UiT
THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Science and Technology

Department of Mathematics and Statistics

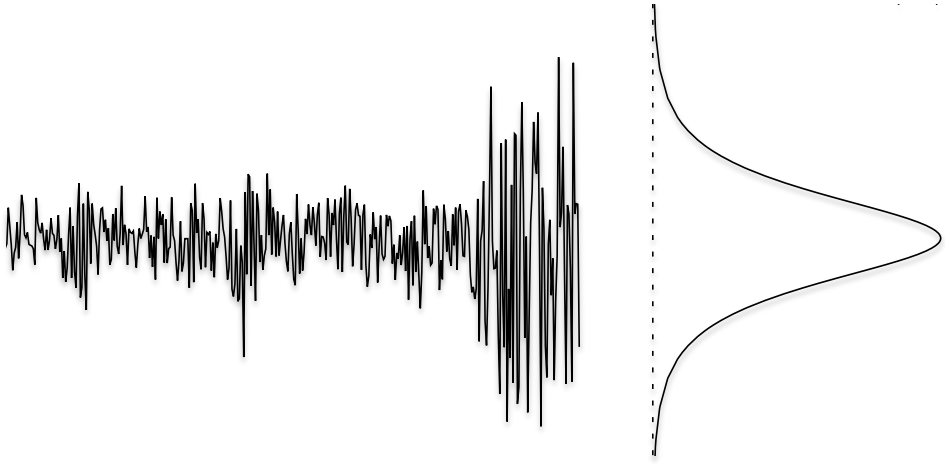
Statistical methods for scale-invariant and multifractal stochastic processes

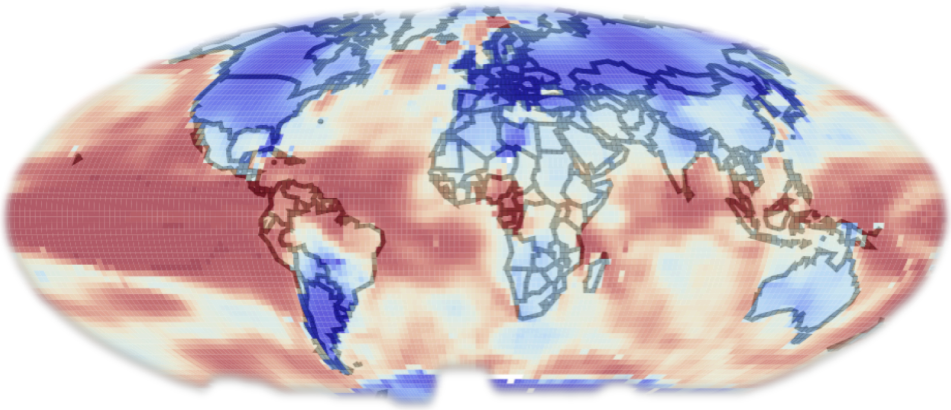
With applications in finance and climate

—

Ola Løvsletten

A dissertation for the degree of Philosophiae Doctor – July 2014





UNIVERSITY OF TROMSØ

Abstract. This thesis focuses on stochastic modeling, and statistical methods, in finance and in climate science. Two financial markets, short-term interest rates and electricity prices, are analyzed. We find that the evidence of mean reversion in short-term interest rates is week, while the "log-returns" of electricity prices have significant anti-correlations. More importantly, empirical analyses confirm the multifractal nature of these financial markets, and we propose multifractal models that incorporate the specific conditional mean reversion and level dependence.

A second topic in the thesis is the analysis of regional ($5^\circ \times 5^\circ$ and $2^\circ \times 2^\circ$ latitude-longitude) globally gridded surface temperature series for the time period 1900-2014, with respect to a linear trend and long-range dependence. We find statistically significant trends in most regions. However, we also demonstrate that the existence of a second scaling regime on decadal time scales will have an impact on trend detection.

The last main result is an approximative maximum likelihood (ML) method for the log-normal multifractal random walk. It is shown that the ML method has applications beyond parameter estimation, and can for instance be used to compute various risk measures in financial markets.

Acknowledgement.

I would like to thank the Department of Mathematics and Statistics for an inspiring and friendly working environment. And, of course my supervisors, Martin and Kristoffer: Thank you for excellent teaching and guidance! I am also grateful to the Economy and Risk Department at SNN, where I shared an office once a week during these years. In particular, thanks to Geir, Solveig, Øystein, Bengt, Svein Ivar, Bjarne and Helene.

Over the last years the physicists have also joined in. Although I will never learn physics, I fully acknowledge the contribution from you guys and girls. So thank you Tine, Hege, Lene, Daniel and Odd Erik. Also, Nick deserves mentioning for showing interest in our work on multifractals, and including us in his network. This is truly appreciated.

Finally, the biggest thanks to the most important people in my life: Mom and dad, Tina, Joar, Bestemor, Anne, Rolf, Øyvind, Eva and the rest of my family and friends.

Contents

1 Introduction

1.1 Overview

This thesis consists of four papers focusing on stochastic modeling and statistical methods, in finance and climate science. Particular emphasis is on stochastic processes with scale invariance in the forms of self-similarity or multifractality. The papers to be defended are:

- Paper 1: Multifractal modeling of short-term interest rates. Joint with M. Rypdal
- Paper 2: Approximated maximum likelihood estimation in multifractal random walks. Joint with M. Rypdal. Published in Physical Review E (2012)
- Paper 3: Modeling electricity spot prices using mean-reverting multifractal processes. Joint with M. Rypdal. Published in Physica A: Statistical Mechanics and its applications (2013)
- Paper 4: Significance of local surface temperature trends from globally gridded data. Joint with M. Rypdal, K. Rypdal and H. B. Fredrikssen.

In the introduction we also refer to the following work, which has also been carried out during the four-year period as a PhD student:

- Paper A1: Assessing market uncertainty by means of a time-varying intermittency parameter for asset price fluctuations. Joint with M. Rypdal, K. Rypdal and E. Sirnes. Published in Physica A: Statistical Mechanics and its applications (2013)
- Paper A2: A multifractal approach towards inference in finance. Joint with M. Rypdal

Other scientific contributions from this period include:

- Invited speaker to "Physical Origins of Correlated Extreme Events", Max Planck Institute for the Physics of Complex Systems. Two presentations:
  - Introductory tutorial: From Bachelier's Brownian motion to Mandelbrot's multifractal model of asset returns and beyond. Joint with M. Rypdal
  - Significance-testing of trends: Long-range dependence and model selection. Joint with M. Rypdal, K. Rypdal and H. B. Fredrikssen.
- Invited speaker to "Aggregation, Inference and Rare Events in the Natural and Socio-economic Sciences", Centre for Complexity Science, University of Warwick. Research talk entitled "Multifractal inference". Joint with M. Rypdal.

1.2 Random walks in finance: from Brownian motions to multifractals

The simplest example of a self-similar process is the Brownian motion. This process was described mathematically by Thiele (1880), Bachelier (1900) and Einstein (1905). Actually, Bachelier, in his thesis entitled "Theory of speculation", proposed Brownian motion as a model for the temporal fluctuations of the prices of financial assets. As observed by Mitchell (1915), the magnitude of price changes often depend on the price level $P(t)$ itself, with higher variability for increasing price levels. The mathematical form of this dependency is well described by the conditional standard deviation $\mathrm{sd}(P(t)|P(t) = x) \propto x^{\gamma}$, where $\gamma > 0$ is a parameter which can estimated from the data (paper 1). For many financial time series one assumes $\gamma \simeq 1$, and in this case a logarithmic transformation will give a process where the increments are independent of the price levels. The model that describes the logarithmic prices as a Brownian motion (with a constant drift term), can be seen as a first refinement of the Bachelier model.

A further improvement is to account for the heavy-tailed distributions of the increments, i.e. the fact that the empirical distributions of the returns are more leptokurtic than Gaussian distributions. Staying within the class of self-similar processes, Mandelbrot (1963) proposed to use Lévy flights. A Lévy flight has increments with α-stable distributions, with $\alpha < 2$. These distributions are characterized by probability density functions (pdfs) $p(x) \sim 1/|x|^{1+\alpha}$ as $x \to \infty$. This implies that statistical moments of higher order than α, do not exist. In particular, the variance is infinite.

Both Brownian motions and Lévy flights have independent increments, while another important "stylized fact" observed across most markets, is that increments are uncorrelated but dependent in time. The property of uncorrelated returns is consistent with the assumption of an arbitrage-free market, but this principle does not imply independence of the increments. While the sign of future price movements are not predictable, the amplitude, to some extent is. Large returns tend to be followed by large returns, and

vice versa for small returns. Among the first models to incorporate time dependence in volatility are the auto-regressive conditional variance (ARCH) models (Engle, 1982). In these processes the volatility (the expected value of squared increments conditioned on the natural filtration generated by the process) follows an auto-regression of lagged squared increments. The model is completed by componding the resulting volatility process with some martingale process, typically with student-t or normal distributed innovations.

A first improvement of the ARCH process is to also include past volatility into the regression. The result is the Generalized ARCH (GARCH) model (Bollerslev, 1986). An alternative is the stochastic volatility (SV) model of Taylor (1982), where, in continuous-time, the logarithm of the volatility is described by an Ornstein-Uhlenbeck (OU) process (Uhlenbeck and Ornstein, 1930). The model is completed by subordinating the resulting volatility process with a Brownian motion. However, there is empirical evidence of long-range dependence (LRD) in the volatility process (e.g. Baillie, 1996) which is not described by simple SV models, since the OU process has an exponentially decaying autocorrelation function (ACF). The natural modification of the simple SV model are the multifractal processes, which include LRD in the volatility process. This was first discovered by Ghashghaie et al. (1996), and in 1997 Mandelbrot, Calvet, and Fisher constructed the Multifractal Model of Asset Return (MMAR), which can be formally written as a subordinated Brownian motion. Here the inner process is the distribution function of a multifractal random measure. This random measure is a product of positive random variables (weights) which are distributed from coarse to fine scales on a b-adic tree. Thus, unlike simple SV models, changes in the volatility occurs at all time scales (up to some large integral scale).

The cascade construction in the MMAR model can be generalized in several ways. One may first note that the inner process has non-stationary increments. However, if one assigns time intervals of random duration in the cascade construction, rather than fixing these, then the result is a stochastic process with stationary increments and multifractal properties. This construction is known as the Poisson multifractal (Calvet and Fisher, 2001). The discrete-time counterpart is known as the Markov Switching Multifractal (MSM, Calvet and Fisher, 2004).

Modeling short-term interest rates

For the MSM model with binomial distributions on the weights (volatility components), the likelihood can be computed exactly, and thus likelihood-based inference can be carried out. This is the approach used in paper 1, where we model short-term interest rates using a version of the MSM model (level-MSM), and compare its performance with GARCH-type of models. A first step in this modeling is to determine the number of levels K to use in the volatility cascade. Actually, this is a test of multifractality since we can determine, by means of likelihood-ratios, if large K are significantly better supported by the data than small values of K. In the paper we conclude that the largest tested value of K (in this case $K = 9$) is significantly more likely than all smaller values of K. In model comparison against alternative models we find that the level-MSM model performs as good, and to

some degree better, than the alternatives considered. The main conclusion is the following:

> *Multifractal models provide accurate statistical descriptions of short-term interest rates.*

Traditionally, interest rates have been modeled using damping terms (Longstaff et al., 1992, and references therein). In our study (paper 1) we included the possibility of mean reversion. However, the estimates of the parameter governing the damping term was not found to be significantly different from what is expected for a random walk. Thus, another conclusion drawn in paper 1 is that there is little evidence of mean reversion in short-term interest rates.

A maximum likelihood estimator for multifractal random walks

While the MSM model is attractive from a statistical point of view, it involves one extra parameter compared to the multifractal random walk (MRW) introduced by Bacry et al. (2001). The random measure in the MRW is constructed by randomizing the cascade in the MMAR model in both the time direction and in the "cascade direction", as opposed to the Poisson multifractal where there is randomization only in time. For a long time, application of the MRW model in stochastic modeling and practical applications (e.g. volatility forecasting) was hampered by the fact that exact likelihood computation is numerical infeasible unless one works with very short time series. In paper 2 we present an algorithm which solves this problem by using an approximation scheme. The first step is to use a Laplace approximation for the n-dimensional integral representing the likelihood (n typically ranges from 10^3 to 10^4). In many statistical problems, using a Laplace approximation offers a superior, in terms of computation speed, alternative to Markov Chain Monte Carlo (MCMC) methods (Rue et al., 2009). In particular, this way of approximating the likelihood has been proposed for simple stochastic volatility models (Skaug and Yu, 2009; Martino et al., 2011). In the Laplace approximation the Jacobian and Hessian of the full likelihood are also computed. In the SV model the Markovian property of the latent AR(1) process implies that these matrices are sparse, while in the MRW model these are dense. We therefore propose a second approximation which involves a truncation in the volatility dependency after τ time lags. The resulting matrices are band-diagonal with bandwidth equal to τ. Note that, a similar approach, i.e. truncating the dependency, is used in estimation of the Hurst exponent using the Haslett-Raftery method (Haslett and Raftery, 1989).

A first application of the approximated likelihood method is to estimate the parameters in the MRW model using maximum likelihood (ML). The model has three free parameters: the integral scale R determines the length of the cascade, i.e. the correlation length, in the time direction. The volatility clustering is determined by the intermittency parameter λ (in the limit $\lambda \to 0$ the MRW converges to a Brownian motion). Finally, the usual scale parameter σ determines the fluctuation level of the outer process. To benchmark the

ML method we compared its performance with the General Methods of Moments (GMM), which has been used in previous studies of the MRW model. Overall, the ML method performs much better (in terms of mean squared error).

The ML estimator for the MRW model is the best estimator (minimum variance) available at present.

Estimation of the integral scale parameter R turns out to be a difficult task. This was previously reported for the GMM estimator. In Muzy et al. (2013) an extension of the MRW model is presented, where the integral scale is removed from the model . We conjecture that an alternative is to make the likelihood smoother as a function of R. The idea comes from asymptotic theory where it is common to impose smoothness conditions on the likelihood, but this conjecture is not yet tested.

The real benefit of the ML estimator comes when one is presented with small sample sizes. This is demonstrated in paper A1. This work concerns high-frequency (tick-by-tick) data from the Oslo Stock Exchange. Using the MLE method for the MRW model we are able to extract meaningful estimates of the intermittency parameter for each month of trading. These estimates are in turn used to test the hypothesis of time-varying intermittency. In this paper we also verify the multifractality of the data using standard non-parametric techniques (due to smaller sample sizes, this is not an easy task when using daily data).

Another application, made possible by the algorithm presented in paper 2, is volatility forecasting, density forecasting and computation of other risk measures (e.g Value at Risk, Expected Shortfall and Unexpected Losses). All these statistics can be computed from the conditional pdf which is, using Bayes rule, proportional to the likelihood. Thus, if we perform a straight forward extension of the ML algorithm for the MRW model, we can compute density forecasts, volatility forecasts and other risk measures. This is the focus of paper A2, where we also apply these methods to real data. In particular, for volatility forecasts, we compare the results of the MRW model with the simple SV model of Taylor. For the smallest time scale (daily in this particular case) the forecasts are almost identical for the two models. However, when we increase the time scale, the forecasts based on these two models diverge. In the SV model the forecast will increase/decrease as a monotonic function of the lead time, while in the MRW model we observe a more complex pattern (e.g. it can increase up to 10 days and then decrease). A similar observation was made for the MSM model by Calvet and Fisher.

In paper 3 we study another financial market; electricity prices and specifically the Nord Pool spot prices. While these time series share some of the stylized facts described so far, namely volatility clustering and leptokurtic distributions, an empirical analysis actually reveals anti-correlations for the increments. Two parsimonious stochastic processes that reproduce this feature are the OU-process and the fractional Brownian motions (fBm) with Hurst exponents $H < 1/2$. While neither of these two alternatives are able to capture the volatility persistence and the strongly non-Gaussian distribution of returns, we can include these properties by combining the models with the MRW. The resulting stochastic

processes are named the damped MRW and fractional MRW. To estimate the parameters we reformulate the ML algorithm to take into account the anti-correlations.

Algorithms for maximum likelihood estimation for the damped and fractional Multifractal Random Walk models are developed.

After estimating the parameters, we perform a Monte Carlo study to see which model that best reproduces the observed statistics of the Nord Pool data. We find that the damped MRW model is the preferred model. This implies that there is a characteristic time scale. Here, one should note that there are several periodicities in the signal (daily, weakly and yearly). Since we consider signals sampled weakly, we are left with only the yearly oscillation. It is checked if this periodicity affects the analysis (i.e., if the choice of the OU process for the conditional mean dynamics is an artifact of the yearly periodicity?) and we conclude that this is not the case. Another interesting result in this paper is that the estimates of the intermittency parameter λ is higher than found in other financial markets (e.g. stock indices, FX and interest rates).

1.3 Statistics and climate

The last paper presented in this thesis (paper 4) concerns stochastic modeling of regional surface temperatures time series. The study is based on gridded temperature records for the time period 1900-2014, and we examine if the warming in the industrial period (after year 1900) can be explained as a realization of a stationary process (climate noise) or if a trend is needed to account for the observed temperature increase. Our hypothesis is that the climate noise has long-range dependence (LRD), and investigation of the LRD hypothesis is by itself of great interest in order to understand the "stylized facts" of climate data.

The idea of long memory in the climate system is not a new one, and many of the most important ideas date back to Harold Hurst's work with rescaled-range analysis on geophysical records in the 1950s. Hurst's discovery of LRD has been confirmed later, using more refined statistical methods, and similar characteristics are found in a wide range of different climatic time series. Nevertheless, there is also a scepticism among climate scientists. For instance, the latest IPCC report (Stocker et al., 2013, chapter 10) states that: *"Although the evidence for long-range dependence in global temperature data remains a topic of debate (Mann, 2011; Rea et al., 2011) ..."*. Unfortunately, the two studies that are cited are both flawed: Mann uses an AR(1) process to represent LRD (which is by definition wrong), while Rea et al. do not take into account that different estimators have different properties (table 1 in this article summarizes 12 estimates of the Hurst exponent, using 12 different methods).

Trend-significance testing with long-range dependent errors

To test the significance of trends in signals with LRD we need to take into account that there are larger pseudo trends for processes with Hurst exponents $H > 0.5$, compared to white noise processes ($H = 0.5$). In fact, for the linear trend model, the ordinary least-square (OLS) slope $\hat{A}_1$ is normal distributed with $\mathrm{sd}(\hat{A}_1) \sim n^{H-2}$. Since the Hurst exponent and scale parameter are estimated, we apply the small-sample correction proposed in Ko et al. (2008), to account for the parameter uncertainty.

To estimate the Hurst exponent we use the ML estimator, augmented with de-trended fluctuation analysis (DFA) of orders two and three, wavelet variance and variograms. The latter three methods give visual information about the scaling properties, and their application can be viewed as robustness tests for the ML estimates. For regional temperature records we observe that the estimated Hurst exponents using these four different methods coincide, when taking estimator uncertainty into account.

The estimates (the scale, trend and Hurst exponent) vary with latitude and longitude. For the Hurst exponents we find significantly higher values in the sea surface compared to over land areas. The trend estimates are all positive, except for a small region in the North-Atlantic. From the estimated trends, fluctuation levels and Hurst exponents we can compute the probability of a fGn producing trend estimates larger than the data estimate. The result is that $\sim 68\%$ ($\sim 47\%$) of the regional time series have signficant trends at the 5% (1%) significance level. The regions where we do not find significant trends are ENSO regions (the tropical Pacific and other regions that correlate highly with the ENSO) and the North-Atlantic. If we assume an AR(1) null hypothesis instead, then the result is that $\sim 94\%$ ($\sim 88\%$) of the time series have significant trends at the 5% (1%) significance level. Again, the locations where we do not find significant trends are mostly the ENSO regions and the North-Atlantic.

The numbers presented above show that significance tests depend crucially on the chosen null model. However, we know that an AR(1) model is wrong for time series where we have power law statistics, and vice versa for a fGn model. From an empirical analysis we find that some geographic regions show poor scaling properties. This finding has previously been reported in several studies (e.g. Huybers and Curry, 2006). To find all such regions we use the likelihood-ratio test statistic. If we at each grid point choose the null model preferred by the model selection test, we find that $\sim 82\%$ ($\sim 73\%$) of the time series have significant trends at the 5% (1%) level. While this result lies between the AR(1) and fGn results, it is closer to the AR(1) result. This reflects the fact that AR(1) is the preferred model in the ENSO region. Based on the previous discussion we conclude that for most regions:

> *Warming trends in surface temperatures, in the time period 1900-2014, are detected.*

The conclusion here should be viewed as a hypothesis, which needs to be further tested. In fact, in the same study (paper 4) we demonstrate that the existence of a LRD on time

scales longer than $\sim$ 20 years will change our conclusions about the significance of linear trends in many regions.

2
Stochastic processes

We review some important notions and mathematical objects central to this thesis. We define the class of self-similar and multifractal stochastic processes, and look at some characteristic properties of these. In papers 2 and 3 the multifractal random walk (MRW) plays a central role. In section 2.4 and 2.5 we review the more general class of infinitely divisible cascades (IDCs) of Bacry and Muzy (2003). A short description of the IDC construction is also given in paper 2.

We recall that a stochastic process $\{X(t), t \in \mathcal{T}\}$ is a family of random variables defined on the *same* probability space $(\Omega, \mathcal{F}, \mu)$. We consider univariate time series, so $X(t) : \Omega \mapsto \mathbb{R}$, where the real line is equipped with the Borel σ-algebra. Initially, we work with processes in continuous time, so the set $\mathcal{T}$ is the real line.

2.1 Correlations and the Hurst exponent

The variance of a cumulative sum of t uncorrelated and identically distributed random variables is proportional to t. As a generalization, consider a centered stochastic process $X(t)$ with stationary increments. If the variogram is a power law, then we define the Hurst exponent H by

$$\mathbb{E}X(t)^2 \propto t^{2H}. \tag{2.1}$$

Uncorrelated and identical distributed increments implies $H = 1/2$. However, many time series are well described by the scaling relation (2.1) with $H \neq 1/2$. We note that for stationary processes, with finite second moments, the variogram is constant. In this case, if the cumulative sum scales as a power law, one can associate the Hurst exponent of the cumulative sum with the stationary process. There are different conventions used in different fields of science, but it is usually clear from the context which definition of the Hurst exponent that is used.

The Hurst exponent determines the correlation at all time scales. Let us first consider the covariances of the process itself. We have

$$2X(t)X(s) = X(t)^2 + X(s)^2 - \{X(t) - X(s)\}^2.$$

The property of stationary increments implies

$$2X(t)X(s) \overset{\mathrm{d}}{\sim} X(t)^2 + X(s)^2 - X(|t-s|)^2,$$

from which we infer

$$\mathbb{E}X(t)X(s) = \frac{1}{2}\left\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right\} \tag{2.2}$$

From (2.2) the auto-correlation function of the increments follows. Defining $x_t = X(t) - X(t-1)$, for $\tau > 0$:

$$\rho(\tau) = \frac{\mathbb{E}x_1 x_{\tau+1}}{\mathbb{E}x_1^2} = \frac{1}{2}\left\{(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}\right\}. \tag{2.3}$$

Observe that $H = 1/2$ implies uncorrelated increments, while for $H \neq 1/2$ we have

$$\rho(\tau) \sim \frac{d^2}{d\tau^2} t^{2H} = 2H(H-1)\tau^{2H-2},$$

as $\tau \to \infty$. Thus, $H \neq 1/2$ implies dependent increments. Choosing $0 < H < 1/2$ results in negatively correlated increments, while for $H > 1/2$ the increments are persistent.

