculated, which was set at $L_{rel} = 100$. We calculated the $\sigma_{rel}(n_{rel})$ for different n, where $n_{rel} = n + L_{rel}/2$, to take into account the σ_{rel} before and after each point. When N_{nsp} is the number of required starting points n_{sp}, the N_{nsp} points with the lowest $|\log(\sigma_{rel})|$ were picked from a certain interval L_{Nsp}. This method proved successful but only in generally excluding the starting points with the largest $|\log(\sigma_{nsp,real})|$ (see Fig. B.1).

The limited success of the application of these principles are discussed in the appendix section (B.2).

B.2 Volatility estimation with bias compensation and multiple starting points

In order to improve the quality of the estimation of the volatility for a power law kernel with higher h_0 and lower φ, we will propose several estimators based on the properties discussed in the last section. Also, their performance will be evaluated and compared.

B.2.1 Introduction of proposed estimators

As $\sigma_{nsp,real}$ had a large influence on R^2 (see 3.1.4), it was proposed in the last section to use multiple starting points n_{sp}. However, multiple n_{sp} would lead to multiple estimations $\sigma_{n,est}$ and the procedure for obtaining a single $\sigma_{n,est}$ has to be defined. A straightforward procedure would be to take the average of multiple estimations. If $\sigma_{n,est}(n, n_{sp})$ is defined as

$$\sigma_{n,est}(n, n_{sp}) = \sigma \exp\left\{ -\frac{h_0}{\sigma} \sum_{i=n_{sp}}^{n-1} d_i (n-i-1)^{-\varphi-1/2} \right\} \tag{B.3}$$

then the estimator that takes the average of multiple estimations is defined as $\bar{\sigma}_{n,est}$:

$$\bar{\sigma}_{n,est} = \frac{1}{N_{sp}} \sum_{j=1}^{N_{sp}} \sigma_{n,est}(n, n_{j,sp}) \tag{B.4}$$

where N_{sp} denotes the number of starting points $n_{j,sp}$ which was set at $N_{sp} = 21$. The starting points $n_{j,sp}$ were selected using the realized volatility (see subsection 3.1.4) from an interval $L_{Nsp} = 1000$ (this interval length showed optimal over shorter interval lengths).

Another option would be to take the median of multiple $\sigma_{n,est}$, defined by the estimator $\widetilde{\sigma}_{n,est}$:

$$\widetilde{\sigma}_{n,est} = \text{median}(\sigma_{n,est}(n, n_{1,sp}), \sigma_{n,est}(n, n_{2,sp}), ..., \sigma_{n,est}(n, n_{N_{sp},sp})) \tag{B.5}$$

Since the volatility $\sigma_{n,est}$ is generally estimated too low, also an operator $\sigma_{n,est}^{min}$ that takes the lowest value out of multiple $\sigma_{n,est}$ was defined:

$$\sigma_{n,est}^{min} = \min(\sigma_{n,est}(n, n_{1,sp}), \sigma_{n,est}(n, n_{2,sp}), ..., \sigma_{n,est}(n, n_{N_{sp},sp})) \tag{B.6}$$

Finally, as a benchmark, we define the previously used operator $\sigma_{n,est}$ (see **??**):

$$\sigma_{n,est} = \sigma_{n,est}(n, n_{sp}) \tag{B.7}$$

This is also the estimator that uses the longest available interval for the estimation (instead of requiring a relatively large interval to select starting points $n_{j,sp}$). For equations B.4, B.5 and B.6 counts that $n \geq max(n_{1,sp}, n_{2,sp}, ..., n_{Nsp,sp})$.

Since a positive bias was observed when estimating $\sigma_{n,est}$ (see 3.1.5), we also investigated for each estimator whether an improvement could be made by removing the average bias. The bias was therefore quantified for different kernel parameter sets as $\sigma_{bias}(n, \varphi, h_0) = \frac{1}{N}\sum_{i=1}^{N} \sigma_{i,est}(n, \varphi, h_0)/\sigma_{i,rel}(n, \varphi, h_0))$ for $N = 4000$ simulations for each different estimator, which resulted in $\bar{\sigma}_{bias}$, $\sigma_{bias,min}$, $\widetilde{\sigma}_{bias}$ and σ_{bias}. As a next step the estimators defined in B.4, B.5, B.6 and B.7 were adjusted to create estimators without bias, $\bar{\sigma}_{n,est}^{bias}$, $\widetilde{\sigma}_{n,est}^{bias}$, $\sigma_{n,est}^{min,bias}$, and $\sigma_{n,est}^{bias}$, defined as:

$$\bar{\sigma}_{n,est}^{bias} = \frac{1}{\bar{\sigma}_{bias}}\bar{\sigma}_{n,est},\ \widetilde{\sigma}_{n,est}^{bias} = \frac{1}{\widetilde{\sigma}_{bias}}\widetilde{\sigma}_{n,est},\ \sigma_{n,est}^{min,bias} = \frac{1}{\sigma_{bias,min}}\sigma_{n,est}^{min},\ \sigma_{n,est}^{bias} = \frac{1}{\sigma_{bias}}\sigma_{n,est} \tag{B.8}$$

In the next section we will compare these 8 proposed estimators.

B.2.2 Performance of proposed estimators

To compare different estimators, the efficiency $e(\sigma_{est}, n)$ was used as a quality indicator. A lower efficiency $e(\sigma_{est}, n)$ indicates a higher quality of estimation. The efficiency was calculated per data point and not over intervals because there was no noise term in the estimator (in contrast with R^2 where the varying order size of σ introduces noise)

$$e(\sigma_{est}, n) = \left(\frac{\sigma_{est}(n)}{\sigma_{real}(n)} - 1\right)^2 \tag{B.9}$$

An example of a typical experiment can be seen below in Figure **??**.

B.2.2.1 Comparison

The performance of the estimators was evaluated using the efficiency $e(\sigma_{est}, n)$ (equation **??**). An example of such a comparison can be seen in Fig. B.2, where $e(\widetilde{\sigma}_{n,est}^{bias}, n)$ indicated by 'median' is most favorable.

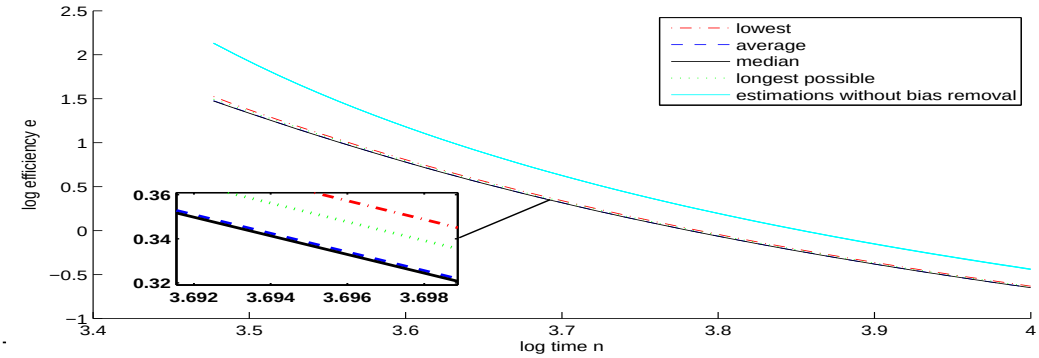


Figure B.2: Efficiency $e(\sigma_{est}, n)$ of different estimators as a function of time. 'Longest possible' indicates $\sigma^{bias}_{n,est}$. Power law kernel with h_0=0.2 and φ=0.01

The estimators with bias compensation performed equally to the estimators without bias compensation only for two parameter sets (φ=0.4 and $h_0 = 0.10, 0.15$) and outperformed them for all other tested parameters. One can therefore conclude that bias compensation proved effective in improving the quality of the volatility estimation.

Out of the estimators with bias compensation, $\sigma^{bias}_{n,est}$ gave the most favorable performance for lower values of h_0 (0.10, 0.15) and all tested φ (0.01-0.4). For higher values of h_0 (0.20, 0.25), and lower values of φ (0.01, 0.04 and 0.1), this was $\widetilde{\sigma}^{bias}_{n,est}$. For $\varphi = 0.4$ and $h_0 = 0.20, 0.25$, again $\sigma^{bias}_{n,est}$ showed the lowest efficiency.