Moreover, in the persistent case, the ACF decays so slowly that the series $\sum_{\tau=-\infty}^{\infty} \rho(\tau)$ diverges. Two classes of stochastic processes with well-defined Hurst exponents are the self-similar and the multifractal processes with existing second moments. The Ornstein-Uhlenbeck process, defined as the solution to stochastic differential equation

$$dX(t) = -\nu X(t)dt + dB(t),$$

where $B(t)$ is Brownian motion and $\nu > 0$, does not satisfy the scaling relation (2.1). However, an Ornstein-Uhlenbeck process scales asymptotically. When $\nu \to 0$, $X(t)$ converges to a Brownian motion, i.e. $H = 1/2$, and as $\nu \to \infty$ the process $X(t)$ is a Gaussian white noise.

2.2 Self-similarity

If we assume that the structure functions $\mathbb{E}|X(t)|^q$ are power laws in t one can define a scaling function $\zeta(q)$ (also known as the zeta-function) by the relation

$$\mathbb{E}|X(t)|^q \propto t^{\zeta(q)}.$$

Here, we usually assume that the relation above holds for all q-values for which the qth moments are finite, and in most examples we will assume finite moments on a half-infinite interval, e.g. $q \in (-1, \infty)$. The zeta-function is always concave (this follows from Hölder's inequality), and a strictly concave scaling function is the property that distinguishes a multifractal from a selfsimilar process.

A stochastic process X(t) is said to be self-similar (or self-affine) if

$$\forall a > 0 : X(at) \stackrel{\mathrm{d}}{=} a^h X(t) \tag{2.4}$$

Notation 1 *The symbol $\stackrel{\mathrm{d}}{=}$ means equality in distribution for stochastic processes, i.e. all finite dimensional marginals coincide, while $\stackrel{\mathrm{d}}{\sim}$ denotes equality in distribution for random variables. We drop the brackets, and denote a stochastic process $\{X(t)\}$ simply as $X(t)$. This notation does not distinguish between a stochastic process and its one dimensional marginals (the random variables $X(t)$), but to which object we are referring should be clear from the context.*

Condition (2.4) defines a large class of stochastic processes, e.g. fractional Brownian motions, fractional Lévy flights and Hermitte-Rosenblatt processes.. From (2.4) it is easily seen that $X(t) \stackrel{\mathrm{d}}{\sim} t^h X(1)$. This implies

$$\mathbb{E}|X(t)|^q \propto t^{hq}, \quad q < \alpha = \sup\{\tau : \mathbb{E}|X(1)|^\tau < \infty\},$$

and, assuming that the probability density functions (pdfs) exist:

$$\forall x \in \mathbb{R} : t^\nu p_{X(t)}\left(x\, t^\nu\right) = p_{X(1)}(x), \tag{2.5}$$

with $\nu = h$. Selfsimilarity implies that the zeta-function is linear:

$$\zeta(q) = hq, \quad q < \alpha.$$

It is seen that if $\alpha \geq 2$, then the Hurst exponent equals the selfsimilarity exponent. From (2.5) we see that the pdfs at different time scales coincide under a suitable rescaling. Note that for a selfsimilar process, the pdfs evaluated at origo is a power-law function in time-scale:

$$p_{X(t)}(0) \sim t^{-\nu}.$$

The properties discussed above are illustrated in figure (2.1).

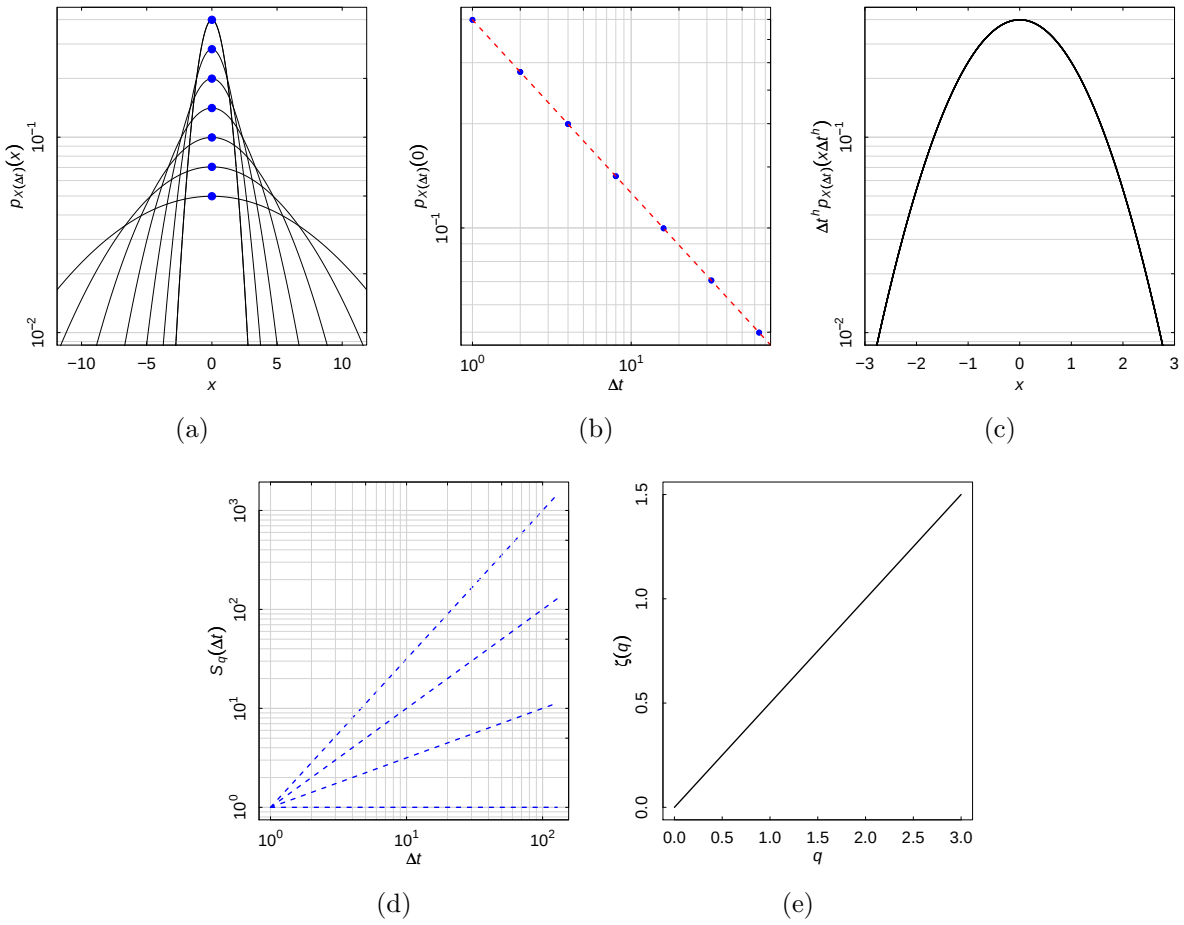


Figure 2.1: Properties of a h-selfsimilar processes X(t): a) Densities of $X(\Delta t)$ at several time scales Δt. b) The power law $p_{X(\Delta t)}(0) \sim \Delta t^{-h}$. c) Rescaled densities $\Delta t^h p_{X(\Delta t)}(x\Delta t^h)$ coincide. d) Structurfunctions $S_q(\Delta t) = \mathbb{E}|X(\Delta t)|^q$ are power laws. e) Linear scaling function $\zeta(q)$.

If we, in addition to self-similarity and stationary increments, requires that the process is Gaussian the result is a fractional Brownian motion (fBm). Alternatively, we can define

a fBm $B_H(t)$ directly as a centered Gaussian process with covariances:

$$\mathbb{E}B_H(t)B_H(s) = \frac{\sigma^2}{2}\left\{|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right\}$$

where $\sigma > 0$ determines the fluctuation level, and $H \in (0,1)$ is the Hurst exponent. The increments of a fractional Brownian motion is known as fractional Gaussian noise (fGn),

2.3 Multifractals

In this thesis we refer to multifractal stochastic processes as processes with well-defined and strictly concave scaling function $\zeta(q)$. For financial time series two stylized facts are uncorrelated and strongly dependent increments, where the dependence is in the volatility rather than the directions of the fluctuations. A process with a strictly concave scaling function can explain the phenomena of volatility clustering, while we can incorporate the stylized fact of uncorrelated returns by imposing the restriction $\zeta(2) = 1$, i.e. a Hurst exponent $H = 1/2$. Some properties of multifractal processes are illustrated in figure (2.2).

2.4 Infinitely divisible multifractal measures

A rigorous treatment of the IDC construction is given in Bacry and Muzy (2003). A special case known as the log-normal multifractal random walk was introduced in Bacry et al. (2001).

Denote by S^+ the upper half plane $\{(x,y) \in \mathbb{R}^2 | y > 0\}$ equipped with the measure $d\mu(dt, dr) = r^{-2}dtdr$. On the measure space (S^+, μ) we define a random measure P. We assume P to be independently scattered and infinitely divisible. We recall that a random variable is infinitely divisible if, for all $n \in \mathbb{N}$, it is equal in distribution to a sum of n iid random variables.

Definition 1 *A real-valued random variable X is infinitely divisible if there for all $n \in \mathbb{N}$ exists a random variable $X^{(1/n)}$ such that*

$$X \overset{\mathrm{d}}{\sim} X_1^{(1/n)} + \ldots + X_n^{(1/n)},$$

where $X_1^{(1/n)}, \ldots X_n^{(1/n)}$ are independent copies of $X^{(1/n)}$.

An infinite divisible variable can be charachertrized by a measure ν, known as the Lévy measure, together with the location and scale parameters a and λ:

Theorem 1 (Lévy Khintchine) *A random variable X infinitely divisible if and only if it can be written on the form*

$$\phi_X(s) = \exp(\varphi(s)),$$

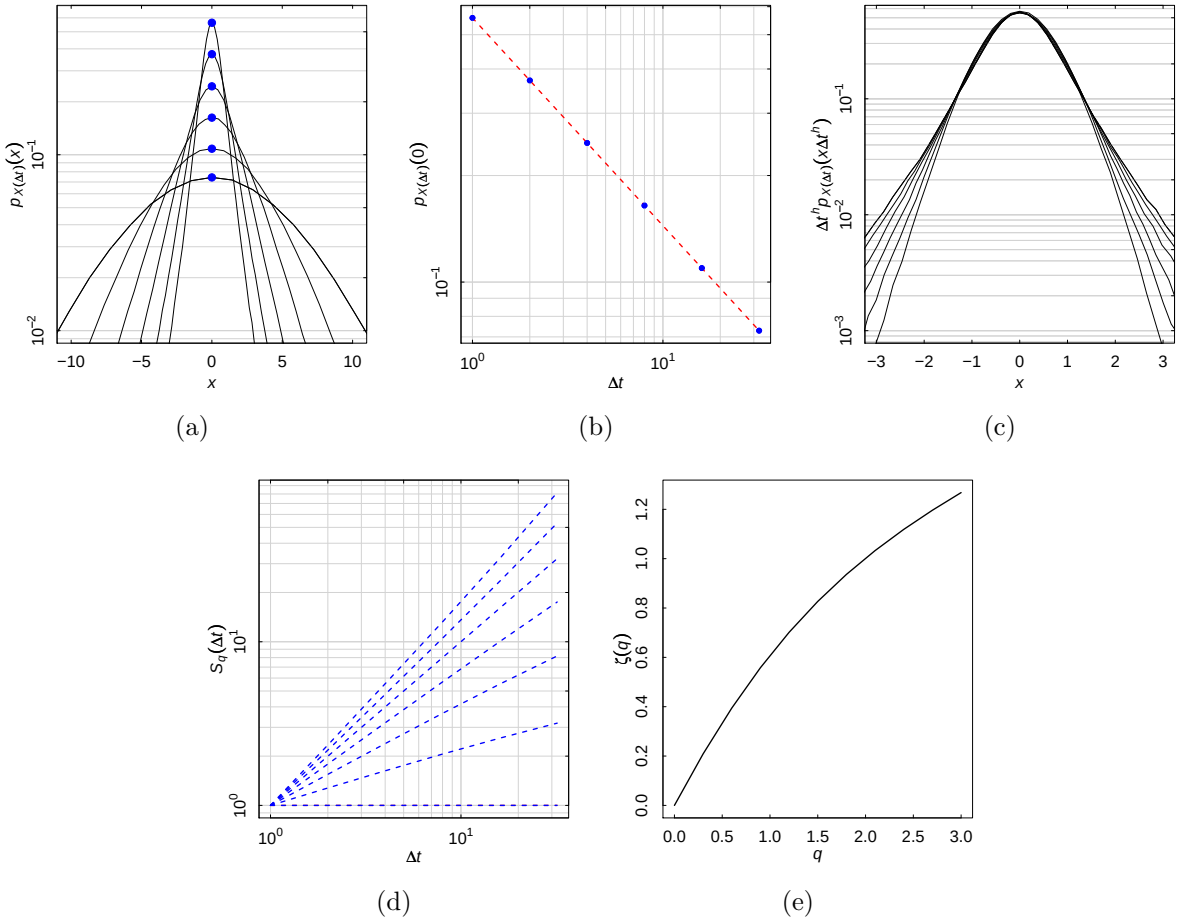


Figure 2.2: Properties of a multifractal processes X(t): a) Densities of $X(\Delta t)$ at several time scales Δt. b) The power law $p_{X(\Delta t)}(0) \sim \Delta t^{-h}$. c) Rescaled densities $\Delta t^h p_{X(\Delta t)}(x\Delta t^h)$ do not coincide. d) Structure functions $S_q(\Delta t) = \mathbb{E}|X(\Delta t)|^q$ are power-laws. e) Non-linear scaling function $\zeta(q)$.

with

$$\varphi(s) = aqi - \frac{\lambda^2 q^2}{2} + \int_{|x|\geq 1} (\exp(iqx) - 1) d\nu(x) + \int_{|x|<1} (\exp(iqx) - 1 - iqx) d\nu(x),$$

$\lambda := \sqrt{\lambda^2} > 0$, $a \in \mathbb{R}$, $\nu(\{0\}) = 0$ *and* $\int \min\{1, x^2\} d\nu(x) < \infty$.

Definition 2 *i)P is independently scattered if, for any disjoint sets* $\mathcal{A}_1, \ldots, \mathcal{A}_n \subseteq S^+$, *the random variables*

$$P(\mathcal{A}_1), \ldots, P(\mathcal{A}_n)$$

are independent.
ii) P is infinite divisible if for any μ*-measurable set* $\mathcal{A} \subseteq S^+$, $P(\mathcal{A})$ *is infinite divisible with*

characteristic function

$$\phi_{P(\mathcal{A})}(s) = \exp\{\varphi(s)\mu(\mathcal{A})\}, \tag{2.6}$$

Definition 3 *For all $r > 0$ with r fixed we define a stochastic process $h_r(t)$ by*

$$h_r(t) = P(\mathcal{A}_r(t)),$$

where

$$\mathcal{A}_r(t) = \left\{(x, y) \in S^+ | y \geq r, |x - t| \leq \frac{\min\{y, R\}}{2}\right\}, \tag{2.7}$$

and $R > 0$ is a parameter in the model.

Definition 4 *Let $dM_r(dt) = \exp(h_r(t))dt$, meaning that*

$$M_r(I) = \int_I \exp(h_r(t))dt \tag{2.8}$$

for all Lebesgue-measureable sets I, and

$$\varphi(-i) = 0. \tag{2.9}$$

The multifractal random measure M is defined as the limit

$$dM(dt) = \lim_{r\to 0} dM_r(dt).$$

The condition (2.9) implies that $\mathbb{E}\exp[h_r(t)] = 1$, which in turn can be used to prove the existence of the multifractal random measure. The choice of cone-like domains in (2.7) will lead to the following exact stochastic scale-invariance:

$$M([0, at]) \overset{\mathrm{d}}{=} W_a M([0, t]). \tag{2.10}$$

Note the generality in the construction. We are free to choose any log-infinitely divisible random variable, or equivalently specify any Lévy measure, under the constraint that the first moment exists. In the sequel we verify (2.10), and for this purpose it is useful to calculate the integrals in $\mu\left(\mathcal{A}_r(t) \cap \mathcal{A}_r(t + \tau)\right)$. This gives the identity, for $\tau > 0$:

$$\rho_r(\tau) := \mu\left(\mathcal{A}_r(t) \cap \mathcal{A}_r(t + \tau)\right) = \begin{cases} \log\frac{R}{r} + 1 - \frac{\tau}{r} & \text{if } \tau \leq r, \\ \log\frac{R}{r} & \text{if } r < \tau \leq R, \\ 0 & \text{if } \tau > R, \end{cases}$$

and $\rho_r(-\tau) = \rho_r(\tau)$.

We consider first the particular case where $\varphi(s)$ is the characteristic function of a normal random variable. In this case $h_r(t)$ are Gaussian processes, and the limiting measures are known as the log-normal multifractal random measures. Condition (2.9) implies that

$$\varphi(q) = -\frac{\lambda^2}{2} q(i + q),$$

which together with (2.6) can be used to find expressions for the mean and covariances of $h_r(t)$. For this purpose, we divide $\mathcal{A}_r(t)$ and $\mathcal{A}_r(s)$ into disjoint sets which enables us to factorize the expectation:

$$\begin{aligned}
&\mathbb{E}\exp[iP(\mathcal{A}_r(t))q_1 + iP(\mathcal{A}_r(s))q_2] \\
&= \mathbb{E}\exp[iP(\{\mathcal{A}_r(t) \setminus \mathcal{A}_r(s)\} \cup \{\mathcal{A}_r(t) \cap \mathcal{A}_r(s)\})q_1 + \\
&\quad iP(\{\mathcal{A}_r(s) \setminus \mathcal{A}_r(t)\} \cup \{\mathcal{A}_r(s) \cap \mathcal{A}_r(t)\})q_2] \\
&= \mathbb{E}\exp[iP(\mathcal{A}_r(t) \setminus \mathcal{A}_r(s))q_1 + \\
&\quad iP(\mathcal{A}_r(s) \setminus \mathcal{A}_r(t))q_2 + iP(A_r(t) \cap \mathcal{A}_r(s)\})(q_1+q_2)] \\
&= \mathbb{E}\exp[iP(\mathcal{A}_r(t) \setminus \mathcal{A}_r(s))q_1] \times \\
&\quad \mathbb{E}\exp[iP(\mathcal{A}_r(s) \setminus \mathcal{A}_r(t))q_2]\mathbb{E}\exp[iP(A_r(t) \cap \mathcal{A}_r(s)\})(q_1+q_2)] \\
&= \exp(\varphi(q_1)\rho_r(0) + [\varphi(q_1+q_2) - \varphi(q_1) - \varphi(q_2)]\rho_r(|s-t|) + \varphi(q_1)\rho_r(0)) \\
&= \exp\left(\rho_r(0)\left[-i\lambda^2/2(q_1+q_2) - \lambda^2/2\left(q_1^2+q_2^2\right)\right] - \rho_r(|s-t|)\lambda^2\right), \qquad (2.11)
\end{aligned}$$

This gives

$$\forall t : \mathbb{E}h_r(t) = -\frac{\lambda^2}{2}\rho_r(0)$$

and

$$\forall (t,s) : \mathrm{Cov}(h_r(t), h_r(s)) = \lambda^2\rho_r(|s-t|).$$

Since the first and second moments are independent of time, $h_r(t)$ are covariance stationary, and combined with the fact that $h_r(t)$ are Gaussian processes, it follows that they are stationarity. If we assume the following scale invariance:

$$M_{ar}([0,at]) \overset{\mathrm{d}}{=} W_a M_r([0,t]), \qquad (2.12)$$

then passing to the limit $r \to 0$ we obtain (2.10). If we assume (2.12), then

$$\begin{array}{c}
M_{ar}([0,at]) \overset{\mathrm{d}}{=} W_a M_r([0,t]) \\
\Updownarrow \\
\displaystyle\int_0^{at} \exp(h_{ar}(t'))dt' \overset{\mathrm{d}}{=} W_a \int_0^t \exp(h_r(t'))dt' \\
\Updownarrow \\
\displaystyle a\int_0^{t} \exp(h_{ar}(at'))dt' \overset{\mathrm{d}}{=} W_a \int_0^t \exp(h_r(t'))dt' \\
\Uparrow \\
h_{ar}(at) \overset{\mathrm{d}}{=} \Omega_a + h_r(t),
\end{array} \qquad (2.13)$$

with $W_a = a\exp(\Omega_a)$. Choosing Ω_a to be normal and independent of $h_r(t)$, implies that both processes in (2.13) are Gaussian. Thus, to prove (2.13), upon specifying the moments of Ω_a, we only need to verify that the means and covariances coincide. Taking expectations,

we see that $\mathbb{E}\,\Omega_a = \lambda/2 \log a$ leads to equality in means. To compare covariances, observe that

$$\rho_{ar}(a\tau) = -\log a + \rho_r(\tau), \quad 0 < \tau < R, a < 1, \tag{2.14}$$

and thus, the choice $\mathrm{var}(\Omega_a) = -\lambda^2 \log a$ leads to the same second-order statistics of the two stochastic processes. This completes the proof of (2.13), (2.12) and (2.10). Note that the Gaussian requirement was only used to verify (2.13). Thus, to generalize the proof to the case of an arbitrary Lévy measure with the constraint (2.9), we only need to deduce the distribution of Ω_a and show (2.13). This is achieved by comparing characteristic functions. This can be done using the following identity:

$$\mathbb{E} \exp\left[i \sum_{k=1}^{N} P(\mathcal{A}_r(t_k)) q_k \right] = \exp\left[i \sum_{k=1}^{N} \sum_{j=1}^{k} \alpha(k,j) \rho_r(|t_k - t_j|) \right], \tag{2.15}$$

where

$$\alpha(j,k) = \varphi(r_{k,j}) + \varphi(r_{k+1,j-1}) - \varphi(r_{k,j-1}) - \varphi(r_{k+1,j})$$

and

$$r_{k,j} = \begin{cases} \sum_{m=k}^{j} q_m & \text{if } k \le j \\ 0 & \text{otherwise} \end{cases}.$$

This is derived in (2.11) for the case $N = 2$ (for $N \in \mathbb{N}$ a proof can be found in Bacry and Muzy (2003, Appendix A)). Let us assume that Ω_a is independent of the process $h_r(t)$. Inserting (2.15) into (2.13) we see, with the help of (2.14), that (2.13) indeed holds if Ω_a is distributed according to

$$\mathbb{E} \exp(i\,\Omega_a q) = \exp(-\varphi(q) \log a). \tag{2.16}$$

Thus, we have shown the following theorem:

Theorem 2 (Exact scaling of MRM) *Let*

$$W_a = a \exp(\Omega_a),$$

where Ω_a is the random variable defined in (2.16)*. The multifractal random measure satisfy the stochastic scale-invariance*

$$\forall a \in (0,1) : M([0,at]) \overset{\mathrm{d}}{=} W_a M([0,t]), \quad 0 < t < R, \tag{2.17}$$

with W_a independent of the process M(t).

The exact scaling of the MRM implies that $M([0,t]) \overset{\mathrm{d}}{\sim} W_{t/R} M([0,R])$, for $0 < t < R$, from which we can easily calculate the q-order moments.

Theorem 3 (Moments of MRM) *Define* $\alpha = \sup\{\tau : \mathbb{E}M([0,R])|^\tau < \infty\}$. *The q-order moments of the multifractal random measure, with* $0 \leq q < \alpha$, *are given by*

$$\mathbb{E}M([0,t])^q = c_q t^{\zeta(q)},$$

with

$$\zeta(q) = q - \varphi(-iq)$$

and

$$c_q = \mathbb{E}M([0,R])^q \, R^{-\zeta(q)}.$$

In particular, note that $\mathbb{E}M([0,t]) = t$. From a given Lévy-measure we can study the zeta-function. For the log-normal MRM measure the scaling function takes the quadratic form $\zeta(q) = q(1+\lambda^2/2) - q^2\lambda^2/2$ (here all non-negative moments exists, i.e. $\alpha = \infty$).