The fact that $\sigma^{min,bias}_{n,est}$ takes an outlier of multiple estimations and that therefore the variability of $\sigma^{min,bias}_{n,est}$ is likely to be higher, can explain that the efficiency $e(\sigma^{min,bias}_{n,est}, n)$ is higher than $e(\widetilde{\sigma}^{bias}_{n,est}, n)$, as an average bias of the outliers is therefore less likely to be accurate for single estimations. Also, it plausible that the variability of a median ($\widetilde{\sigma}^{bias}_{n,est}$) is lower than the variability of an average ($\bar{\sigma}^{bias}_{n,est}$), and that therefore the bias was more accurately estimated to result in lower efficiency. This can explain our observation that $e(\widetilde{\sigma}^{bias}_{n,est}, n)$ is slightly lower than $e(\bar{\sigma}^{bias}_{n,est}, n)$ for all kernel parameters.

$\sigma^{bias}_{n,est}$ with a single n_{sp} showed lower efficiency than $\widetilde{\sigma}^{bias}_{n,est}$ and all other estimators using multiple starting points n_{sp} at most kernel parameter intervals . The results therefore demonstrate that the estimators $\bar{\sigma}^{bias}_{n,est}$, $\widetilde{\sigma}^{bias}_{n,est}$ and $\sigma^{min}_{n,est}$ were not sufficiently successful in either avoiding starting points with a high $|\log(\sigma_{nsp,real})|$ or reducing its influence, in order to justify the use of a shorter interval to estimate σ_{est}, in these kernel parameter intervals.

Aside from its overall lowest efficiency $e(\sigma^{bias}_{n,est}, n)$, $\sigma^{bias}_{n,est}$ is also considerably simpler and easier to compute than $\widetilde{\sigma}^{bias}_{n,est}$. It was therefore concluded that overall $\sigma^{bias}_{n,est}$ is the most suitable estimator. In the next paragraph, the performance of this estimator in terms of R^2 and the quantiles of the volatility with a confidence interval of 95% is displayed.

B.2.2.2 Most accurate results

Figure B.3 (Left) shows isoplots of R^2 for different h_0 and φ for $\sigma^{bias}_{n,est}$. Figure B.3 (Right) displays the percentage of simulations that reached $R^2 = 0.8$ within 8000 data points for different kernel parameter sets. Using the new estimator, indeed a significantly higher percentage of success could be reach for all parameter sets (except for φ= 0.1 and $h_0 = 0.25$, probably because of the instability due to the high h_0). However the increase in quality was not enough to elevate the quality of new kernel parameter sets (with $R^2 < 0.8$ for the old operator) above the $R^2 = 0.8$ threshold.

Fig. B.4 shows the quantiles for the relative estimated volatility ($\sigma_{est}/\sigma_{rel}$) with a confidence level of 95%

also after bias compensation. These results are significantly better than those without bias compensation (see section 3.1.3).

However the proposed estimator can not be applied to real financial data. As we discuss in section **??**, parameter ranges are found for which Monte Carlo simulations are not possible. Therefore quantification of the bias is not possible. Furthermore, the innovations of the real financial time series behave differently (not i.i.d. nor a Gaussian distribution) than proposed for the process.

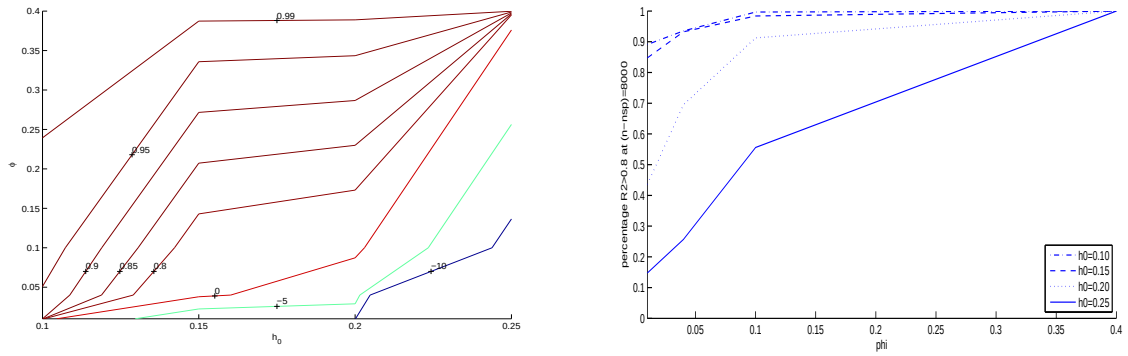


Figure B.3: **Left**:Isoplots of the average R^2 after bias compensation for different φ and h_0. **Right**: Dependence of $P_{R2\geq 0.8}$ on φ for various h_0 with $T_{R2} = 8000$ for a power law kernel for $\sigma^{bias}_{n,est}$. Each point is based on 4000 simulations.

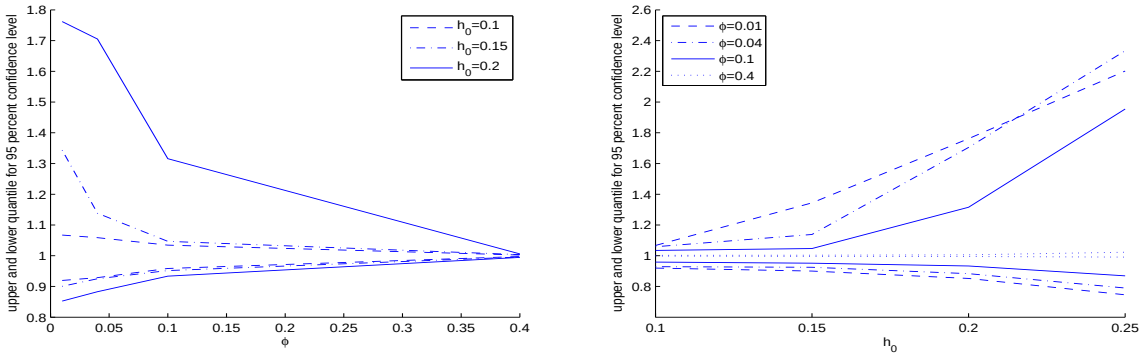


Figure B.4: **Left**:Upper and lower quantiles as a function of φ. **Right**: Upper and lower quantiles as a function of h_0 . Each point is based on 4000 simulations.

Appendix C

Alternative methods for kernel type and parameters estimation

C.1 Comparing distribution functions to estimate kernel parameters

Because the kernel type and parameters influence the distribution of the returns of the process, the final proposal for parameter estimation is based on the probability distribution function (pdf) of the observed process.

C.1.1 Minimizing Cramér-von-Mises distance

Several methods, such as Kolmogorov-Smirnov, Crámer-von-Mises and Anderson-Darling, can be used to test the similarity of different (empirical) distributions. In this research, it is chosen to minimize the squared distance between different distributions, which is known as the Crámer-von-Mises distance and is expressed as [3]:

$$\omega^2 = int_{-\infty}^{\infty} \left[F_N(x) - F(x)\right]^2 dF(x) \tag{C.1}$$

where $F(x)$ is the theoretical cumulative distribution function and $F_N(x)$ is the empirical distribution function.

It was derived by Anderson [3] for two empirical distribution functions this distance could be expressed as

$$U = N \sum_{i=1}^{N} (r_i - i)^2 + M \sum_{j=1}^{M} (s_j - j)^2 \tag{C.2}$$

where r_i and s_j are the ranks of two different empirical samples in the pooled $(r + s)$sample.

Because an analytical expression for the cdf of the process in terms of the parameters was not found nor approximated, a collection of returns $r(\theta)$ was created using Monte Carlo simulations with known kernel type and parameters. For an empirically observed set of returns s, it was then tested for which parameters the Cramér-von-Mises distance was the smallest.

C.1.2 Influence of the kernel parameters on the distribution of the returns

To obtain the cumulative distribution functions (pdf) for different parameter sets, we simulate 10^3 time series with length 10^4 for different kernel type and parameter sets ($\varphi_{exp} = 0.01, 0.05$, $h_{0,exp} = 0.04 - 0.10$) and ($\varphi_{pow} = (0.01, 0.4)$, $h_{0,pow} = 0.10, 0.20$). The resulting collection of absolute returns is distributed over a histogram with logarithmic bin scale. The histogram is normalized for bin size and the total number of returns and displayed in Fig. C.1 as a function of $|r|$. The resulting pdf has the same shape as the pdf published in [9] (different parameters are used for the power law kernel).