2.5 Multifracal random walks

Let $\Theta(t) = M([0,t])$ be the distribution function of the multifractal random measure, and $B(t)$ a Brownian motion independent of $\Theta(t)$. A multifractal random walk (MRW) $X(t)$ is defined by

$$X(t) = B(\Theta(t)).$$

Let $\tilde{M}_r([0,t])$ be the measures defined in (2.8) and $\tilde{B}(t)$ a Brownian motion independent of the processes $\tilde{M}_r([0,t])$. A MRW has the Wiener-integral representation

$$\tilde{X}(t) = \lim_{r\to 0} \int_0^t \tilde{M}_r([0,t]) d\tilde{B}(t),$$

in the sense that $\tilde{X}(t) \stackrel{\mathrm{d}}{=} X(t)$. Theorem 2, combined with the selfsimilarity of the Brownian motion, implies the scale invariance

$$X(at) = W_a^{1/2} X(t),$$

for $t \leq R$ and $0 < a < 1$. The scaling function $\zeta_X(q)$ for the MRW process is given by

$$\zeta_X(q) = \zeta_\Theta(q/2),$$

where $\zeta_\Theta(q)$ is the scaling function for the distribution function of the corresponding multifractal random measure. For the log-normal MRW, we have

$$\zeta_X(q) = q/2(1+\lambda^2/2) - q^2\lambda^2/8.$$

3 Statistical inference LRD

In this chapter we describe the statistical methods that are applied in paper 4. In particular, we consider significance testing of linear trends under a fractional Gaussian noise (fGn) null hypothesis. A more general treatment of statistical inference for long-range dependent processes can be found in Palma (2007) and Beran et al. (2013) .

3.1 Linear trend model with fGn

Consider n observations from the linear trend model

$$Y_t = a_0 + a_1 t + X_t, \tag{3.1}$$

where the error term process X_t is a fGn with scale parameter $\text{var}(X_1) = \sigma^2$ and Hurst exponent H. The random vector $X = (X_1, \ldots, X_n)^T$ is multivariate normal distributed

$$X \sim \mathcal{N}(0, \Gamma),$$

where the $(n \times n)$ covariance matrix Γ is the Toeplitz matrix of the auto-covariances $(\gamma(0), \ldots, \gamma(n-1))$. The Toeplitz property means that the elements (i, j) of Γ are on the form $\gamma(|i-j|)$. Denote by R_H the correlation matrix of X and note that $\Gamma = \sigma^2 R_H$. It is convenient to write the linear trend model on vector form:

$$Y = B^T a + X,$$

where $a = (a_0, a_1)^T$ and B is the $(2 \times n)$ design matrix with 1s on the first row and the sampling times $(1, 2, \ldots, n)$ as the second row.

3.1.1 Ordinary least squares

The ordinary least square (OLS) estimator of a can then be written as

$$\hat{A} \stackrel{\text{def}}{=} \left(\hat{A}_0, \hat{A}_1\right)^T = \left(BB^T\right)^{-1} BY\,. \tag{3.2}$$

Since Gaussian vectors are closed under linear transformations, the estimator $\hat{A}$ is (bivariate) normal distributed. This observation is useful for more complex mean specifications since linear estimators of the mean, e.g. weighted least squares and OLS, will be normal distributed as long as the mean specification is linear in the parameters. An example is a nth order polynomial trend.

Computation of the first two moments shows that the OLS estimator is unbiased, i.e. $\mathbb{E}\hat{A} = a$, with covariance matrix

$$\mathrm{Cov}(\hat{A}) = C(H)\sigma^2, \tag{3.3}$$

with

$$C(H) = \left(BB^T\right)^{-1} BR_H B^T \left(BB^T\right)^{-1}.$$

If we define $c(H)$ to be element (2,2) of the correlation matrix $C(H)$, then the estimator for the slope is distributed as $\hat{A}_1 \stackrel{\text{d}}{\sim} \mathcal{N}\left(a_1, \sigma^2 c(H)\right)$, i.e.

$$T_{a_1}(H, \sigma, \hat{A}_1) \stackrel{\text{def}}{=} \frac{\hat{A}_1 - a_1}{\sigma c(H)^{1/2}} \stackrel{\text{d}}{\sim} \mathcal{N}(0,1). \tag{3.4}$$

A closed-form expression for the factor $c(H)$ is given by (Lee and Lund, 2004)

$$c(H) = 1 + 2\frac{\sum r_H(j) w_j}{\sum (t - \bar{t})^2}, \tag{3.5}$$

where $\bar{t} = (n+1)/2$ and

$$w_j = (n-j)\frac{n^2 - 2jn - 2j^2 - 1}{n(n+1)(n-1)}.$$

We also have the asymptotic result (e.g. Baillie and Chung, 2002)

$$c(H)^{1/2} \sim n^{H-2}.$$

3.1.2 BLUE

For Gaussian white noise the maximum likelihood (ML) and OLS estimator coincide. Moreover, they have the lowest possible variance in the class of unbiased estimators. However, for $H \neq 1/2$ the OLS estimator is no longer the best linear unbiased estimator (BLUE), while the MLE *is*. Here, the maximum likelihood estimator takes the form

$$\tilde{A} = C_{\mathrm{ML}}(H) B R_H^{-1} Y, \quad C_{\mathrm{ML}}(H) = \left(BR_H^{-1}B^T\right)^{-1}. \tag{3.6}$$

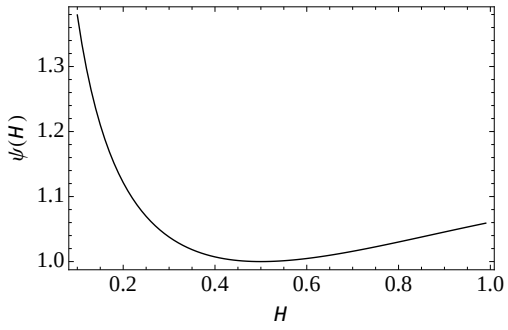


Figure 3.1: The factor $\psi(H)$ is the increased uncertainty, measured by the standard deviation, in the slope coeffisient by using OLS compared to MLE. Sample length is $n = 2000$ in this figure.

As for OLS the MLE of a is unbiased and Gaussian, but the maximum likelihood method takes into account the correlations of the noise process in estimating the parameters. The covariance matrix of $\tilde{A}$ is given by $\mathrm{Cov}(\tilde{A}) = C_{\mathrm{ML}}(H)\sigma^2$, and in particular $\mathrm{var}(\tilde{A}_1) = \sigma^2 c_{\mathrm{ML}}(H)$, where $c_{\mathrm{ML}}(H)$ is element (2,2) of the correlation matrix $C_{\mathrm{ML}}(H)$. By using OLS, compared to ML, the uncertainty in the slope coefficient is increased by the factor

$$\psi(H) = \left(\frac{c(H)}{c_{\mathrm{ML}}(H)} \right)^{1/2}. \tag{3.7}$$

The function $\psi(H)$ is plotted in figure 3.1. The sample length is chosen to be $n = 2000$ and the function is evaluated for $0.01 \leq H \leq 0.99$. Since the OLS and MLE coincide for white noise we have $\psi(1/2) = 1$. For $H > 1/2$ we observe that $\psi(H)$ is increasing monotonically, and the value at the endpoint is $\psi(0.99) \simeq 1.06$. When the noise process is anti-persistent, the difference in efficiency is increased for $H \ll 0.5$ compared to the long-range dependent (LRD) case. Since our primary concern is $H > 1/2$, there is, in our opinion, little efficiency gained from using the computational more demanding ML estimator for the slope parameter. Hence, we will restrict our attention to the OLS estimator.

3.1.3 MLE scale and Hurst exponent

To estimate the Hurst-exponent and scale parameter we use the maximum likelihood (ML) method. For the random fGn vector X the likelihood $L(x|H, \sigma)$ is given by

$$\log L(H, \sigma|x) = -n \log \sigma - \frac{1}{2} \log |R_H| - \frac{1}{2\sigma^2} x^T R_H^{-1} x.$$

The ML estimates are defined as the global maxima of L, and it is easy to see that the ML estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} x^T R_{\hat{h}}^{-1} x, \tag{3.8}$$

where $\hat{h}$ is the ML estimate of the Hurst exponent:

$$\hat{h} = \arg\max_H \left\{ -\frac{n}{2} \log x^T R_H^{-1} x - \frac{1}{2} \log |R_H| \right\}.$$

Typically $\hat{h}$ is found by a numerical optimization. In Dahlhaus (1989, 2006) it is shown that the ML estimator $\hat{H}$ is asymptotically efficient .

A small-sample Monte Carlo study of the ML estimator $\hat{H}$ is summarized in table 3.1. Here we have considered three mean specifications: a constant mean, linear and quadratic trend. The table summarizes the bias and uncertainty. An important observation is that the ML estimator is very accurate for small sample lengths. We also note that the bias can be corrected by using that $H \mapsto \mathbb{E}\hat{H} = f(H)$ is strictly increasing and thus invertible. A corrected estimator $\hat{H}_{\mathrm{c}}$ is found by the inverse map

$$\hat{H}_{\mathrm{c}} = f^{-1}(\hat{H}).$$

3.1.4 Trend significance

Consider again the linear trend model (3.1). Based on the sample y we are interested in rejecting one of the following hypotheses:

$$\mathrm{H}_0 : a_1 = 0$$
$$\mathrm{H}_A : a_1 \neq 0$$

The significance of the estimated slope can be assessed by means of p-values. We can think of these as the probability of a fGn, with scale parameter σ and Hurst exponent H, having trend estimates larger than the observed estimate $\hat{a}_1$. More precisely, using the test statistic in (3.4), we have

$$p = 2\mathbb{P}(T_0(H, \sigma, \hat{A}_1) > |t_0(H, \sigma, \hat{a}_1)|) = 2(1 - \Phi(|t_0(H, \sigma, \hat{a}_1)|), \tag{3.9}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Replacing (H, σ) with estimators $(\hat{H}, \hat{\Sigma})$ implies that the test statistic

$$T_0(\hat{H}, \hat{\Sigma}, \hat{A}_1)$$

is not normal. A similar situation arises if we assume $H = 0.5$, in which case it is well known that the resulting distribution is student-t. However, the distribution of $T_0(\hat{H}, \hat{\Sigma}, \hat{A}_1)$ is unknown and depends on H in a non-trivial way. A possible approximation is simply to use the normal distribution. This approach can be justified based on asymptotic properties. If $\{(\hat{H}, \hat{\Sigma})\}_n$ is a consistent sequence of estimators, then as $n \to \infty$, the test statistic T converges to a standard normal distribution. To take into account parameter uncertainty one can use the bootstrap method, see e.g. Bølviken (2014).

H	const	lin	quad
		$n = 200$	
0.5	0.49 (0.047)	0.48 (0.049)	0.47 (0.05)
0.6	0.59 (0.048)	0.57 (0.05)	0.56 (0.05)
0.7	0.68 (0.049)	0.67 (0.053)	0.66 (0.052)
0.75	0.73 (0.05)	0.72 (0.051)	0.71 (0.055)
0.8	0.78 (0.051)	0.77 (0.052)	0.76 (0.053)
0.85	0.83 (0.048)	0.82 (0.05)	0.8 (0.052)
0.9	0.87 (0.045)	0.86 (0.049)	0.85 (0.05)
0.95	0.92 (0.041)	0.9 (0.044)	0.9 (0.047)
0.99	0.94 (0.033)	0.94 (0.037)	0.92 (0.042)
		$n = 500$	
0.5	0.49 (0.028)	0.49 (0.029)	0.49 (0.03)
0.6	0.59 (0.029)	0.59 (0.03)	0.58 (0.031)
0.7	0.69 (0.031)	0.69 (0.031)	0.68 (0.031)
0.75	0.74 (0.03)	0.74 (0.031)	0.73 (0.032)
0.8	0.79 (0.03)	0.79 (0.032)	0.78 (0.032)
0.85	0.84 (0.031)	0.84 (0.03)	0.83 (0.031)
0.9	0.89 (0.03)	0.88 (0.032)	0.88 (0.031)
0.95	0.93 (0.027)	0.93 (0.028)	0.93 (0.029)
0.99	0.96 (0.02)	0.96 (0.022)	0.96 (0.023)
		$n = 1000$	
0.5	0.5 (0.02)	0.49 (0.02)	0.49 (0.021)
0.6	0.6 (0.021)	0.59 (0.021)	0.59 (0.021)
0.7	0.7 (0.021)	0.69 (0.022)	0.69 (0.022)
0.75	0.75 (0.021)	0.74 (0.021)	0.74 (0.022)
0.8	0.8 (0.022)	0.79 (0.022)	0.79 (0.022)
0.85	0.84 (0.021)	0.84 (0.021)	0.84 (0.022)
0.9	0.89 (0.021)	0.89 (0.021)	0.89 (0.021)
0.95	0.94 (0.02)	0.94 (0.021)	0.94 (0.021)
0.99	0.97 (0.014)	0.97 (0.015)	0.97 (0.016)

Table 3.1: Mean value for maximum likelihood estimates of the Hurst-exponent. In parentheses the standard deviations. In the columns *const*, *lin* and *quad* the sample mean, a linear and quadratic trend, respectively, is subtracted before estimating the Hurst exponent. The data-generating process is fGn with sample length n and Hurst exponent H listed in the first column. The number of Monte Carlo runs is $n_{\text{mc}} = 2000$. Note that the estimator is highly skewed for H close to 1.

Another alternative is the small-sample correction proposed in Ko et al. (2008). In this case $\hat{H}$ is a biased-corrected estimate of the Hurst exponent, and $\hat{\Sigma}$ is the corresponding ML estimate. This means that we are using (3.8) with the bias-corrected Hurst exponent.

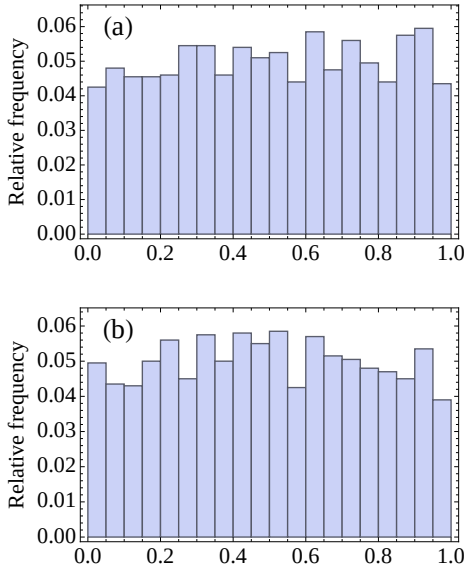


Figure 3.2: a)p-values of slope parameter. The data-generating process is a fGn with sample length $n = 340$ and Hurst-exponent $H = 0.85$. The number of Monte Carlo runs is $n_{\mathrm{mc}} = 2000$. b) $n_{\mathrm{mc}} = 2000$ realizations from the uniform $[0, 1]$ distribution.

The p-values are now computed as

$$p = 2\mathbb{P}(T_0(\hat{H}, \hat{\Sigma}, \hat{A}_1) > |t_0(\hat{h}, \hat{\sigma}, \hat{a}_1)|) \simeq 2(1 - \Phi(|t_0(\hat{h}, \hat{\sigma}, \hat{a}_1)|, \hat{n}_e - 2), \tag{3.10}$$

where $\Phi(\cdot, n)$ is the cumulative student-t distribution with n degrees of freedom, and

$$\hat{n}_e = n\frac{c(1/2)}{c(\hat{h})} \tag{3.11}$$

In Ko et al. (2008) the error term model is a fractional differenced noise, while we use a fGn model. A small-sample Monte Carlo study, summarized in table 3.2 verifies that this methodology also works for fGns. In table we 3.2 report the results in terms of confidence intervals. Alternatively, we can use the result that p-values are uniformly distributed on $[0, 1]$, or approximately so (Casella and Berger, 2002). An example is shown in figure 3.2, which again confirms the method.

3.2 Noise misspecification and robustness

3.2.1 Semi-heavy tails

A first guard against model misspecification is a comprehensive empirical analysis, which will typically restrict the class of stochastic models we use. Nevertheless, there is a likelihood of missing, or ignoring, some features in the data in a preliminary analysis. Assume

Maximum likelihood

	$n = 200$				$n = 500$				$n = 1000$			
H	0.8	0.9	0.95	0.99	0.8	0.9	0.95	0.99	0.8	0.9	0.95	0.99
0.5	0.79	0.90	0.95	0.99	0.79	0.89	0.94	0.98	0.80	0.90	0.95	0.99
0.55	0.77	0.88	0.93	0.98	0.80	0.89	0.95	0.99	0.80	0.89	0.94	0.98
0.6	0.81	0.89	0.94	0.98	0.80	0.89	0.94	0.98	0.82	0.91	0.96	0.99
0.65	0.79	0.89	0.94	0.98	0.80	0.89	0.94	0.99	0.81	0.90	0.95	0.99
0.7	0.80	0.90	0.94	0.99	0.79	0.89	0.94	0.98	0.80	0.90	0.95	0.99
0.75	0.80	0.90	0.94	0.99	0.81	0.90	0.94	0.99	0.81	0.91	0.95	0.99
0.8	0.81	0.90	0.94	0.99	0.80	0.90	0.95	0.99	0.81	0.91	0.95	0.99
0.85	0.80	0.90	0.95	0.99	0.81	0.91	0.95	0.99	0.82	0.91	0.95	0.99
0.9	0.81	0.90	0.95	0.99	0.82	0.91	0.95	0.99	0.80	0.91	0.95	0.99
0.95	0.80	0.90	0.95	0.99	0.82	0.92	0.96	1.00	0.81	0.91	0.95	0.99
0.99	0.77	0.88	0.94	0.99	0.78	0.88	0.94	0.98	0.79	0.89	0.94	0.98

Table 3.2: Coverage probabilities of slope parameter for nominal confidence levels 80%, 90%, 95% and 99% and various Hurst exponents (1. column). The data-generating process is a fGn with sample length $n = 200$. The number of Monte Carlo runs is $n_{\text{mc}} = 2000$ for each n and H.

that the second-order statistics are power laws, but the 1-d pdfs have heavier tails than the normal distribution. For instance we can consider student-t distributed innovations with ν degrees of freedom. Denote by X_t a fGn and $\mathcal{F}_t = \Sigma(\{X_s\}_{s=1}^t)$, with $\mathcal{F}_0 = \emptyset$, the natural filtration generated by the process. If we define $m_t = \mathbb{E}[X_t|\mathcal{F}_{t-1}]$ and $P_t = \mathrm{var}(X_t|\mathcal{F}_{t-1})$, a fGn can be written on the form

$$X_t = m_t + P_t^{1/2}\, U_t, \tag{3.12}$$

where U_t is standard (unit variance) Gaussian white noise. We introduce non-Gaussian distributions by replacing the distributional assumption on the sequence U_t. Let

$$U_t = \sqrt{\frac{\nu-2}{\nu}}\varepsilon_t, \tag{3.13}$$

with ε_t being a sequence of independent student-t distributed innovations with $\nu > 2$ degrees of freedom. The pre-factor in (3.13) makes σ the standard deviation of the process. Let us note that as $\nu \to \infty$, the process is again a fGn. The (excess) kurtosis is a measure of how leptokurtic the pdfs are. For the student-t distribution the kurtosis is $6/(\nu-4)$ for $\nu > 4$, which can be compared to the zero kurtosis for the normal distribution. Sample paths from the fractional noise with student-t innovations are shown in figure 3.3. Figure 3.4 shows the p-values of the slopes under the (incorrect) assumption of a fGn, when the data-generating process is in fact a fractional student-t noise. It is seen that in this case, the deviance from normality has a negligible effect with respect to trend significance.

3.2.2 Short-range memory statistics and model selection

To take into account possible short-memory statistics, consider a first order auto-regressive (AR(1)) process z_t, defined as the stationary and causal solution to the stochastic difference equation

$$z_t = \phi\, z_{t-1} + v_t, \tag{3.14}$$

where v_t is Gaussian white noise with $\mathrm{var}(v_t) = \sigma_v^2$. Stationarity implies $|\phi| \neq 1$ and with the additional requirement of causality we have $\phi \in (-1, 1)$. We will restrict our attention to the persistent regime, i.e. $\phi \in (0, 1)$, and the aim here is to separate between an AR(1) and a fGn representation of the error term. For this purpose we suggest to use the likelihood-ratio (LR) test statistic. Let $p_g(x_t|x_1 \ldots, x_{t-1}, \gamma)$ be the conditional density of $z_t|\{z_1 = x_1, \ldots, z_{t-1} = x_{t-1}\}$ with parameters $\gamma = (\phi, \sigma_v)$, and let $p_f(x_t|x_1 \ldots, x_{t-1}, \theta)$ be the corresponding conditional density for a fGn with parameters $\theta = (H, \sigma)$. Denote by $\hat{\gamma}$ and $\hat{\theta}$ the ML estimators for the random vector $Y = (Y_1, \ldots, Y_n)$ and

$$a_t = \log \frac{p_f(Y_t|Y_1 \ldots, Y_{t-1}, \hat{\theta})}{p_g(Y_t|Y_1 \ldots, Y_{t-1}, \hat{\gamma})}. \tag{3.15}$$

The standard LR-statistic is the sum over the terms a_t. However, we find it useful to normalize to a more familiar scale, so we define the LR statistic by

$$\mathcal{L}(Y) = \frac{2}{\sqrt{n}\,\hat{\sigma}_*} \sum_{t=1}^{n} a_t, \tag{3.16}$$

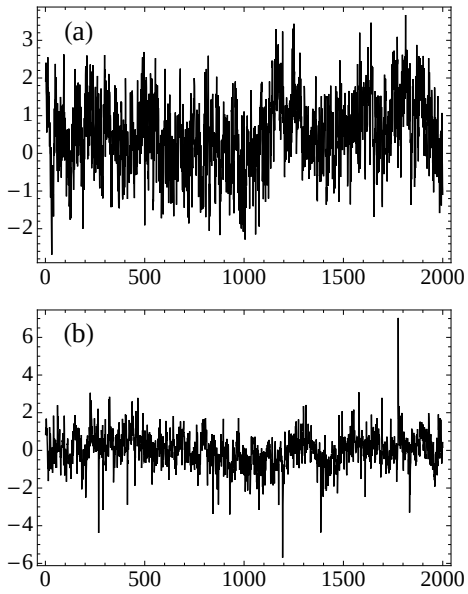


Figure 3.3: a) Sample path of fractional Gaussian noise with Hurst exponent $H = 0.85$. b) Sample path of fractional student noise with Hurst exponent $H = 0.85$ and $\nu = 3$ degrees of freedom. The same random seed is used in (a) and (b).

where $\hat{\sigma}_*$ is an estimator of $\sigma_* = \text{var}(a_t)^{1/2}$. From the assumption of Y_t being stationary (with a finite second moment), it follows that a_t is stationary. In particular, $\text{var}(a_t)$ is time invariant. Let $\hat{\sigma}_*$ be the sample standard deviation. When the data-generating process is a white noise $\mathcal{L}(Y)$ converges to a standard normal distribution. We can use this to determine the significance of the observed LR statistic. From the definition of the addends we see that a large likelihood-ratio points towards a fGn, while large negative values points towards an AR(1) process. Coupled with the asymptotic result under white noise, we suggest to use quantiles from the normal distribution to assess the significance.

Table 3.3 summarizes a Monte Carlo study using the proposed method. Note that the OLS linear trend is subtracted prior to estimating the remaining quantities. It is seen that distinguishing the two models requires more data when the persistence is low. Here one should note that even with long time series one can have something in between an AR(1) and a fGn, in the sense of Vuong (1989). Let us finally note that an alternative test is to include both an AR(1) and a fGn in the error term. This can for instance be done using equation (3.14) with v_t being a fGn.

3.3 Missing data

It is straight forward to adjust the linear trend model to handle gaps in the time series. To estimate the parameters in ARMA models (AR(1) in this study) we use the Kalman filter (e.g. Shumway and Stoffer, 2010). The biggest challenge is to estimate the Hurst

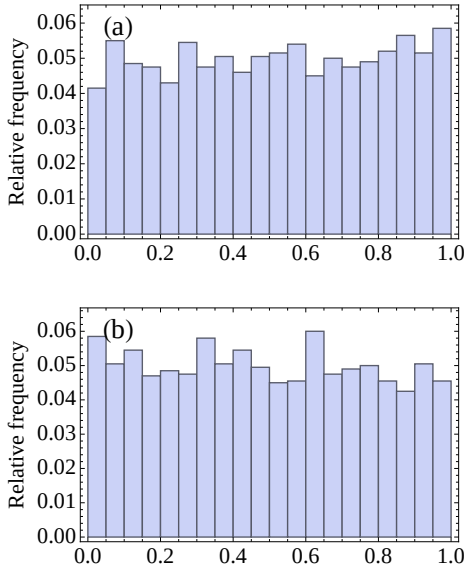


Figure 3.4: a) p-values of slope parameter computed from a fGn null hypothesis. Data-generating process is fractional student noise with $\nu = 3$ degrees of freedom, Hurst-exponent $H = 0.85$ and sample length $n = 340$. Number of Monte Carlo runs is $n_{\text{mc}} = 2000$. b) $n_{\text{mc}} = 2000$ realizations from the uniform $[0, 1]$ distribution.

	$H = 0.5$	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
		$n = 340$			
fGn	0.02	0.33	0.60	0.77	0.85
AR(1)	0.03	0.05	0.02	0.01	0.01
		$n = 680$			
fGn	0.02	0.48	0.79	0.92	0.97
AR(1)	0.03	0.03	< 0.01	< 0.01	< 0.01
		$n = 1360$			
fGn	0.03	0.65	0.94	0.99	1.00
AR(1)	0.03	0.01	< 0.01	< 0.01	< 0.01

Table 3.3: Likelihood-ratio test of fGn versus AR(1). The decision is based on 95% standard normal confidence interval, i.e. we choose a fGn if LR $> z_{0.975}$, and a AR(1) if LR $< z_{0.025}$. The test is undecided otherwise. Here z_α is the α-quantile of the standard normal distribution. The LR statistic is calculated from residuals of a linear trend model. The data-generating process is a fGn with Hurst exponent H.

exponent. We note that the fGn vector (found by removing entries that are missing) is still multivariate normal with covariance matrix Γ_{obs}. The matrix Γ_{obs} is found by removing

	$\phi = 0$	$\phi = 0.2$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
		$n = 340$			
fGn	0.02	0.02	< 0.01	< 0.01	< 0.01
AR(1)	0.03	0.45	0.72	0.84	0.95
		$n = 680$			
fGn	0.02	0.01	< 0.01	< 0.01	< 0.01
AR(1)	0.03	0.45	0.72	0.84	0.95
		$n = 1360$			
fGn	0.03	< 0.01	< 0.01	< 0.01	< 0.01
AR(1)	0.03	0.77	0.98	1.00	1.00

Table 3.4: Same as in 3.3, except here the data-generating process is AR(1) with parameter ϕ.

the rows and columns in Γ where observations are missing. As a result the matrix Γ_{obs} is no longer Toeplitz. The Toeplitz property is used to create an efficient algorithm in the "gap-free" case (McLeod et al., 2007). However, as shown in Bos et al. (2012), there is a simple relation between Γ and Γ_{obs}, which can be used to create an efficient algorithm for ML estimation of the Hurst exponent. For the wavelet variance, methods for handling missing data are presented in Mondal and Percival (2008) and Craigmile and Mondal (2013), and these results are easily adopted to the variogram method.

3.4 Not significant versus no trend

3.4.1 Type II error

By construction we control the probability of comiting a type I error, i.e. rejecting the null hypothesis given that the null is true. We do not control the other error (type II) involved; keeping the null when the alternative is hypothesis is true. However, one can study the probability of making type II errors. For this purpose, one can use the so-called (statistical) power functions. A power function gives the probability of the type II error as a function of the parameter value(s) under the alternative hypothesis. We observe that, for the linear trend model with fGn, the statistical power can be computed from (3.4) for a known Hurst exponent and a known scale parameter. When these parameters are estimated we use (3.10) instead, adjusted such that $a_1 \neq 0$.

In the case of Gaussian white noise, where the Hurst exponent is estimated, Cohn and Lins (2005) found that the statistical power was not reduced substantially compared to an a priori assumption of $H = 0.5$. In that study, the likelihood-ratio test was used. We report a similar finding (not shown) for the OLS test statistic.

3.4.2 Effective sample size

Another way to study the impact of LRD in statistical inference is to calculate effective sample sizes, a measure of the amount of reduced information about the mean term that is caused by positive serial correlation (Lee and Lund, 2008, see also Zwiers and von Storch (1995); Lee and Lund (2004); Ko et al. (2008)). It is easiest to motivate this term in the case of the sample mean:

Let $B_H(t)$ be a fractional Brownian motion (fBm) with Hurst exponent H and scale-parameter $\sigma = (\mathbb{E}B_H(1)^2)^{1/2}$. Denote by $X_t = B_H(t) - B_H(t-1)$ the corresponding fGn. The sample mean $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ is a scaled, discrete-time fractional Brownian motion, i.e.

$$\bar{X} \stackrel{\mathrm{d}}{=} n^{-1} B_H(n). \tag{3.17}$$

Thus, the one-dimensional marginals are given by

$$\bar{X} \sim \mathcal{N}\left(0, \sigma^2 n^{2H-2}\right) \Leftrightarrow n^{1-H}\frac{\bar{X}}{\sigma} \sim \mathcal{N}(0,1). \tag{3.18}$$

The effective sample size is defined as

$$n_e \equiv n\frac{\mathrm{var}(\bar{X}_{\mathrm{IID}})}{\mathrm{var}(\bar{X})} = n^{2-2H}. \tag{3.19}$$

Observe that, if n_e is an integer, the variance of the sample mean for n_e independent random variables equals the the variance of sample mean for a fGn with length n. So one only needs n_e data points from a white noise process to produce the same uncertainty in the sample mean as a fGn with sample length n. As an example, for $H = 0.9$ and $n_e = 10$ we need a sample length of $n = 10^5$ from the fGn. Increasing the Hurst exponent to $H = 0.99$ results in $n = 10^{50}$, reflecting that the sample mean of a fractional Brownian motion is not consistent. The effective sample size for the OLS estimator of the slope is defined similar to (3.19), see (3.11).

4
Paper 1

Abstract. We propose a multifractal model for short-term interest rates. The model is a version of the Markov-Switching Multifractal (MSM), which incorporates the well-known level effect observed in interest rates. Unlike previously suggested models, the level-MSM model captures the power-law scaling of the structure functions and the slowly decaying dependency in the absolute value of returns. We apply the model to the Norwegian Interbank Offered Rate with three months maturity (NIBORM3) and the U.S. Treasury Bill with three months maturity (TBM3). The performance of the model is compared to level-GARCH models, level-EGARCH models and jump-diffusions. For the TBM3 data the multifractal out-performs all the alternatives considered.

4.1 Introduction

Interest rates play an important role for financial institutions, for instance in risk management, where the interest-rate risk needs to be assessed. Often this is done by a simple stress test, where one considers a parallel shift in the yield curve, typically 100 or 200 basis points, and report the change in the value of the portfolio. This approach can be improved by stochastic modeling of future movements in the interest-rate yield curve. Vasicek (1977) and Cox et al. (1985) argue that the yield curve is given by the spot rate alone. Short-term interest rates, such as NIBORM3 and TBM3, are frequently used as proxies for the spot rate, and hence accurate modeling of these time series is potentially very important. To our knowledge this is the first published paper considering the Norwegian rate, while the TBM3 data are much studied (see e.g. Andersen and Lund (1997), Chapman and Pearson (2001), Durham (2003), Johannes (2004), Bali and Wu (2006)).

Traditionally models for short-term rates have been modeled by Itô stochastic differential equations on the form

$$dR(s) = f(R(s))ds + cR(s)^{\gamma}dB(s), \tag{4.1}$$

where s is the time variable, $R(s)$ denotes the risk-free rate and $B(s)$ is a Brownian motion. If we discretize this equation, by letting $s = t\Delta s$ and $r_t = R(s)$, then we obtain a stochastic difference equation on the form $r_t = r_{t-1} + f(r_{t-1})\Delta t + c\Delta s^{1/2} r_{t-1}^{\gamma} w_t$, where w_t are independent and normal distributed random variables with unit variance. It is convenient to write this equation on the form

$$r_t = \mu_t + \sigma_t, \tag{4.2}$$

with $\mu_t = r_{t-1} + \Delta s f(r_{t-1})$ and $\sigma_t = r_{t-1}^{\gamma} x_t$. Here $x_t = \sigma w_t$ with $\sigma = c\Delta s^{1/2}$. Throughout the rest of the paper we will consider linear drift terms on the form $f(r) = A_0 + A_1 r$. If $A_1 < 0$, then (4.1) has a stationary solution, and the discretized equation has a stationary solution for sufficiently small $\Delta s > 0$. For a fixed Δs we write $\mu_t = r_{t-1} + \alpha_0 + \alpha_1 r_{t-1}$. In this case we have stationarity for $-2 < \alpha_1 < 0$.

The number $\gamma \geq 0$ is called the Constant Elasticity Variance (CEV) parameter. For $\gamma \neq 0$ the models feature the so-called level effect. It is generally accepted that this effect is present in interest-rate data, see for instance (Longstaff et al., 1992). The level-effect introduces volatility persistence, i.e. strong dependence between the absolute values of the increments $\Delta r_t = r_t - r_{t-1}$. However, if $B(s)$ is a Brownian motion, then the variables

$$\frac{\Delta r_t - (\alpha_0 + \alpha_1 r_{t-1})}{r_{t-1}^{\gamma}} \tag{4.3}$$

are i.i.d. and Gaussian, and hence the volatility persistence will vanish under a simple transformation of the increments. Such processes are called pure-level models and have been studied by Cox et al. (1985) and Longstaff et al. (1992). Using time-series data for short-term interest rates we can optimize the likelihood to determine the parameters α_0, α_1 and γ. These results are shown in table 4.1. We can then transform the data according to

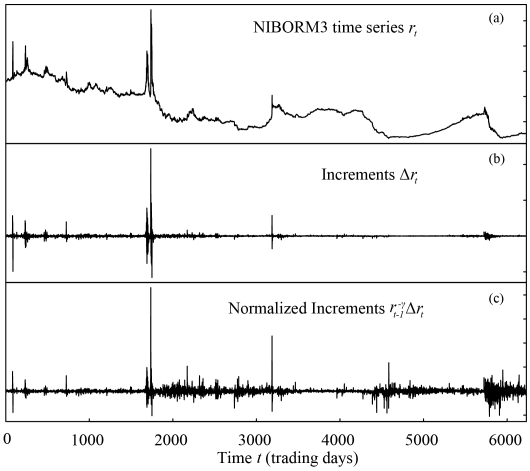


Figure 4.1: (a): The NIBORM3 data r_t for the time period 1986-01-02 to 2010-09-24. (b): The one-day increments $\Delta r_t = r_t - r_{t-1}$. (c): The normalized one-day increments $\Delta r_t / r_{t-1}^{\gamma}$. The value of the CEV parameter is $\gamma = 1.61$.

(4.3). The resulting time series (which are plotted in figures 4.1(c) and 4.2(c)) are realizations of the so-called normalized increment process. Just from inspection of figures 4.1(c) and 4.2(c)) we observe that the resulting time series exhibit strong volatility persistence. This is confirmed in figures 4.3(a) and 4.3(b), where we have plotted the autocorrelation functions for the absolute values of the normalized increments. As a consequence of these observations we conclude that pure-level models are insufficient, and we will therefore replace the process w_t with dependent variables x_t.

In our empirical investigations we find that the dependency structure in the variables x_t resembles the stylized facts of logarithmic returns of asset prices, namely that the variables themselves are uncorrelated (or weakly correlated) whereas their absolute values have slowly decaying autocorrelation functions. To describe this dependency several authors

Innovations	α_0	γ	σ	ν	$\log L$
		U.S. Treasury Bill			
Normal	0.0009709	0.3755	0.04788		15707.94
t_ν	−0.0002434	0.5377	0.2186	2.01	20688.54
		NIBORM3			
Normal	0.0002863	1.61	0.006298		4771.78
t_ν	0.0002419	0.988	0.0921	2.01	8041.42

Table 4.1: The constant elasticity volatility (CEV) model of Longstaff et al. (1992) with normal and student-t innovations.

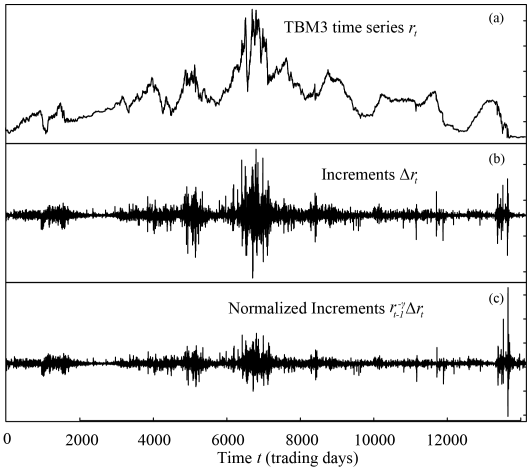


Figure 4.2: (a): The TBM3 data r_t for the time period 1954-01-04 to 2010-09-22. (b): The one-day increments $\Delta r_t = r_t - r_{t-1}$. (c): The normalized one-day increments $\Delta r_t / r_{t-1}^{\gamma}$. The value of the CEV parameter is $\gamma = 0.38$.

(e.g. Brenner et al. (1996) and Koedijk et al. (1997)) have suggested using a generalized autoregressive conditional heteroskedasticity (GARCH) model for the process x_t. The corresponding process r_t is then called a level-GARCH model. This model can be further improved by letting x_t be an Exponential GARCH (EGARCH) model. The level-EGARCH model was introduced by Andersen and Lund (1997) and has been reported to perform better than the standard GARCH model on the TBM3 data.

In this work we propose a multifractal model, specifically a level-MSM model, as an alternative to level-GARCH models and level-EGARCH models for short-term interest rates. This model is a slight modification of the standard MSM model constructed by Calvet and Fisher (2004). The motivation for introducing multifractal models for short-term interest rates is similar to the motivation for multifractal modeling of asset prices, namely that these models capture the dependency structure and scaling properties of the process x_t. Secondly, encouraging results (in and out-of-sample) have been obtained in a preliminary study of the NIBORM3 data (Løvsletten, 2010).

We use a HAC-adjusted version of the Vuong-test (Vuong, 1989) for model selection (Calvet and Fisher, 2004), and the main result of this work is that the level-MSM model performs well compared against level-GARCH models, level-EGARCH models and jump-diffusions. The paper is organized as follows: In section 4.2 we define the level-MSM model and consider some basic properties. A brief description of the alternative models are presented in section 4.3, and in section 4.4 the results of the Vuong-test are presented. In section 4.5 we give some concluding remarks.

Remark. *We note that the data analyzed in this paper is freely available online. Both data sets are given with daily resolution. The TBM3 data is taken from the period 1951-*

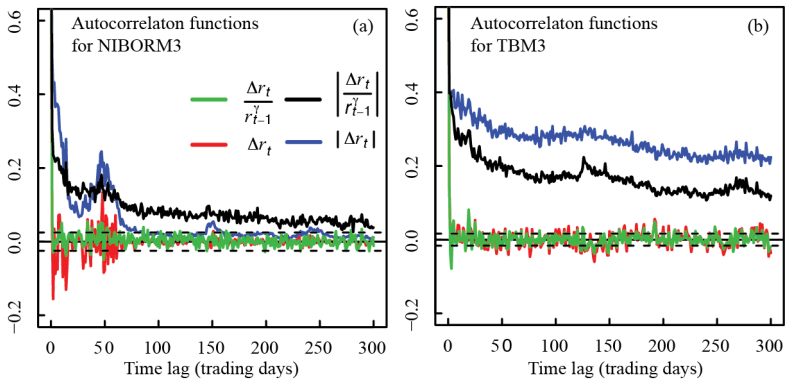


Figure 4.3: (a): Autocorrelation functions for one-day increments in the NIBORM3 data. The variables considered are the standard increments $\Delta r_t = r_t - r_{t-1}$, the normalized increments $\Delta r_t / r_{t-1}^{\gamma}$ and the absolute values of these two time series. The value of the CEV parameter is $\gamma = 1.61$. (b): Same as (a), but now for the TBM3 data. The value of the CEV parameter is $\gamma = 0.38$. In both figure the dotted lines represent a 0.95 confidence interval for the autocorrelation function assuming independence.

01-04 to 2010-09-22, and consists of 14172 data points. The NIBORM3 data is taken from the period 1986-01-02 to 2010-09-24, and consists of 6231 data points. The TBM3 data contains several days with zero value, something that causes certain technical problems in the stochastic modeling. The problem is resolved by a simple variable shift $r \mapsto r + b$. In order to be consistent with our set of models, we estimate the constant $b > 0$ by estimating the conditional standard deviations $\mathrm{sdv}[\Delta r_t | r_{t-1} = r]$ *(for various r) and fitting a function $cr^{\gamma} - b$ to this data set. Using this model we find $b = 0.03$.*

4.2 Multifractal models

The application of multifractal processes to finance was introduced by Mandelbrot et al. (1997). We recall that a process $X(t)$ is multifractal if the structure functions $S_q(t) = \mathbb{E}[|X(t)|^q]$ are power laws in t. One can then define a scaling function $\zeta(q)$ through the relation $S_q(t) \sim t^{\zeta(q)}$ as $t \to 0$, i.e.

$$\zeta(q) = \lim_{t \to 0} \frac{\log S_q(t)}{\log t} .$$

Using Hölder's inequality it is easy to show that the scaling function $\zeta(q)$ is concave, and we also note that if the process $X(t)$ is h-self similar, i.e. $X(at) \stackrel{\mathrm{d}}{=} a^h X(t)$, then $\zeta(q) = hq$. We are therefore interested in the situations where $\zeta(q)$ is strictly concave. In this case we

see that

$$\frac{\mathbb{E}[X(t)^4]}{\mathbb{E}[X(t)^2]^2} \sim t^{\zeta(4)-2\zeta(2)} \to \infty \quad \text{as} \quad t \to 0\,.$$

If we assume that $X(t)$ also has stationary increments, then this implies that the Δt-lagged increments $X(t+\Delta t) - X(t)$ are more leptokurtic for small Δt than for larger Δt. In particular, this shows that $X(t)$ can not be Gaussian.

In financial time series one will most often want the process $X(t)$ to have uncorrelated and stationary increments. In order for this to be satisfied we must impose the condition $\zeta(2) = 1$, because otherwise the variables $\Delta X(t) = X(t+1) - X(1)$ have slowly decaying autocorrelation. In fact,

$$\mathrm{Cov}(\Delta X(0), \Delta X(t)) \sim t^{-(2H_d-2)}\,,$$

where $H_d := \zeta(2)/2$. A simple example of such a process is a Brownian motion, which is h-selfsimilar with $h = H_d = 1/2$. As opposed to the Brownian motion, a multifractal process with strictly concave scaling function has strongly dependent increments, even for $H_d = 1/2$. For instance, in the MSM model[1] we have $\zeta(2) = 1$ and

$$\mathrm{Cov}(|\Delta X(0)|^q, |\Delta X(t)|^q) \sim t^{-(\zeta(2q)-2\zeta(q))}\,. \tag{4.4}$$

See proposition 1 of (Calvet and Fisher, 2004). This inherent long-range volatility persistence serves as our motivation for modeling short-rate interest rates using multifractals.

In the next subsection we will present several examples of multifractal processes and look at some basic properties. This part of the text is included for completeness, and readers who are familiar with multifractal models may skip to section 4.2.2.

4.2.1 Mandelbrot's MMAR processes

The multifractal models introduced by Mandelbrot et al. (1997) are stochastic processes on the form

$$X(t) = \sigma\, T^H\, B_H\big(\Theta(t)\big)\,, \tag{4.5}$$

defined for $0 \le t \le T$, where $\Theta(t) = \mu([0,t])$ is a random probability measure on $[0,T]$ and $B_H(t)$ is a fractional Brownian motion with $\mathbb{E}[B_H(1)^2] = 1$. The process $\Theta(t)$ is itself a multifractal process with scaling function $\zeta_\Theta(q)$, and if μ and $B_H(\cdot)$ are independent, then $\zeta_X(q) = \zeta_\Theta(Hq)$. By construction, the measures we will mention in this paper satisfy $\mathbb{E}[\mu([0,t])] \propto t$, i.e. $\zeta_\Theta(1) = 1$. It follows that $\zeta_X(1/H) = 1$, and in order to satisfy the condition $H_d = 1/2$ one must have $H = 1/2$. However, since processes with correlated increments are useful in other applications (see e.g. Rypdal and Rypdal (2010)), there is no reason to restrict ourselves to the case $H = 1/2$ in the definition of the models.

[1] It is pointed out by Lux (2006) that the MSM model is only a finite-level approximation, and so equation (4.4) is only valid on time scales up to b^K. Here $b > 0$ and $K \in \mathbb{N}$ are parameters in the MSM model. See section 4.2.2.

There are several choices for the measure μ, and below we give three examples:

Example 1: In the simplest case the measure μ is a randomized dyadic Bernoulli measure with probabilities $p_1 = p$ and $p_2 = 1-p$. This measure is constructed through an iterative procedure, where we in the first step divide the interval $[0,T]$ in two pieces Δ_0 and Δ_1 of equal length. One of the intervals is chosen at random and given a mass $p_1\,\mu([0,T])$, while the other interval is given the mass $p_2\,\mu([0,T])$. This procedure is then repeated recursively, i.e. the interval Δ_0 is divided into the equally sized intervals Δ_{00} and Δ_{01}. One of these intervals is given the mass $p_1^2\,\mu([0,T])$ and the other is given the mass $p_1p_2\,\mu([0,T])$. Formally the measure can be defined by letting f be a random bijection[2] $\{0,1\} \to \{p_1,p_2\}$ and $f_{i_1,\dots,i_k}$ be independent copies of f. The measure is constructed by assigning the mass $\mu(\Delta_{i_1,i_2,\dots,i_k}) = f(i_1)f_{i_1}(i_2)\cdots f_{i_1,i_2,\dots,i_{k-1}}(i_k)$ to the dyadic intervals

$$\Delta_{i_1,\dots,i_k} = \left[T\cdot(0.i_1\cdots i_k)_2, T\cdot(0.i_1\cdots i_k)_2 + T\cdot 2^{-k}\right], \quad i_n \in \{0,1\}\,.$$

A simple combinatory argument shows that for $t = T\cdot 2^{-k}$, the random variable $\Theta(t)$ has density

$$p_{\Theta,t}(x) = \frac{1}{2^k}\sum_{n=0}^{k}\binom{k}{n}\delta(x - p^n(1-p)^{k-n})\,.$$

When the processes $B_H(t)$ and $\Theta(t)$ are independent, we obtain the density of $X(t)$:

$$p_{X,t}(x) = \int p_{\sigma\,T^H\,B_H,s}(x)\,p_{\Theta,t}(s)\,ds = \frac{1}{2^k}\sum_{n=0}^{k}\binom{k}{n}\frac{\exp\left(-\frac{x^2}{2\sigma^2\left(Tp^n(1-p)^{k-n}\right)^{2H}}\right)}{\sqrt{2\pi\sigma^2\left(Tp^n(1-p)^{k-n}\right)^{2H}}}\,.$$

From this density we can easily calculate the structure functions $S_q(t) = \mathbb{E}[|X(t)|^q]$ and see that $S_q(t) \sim t^{\zeta(q)}$ as $t \to 0$, where $\zeta(q) = 1 - T(Hq)$ and $T(q) = \log_2(p^q + (1-p)^q)$.