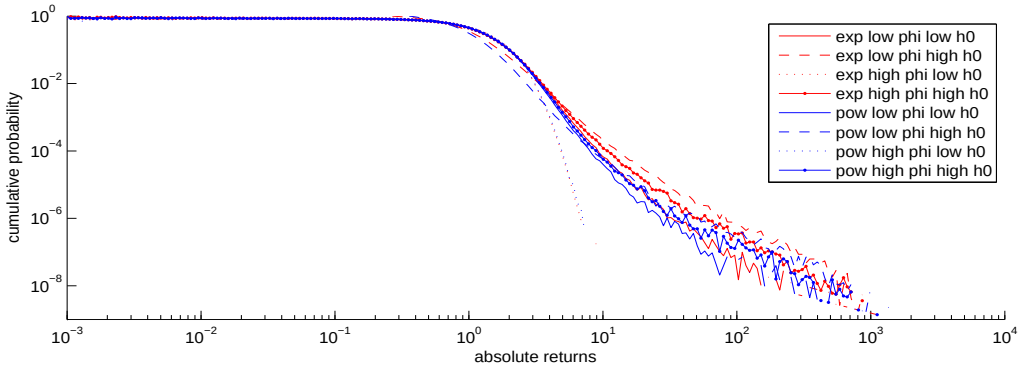


Figure C.1: Pdf of the absolute returns for different kernel types and parameters

We see that the distributions are very similar. The differences between all kernel and parameter sets are the largest in the tails of the distributions, when $|r| > 10^1$. However, because a typical sample size has only around 10^4 data points, the distribution in the range $|r| < 5 * 10^1$ is most relevant. We see that low φ and high h_0 has a slightly distinctive shape for both kernel types in the range $|r| = 10^0 - 10^1$. Furthermore high φ and high h_0 has a quite distinctive shape in the range $|r| = 5 * 10^0 - 10^1$.

C.1.3 Results of minimizing Cramér-von-Mises distance

As a next step we apply Anderson's equation for computing Cramér-von-Mises distance (see equation C.2) for Monte Carlo simulations. We simulate 20 time series for each of the earlier defined parameter sets to obtain the returns r. For the sample s we use the 10^3 time series that were used to compute the cdf above. We then calculate U for each of the eight available parameter sets. The parameter set for which U is the lowest, is the final result. The fraction of simulations that pointed to each of the eight parameter sets can be seen in Table C.1.

Above: resulting parameters. Below: Experiment parameters	Exp: low φ_{est} low $h_{0,est}$	Exp: low φ_{est} high $h_{0,est}$	Exp: high φ_{est} low $h_{0,est}$	Exp: high φ_{est} high $h_{0,est}$	Pow: low φ_{est} low $h_{0,est}$	Pow: high φ_{est} low $h_{0,est}$	Pow: high φ_{est} low $h_{0,est}$	Pow: high φ_{est} high $h_{0,est}$	Percentage of the results that points to right set
Exp: low φ low h_0	50%		5%		5%		10%	30%	50%
Exp: low φ high h_0		90%				10%			90%
Exp: high φ low h_0			50%		5%		45%		50%
Exp: high φ high h_0	30%			45%	5%			20%	45%
Pow: low φ low h_0	20%	5%	10%				30%	35%	0%
Pow: low φ high h_0		40%				60%			60%
Pow: high φ low h_0			55%				45%		45%
Pow: high φ high h_0	15%			15%			5%	65%	65%
Percentage of results that is correct	44%	67%	42%	75%	0%	86%	33%	43%	51%

Table C.1: Results of minimizing Cramér-von Mises distance using Anderson's ranking algortihm

We see that on average the method has an accuracy of indicating the same parameter set by which the time series was generated as the optimal set of around 50%. This result is worse than the results for CML and MIC (particularly for an exponential kernel, for which the accuracy is quite high for both methods). Furthermore we observe biases for several parameter sets, for example the exponential kernel with low φ and high h_0 is indicated frequently as the optimal set by time series generated with different parameters or even kernel type. The same counts for a power law kernel with high φ and high h_0.