Remark. *The function $T(q)$ is sometimes called the scaling function of the random measure. The dyadic measure defined by $\mu(\Delta_{i_1,\dots,i_k}) = p_{i_1}\cdots p_{i_k}$ has a multifractal spectrum $f(\alpha)$ given by the Legendre transform of $T(q)$: $f(\alpha) = \inf_{q\in\mathbb{R}}\{\alpha q - T(q)\}$, and the Hentchel-Procaccia dimension spectrum is $D_q = (q-1)^{-1}T(q)$. See e.g. Pesin (1997) for an account of the relation between scaling functions and spectra of fractal dimensions. The function $\alpha \mapsto f(H\alpha)$ is often called the singularity spectrum of the process $X(t)$.*

Example 2: A different class of multifractal measures consists of the b-adic random multiplicative cascades (see e.g. Mandelbrot et al. (1997) for a more detailed account). For an integer $b \geq 2$ and $i_n \in \{0,1,\dots,b-1\}$ we define the b-adic subintervals of $[0,T]$ as

$$\Delta_{i_1,\dots,i_k} = [T\cdot(0.i_1\dots i_k)_b\,,\; T\cdot(0.i_1\dots i_k)_b + T\cdot b^{-k}]\,.$$

[2]with probability 1/2 for each of the two outcomes $(0,1)\mapsto(p_1,p_2)$ and $(0,1)\mapsto(p_2,p_1)$

For a positive random variable M, with $\mathbb{E}[M] = b^{-1}$, let

$$\mu(\Delta_{i_1,\ldots,i_k}) = M_{i_1} M_{i_1 i_2} \cdots M_{i_1 i_2 \cdots i_k} \, \Omega_{i_1,\ldots,i_k} \, ,$$

where $M_{i_1,\ldots,i_n}$ are independent copies of M and

$$\Omega_{i_1,\ldots,i_k} = \lim_{n\to\infty} \sum_{j_1,\ldots,j_n} M_{i_1} \cdots M_{i_1 i_2 \cdots i_k} M_{i_1 i_2 \cdots i_k j_1} \cdots M_{i_1 i_2 \cdots i_k j_k \cdots j_n} \, .$$

Again a stochastic process is constructed according to equation (4.5) with $\Theta(t) = \mu([0,t])$. This process has a scaling function $\zeta(q) = 1 - T(Hq)$, where $T(q) = \log_b \mathbb{E}[M^q]$. Popular choices for the multiplier M are log-normal distributions, which give quadratic scaling functions, or other log-infinitely divisible distributions. Almost every realization of the measure μ has the multifracal spectrum $f(\alpha) = \inf_{q\in\mathbb{R}}\{\alpha q - T(q)\}$.

Example 3: The Poisson multifractal measure on $[0,T]$ generalizes the b-adic multiplicative cascade by introducing randomness in the construction of the intervals $\Delta_{i_1,i_2,\ldots,i_k}$. In the original multifractal models, each interval on level k are divided into b pieces of equal length at level $k+1$. As a result the interval length decreases as b^{-k} with the level k. In the Poisson multifractal measure, the splitting of an interval at level k is preformed by drawing the lengths of the new pieces randomly from an exponential distribution with rate l_{k+1}. This means that the mean and median length of an interval at level $k+1$ are $1/l_{k+1}$ and $\log(2)/l_{k+1}$ respectively, so to maintain exponential decay of interval lengths (as a function of level), one chooses $l_k = b^{k-1} l_1$. Note that b no longer is restricted to the integers.

Formally, the Poisson multifractal measure on the interval $[0,T]$ is defined via sequence of measures μ_k specified on randomly generated intervals

$$\Delta_{i_1,i_2,\ldots,i_k} = [t_{i_1,i_2,\ldots,i_k}, t_{i_1,i_2,\ldots,(i_k+1)}] \, ,$$

where the numbers $t_{i_1,\ldots,i_k}$ are defined by the following recursive construction:

Let $\{\tau_{i_1,\ldots,i_k}(j)\}_{j\in\mathbb{N}}$ denote independent and exponentially distributed random variables with rates $l_k = b^{k-1} l_1$. Define

$$N_{i_1,\ldots,i_{k-1}} = \max\left\{ m : \sum_{j=1}^{m} \tau_{i_1,\ldots,i_{k-1}}(j) < \operatorname{diam}(\Delta_{i_1,i_2,\ldots,i_{k-1}}) \right\} \tag{4.6}$$

and

$$t_{i_1,\ldots,i_k} = \begin{cases} t_{i_1,\ldots,i_{k-1}} & \text{if} \quad i_n = 0 \\ t_{i_1,\ldots,i_{k-1}} + \sum_{j=1}^{i_k} \tau_{i_1,\ldots,i_{k-1}}(i_k) & \text{if} \quad 1 \le i_k \le N_{i_1,\ldots,i_{k-1}} \\ t_{i_1,\ldots,(i_{k-1}+1)} & \text{if} \quad i_k = N_{i_1,\ldots,i_{k-1}} + 1 \end{cases} . \tag{4.7}$$

This means that the interval $\Delta_{i_1,\ldots,i_k}$ is divided into $N_{i_1,\ldots,i_k}$ subintervals by the cuts made by a Poisson process with rate l_{k+1}. We start with an interval $\Delta = [0,T]$.

A sequence μ_k of measures can now be defined via the formula

$$\mu_k(\Delta_{i_1,i_2,\ldots,i_k}) = T^{-1} \operatorname{diam}(\Delta_{i_1,i_2,\ldots,i_k}) \, M_{i_1} M_{i_1,i_2} \cdots M_{i_1,\ldots,i_k} \, ,$$

where $M_{i_1 \cdots i_k}$ are independent copies of a positive random variable M satisfying $\mathbb{E}[M] = 1$. If $\mathbb{E}[M^2] < b$, then μ_k converges weakly to a Borel measure μ on $[0, T]$. With this choice of random measure the model given by equation (4.5) has scaling function

$$\zeta(q) = 1 - T(Hq)\,, \quad \text{where} \quad T(q) = 1 - q + \log_b \mathbb{E}[M^q]\,. \tag{4.8}$$

Calvet and Fisher constructed their MSM model by discretizing the time interval $[0, T]$ and assigning discrete geometric (rather than an exponential) distributions on the waiting times τ. More precisely one will fix an integer $K > 1$ determining the number of levels that are to be included in the discrete model, and consider the integer values $\{0, 1, \ldots, \tilde{T}\}$, where $\tilde{T} = m^K$ for some positive integer m. As for the construction of the Poisson multifractal one makes random partitions $\Delta_{i_1,i_2,\ldots,i_k} = [t_{i_1,i_2,\ldots,i_k}, t_{i_1,i_2,\ldots,(i_k+1)}]$ of the interval $[0, \tilde{T}]$. These are constructed using the recursive procedure described by equations (4.6) and (4.7), where the variables $\tau_{i_1,\ldots,i_k}$ are replaced by discrete random variables $\tilde{\tau}_{i_1,\ldots,i_k}$ with geometric distributions $\mathbb{P}(\tau_{i_1,\ldots,i_k} = \tau') = \lambda_k\,(1-\lambda_k)^{\tau'-1}$, with $1 - \lambda_k = e^{-lb^{k-1}\tilde{T}/T}$. This choice of parameters λ_k implies that the median of $\tilde{\tau}_{i_1,\ldots,i_k}$ is proportional to $l^{-1}b^{-n}$. Note that the parameters λ_k are given by λ_1 through the formula

$$\lambda_k = 1 - (1 - \lambda_1)^{b^{n-1}}\,. \tag{4.9}$$

A measure $\tilde{\mu}$ is defined by specifying that if t is an integer and $[t-1, t] \subseteq \Delta_{i_1,\ldots,i_K}$, then

$$\tilde{\mu}([t-1, t]) = m^{-K} M_{i_1} M_{i_1,i_2} \cdots M_{i_1,\ldots,i_K}\,,$$

where $M_{i_1,\ldots,i_k}$ are independent copies of a positive random variable M with $\mathbb{E}[M] = 1$. Again we define $\tilde{\Theta}(t) = \tilde{\mu}([0, t])$, and by taking the composition with a fractional Brownian motion we get a discrete time stochastic process $\tilde{X}(t) = C\,B_H(\tilde{\Theta}(t))$. We denote $x_t = \tilde{X}(t) - \tilde{X}(t-1)$, and observe that

$$x_t \stackrel{\mathrm{d}}{=} C\,\tilde{\mu}([t, t-1])^H \Big(B_H(t) - B_H(t-1)\Big) = \sigma\,(M_{i_1} M_{i_1,i_2} \cdots M_{i_1,\ldots,i_K})^H\,\varepsilon_t\,,$$

where $\sigma = C\,m^{-K}$ and $\varepsilon_t = B_H(t) - B_H(t-1)$ is a discrete version of a fractional Brownian noise.

We can simplify notations by denoting $M_{k,t} = M_{i_1,\ldots,i_k}$ for $t \in \Delta_{i_1,\ldots,i_k}$. For $H = 1/2$ the result is the MSM model described in the next section.

4.2.2 The MSM model

The Markov Switching Multifractal model of order $K > 1$ is given by

$$x_t = \sigma g(\mathbf{M}_t)^{1/2} \varepsilon_t\,, \tag{4.10}$$

where $\mathbf{M}_t = (M_{1,t}, \ldots, M_{K,t})$, and the function $g : \mathbb{R}^K \to \mathbb{R}$ is the product of the vector components. The innovatons ε_t are normal distributed with unit variance, and in this

setting also assumed to be independent. The process $\mathbf{M}_t$ is a Markov-chain defined by the following updating scheme:

$$M_{k,t} = \begin{cases} \text{drawn from } M & \text{with probability } \lambda_k, \\ M_{k,t-1} & \text{with probability } 1-\lambda_k. \end{cases} \tag{4.11}$$

Components are updated independently of all previous updates, and the frequencies λ_k are related to each other through (4.9).

For simplicity we choose a two-point distribution M:

$$\mathbb{P}(M = m_0) = \mathbb{P}(M = m_1) = \frac{1}{2}, \text{ with } \; m_0 \in (1,2) \; \text{ and } \; m_1 = 2 - m_0. \tag{4.12}$$

This version of the MSM is known as the binomial multifractal (Calvet and Fisher, 2004). To higlight the number of multipliers K we will use the notation MSM(K). Note that, unlike GARCH(p,q), the number of parameters remains unchanged with increasing order.

Let $\{\mathbf{m}^i \,|\, i = 1, \ldots, d := 2^K\}$ denote the sample space of $\mathbf{M}_t$. We also define vectors $\omega(x) : \mathbb{R} \to \mathbb{R}^d$ and $\mathbf{p}_t : \mathbb{N} \to \mathbb{R}^d$ with components $\omega_k(x) = n(x|\sigma^2 g(\mathbf{m}^k))$ and $p_{k,t} = \mathbb{P}(\mathbf{M}_t = \mathbf{m}^k|\mathbf{x}_t)$ respectivly. Here $\mathbf{x}_t = (x_2, \ldots, x_t)$, and $n(\cdot|\sigma)$ is the density of the normal distribution with zero mean and variance σ^2. With this notation the density of $x_t|\mathbf{x}_{t-1}$ is given by

$$p(x_t|\mathbf{x}_{t-1}) = \mathbf{p}_{t-1}^T A\, \omega(x_t), \tag{4.13}$$

where A is the transitition matrix of the Markov chain, that is $A_{j,k} = \mathbb{P}(\mathbf{M}_t = \mathbf{m}^k|\mathbf{M}_{t-1} = \mathbf{m}^j)$. This is seen by conditioning on the underlying Markov-chain at time $t-1$ and t, and using Bayes' rule. It is easily seen, once again using Bayes' rule, that the vectors $\mathbf{p}_t$ follow the recursion

$$\mathbf{p}_t^T = \frac{1}{p(x_t|\mathbf{x}_{t-1})} \omega^T(x_t) * (\mathbf{p}_{t-1}^T A), \tag{4.14}$$

where $*$ denotes the Hadamard-product. We start the recursion with the limiting probabilities of $\mathbf{M}_t$.

4.2.3 The level-MSM model for interest rates

We propose the following level-MSM model for the short-rate

$$r_t - r_{t-1} = \alpha_0 + \alpha_1 r_{t-1} + r_{t-1}^{\gamma} x_t, \tag{4.15}$$

where x_t is the binomial MSM model defined above. The likelihood

$$L = \sum_{t=2}^{n} \log f(r_t|\mathbf{r}_{t-1}),$$

for data $r_2, \ldots, r_n$ with r_1 taken as a pre-sample value, now follows from equations (4.13) and (4.14) together with the relation

$$f(r_t|\mathbf{r}_{t-1}) = \frac{1}{r_{t-1}^{\gamma}} p(x_t|\mathbf{x}_{t-1}),$$

Model	$10^3 \times \alpha_0$	$10^4 \times \alpha_1$
U.S. Treasury Bill M3		
MSM(9)	0.1214 (0.354)	−0.1905 (1.04)
GARCH	0.2821 (0.386)	−0.6731 (1.10)
EGARCH	0.3595 (0.310)	−1.3524 (0.92)
Jump-diffusion	0.7074 (0.397)	−0.8695 (1.15)
NIBORM3		
MSM(9)	2.228 (0.726)	−3.201 (1.60)
GARCH	2.424 (0.746)	−3.986 (1.66)
EGARCH	2.515 (0.430)	−3.626 (0.87)
Jump-diffusion	2.621 (0.788)	−4.053 (1.77)

Table 4.2: A linear term in the drift term was added and the parameters estimated using ML. In the TBM3 none of the parameters α_1 are significantly different from zero. Standard deviations are reported in parentheses.

where f is the density of $r_t|\mathbf{r}_{t-1}$.

We have fitted the level-MSM models to the two time series under consideration for $K = 2, \ldots, 9$. For the case $K = 9$, the maximum likelihood (ML) estimates for the parameters α_0 and α_1 are reported in table 4.2. In this table the estimates of α_0 and α_1 for the alternative models are also included. As expected the estimates for the parameter α_1 are negative, but we observe that none of these ML estimates are significantly different from zero for the TBM3 data. The estimates for α_1 are also very small for the NIBORM3 data, and we will therefore consider the model defined by (4.15) with $\alpha_1 = 0$. The corresponding ML estimates for the other parameters are presented in tables 4.3 and 4.4. From these results we also see that the likelihoods increase monotonically with K. Using a Vuong-test we compare the level-MSM of order $K = 9$ against level-MSM models of lower order. When comparing two models with this method, the null-hypothesis is that both models are equally far from the data-generating process measured by the Kullback-Leibler distance. Hence small p-values indicate that the level-MSM model of order $K = 9$ is significantly closer to the data generating process than the level-MSM models of order $K = 7$ or lower.

The same model-selection-test is used to compare the level-MSM model with various alternative models, and more details on the test are presented in section 4.4.

4.3 Alternative models

In this section we briefly discuss the processes which we compare level-MSM model to. As for the level-MSM model we have first considered these models with drift terms on the form $\mu_t = r_{t-1} + \alpha_0 + \alpha_1 r_{t-1}$. The ML estimates for the parameters α_0 and α_1 are reported in table 4.2. Again we see that the contribution from the parameter α_1 is negligible. Consequently we will consider models with drift terms on the form $\mu_t = r_{t-1} + \alpha_0$.

K	$10^3 \times \alpha_0$	γ	m_0	b	λ_K	σ	$\log L$	BIC	Vuong	HAC-adj.
					U.S. Treasury Bill M3					
2	0.2752	0.2824	1.808	23.85	0.1391	0.07037	22309.49	-3.1446	13.573	9.091
	(0.247)	(0.0106)	(0.00471)	(4.92)	(0.011)	(0.00162)			(< 0.001)	(< 0.001)
3	0.2052	0.1913	1.73	10.62	0.1660	0.08078	22676.53	-3.1964	11.618	10.138
	(0.234)	(0.0175)	(0.0062)	(1.45)	(0.0163)	(0.00336)			(< 0.001)	(< 0.001)
4	0.2276	0.2269	1.661	9.643	0.2956	0.07202	22818.9	-3.2165	8.4076	7.562
	(0.226)	(0.032)	(0.00674)	(1.23)	(0.0462)	(0.00357)			(< 0.001)	(< 0.001)
5	0.0643	0.2280	1.609	9.804	0.6761	0.06865	22901.03	-3.2281	5.036	4.805
	(0.222)	(0.0191)	(0.00667)	(1.04)	(0.0597)	(0.00265)			(< 0.001)	(< 0.001)
6	0.0375	0.2115	1.559	6.577	0.778	0.06368	22920.57	-3.2308	4.346	4.335
	(0.219)	(0.0282)	(0.00744)	(0.564)	(0.0541)	(0.00360)			(< 0.001)	(< 0.001)
7	0.0660	0.1744	1.527	5.713	0.7928	0.08208	22933.55	-3.2326	3.4883	3.112
	(0.223)	(0.0291)	(0.00752)	(0.465)	(0.0561)	(0.00441)			(< 0.001)	(0.001)
8	0.0839	0.1575	1.503	4.932	0.8474	0.06956	22951.89	-3.2352	1.131	0.938
	(0.215)	(0.0260)	(0.00815)	(0.355)	(0.0509)	(0.0036)			(0.129)	(0.174)

Table 4.3: ML-estimates for the binomial multifractal with standard deviations in parentheses. The Vuong-column reports the likelihood ratio statistic with corresponding p-value in parentheses (Vuong, 1989). The nullhypothesis is that MSM(9) and MSM(K) have equal Bayesian Information Criteria, with the alternative hypothesis being that the MSM(9) is closer to the data generating process. The HAC-adjusted version of the Vuong-test corrects for heteroschedacity and autocorrelation in the addends (Calvet and Fisher, 2004).

K	$10^3 \times \alpha_0$	γ	m_0	b	λ_K	σ	$\log L$	BIC	Vuong	HAC-adj.
					NIBORM3					
2	1.156	1.805	1.843	16.23	0.1016	0.006092	8514.45	-2.7250	5.219	5.298
	(0.408)	(0.04)	(0.00547)	(4.68)	(0.0116)	(0.000516)			(< 0.001)	(< 0.001)
3	1.041	1.761	1.772	18.37	0.4259	0.005674	8671.87	-2.7755	3.054	3.093
	(0.378)	(0.0391)	(0.00685)	(3.07)	(0.0459)	(0.000463)			(0.001)	(< 0.001)
4	1.020	1.482	1.717	12.43	0.3939	0.01401	8760.38	-2.8039	3.424	3.588
	(0.378)	(0.043)	(0.00794)	(1.93)	(0.0483)	(0.00129)			(< 0.001)	(< 0.001)
5	1.13	1.933	1.642	7.495	0.5455	0.004534	8779.88	-2.8102	3.292	3.265
	(0.381)	(0.0635)	(0.0100)	(1.04)	(0.086)	(0.000555)			(< 0.001)	(0.001)
6	1.079	1.620	1.590	5.347	0.6584	0.00897	8789.49	-2.8133	2.702	2.866
	(0.379)	(0.0481)	(0.0107)	(0.542)	(0.0726)	(0.000841)			(0.003)	(0.002)
7	1.006	1.934	1.537	4.431	0.9000	0.003719	8798.81	-2.8162	1.809	1.727
	(0.379)	(0.0685)	(0.00926)	(0.448)	(0.0748)	(0.000464)			(0.0352)	(0.042)
8	1.023	1.985	1.500	3.536	0.9173	0.003220	8802.66	-2.8175	1.058	0.940
	(0.380)	(0.063)	(0.00933)	(0.292)	(0.0624)	(0.000388)			(0.145)	(0.173)
9	0.998	1.882	1.464	2.964	0.949	0.003706	8804.44	-2.8181		
	(0.379)	(0.112)	(0.00927)	(0.216)	(0.0484)	(0.00102)				

Table 4.4: ML-estimates for the binomial multifractal with standard deviations in parentheses. The Vuong-column reports the likelihood ratio statistic with corresponding p-value in parentheses (Vuong, 1989). The nullhypothesis is that MSM(9) and MSM(K) have equal Bayesian Information Criteria, with the alternative hypothesis being that the MSM(9) is closer to the data generating process. The HAC-adjusted version of the Vuong-test corrects for heteroschedacity and autocorrelation in the addends (Calvet and Fisher, 2004).

$10^3 \times \alpha_0$	γ	$10^5 \times a_0$	a_1	b	ν
U.S. Treausury Bill M3					
0.0904	0.1699	0.7092	0.1301	0.8915	3.7995
(0.225)	(0.0496)	(0.139)	(0.0092)	(0.0062)	(0.134)
NIBORM3					
0.9104	0.9717	0.4889	0.20993	0.80838	3.3305
(0.397)	(0.0719)	(0.158)	(0.0209)	(0.0135)	(0.155)

Table 4.5: ML estimates for the parameters in the level-GARCH model with standard deviations in parentheses.

4.3.1 The level-GARCH model

When comparing to MSM models, we prefer the standard level-GARCH model proposed by Koedijk et al. (1997)[3]. This model is defined by (4.15), with x_t a GARCH(1,1) and $\alpha_1 = 0$. As is common for this model, we include student-t distributed innovations. The model then reads

$$r_t = \alpha_0 + r_{t-1} + w_t, \quad w_t = h_t^{1/2} r_{t-1}^{\gamma} \varepsilon_t, \quad h_t = a_0 + a_1 \left(\frac{w_{t-1}}{r_{t-2}^{\gamma}} \right)^2 + b\, h_{t-1}, \tag{4.16}$$

where the innovations ε_t are i.i.d. student-t distributed with unit-variance and ν degrees of freedom. For $\gamma = 0$ we have the pure-GARCH(1,1) model, and for $a_1 = b = 0$ we have the standard level model.

The estimated parameters for the NIBORM3 and TBM3 data are presented in table 4.5. We observe that simulated paths with these exponents have far to wild fluctuations compared to the real data, indicating that this model fails to accurately describe the interest rate fluctuations. In addition, we know that the GARCH models have exponentially decaying autocorrelation functions for the absolute values of the increments. This means that the long-range volatility persistence observed in the short-term interest rates is not inherent in these models.

4.3.2 The level-EGARCH model

Instead of using (4.16), Andersen and Lund (1997) propose using the EGARCH model. They find that this model gives an adequate fit to the TBM3 data, and a better fit compared to the standard GARCH model. In the EGARCH model the logarithm of the conditional variance replaces the conditional variance. The variance-recursion is then

$$\log h_t = a_0 + a_1 \varepsilon_{t-1} + a_2 |\varepsilon_{t-1}| + b \log h_{t-1}. \tag{4.17}$$

[3]We use a version of the model in Koedijk et al. (1997), where the quadratic term in the conditional means are set to zero.

$10^3 \times \alpha_0$	γ	a_0	a_1	a_2	b	ν
			U.S. Treausury Bill M3			
0.1039	0.3441	−0.2360	−0.03127	0.2322	0.9888	3.866
(0.212)	(0.0648)	(0.0165)	(0.00634)	(0.0123)	(0.00154)	(0.138)
			NIBORM3			
1.120	0.9101	−0.4024	0.02930	0.2664	0.9741	3.358
(0.396)	(0.0887)	(0.0372)	(0.0104)	(0.0188)	(0.00333)	(0.156)

Table 4.6: ML estimates for the parameters in the level-EGARCH model with standard deviations in parentheses.

The extra parameter a_1 controls potentially different responses to positive and negative returns. The use of the logarithm guarantees positive values of the volatility for all parameter values. In addition, this model provides the extra flexibility by letting the conditional distribution be non-symmetric. As we will see from the results presented in the next section, the EGARCH model gives better results than the GARCH model for both of the time series considered in this paper. This confirms the results in Andersen and Lund (1997).

The ML estimates for the parameters in the EGARCH model are presented in table 4.6.

4.3.3 Jump-diffusions

The last class of alternative models considered are the jump-diffusions. Both Johannes (2004) and Das (2002) propose these processes in order to describe the large spikes seen in interest-rate data. As a discretized jump-diffusion model we use the following specification:

$$r_t = \alpha_0 + r_{t-1} + r_{t-1}^{\gamma}(x_t + J_t z_t),$$

where J_t is the jump-indicator assumed to follow a Bernoulli-distribution. The probability of a jump taking place at time t is then given by

$$\left(1 + \exp(-c - d r_{t-1})\right)^{-1}.$$

By letting $x_t = h_t^{1/2}\varepsilon_t$ follow a GARCH-process, conditional heteroschedacity in levels are also accomodated (Das, 2002; Hong et al., 2004). The innovations are distributed as $z_t \overset{\mathrm{d}}{\sim} \mathcal{N}(0, \tau^2)$ and $\varepsilon_t \overset{\mathrm{d}}{\sim} \mathcal{N}(0, 1)$.The variance recursion in the GARCH process is given by

$$h_t = a_0 + a_1 \left(\frac{\Delta r_{t-1} - \alpha_0}{r_{t-2}^{\gamma}} \right)^2 + b\, h_{t-1},$$

where $\Delta r_t = r_t - r_{t-1}$. The ML estimates for the parameters in the jump-diffusion model are presented in table 4.7.

$10^3 \times \alpha_0$	γ	$10^6 \times a_0$	a_1	b	c	d	τ
		U.S. Treasury Bill M3					
0.464	0.1670	5.131	0.0931	0.8773	-3.751	0.2761	0.08622
(0.232)	(0.0374)	(0.867)	(0.0055)	(0.0062)	(0.170)	(0.0271)	(0.00624)
		NIBORM3					
1.085	1.872	0.0436	0.1291	0.8301	-1.883	-0.2119	0.01127
(0.417)	(0.0892)	(0.0198)	(0.0099)	(0.0111)	(0.328)	(0.0417)	(0.00216)

Table 4.7: ML estimates for the parameters in the jump-diffusion model with standard deviations in parentheses.

4.4 In-sample comparision

To test the binomial MSM model against the alternative models we use a version of the Vuong-test which is adjusted for heteroschedacity and autocorrelation (HAC) (Calvet and Fisher, 2004). For each of the alternative models the null-hypothesis is that this model and the level-MSM model of order $K = 9$ are equidistant from the data-generating process, measured in the Kullback-Leibler distance (Kullback and Leibler, 1951). In the classical Vuong-test it is assumed that the data-generating process is i.i.d., and then the log-ratio of the likelihoods for the two models will converge to a normal distribution with zero mean, and with a variance which is consistently estimated by the sample variance for the log-ratio of the likelihoods. Using the corresponding normal distribution one can then easily calculate a p-value under the null-hypothesis. In the HAC-adjusted version of this test, the data-generating process may have dependence, but the variables r_t should be identically distributed. This is satisfied if we assume that the data-generating process r_t is stationary.

The results of this test are presented in table 4.8. We observe that for the TBM3 data, the level-MSM model of order $K = 9$ is significantly closer to the data-generating process than any of the alternative models. For the NIBORM3 data the EGARCH model performs best, whereas the level-MSM model performs better than the standard level-GARCH model and jump-diffusions.

4.5 Concluding remarks

In this paper we have introduced a multifractal model for short-term interest rates. The model combines the well-established level effect described in Longstaff et al. (1992) with the discretized multifractal model of Calvet and Fisher (2001). In a comparison with level-GARCH, level-EGARCH and jump-diffusions, we find that this model well describes the fluctuations of the TBM3 and NIBORM3 time series. The main result of this work is that the level-MSM outperforms all alternatives for the TBM3 data. This motivates further research on multifractal modeling of short-term interest rates, in particular an out-of-sample analysis of the level-MSM model.

Model	dim(θ)	log L	BIC	Vuong	Hac-adj.
			U.S. Treasury Bill M3		
MSM(9)	6	22955.80	-3.2358		
GARCH	6	22873.03	-3.2241	$-4.064(< 0.001)$	$-3.477(< 0.001)$
EGARCH	7	22936.53	-3.2324	$-1.420(0.078)$	$-1.361(0.087)$
Jump-diffusion	7	22597.89	-3.1839	$-9.878(< 0.001)$	$-9.240(< 0.001)$
			NIBORM3		
MSM(9)	6	8804.44	-2.8181		
GARCH	6	8783.27	-2.8113	$-1.099(0.136)$	$-1.086(0.139)$
EGARCH	7	8828.90	-2.8245	1.231(0.891)	1.470 (0.929)
Jump-diffusion	7	8579.61	-2.7430	$-6.964(< 0.001)$	$-5.765(< 0.001)$

Table 4.8: The Vuong-column reports the test-statistic for differences in BIC. The nullhypothesis is that the multifractal and the alternative model are equally good, with the alternative hypothesis being that the multifractal model performs best. The HAC-adjusted column adjusts for heteroscedasticity and autocorrelation in the addends. Corresponding p-values are in parentheses.

It is also interesting to note in the level-GARCH model the parameter estimates (for both the TBM3 data and the NIBORM3 data) fall outside the covariance-stationarity region. As a result the level-GARCH model has a wild volatility pattern, which does not seem to be an accurate description of the interest-rate data. This confirms the results of e.g. Andersen and Lund (1997).

Acknowledgment. This project was partly funded by *Sparebank 1 Nord-Norge* and the Norwegian Research Council.

5
Paper 2

Abstract. We present an approximated maximum likelihood method for the multifractal random walk processes of Bacry et al. (2001). The likelihood is computed using a Laplace approximation and a truncation in the dependency structure for the latent volatility. The procedure is implemented as a package in the R computer language. Its performance is tested on synthetic data and compared to an inference approach based on the generalized method of moments. The method is applied to estimate parameters for various financial stock indices.

5.1 Introduction

Multifractal models were first introduced in the 1960s by the so-called Russian school in turbulence theory (Obukhov, 1962; Kolmogorov, 1962). In turbulence, multifractality can be conceived as a weakening of the spatial self-similarity in the velocity field implicitly assumed in Kolmogorovs 1941 theory. This generalization is called the Kolmogorov-Obukhov model and entails modeling the spatial variability of the energy dissipation rate as a random measure with certain multiscaling properties. The Kolmogorov-Obukhov model is treated rigorously by Kahane (1985) and this construction is known as Gaussian multiplicative chaos. In recent years, multifractal random processes and multifractal random measures have received increased attention and are widely used in physics, geophysics, and complex systems theory. Examples include phenomena as diverse as internet traffic (Riedi et al., 1999), geomagnetic activity (Rypdal and Rypdal, 2010, 2011), and rainfall patterns (Pathirana and Herath, 2002). In addition, multifractal processes provide natural models for the long-range volatility persistence observed in financial time series. This was discovered by Ghashghaie et al. (1996) and Mandelbrot et al. (1997), and since the late 1990s much work has been done on multifractal modeling of financial markets (Calvet and Fisher, 2001; Di Matteo, 2007). Logarithmic returns of assets are modeled as $x_t = X(t + \Delta t) - X(t)$, where $X(t)$ are continuous-time processes with stationary increments and multifractal scaling. The latter means that the moments of $X(t)$ are power laws as functions of time,

$$\mathbb{E}[|X(t)|^q] \sim t^{\zeta(q)}, \tag{5.1}$$

either in some interval $t \in (0, R)$ or asymptotically as $t \to 0$. Here $\mathbb{E}$ denotes expectation value. The scaling function $\zeta(q)$ is linear for self-similar processes, but may in general be concave. Processes satisfying (5.1) with strictly concave scaling functions are generally referred to as multifractal. Two well-known stylized facts of financial time series are that log returns are uncorrelated and non-Gaussian. Based on this, Mandelbrot (1963) deduced that if prices are described as self-similar processes, then these processes must be so- called Lévy flights, i.e., α-stable Lévy processes with $\alpha < 2$. However, if one allows nonlinear scaling functions, then one can maintain uncorrelated log returns by simply imposing the condition $\zeta(2) = 1$. The concave shape of $\zeta(q)$ implies that the variables $X(t)$ are increasingly leptokurtic with decreasing t, and consequently non-Gaussian. Moreover, as opposed to Lévy flights, multifractal processes have strongly dependent increments and can therefore describe a third stylized fact of financial time series, namely, volatility clustering. Notwithstanding that multifractal processes provide accurate and parsimonious descriptions of temporal financial fluctuations, the models are rarely implemented for forecasting and risk analysis in financial institutions. This is partially due to a lack of accurate, stable, and efficient inference methods for multifractal processes. Parameter estimation has so far mostly been made using various moment-based estimators, such as the generalized method of moments (GMM). Alternatively, one can fit the estimated scaling functions to theoretical expressions of $\zeta(q)$. However, as pointed out, e.g., in Lux (2003) and Chapman et al. (2005), the standard estimators of scaling exponents have large mean square errors for

time series of length comparable to those typically available in econometrics. An exception to the statements above is the Markov switching multifractal (MSM) model (Calvet and Fisher, 2001) where maximum likelihood estimation is feasible. In discrete time, the MSM model implies that the increments x_t are described by a stochastic volatility model of the form

$$x_t = \sigma\sqrt{M_t}\varepsilon_t. \tag{5.2}$$

Here, $\varepsilon_t \overset{\mathrm{d}}{\sim} \mathcal{N}(0,1)$ are independent variables and the volatility is a product of the form

$$M_t = M_{t,1}M_{t,2}\cdots M_{t,K},$$

where (for each time step t) $M_{t,k}$ are updated from a distribution M with a probability $\gamma_k = 1-(1-\gamma_1)^{b^{k-1}}$. In this model, however, maximum likelihood estimation is only possible in the case where M is defined on a discrete state space, and there is a limitation on the magnitude of K which should not exceed ≈ 10 (Lux, 2008). These restrictions not only limit flexibility with respect to the distribution of returns, but also the possible range in the volatility dependency. This paper concerns parametric inference for the multifractal random walk (MRW) introduced by Bacry et al. (2001). The increment process $x_t = X(t+\Delta t) - X(t)$ is still a discrete-time stochastic process described by (5.2), but now the volatility is modeled as $M_t = c\exp(h_t)$, where h_t is a stationary and centered Gaussian process with the covariance structure

$$\mathrm{Cov}(h_t, h_s) = \lambda^2 \log^+ \frac{T}{(|t-s|+1)\Delta t}\,, \tag{5.3}$$

where $\log^+ a = \max\{\log a, 0\}$. Here, $T > 0$ is called the correlation range [1] and $\lambda > 0$ is called the intermittency parameter. The constant c ensures normalization and is chosen so that $1/c = \mathbb{E}[\exp(h_t)]$. We denote $R = T/\Delta t$. Let $\theta = (\lambda, \sigma, R)$ denote the parameter vector and $y = (y_1, ..., y_n) \in \mathbb{R}^n$ denote a fixed time series. The main result of this paper is the development of a method for efficiently computing approximations to the likelihood function

$$\mathcal{L}(\theta|y) = p_x(y|\theta),$$

where $p_x(\cdot|\theta)$ is the probability density function of a random vector $x = (x_1, ..., x_n)$ produced by the MRW model with parameters θ. Using the likelihood function, parameters can be determined by means of the maximum likelihood (ML) estimator:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta|y).$$

Our method exploits that the discrete MRW model has a construction similar to simple volatility (SV) models. The distinguishing feature is that the processes h_t are autoregressive in SV models. By truncating the dependency structure in the logarithmic volatility h_t, the computation of the likelihood function is mapped onto a similar problem for SV models,

[1] In turbulence, T corresponds to the integral scale.

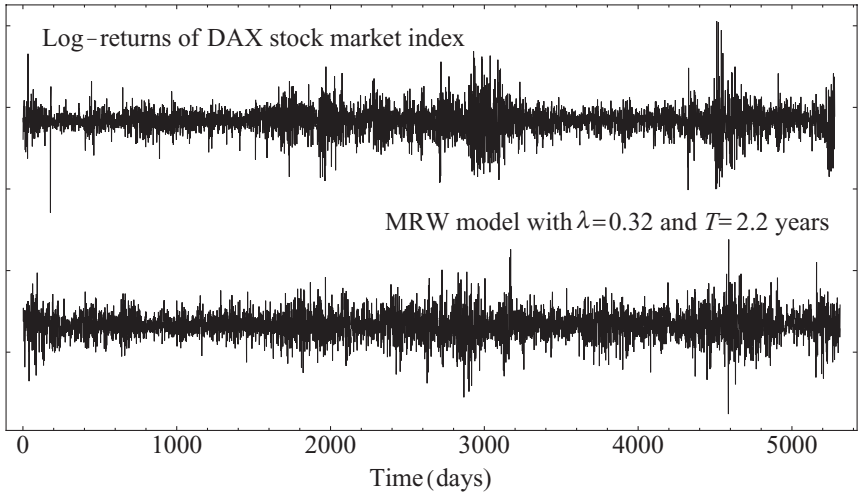


Figure 5.1: The top figure shows the daily log returns of the German DAX index for the time period 26 November 1990 to 25 November 2011. The standard deviation of the data is normalized to unity. For $\tau = 500$, the ML estimates are $\lambda = 0.32$ and $T = 2.2$ years. The lower figure shows a simulation of the MRW model x_t with the estimated parameters.

and hence existing techniques for further approximations are available. The present paper presents results on ML estimation for multifractal models with continuous state spaces for the volatility. Such estimates may be of great practical importance, since accurate parameter estimation is essential for volatility forecasts and risk estimates. In the MRW model, this degree of accuracy is particularly important for the intermittency parameter λ, which determines the peakedness of the return distributions on all time scales. In applications other than finance, accurate estimates of λ can be used as supplements to the empirical scaling functions, and thereby the ML estimator can provide a tool for quantifying multifractality in data.

The paper is organized as follows: In section 5.2, we briefly explain the construction of the continuous-time process $X(t)$ for which the model given by (5.2) and (5.3) is a discretization. There exists a large class of multifractal processes which are related to a construction known as infinitely divisible cascades (IDCs). In general, the random walk models associated with IDC processes have logarithmic volatility with infinitely divisible distributions, and the MRW model considered in this paper is a discrete approximation to the random walk model obtained in the special case when the logarithmic volatility is Gaussian. Section 5.3 contains the procedure for approximated ML estimation in the MRW model. In section 5.4, we test the estimator by first applying it to various stock market indices, and then by running a small Monte Carlo study. The results are compared with the GMM method used in Bacry et al. (2008). We finally remark that the methods presented in this paper have been implemented in a package for the R statistical software. This package is available online (Løvsletten and Rypdal, 2012b).

5.2 Motivation of the model

There exist several popular models for multifractal stochastic processes with uncorrelated increments. All of these models can be written either of the form $X(t) = B(A(t))$, where $B(t)$ is a Brownian motion and $A(t) = m([0,t])$ is the distribution function of a multifractal random measure m on the time axis, or as

$$X(t) = \lim_{r\to 0} \int_0^t \sqrt{A_r(t)} dB(t),$$

where $A_r(t) \to A(t)$ as $r \to 0$. The meaning of $A_r(t)$ is discussed below. These two types of constructions are equivalent as long as $B(t)$ is a Brownian motion. (This is not the case for fractional Brownian motions with $H \neq 1/2$.) The differences between the various multifractal models are then related to the construction of the random measure m. The log-normal MRW model is, on one hand, based on a particular construction of m known as multiplicative chaos, and, on the other hand, it can be seen as a special case of the more general IDC constructions. In multiplicative chaos, which was developed rigorously in Kahane (1985), one considers a sequence m_n of measures defined via random densities of the form

$$dm_n(t) = c_n \exp(h_n(t))dt,$$

where $1/c_n = \mathbb{E}[\exp(h_n(t))]$, and $h_n(t)$ are centered Gaussian processes with covariance structures $g_n(t,s) = \mathrm{cov}[h_n(t), h_n(s)]$ that converge to some expression $g(t,s)$ in the limit $n \to 0$. Kahane (1985) showed that if g is σ-positive, meaning that there are positive and positive definite functions $K_m(t,s)$ such that

$$g_n(t,s) = \sum_{m=1}^{n} K_m(t,s)\,,$$

then the sequence m_n converges weakly to a Borel measure m which depends only on the function $g(t,s)$. One can therefore informally think of m as being of the form $dm(t) = c\exp(h(t)dt$, where $1/c = \mathbb{E}[c\exp(h(t)]$, and $h(t)$ is a Gaussian process with covariance $g(t,s)$. Then, if one makes the choice

$$g(t,s) = \lambda^2 \log^+ \frac{R}{|t-s|}, \tag{5.4}$$

then one easily obtains the relation $h(at) \overset{\mathrm{d}}{=} h(t) + \Omega(a)$, where $\Omega(a)$ are independent of $h(t)$ and distributed according to $\Omega(a) \overset{\mathrm{d}}{\sim} \mathcal{N}(0, -\lambda^2 \log a)$. It follows that for $t < R$ and $0 < a < 1$, we have the scaling relation

$$m\left([0,at]\right) \overset{\mathrm{d}}{=} M(a)m([0,t]), \tag{5.5}$$

with $\log M(a) \overset{\mathrm{d}}{\sim} \mathcal{N}[(1+\lambda^2/2)\log a, -\lambda^2 \log a]$. See proposition 3.3 in Robert and Vargas (2010) for a rigorous proof of (5.5), and see example 2.3 of the same paper for a verification

that the function $g(t,s)$ in (5.4) is σ-positive. By using the well-known formula for the q-th moments of log-normal variables together with (5.5), we easily verify the multifractality of the process $A(t) = m([0,t])$: Denote $C_q = \mathbb{E}|m([0,1])|^q$ and observe that

$$\mathbb{E}|m([0,t])|^q = C_q \mathbb{E}M(t)^q = C_q t^{\zeta_A(q)},$$

where $\zeta_A(q) = q(1+\lambda^2/2) - \lambda^2 q^2/2$. Since a Brownian motion is self-similar with $H = 1/2$, the scaling function of $X(t)$ is given by $\zeta(q) = \zeta_A(q/2)$.

Alternatively, the model defined by (5.2) and (5.3) can be motivated by considering the more general class of IDC models. Here we briefly mention the main ideas and results in this theory, and we refer to (Bacry and Muzy, 2003; Muzy and Bacry, 2002) for details. At the base of this construction is an object called an independently scattered, infinitely divisible, random measure $P(dt,dr)$ defined on the half plane $S^+ = \{(t,r) \in \mathbb{R}^2 | r > 0\}$. The defining properties of the random measure P are (1) for any measurable set $A \subset S^+$, the random variable $P(A)$ is infinitely divisible with characteristic function

$$\varphi_{P(A)}(q) = \exp(\psi(q)\mu(A)),$$

where $\mu(dt,dr) = r^{-2}dtdr$; and (2) for any finite sequence $A_k \subset S^+$ of disjoint and measurable sets, the corresponding random variables $P(A_k)$ are independent. If we assume that $\psi'(0) = 0$, then the random measure P induces a family of centered and stationary stochastic processes through the equation

$$h_r(t) = P(A(r,t)),$$

where $A(r,t)$ are conelike domains defined by

$$A(r,t) = \{(t',r') \in S^+ | r' \geq r, |t'-t| \leq f(r)/2\},$$

with $f(r) = r$ for $r \leq R$, and $f(r) = R$ for $r > R$. The time correlations in the processes $h_r(t)$ are characterized by the functions

$$\rho_r(t) := \mu\left[A(0,r) \cap A(t,r)\right] = \begin{cases} \log\frac{R}{r} + 1 - \frac{t}{r} & \text{if } t < r, \\ \log^+\frac{R}{t} & \text{if } t \geq r. \end{cases}$$

In fact, the covariance of $h_r(t)$ is given by

$$\mathrm{cov}[h_r(t), h_r(s)] = \lambda^2 \rho_r(|t-s|),$$

where $\lambda^2 = -\psi''(0)$. Random measures are defined by $dm_r(t) = c_r \exp(h_r(t))dt$, where $1/c_r = \mathbb{E}[\exp(h_r(t))]$. The corresponding distribution functions are $Ar(t) = m_r([0,t])$ and corresponding random walks are

$$X_r(t) = \int_0^t \sqrt{A_r(t)} dB(t),$$

By using the relation $\rho_{ar}(at) = -\log a + \rho_r(t)$, one can show that

$$h_{ar}(t) \stackrel{\mathrm{d}}{=} h_r(t) + \Omega(a),$$

for $a \in (0,1)$ and $t \leq R$, where $\Omega(a)$ are independent of $h_r(t)$ and have characteristic functions $\varphi_{\Omega(a)}(q) = \exp(-\psi(q)\log a)$. Consequently, the limit process $X(t) = \lim_{r\to 0} X_r(t)$ has scaling function

$$\zeta(q) = (1 + \psi(-i))q/2 - \psi(-iq/2).$$

In the case that $h_r(t)$ are Gaussian, i.e., $\psi(q) = -\lambda^2 q^2/2$, the covariance is of the form

$$\mathrm{Cov}(h_r(t), h_r(s)) = \lambda^2 \log^+ \frac{R}{|t-s|}$$

for $|t-s| > r$, and hence it can be approximated by the process defined by (5.3). In this case, the scaling function is

$$\zeta(q) = (1 + \lambda^2/2)q/2 - \lambda^2 q^2/8.$$

We note that for $\lambda = 0$, the process $X(t)$ is reduced to a Brownian motion and $\zeta(q) = q/2$.

We point out that this paper presents an ML estimator for the discrete-time process x_t defined by (5.2) and (5.3). This is sufficient for the purpose of modeling and forecasting volatility in financial time series, since the discrete-time MRW model is directly comparable to generalized autoregressive conditional heteroskedasicity (GARCH)-type models. In other applications, such as modeling the velocity field in turbulence, one is interested in the continuous-time process $X(t)$. Since x_t is an approximation to the continuous-time process $X(t)$, our method can also be interpreted as an estimator for this process. In this case, one must be aware that the increment process $X(t+\Delta t) - X(t)$ is not proportional (in law) to $\exp(h_{\Delta t}(t))\varepsilon_t$, and that this is only an approximation in the limit $\lambda^2 \ll 1$; see Appendix A.1 in Bacry et al. (2008). In the case of strong intermittency, the estimator for the continuous-time process $X(t)$ may therefore depend significantly on the time scale Δt for which the data is sampled. An analysis of how our method performs as an estimator for $X(t)$ will require extensive Monte Carlo simulations (with varying λ and Δt), and this is beyond the scope of this paper. We also remark that in some applications it is relevant to estimate the parameters of the measure $dm(t)$, for instance when modeling the energy dissipation fields in turbulence. In the discrete-time approximation, this corresponds to the process $\exp(h_t)$, where h_t is described by (5.3). Since h_t is Gaussian, this problem is much easier than the one considered in this paper. The ML estimator for $\exp(h_t)$ can be constructed using standard methods (McLeod et al., 2007) and no approximations are required.