The method performs particularly bad in appointing the right kernel type (around 50%). The average result of appointing the right parameter can be improved up to around 75% if the kernel type is known in advance.

The accuracy of the method can not be improved by using a larger sample for s, as we see that the accuracy does not increase with sample size after a sample size of around $2 * 10^2$ is used.

Another complication of the method are the long computation times and the difficulty of going to finer grid scales. Analytical expressions for the distributions are unlikely to improve the accuracy of the method as the exact distribution must be simplified with regards to the currently used distribution.

Therefore, we conclude that this method is inferior for kernel type and parameter estimation with respect to MIC and especially CML.

C.2 Kendall tau coefficient

The Kendall tau rank correlation coefficient τ is a measure of correlation in rank between two different time series x and y. It is expressed as:

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\frac{1}{2}n(n-1)} \tag{C.3}$$

Where (x_i, x_j) and (y_i, y_j) is a concordant pair if

$$x_i > x_j \text{ and } y_i > y_j \text{ or } x_i < x_j \text{ and } y_i < y_j \tag{C.4}$$

and discordant otherwise. Because the proposes SEMF process is self organizing, it is supposed that the correlation in rank between d and $\xi(\theta)$ is minimum. Therefore, we tested whether there is any negative correlation between optimal parameter set and Kendall tau coefficient and whether it is possible to use this method for parameter estimation.

However it was found that the Kendall-tau coefficient has no correlation to the accuracy of the estimated parameters. We simulated several time series for an exponential kernel and a power law kernel and estimated the innovations with different φ_{est} and $h_{0,est}$. For the resulting innovations the Kendall tau coefficient was computed with the returns. Fig. C.2 shows the resulting values for the estimation parameters.

Kendall tau coefficient: Exponential kernel	$h_{0,est} = 0.04$	$h_{0,est} = 0.10$	$h_{0,est} = 0.15$	Kendall tau coefficient: Power law kernel	$h_{0,est} = 0.10$	$h_{0,}$
$\varphi_{est} = 0.005$	0.93	0.81	0.72	$\varphi_{est} = 0.01$	0.93	
$\varphi_{est} = 0.01$	0.95	0.85	0.77	$\varphi_{est} = 0.04$	0.94	
$\varphi_{est} = 0.02$	0.96	0.90	0.83	$\varphi_{est} = 0.1$	0.95	

Table C.2: Kendall tau coefficient of the returns and the innovations for a time series generated by an exponential kernel with $\varphi = 0.01$ and $h_0 = 0.10$ fitted by both kernel types with varying parameters. $\sigma = 1$ during the experiment.

We see that the Kendall tau coefficient correlated more with the parameters used for estimation that the accuracy of the parameters. E.g., the value for the correct kernel type and parameters, 0.85, is around the average of the total sample.

Therefore we concluded that using the Kendall tau coefficient is not a promising method for kernel type and parameters estimation.

Bibliography

[1] Lux, Thomas. 2009. Stochastic Behavioral Asset-Pricing Models and the Stylized Facts. HANDBOOK OF FINANCIAL MARKETS: Dynamics and Evolution. First Edition. Elsevier Inc. doi:10.1016/B978-0-12-374258-2.50007-5.

[2] Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1, no. 2 (February 1): 223-236. doi:10.1080/713665670.

[3] G. Soros, The Alchemy of Finance (Simon & Schuster, 1988), 1st ed.

[4] Bouchaud, Jean-philippe. 2009. How Markets Slowly Digest Changes in Supply and Demand. HANDBOOK OF FINANCIAL MARKETS: Dynamics and Evolution. First Edition. Elsevier Inc. doi:10.1016/B978-0-12-374258-2.50006-3. http://dx.doi.org/10.1016/B978-0-12-374258-2.50006-3.

[5] Shiller, R.J. 2011. Do Stock Prices Move Too Much to be Justied by Subsequent Changes in Dividends? The American, and Economic Review. 71 (3): 421-436.

[6] David, M, M James, and H Lawrence. 1989. What Moves Stock Prices ? Spring.