5.3 Approximated maximum likelihood

In this section, we explain our method of approximated maximum likelihood estimation. Let x_t and h_t be the processes defined by (5.2) and (5.3). Denote $x = (x_1, ..., x_n)$ and

$h = (h_1, ..., h_n)$. The first step is to write

$$p_x(x) = \int_{\mathbb{R}^n} p_{x,h}(x,h)\, dh = \int_{\mathbb{R}^n} p_{x|h}(x|h) p_h(h) dh. \tag{5.6}$$

The first factor $p_{x|h}(x|h)$ in the integrand is computed by noting that when conditioned on h, the variables $x_1, \ldots, x_n$ are independent and Gaussian. In fact,

$$\begin{aligned} \log p_{x|h}(x,h) &= \sum_{t=1}^{n} \log p_{x_t|h_t}(x_t, h_t) \\ &= -n \log \sqrt{2\pi}\sigma + \sum_{t=1}^{n} \Big(-\frac{h_t}{2} - \frac{x_t^2}{2\sigma^2 \exp(h_t)} \Big) . \end{aligned} \tag{5.7}$$

For the second factor $p_h(h)$, we use that h_t is a centered Gaussian process with a specified covariance structure, $\mathrm{cov}(h_t, h_s) = \gamma(|t-s|)$. First, we decompose the density into one-dimensional marginals,

$$p_h(h) = \prod_{t=1}^{n} p_{h_t|h_{1:t-1}}(h_t)\,, \tag{5.8}$$

where we have used the notation

$$h_{n:m} = \begin{cases} (h_n, h_{n+1}, \ldots, h_m) & \text{for } m > n \\ (h_n, h_{n-1}, \ldots, h_m) & \text{for } n > m \end{cases} .$$

Denote by Γ_t the covariance matrix of the vector $h_{1:t}$, and let $\gamma_{1:t} = [\gamma(1), \ldots, \gamma(t)]$. The covariance matrix can be written of the block form

$$\Gamma_t = \begin{pmatrix} \Gamma_{t-1} & \gamma_{1:t-1} \\ \gamma_{1:t-1}^T & \gamma(0) \end{pmatrix} .$$

By performing standard computations of conditional marginals in multivariable Gaussian distributions, we deduce that $h_t|h_{1:t-1}$ is a Gaussian with mean $m_t = \gamma_{1:t-1}\Gamma_{t-1}^{-1} h_{(t-1):1}^T$ and variance $S_t^2 = \gamma(0) - \gamma_{1:t-1}\Gamma_{t-1}^{-1}\gamma_{1:t-1}^T$. As usual, it is convenient to introduce vectors $\phi^{(t)}$ defined by $\phi^{(t)}\Gamma_t = \gamma_{1:t}$. This allows us to write the mean as $m_t = \phi^{(t-1)} h_{(t-1):1}^T$ and the variance as $S_t^2 = \gamma(0) - \phi^{(t-1)}\gamma_{1:t-1}^T$. Then from equation (5.8) we have

$$\log p_h(h) = -n \log \sqrt{2\pi} - \sum_{t=1}^{n} \log S_t - \sum_{t=1}^{n} \frac{(h_t - \phi^{(t-1)} h_{(t-1):1}^T)^2}{2S_t^2} . \tag{5.9}$$

Combining equation (5.9) with equation (5.7) we get an expression for the full likelihood:

$$\begin{aligned} p_{x,h}(x,h) &= -n \log 2\pi\sigma + \sum_{t=1}^{n} \Big(-\frac{h_t}{2} - \frac{x_t^2}{2\sigma^2 e^{h_t}} \Big) \\ &- \sum_{t=1}^{n} \log S_t - \sum_{t=1}^{n} \frac{(h_t - \phi^{(t-1)} h_{(t-1):1}^T)^2}{2S_t^2} . \end{aligned} \tag{5.10}$$

We keep in mind that c depends on R and λ through the relation $1/c = \mathbb{E}[\exp(h_t)] = R^{\lambda^2/2}$.

Approximation 1: By comparing covariances, the process h_t can be written as

$$h_t = \phi_1^{(t-1)} h_{t-1} + \cdots + \phi_{t-1}^{(t-1)} h_1 + w_t \,, \tag{5.11}$$

where w_t are independent Gaussian variables with zero mean and variances equal to S_t^2. As approximations to h_t, we can consider processes obtained by truncating the sum in (5.11). We fix a parameter $\tau \in \mathbb{N}$, and for $t >$ we replace (5.11) with

$$h_t = \phi_1^{(t-1)} h_{t-1} + \cdots + \phi_\tau^{(t-1)} h_{t-\tau} + w_t^{(\tau)} \,, \tag{5.12}$$

where $w_t^{(\tau)}$ are independent Gaussian variables with zero mean and variances equal to $S_{\tau+1}^2$. Note that $h_t | h_{t-1:t-\tau}$ in (5.11) has the same distribution as obtained from (5.12), namely, a Gaussian with mean $m_t = \phi^{(\tau)} h_{(t-1):t-\tau}^T$ and variance $S_{\tau+1}^2$. In effect, we have approximated the distribution of $h_t | h_{t-1:1}$ by truncating the dependency after a lag τ. As a result of this, the approximation given by (5.9) becomes

$$\begin{aligned} \log p_h(h) &= -n \log \sqrt{2\pi} - \sum_{t=1}^{n} \log S_t - \sum_{t=1}^{\tau} \frac{(h_t - \phi^{(t-1)} h_{(t-1):1}^T)^2}{2 S_t^2} \\ &+ \sum_{t=\tau+1}^{n} \frac{(h_t - \phi^{(\tau)} h_{t-1:\tau}^T)^2}{2 S_\tau^2} \,. \end{aligned} \tag{5.13}$$

In order to compute the expression in (5.13), we need to solve the equations

$$\phi^{(t)} \Gamma_t = \gamma_{1:t} \,, \;\; t = 1, \ldots, \tau \,.$$

This is done efficiently using the Durbin-Levinson algorithm (Levinson, 1946; Trench, 1964). We remark that for $\tau = n$, the expression in (5.13) is exact.

Approximation 2: The second approximation is the so-called Laplace's method, which is frequently used for the approximation of likelihoods in SV models; see, e.g., (Skaug and Yu, 2009; Martino et al., 2011). We write (5.6) of the form

$$p_x(x) = \int_{\mathbb{R}^n} e^{n f_x(h)} dh \,, \tag{5.14}$$

where

$$\begin{aligned} f_x(h) &= \frac{1}{n} \log p_{x,h}(x, h) \\ &= \frac{1}{n} \sum_{t=1}^{n} \log p_{x_t|h_t}(x_t) + \frac{1}{n} \sum_{t=1}^{n} \log p_{h_t|h_{1:t-1}}(h_t) \,. \end{aligned} \tag{5.15}$$

The Laplace method is to assume that $f_x(h)$ has a global maximum in $\mathbb{R}^n$, which we denote by h^*. When n is large, the contribution to the integral in (5.14) is concentrated around h^*, and therefore we make a second order Taylor approximation to $f_x(h)$ around this point. Since h^* is also a local maximum, we have

$$f_x(h) \approx \frac{1}{n} \log p_{x,h}(x, h^*) + \frac{1}{2n} (h - h^*) \, \Omega_x \, (h - h^*)^T \,,$$

where

$$\Omega_x = \frac{\partial^2 \log p_{x,h}(x, h^*)}{\partial h \, \partial h^T}$$

is the Hessian matrix of $f_x(h)$ at the point h^*. The approximation now reads

$$\begin{aligned} p_x(x) &\approx e^{f_x(h^*)} \int_{\mathbb{R}^n} e^{\frac{1}{2}(h-h^*)\,\Omega_x\,(h-h^*)^T} dh \\ &= (2\pi)^{n/2} \, |\det \Omega_x|^{1/2} \, p_{x,h}(x, h^*) \, . \end{aligned}$$

The maximum h^* is found by computing the partial derivatives of $f_x(h)$ with respect to h, setting them equal to zero, and solving the corresponding system of equations numerically using the algorithm df-sane (La Cruz et al., 2006), which is implemented in the R package BB (Varadhan and Gilbert, 2009). The matrix Ω_x is obtained using analytical expressions for the second derivatives. This matrix is band diagonal, with the bandwidth equal to the truncation parameter τ, and, in the R software, such matrices are efficiently stored and manipulated using the package MATRIX (Bates and Maechler, 2014).

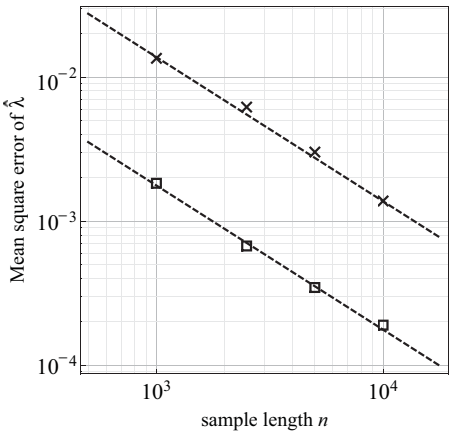


Figure 5.2: Double-logarithmic plot of the mean square errors as functions of sample length n for the ML estimator with $\tau = 100$ (squares) and the GMM estimator (crosses). The dotted lines have slopes equal to -1, i.e., the mean square errors decay roughly as $1/n$ for both estimators.

5.4 Estimator comparisons

In this section, the ML estimator is compared with a GMM approach which is similar to the one used in Bacry et al. (2008). This GMM version is essentially a least-squares fitting of the autocorrelation function for the logarithmic volatility, and we briefly explain this method in the following: Denote $m_t = \log x_t^2$, and observe that

$$m_t = h_t + y_t,$$

where $y_t = \log c + \log \varepsilon_t^2$ are independent and identically distributed. We can use the sample standard deviation to normalize m_t so that it has unit variance. Then, if we let $\mu_m = \mathbb{E}[m_t] = \mathbb{E}[y_t]$ denote the mean of m_t, the autocorrelation function (ACF) of m_t has the form

$$\begin{aligned} \mathrm{ACF}_m(t) &= \mathbb{E}[(m_1 - \mu_m)(m_{t+1} - \mu_m)] \\ &= \mathbb{E}h_1 h_{t+1} = \lambda^2 \log^+ \tfrac{R}{t+1}. \end{aligned}$$

For $t \leq R$, we have

$$\mathrm{ACF}_m(t) = \lambda^2 \log R - \lambda^2 \log(t+1)$$

and $\log R$ and λ can be found by linear regression of the autocorrelation function versus $\log(t+1)$. We begin testing the approximated ML estimator by applying it to various stock market indices. An example is shown in figure 5.1, where we have plotted a realization of the MRW model with parameters estimated from the German DAX index. For comparison, the data are plotted together with the log-returns of the DAX index. In table 5.1 we present estimates for six different stock indices. We use daily log returns and in all of the estimates the truncation parameter is set to $\tau = 500$ days. We observe that the intermittency parameter τ varies from 0.29 to 0.37 for the different indices and time periods. We also observe that the correlation range parameter T varies by roughly one order of magnitude, in the range $1.4 - 12.2$ years. If we compare with the GMM, we see that for all of the indices, the estimates of λ are lower using the ML method. For the parameter T, the estimates using ML and GMM are more or less consistent, but with quite large variations between the two estimators.

To further test the performance of the proposed ML estimator, we run a small-sample Monte Carlo study. We have used three different sample lengths, $n \in \{2500, 5000, 10000\}$, and for each sample length n we simulated 500 sample realizations. The parameter vector considered is $\lambda = 0.35, \sigma = 1$, and $R = 2000$. For the truncation parameter, we have

		ML		GMM	
Index	Time period	λ	T (years)	λ	T (years)
CAC 40	$1990 - 2011$	0.29	2.8	0.36	2.5
S & P 500	$1950 - 2011$	0.32	12.2	0.36	10.2
DAX	$1990 - 2011$	0.32	2.2	0.44	3.1
Nikkei 225	$1984 - 2011$	0.36	1.4	0.40	3.0
Hang Seng	$1986 - 2011$	0.37	2.8	0.44	2.5
FTSE 100	$1984 - 2011$	0.28	4.2	0.36	2.9

Table 5.1: Estimated parameters for the log returns of various stock market indices. Prior to the analysis, the sample standard deviation of each data set is normalized to unity. All ML estimates are run with $\tau = 500$ and the GMM estimates are performed with a maximum time lag $t_{\max} = 500$ days in the autocorrelation function of $m_t = \log x_t^2$. The analyzed data is retrieved from http://finance.yahoo.com/.

Table 5.2: The results of a Monte Carlo study of the ML and GMM estimator. The parameters in the simulations are $\lambda = 0.35$ and $R = 2000$ (i.e. $\log R = 7.6$). In the GMM estimator, we have we have used a maximum time lag $t_{\max} = 500$ days in the autocorrelation function of $m_t = \log x_t^2$. The reported values are the mean estimates together with the standard deviations (in brackets).

		ML			GMM		
n	τ	λ	$\log R$	σ	λ	$\log R$	σ
2500	10	0.31 (0.03)	6.87 (3.41)	0.97 (0.19)	0.34 (0.08)	6.11 (0.76)	0.97 (0.19)
	50	0.34 (0.03)	6.47 (1.73)	0.97 (0.19)			
	100	0.34 (0.03)	6.35 (1.67)	0.97 (0.19)			
5000	10	0.30 (0.03)	5.58 (2.18)	0.98 (0.14)	0.35 (0.05)	6.69 (0.96)	0.981 (0.15)
	50	0.34 (0.02)	7.02 (1.44)	0.98 (0.14)			
	100	0.34 (0.02)	6.87 (1.31)	0.97 (0.14)			
10000	10	0.30 (0.02)	9.10 (1.80)	0.98 (0.10)	0.35 (0.04)	7.11 (0.92)	0.98 (0.10)
	50	0.34 (0.01)	7.37 (1.24)	0.98 (0.10)			
	100	0.34 (0.01)	7.21 (1.16)	0.98 (0.10)			

considered the cases $\tau \in \{10, 50, 100\}$, and in the GMM method, we use a maximum time lag $t_{\max} = 500$ days in the autocorrelation function of $m_t = \log x_t^2$. The results are presented in table 5.2. For both the GMM and the ML methods, the estimates of R are highly unstable. This is also pointed out in Bacry et al. (2008). However, the processes x_t only depend on the R through expressions of the form $\lambda^2 \log R$. Therefore, in order to have an estimator which is comparable to λ, we should consider the variable $\log R$. The estimators of $\log R$ behave reasonably well, even though there are significant mean square errors and some bias. We see that both the ML and GMM methods underestimate $\log R$ and that the errors are roughly the same for the two estimators. On the other hand, we observe that the ML estimates of λ have standard deviations which are much smaller than the corresponding standard deviation for the GMM estimate, especially for $\tau = 100$. This can also be seen from figures 5.3 and 5.4, where the probability density functions (PDFs) for the GMM estimates and the ML estimates are presented. Based on this, we conclude

that the ML estimator performs better than the GMM. Moreover, if one allows longer computing times, then the truncation parameter τ can be increased to obtain even more accurate estimates. For a time series of $n = 10^4$ data points, an ML estimate with $\tau = 500$ takes a few minutes on a personal computer. In figure 5.2, we have plotted the mean square errors (MSEs) $\mathbb{E}[(\hat{\lambda} - \lambda)^2]$ for the ML estimator, with $\tau = 100$, and the GMM estimator. We see that for both of the estimators, the MSE is roughly inversely proportional to the sample length. However, from table 5.2, we see that there is a slight negative bias in $\hat{\lambda}$ for the ML estimator. This bias decreases with increasing τ, and we suspect the estimator to be asymptotically unbiased in the limit $\tau = n \to \infty$.

5.5 Concluding remarks

In this paper, we have presented an approximate ML estimator for MRW processes. The method is implemented and tested in a Monte Carlo study, and the results show significant improvements over existing methods for the intermittency parameter λ. The methods of this paper represent a suitable starting point for two important generalizations. The first generalization is to allow for correlated innovations, for instance by letting ε_t be a fractional Gaussian noise. This has several important applications, for instance in the modeling of geomagnetic activity and electricity spot prices (Malo, 2006). Another interesting generalization is to consider the non-Gaussian IDC models referred to in section 5.2. We also point out that the techniques presented in section 5.3 can be used to calculate conditional densities of the form $p(x_{t+1}, ..., x_{t+s}|x_1, ..., x_t)$. At time t, such an expression provides a complete forecast over the next s time steps. Forecasting and risk analysis based on the MRW model and the methods in this paper is a promising topic that will be pursued in future work.

Acknowledgment

This project was supported by Sparebank 1 Nord-Norge and the Norwegian Research Council (Project No. 208125). We thank K. Rypdal and the anonymous referee for useful comments and suggestions.

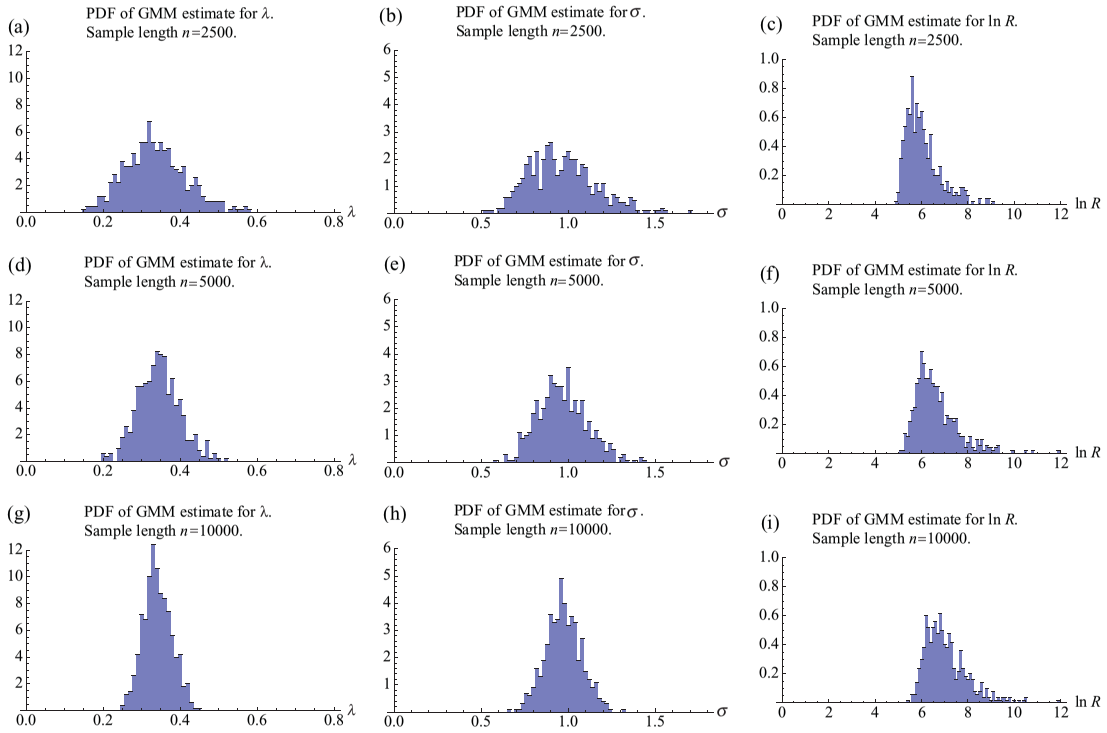


Figure 5.3: The results of the Monte Carlo study for the GMM estimator explained in section 5.3. The figures show the estimated probability density functions (PDFs) for the estimators based on 500 realizations of the process. The parameters are $\lambda = 0.35$, $\sigma = 1$, and $R = 2000$ (i.e., $\log R = 7.6$). In (a)-(c), the sample lengths are $n = 2500$; in (d)-(f), the sample lengths are $n = 5000$; and in (g)-(i), the sample lengths are $n = 10000$. The means and standard deviations of the estimators are reported in table 5.2.

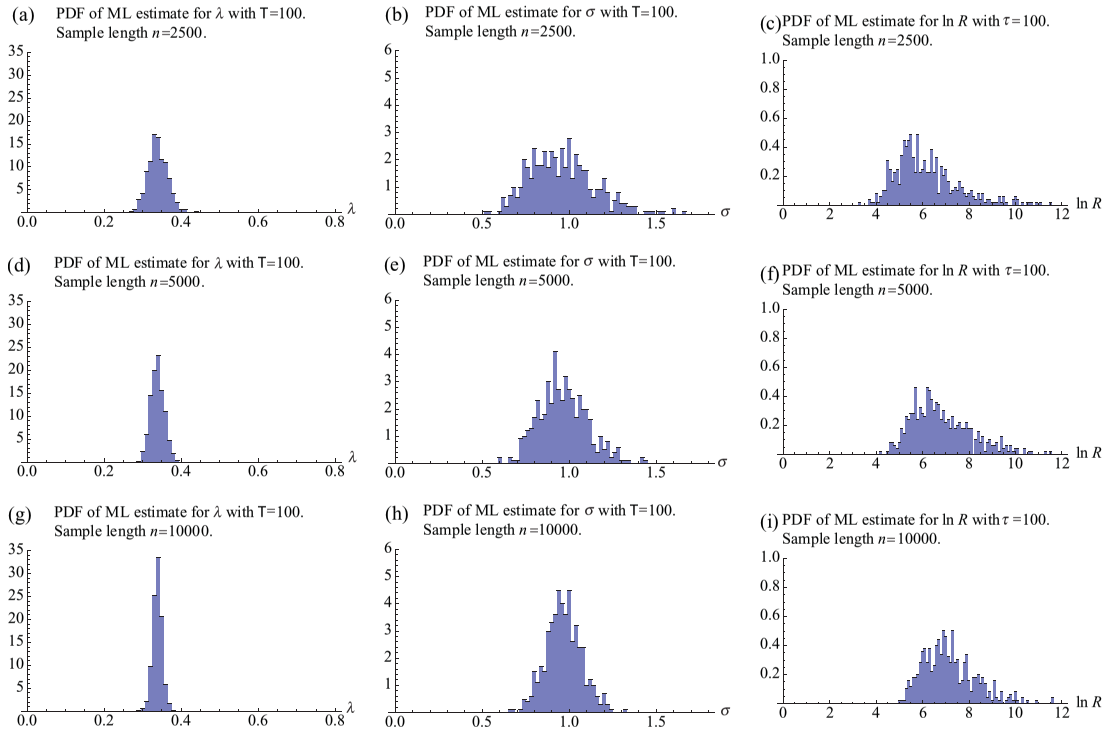


Figure 5.4: The results of the Monte Carlo study for the ML estimator with $\tau = 100$. The figures show the estimated probability density functions (PDFs) for the estimators based on 500 realizations of the process. The parameters are $\lambda = 0.35$, $\sigma = 1$, and $R = 2000$ (i.e., $\log R = 7.6$). In (a)-(c), the sample lengths are $n = 2500$; in (d)-(f), the sample lengths are $n = 5000$; and in (g)-(i), the sample lengths are $n = 10000$. The means and standard deviations of the estimators are reported in table 5.2.

6
Paper 3

Abstract. We discuss stochastic modeling of volatility persistence and anti-correlations in electricity spot prices, and for this purpose we present two mean-reverting versions of the multifractal random walk (MRW). In the first model the anti-correlations are modeled in the same way as in an Ornstein-Uhlenbeck process, i.e. via a drift (damping) term, and in the second model the anti-correlations are included by letting the innovations in the MRW model be fractional Gaussian noise with $H < 1/2$. For both models we present approximate maximum likelihood methods, and we apply these methods to estimate the parameters for the spot prices in the Nordic electricity market. The maximum likelihood estimates show that electricity spot prices are characterized by scaling exponents that are significantly different from the corresponding exponents in stock markets, confirming the exceptional nature of the electricity market. In order to compare the damped MRW model with the fractional MRW model we use ensemble simulations and wavelet-based variograms, and we observe that certain features of the spot prices are better described by the damped MRW model. The characteristic correlation time is estimated to approximately half a year.

6.1 Introduction

Since the 1990s, several of the world's electricity markets have been deregulated and re-organized in order to introduce competition and increase efficiency (Bye and Hope, 2005). In the de-regularized electricity markets there is trading of contracts for physical delivery of electric energy at a certain hour the next day. The price of such a contract is called the electricity spot price, and it is recorded for every hour of the year. The records of historical spot prices are extremely interesting from a scientific point of view, and a lot of effort has been devoted to describing and modeling their dynamics.

In this paper we analyze data from the Nordic electricity spot market (Nord Pool), which is known to exhibit a daily periodicity, a weekly periodicity and a one-year periodicity. These periodicities can be understood from a simple analysis of supply and demand. The consumption of electric energy is generally lower at night than during the day, and this causes a daily cycle in price. In the same way, the industry's demand for electric energy is lower during the weekend, and in the Nordic countries the demand for electric energy increases in winter due to the need for heating. In addition, the Nord Pool market is largely based on hydroelectric energy which supply has a seasonal variation.