[7] Joulin, Armand, Augustin Lefevre, Daniel Grunberg, and Jean-philippe Bouchaud. 2008. "Stock price jumps : news and volume play a minor role": 1-14.

[8] Sornette, Didier, and Yannick Malevergne. 2003. What causes crashes ? Power (February).

[9] Bollerslev, Tim. 1986. GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY. Journal of Econometrics 31: 307-327.

[10] Nelson, D.B. 1991. Conditional Hetereoskedasticity in Asset Returns: A New Approach. Econometrica, Vol. 59, No. 2 (Mar., 1991), 347-370.

[11] Mandelbrot, B.B. and van Ness, J.W. 1968. Fractional Brownian motions, fractional noises and applications. SLAM review Vol. 10, No. 4, oktober 1968

[12] Calvet, L. Fischer, A. 2001. Forecasting multifractal volatility. Journal of Econometrics 105, no. 1 (November): 27-58. doi:10.1016/S0304-4076(01)00069-0.

[13] Bacry, E., J. Delour, and J. Muzy. 2001. Multifractal random walk. Physical Review E 64, no. 2 (July): 2-5. doi:10.1103/PhysRevE.64.026103.

[14] Saichev, A.I. and V.A. Filimonov, 2007. On the Spectrum of Multifractal Diffusion Process. Journal of Experimental and Theoretical Physics, 2007, Vol. 105, No. 4, pp. 1085–1093

[15] Saichev, A, and D Sornette. 2006. Generic multifractality in exponentials of long memory processes 1: 1-11. doi:10.1103/PhysRevE.74.011111.

[16] Filimonov, V., and D. Sornette. 2011. Self-excited multifractal dynamics. EPL (Europhysics Letters) 94 (4) (May 1): 46003. doi:10.1209/0295-5075/94/46003. http://stacks.iop.org/0295-5075/94/i=4/a=46003?key=crossref.600715f9cf6f13ceaa89222d3a69d18b.

[17] Gordon, M.J. 2011. DIVIDENDS , EARNINGS , AND STOCK PRICES. Press, The M I T, and The Review. Review Literature And Arts Of The Americas 41, no. 2: 99-105.

[18] Mantegna R.N., Stanley H.E. 2000. An introduction to econophysics: correlations and complexity in nance. Cambridge, England: Cambridge University Press

[19] Johansen, Anders, Olivier Ledoit, and Didier Sornette. Crashes as Critical Points.

[20] Taleb, Nicolas Nassem. 2007. The Black Swan: Impact of the Highly Improbable. Random House.

[21] Huang, Da, Hansheng Wang, and Qiwei Yao. 2008. Estimating GARCH models: when to use what? The Econometrics Journal 11, no. 1 (March): 27-38. doi:10.1111/j.1368-423X.2008.00229.x. http://doi.wiley.com/10.1111/j.1368-423X.2008.00229.x.

[22] Mandelbrot, Benoit B., Fisher, Adlai J. and Calvet, Laurent E., 1997. A Multifractal Model of Asset Returns Cowles Foundation Discussion Paper No. 1164; Sauder School of Business Working Paper. Available at SSRN: http://ssrn.com/abstract=78588

[23] Anderson, T.W. 1962. On the distribution of the two-sample Cramér-von Mises Criterion. Ann. Math. Statist. Volume 33, Number 3 (1962), 1148-1159.

[24] http://www.wessa.net/stocksdata.wasp

[25] Bloomberg

[26] Bloomberg

[27] Kantelhardt, Jan W, Stephan A Zschiegner, and H Eugene Stanley. 2002. "Multifractal detrended uctuation analysis of nonstationary time series." Physica A 316: 87 - 114.

[28] Baillie, Richard T, Tim Boi, and Hans Ole. "Econometrics Fractionally integrated generalized autoregressive conditional heteroskedasticity". Journal of Econometrics 6: 3-30.

[29] Lux, Thomas, and Taisei Kaizoji. "Forecasting Volatility and Volume in the Tokyo Stock Market : Long Memory , Fractality and Regime Switching *." Science i (516446): 1-36.

[30] J.C. Lagarias, J.A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder - Mead simplex method in dimensions. SIAM Journal of Optimization, 9:112-147, 1998