On top of the periodic variations, the electricity spot prices have more unpredictable changes which are related to factors such as the weather, the distribution network, the level of industrial activity and general market dynamics. The aim of this paper is to present models for these non-periodic fluctuations. This task is important for several reasons. For instance, accurate quantification of the variability of spot prices is essential for correct pricing of futures and other electricity-based derivatives. It is also interesting to compare the characteristics of electricity spot prices with the prices of other commodities. It is observed that some of the "stylized facts" of electricity spot prices are similar to what is seen for stock and currency, whereas other properties are quite different. See e.g. Weron (2006).

Among the "stylized facts" of spot prices that *are* similar to other financial time series, are non-Gaussian distributions of log-returns and long-range volatility dependence. The latter is related to the term "volatility clustering", which refers to the observation of Mandelbrot (1963), that small price changes tend to be followed by small price changes, and that large price changes tend to be followed by large price changes. In stock markets it is known that volatility is closely tied to the trading volume (Lobato and Velasco, 2000), and it is reasonable to assume that the same is true for the electricity market. However, since electric energy is expensive to store, the delivered volume must equal the consumption. It is therefore somewhat surprising that electricity spot prices have such clear memory effects in volatility, and it shows that volatility clustering can be present even in markets with limited room for speculative behavior. On the other hand, the (non-periodc component of the) demand for electric energy is obviously not constant. It depends on a range of physical and economic factors, which certainly may contain long-range memory effects.

The "stylized facts" mentioned above can be described in a parsimonious way using multifractal models, and it has already been suggested by some authors (Norouzzadeh et al., 2007; Malo, 2006, 2009) to apply multifractal modeling to electricity spot prices.

Within this framework the logarithmic returns are modeled as $x_t = X(t + \Delta t) - X(t)$, where $X(t)$ is a multifractal process with stationary increments. Multifractality means that $X(t)$ is characterized by power-law scaling of its moments, i.e. $\mathbb{E}[|X(t)|^q] \sim t^{\zeta(q)}$, in some range $0 < t < T$ or asymptotically as $t \to 0$. For self-affine processes with finite variances, such as Brownian motion and fractional Brownian motion, the scaling function $\zeta(q)$ is linear and its slope equals the self-affinity exponent. However, in general, the scaling functions are concave and in the following we will refer to multifractal processes as those with strictly concave scaling functions.

The stochastic modeling presented in this paper is based on the (log-normal) MRW model. This model was introduced by Bacry et al. (2001), and it is often preferred over other multifractal models because of its simplicity and its desirable theoretical properties. For the purpose of modeling financial time series, the most important property of the MRW model (and other multifractal models) is that, even if the log-returns x_t are uncorrelated, the auto-correlation functions for their amptitudes $|x_t|$ decay as power-laws as functions of the time lag τ:

$$R_{|x|}(\tau) \stackrel{\text{def}}{=} \frac{\mathbb{E}[|x_t x_{t+\tau}|] - \mathbb{E}[|x_t|]^2}{\text{Var}[|x_t|]} \sim \tau^{-\lambda^2/4} \,. \tag{6.1}$$

Here λ is called the intermittency parameter. This property allows us to model volatility clustering without imposing any particular type of correlations between the returns themselves. Simultaneously, the concave shape of the scaling function $\zeta(q)$ implies that the kurtosis of $X(t)$ increases as $t \to 0$. This means that the return distributions are increasingly leptokurtic on shorter time scales, and consequently non-Gaussian.

In addition to "fat-tailed" distributions and slowly decaying volatility dependence, electricity spot prices have anti-correlated returns. This was first discovered by Weron (2000) and has since been confirmed by several authors. Anti-correlations are atypical in financial time series and would normally lead to arbitrage possibilities. (See e.g. Samuelson (1965) or chapter 2 in Mantegna and Stanley (2000).) However, since electricity is expensive to store, such arbitrage possibilities are hard to exploit, and hence anti-correlated returns may exist. There are mainly two ways in which these anti-correlations are modeled: The simplest approach is to consider models similar to Ornstein-Uhlenbeck (OU) processes (Uhlenbeck and Ornstein, 1930). The standard OU processes are defined via stochastic differential equations on the form

$$dX(t) = -\nu \big(X(t) - m\big)\, dt + \sigma\, dB(t) \,, \tag{6.2}$$

where $B(t)$ is a Brownian motion. The first term on the right-hand side of equation (7.3) is called the drift term (or the damping term), and for $\nu > 0$ this causes anti-correlations since it prevents $X(t)$ to diffuse far from its mean value m. One can choose initial conditions for $X(t)$ such that the OU process is stationary, and in this case the auto-correlation function for the returns $x_t = X(t + \Delta t) - X(t)$ has an exponential decay with a characteristic time scale $1/\nu$:

$$R_x(\tau) \stackrel{\text{def}}{=} \frac{\mathbb{E}[x_t x_{t+\tau}] - \mathbb{E}[x_t]^2}{\text{Var}[x_t]} \sim -\nu^2 e^{-\nu\tau} \,. \tag{6.3}$$

To include the effects of non-Gaussian log-returns and volatility clustering, the white noise $dB(t)$ can be replaced by other types of (uncorrelated) noise processes, such as jump processes with regime switching or (as in this paper) multifractal processes. Examples of non-Gaussian OU-type models for spot prices are given by Benth et al. (2007), Erlwein et al. (2010) and Weron et al. (2003).

An alternative to including mean reversal via a drift (damping) term is to describe the anti-correlations using Hurst exponents. This assumes that the logarithm (of the non-periodic component) of the spot price can be described by a process $X(t)$ with stationary increments and power-law scaling of the variogram. The latter means that the second moment of the differences $\delta X_\tau(t) = X(t+\tau) - X(t)$ is a power-law as a function of the lag τ, in which case H is defined by

$$\mathbb{E}[\delta X_\tau(t)^2] \propto \tau^{2H} \,. \tag{6.4}$$

For self-affine processes with stationary increments and finite variances, such as fractional Brownian motions, the Hurst exponents are equal to the self-affinity exponents.

This follows from the definition of self-affinity[1]: If $X(at) \overset{\mathrm{d}}{=} a^H X(t)$ for all $a > 0$, then $\delta X_\tau(t) \overset{\mathrm{d}}{=} X(\tau) \overset{\mathrm{d}}{\sim} \tau^H X(1)$. Taking the second moment gives equation (6.4). However, equation (6.4) does not imply that the qth moments, $\mathbb{E}[|\delta X_\tau(t)|^q]$, are power laws in τ. In particular, we do not need to assume self-affinity to use Hurst exponents. If the Hurst exponent is well-defined, $H \neq 1/2$ and $X(t)$ has stationary increments, then the auto-correlation function of the log-returns x_t decays as

$$R_x(\tau) \sim 2H(2H-1)\,\tau^{2H-2} \,. \tag{6.5}$$

See e.g. appendix A of Rypdal and Rypdal (2012). We note that the case $H < 1/2$ corresponds to algebraically decaying anti-correlations.

By the Wiener-Khinchin theorem (assuming stationarity of x_t), the auto-correlation function $R_x(\tau)$ and the power spectrum of x_t are related via a Fourier transform. Therefore, equation (6.4) implies that the power spectrum of $X(t)$ is a power law. In principle, the power-law exponents of the variogram, the auto-correlation function and the power spectrum contain the same information, namely the characteristics of the second-order statistics. On the other hand, the *estimators* corresponding to these quantities behave quite differently.

There exist several methods for estimating Hurst exponents, and various authors have reported different estimates in different electricity markets using different techniques. However, all reported values of H are in the interval $H < 1/2$. Weron and Przybylowicz (2000) applied R/S analysis to daily averaged prices from the California Power Exchange (CalPX), and reported a Hurst exponent around $H = 0.42$. Using an average wavelet coefficient (AWC) method, Simonsen (2003) has estimated a Hurst exponent $H = 0.41$ for the Nord Pool market. Anti-persistence in the Nord Pool data is also found by Erzgräber

[1]The notation $\overset{\mathrm{d}}{\sim}$ denotes equality in distribution for random variables, while $\overset{\mathrm{d}}{=}$ denotes equality in distribution for stochastic processes, i.e. that all finite-dimensional marginals are equal in distribution.

Authors	Market	Period
Weron and Przybylowicz (2000)	CalPX	Apr. 1998 - Jan. 2000
Simonsen (2003)	NordPool	May 1992 - Dec. 2000
Erzgräber et al. (2008)	NordPool	May 1992 - Dec. 1998
Norouzzadeh et al. (2007)	OMEL	Jan. 1998 - May 2006
	Method	Estimate
Weron and Przybylowicz (2000)	R/S	$H = 0.42$
Simonsen (2003)	AWC	$H = 0.41$
Erzgräber et al. (2008)	R/S	$H = 0.44$
Norouzzadeh et al. (2007)	DFA	$H = 0.16$

Table 6.1: A summary of some of the previously obtained results that have been obtained using Hurst analysis on electricity spot prices. The table includes the method of estimation, the time series analyzed and the obtained estimate of H. Further details are found in the cited references.

et al. (2008) by R/S-analysis. Using de-trended fluctuation analysis Norouzzadeh et al. (2007) have estimated the Hurst exponent of the Spanish electricity exchange, Compania O Peradora del Mercado de Electricidad (OMEL), to be $H = 0.16$. These results are summarized in table 6.1.

In this work we present two stochastic models for electricity spot prices. In the first model, which we will refer to as a damped MRW model, we consider a process of OU type, but where we have introduced stochastic volatility in the same way as in a standard MRW model. This type of model (which was first introduced in Rypdal and Rypdal (2011) to describe magnetic field fluctuations in the turbulent solar wind) has exponentially decaying anti-correlation of returns and algebraically decaying dependence in volatility.

The second model is referred to as a fractional MRW process. Here we use the same stochastic volatility as in the standard MRW model, but instead of a white noise, the process is driven by a fractional Gaussian noise with Hurst exponent $H < 1/2$.

The main results of this paper are presented in sections 6.3 and 6.4. In section 6.3 we discuss methods for modeling spot prices using mean-reverting multifractal processes, and in section 6.4 we derive approximate ML estimators. These methods are generalizations of a recently developed method for inference on standard (uncorrelated) MRW models (Løvsletten and Rypdal, 2012a). In section 7.4 we apply the ML estimators to the Nord Pool data and show that the estimates are consistent with the preliminary analysis presented in section 6.2.

As a part of our further analysis we compare our two models in order to determine which provides the better description of the Nord Pool data. To do this we construct ensembles of synthetic signals from the two models, with periodic components added, and compute wavelet-based variograms. These variograms are then compared to the corresponding variogram for the Nord Pool data. The details and results of this test are described in section

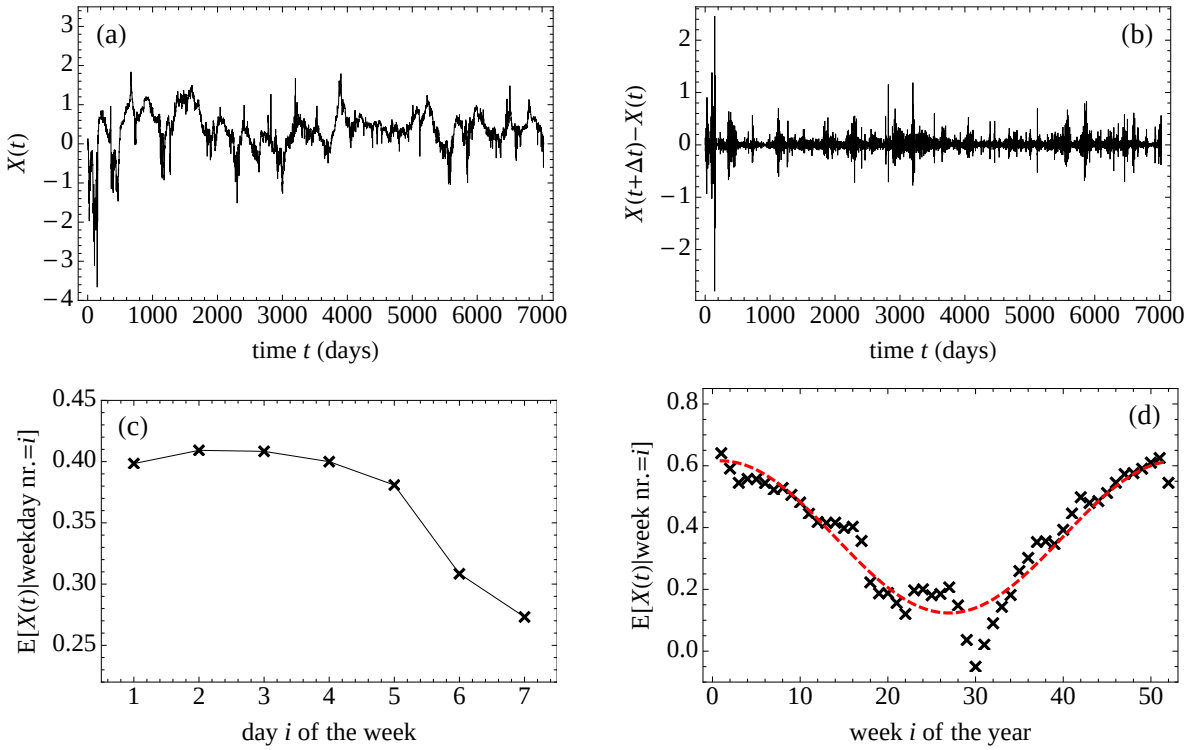


Figure 6.1: (a): The time series obtained from the spot prices by taking the logarithm and subtracting a linear trend. The plotted signal consists of daily means from May 4th 1992 to August 27th 2011. (b): The increments of the time series plotted in (a). (c): For the signal in (a) we plot the mean value conditioned on the weekday. This means that the first point is is obtained from the signal in (a) by taking the mean value over all Mondays, the second is obtained by taking the mean over all Tuesdays, and so on. (d): The weekly means of the signal in (a) conditioned on the week of the year. The dotted line shows a fitted sinusoidal oscillation with a period of one year and an amplitude 0.25.

6.6. In section 6.7 we give some concluding remarks.

6.2 Description of data and preliminary analysis

The data analyzed in this paper are received from the Data Administrator at Nord Pool Spot (`http://www.nordpoolspot.com/`) upon request. The data set consists of hourly spot prices measured in Norwegian Kroner (NOK) from May 4th 1992 to August 27th 2011. The one-day period and the seven-day period show up as peaks in the estimated power spectrum, as do their higher harmonics. This can be seen in figure 6.2(c).

To eliminate the effect of the strong daily periodicity we consider the signal $P(t)$ of daily mean prices, and we denote $X(t) = -\mu\, t + \log P(t)$, where μ is the mean daily log-return,

i.e.

$$\mu = \mathbb{E}\Big[\log \frac{P(t)}{P(t-\Delta t)}\Big],$$

where $\Delta t = 1\,\text{day}$ and "log" denotes the natural logarithm. The resulting signal $X(t)$ is shown in figure 6.1(a). We will refer to the time series $X(t+\Delta t) - X(t)$ as the daily log-returns, and they are shown in figure 6.1(b). The sample standard deviation of this signal is $\sigma = 0.12$.

Even if we study daily mean values, there are still seven-day and one-year periodicities in the signal. This is illustrated in figures 6.1(c) and 6.1(d). In figure 6.1(c) we have plotted the conditional mean of $X(t)$ given that the signal is sampled only on a specific day of the week. We see from the figure that the spot price is lower for Saturdays and Sundays. The conditional mean of $X(t)$ has a variation of about 0.15 from weekdays to the weekend. This variation is rather large, roughly equal to the standard deviation of the log-returns, and therefore the seven-day periodicity exerts substantial influence on certain estimates. For instance, the estimated auto-correlation function of the daily log-returns has a strong seven-day period which makes it hard to analyze the correlation decay in the non-periodic component of the signal.

Due to the strong weekly periodicity we will for most of the remaining analysis consider a discretization $y_k = X(k\Delta t)$, where $\Delta t = 1\,\text{week}$. This simply means that we consider the signal $X(t)$ sampled once a week, for instance every Monday. This gives us seven (dependent) time series, and as we will see in section 7.4, the parameter estimates vary little between these time series, with some exceptions for Saturdays and Sundays.

In the weekly sampled signals there is still a one-year periodicity. This is illustrated in figure 6.1(d), where we have plotted the weekly mean of $X(t)$ conditioned on the week number of the year. The seasonal dependence can very roughly be described by a sinusoidal oscillation with amplitude 0.25. This amplitude is greater than the the weekly oscillation, but since the period is about 50 times greater, the one-year oscillation exerts much less influence on the analysis. We make no attempt to de-trend this seasonality, but we have crudely tested the effect that a sinusoidal oscillation with amplitude of 0.25 has on the estimates that we perform. For the ML estimators presented in section 6.4 we observe that the addition of the sinusoidal signal slightly decreases the estimate of H in the fractional MRW model, and slightly increases the time scale $1/\nu$ in the damped MRW model. The relative changes in H and $1/\nu$ are typically $< 10\%$, whereas the intermittency parameter is almost unchanged. Other estimates, such as the auto-correlation function for the fluctuation amplitudes $|X(t+\Delta t) - X(t)|$, are more influenced by the seasonal variation. We have plotted this correlation function in figure 6.2(b), and we clearly see a one-year oscillation on top of the decay. Simonsen Simonsen (2005) has found that this correlation function decays algebraically as $1/\tau^{0.07}$, and if we compare with equation (6.1), this corresponds to $\lambda = 0.53$. This value of the intermittency parameter is slightly higher than what is estimated for stock markets, which typically are in the range 0.3-0.4 (Løvsletten and Rypdal, 2012a; Bacry et al., 2008). This difference between stock markets and electricity spot markets is confirmed by the ML estimates presented in section 7.4.

The estimated correlation functions for the log-returns themselves are plotted in figure

6.2(a). It is the average of the correlation functions for the seven time series obtained by sampling with weekly intervals on a given weekday. It might be possible to detect some anti-correlation from this figure, but it is impossible to distinguish between an exponential and an algebraic decay, i.e. between the expressions in equations (6.3) and (6.5). The correlations of returns are better analyzed using a wavelet-based variogram $V(a) = \mathbb{E}[|W(t,a)|^2]$, where

$$W(t,a) = \frac{1}{\sqrt{a}} \int X(t')\, \psi\Big(\frac{t'-t}{a}\Big)\, dt' \,, \tag{6.6}$$

is the wavelet transform of $X(t)$ with respect to the mother wavelet ψ. The wavelet transform scales as (Muzy et al., 1991; Simonsen et al., 1998):

$$W(t,a) \sim \sqrt{a}\, \Big(X(t+a) - X(t)\Big) \,, \tag{6.7}$$

and hence, if the Hurst exponent of $X(t)$ is well defined[2] we have

$$V(a) \sim a^{2H+1} \,. \tag{6.8}$$

In figure 6.2(d) we show the wavelet-based variograms for the weekly sampled data estimated using a wavelet ψ that is the first derivative of a Gaussian. The dotted line above the variograms has slope 1.8 in the double-logarithmic plot, corresponding to a Hurst exponent $H = 0.4$.

A striking feature in figure 6.2(d) is the sharp "breaks" in the variograms in the range 20-50 weeks. This represents a characteristic time scale, and the existence of characteristic scales is not consistent with an algebraic decay of the auto-correlation function. In fact, this feature of the spot price signal suggests a process of OU-type, for which the correlation decay has a characteristic time scale $1/\nu$. To support this statement we have simulated an ensemble of OU processes with $1/\nu = 20$ weeks and estimated the wavelet-based variograms using the same method as for the spot-price data. The results of this analysis are shown as the bottom curves in figure 6.2(d). We see from these plots that the OU-process captures the flattening of the variogram on long time scales. The same feature is seen in the power spectrum, which is plotted in figure 6.2(c). As opposed to the pure power-law spectrum, a Lorentzian spectrum, which scales as $\sim 1/f^2$ for $f \gg \nu$, captures the flattening of the spectrum at low frequencies. This seems to indicate that a model of OU type is preferable over a fractional model. However, we must take into account that the observation of a characteristic time scale may be an effect of the one-year periodicity, and that a scale-invariant description of the non-periodic component may nevertheless be appropriate. These questions are discussed in more detail in section 6.6.

6.3 Modeling anti-correlations and intermittency

In econometrics, the "stylized facts" of asset price fluctuations are often described by discrete-time models, such as generalized autoregressive conditional heteroskedasticity (GARCH)

[2]The Hurst exponent is well defined if equation (6.4) holds, and we do not need to assume that $X(t)$ is self-affine nor multifractal. Equation (6.8) follows from (6.7) by the definition of H.

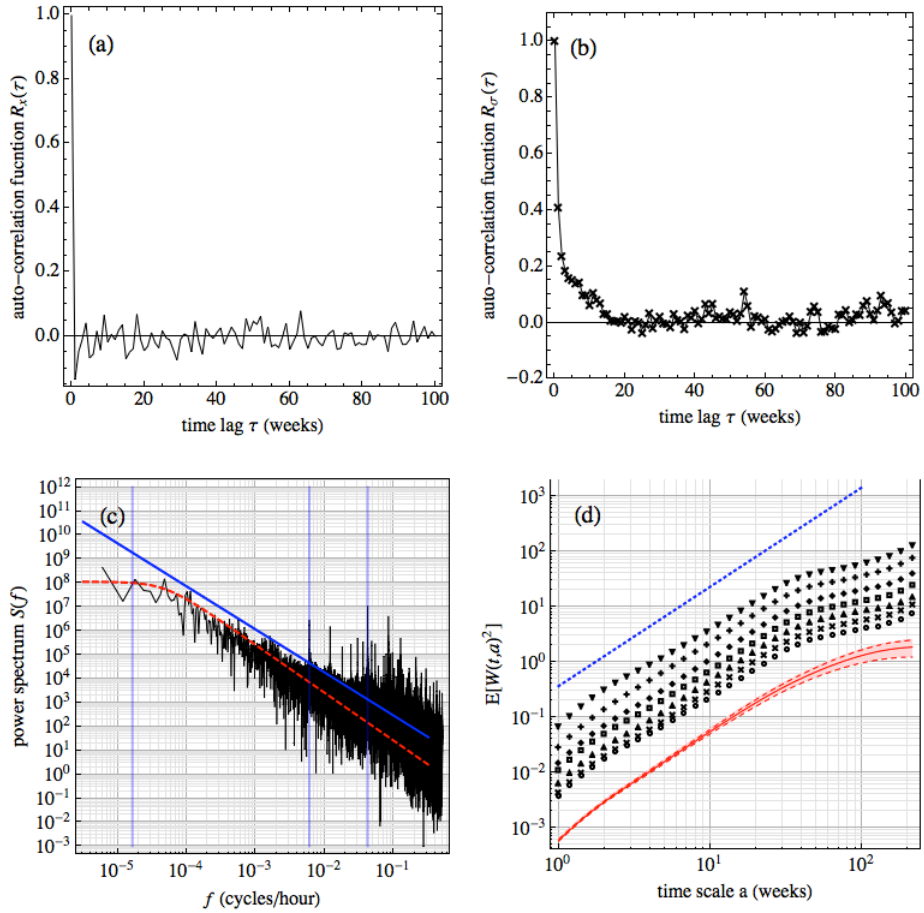


Figure 6.2: (a): The auto-correlation function of the daily mean log-returns sampled at weekly intervals on a given weekday, and then averaged over the seven weekdays. (b): The auto-correlation function for the absolute values of the daily mean log-returns sampled on Mondays. (c): Double logarithmic plot of the power spectrum of the daily mean spot prices. The vertical lines correspond to periods of one year, one week and one day. The solid line represents a power law $1/f^{2H+1}$ with $H = 0.4$, whereas the dotted line is a Lorentzian spectrum $\nu/(\nu^2 + (2\pi f)^2)$, with $1/\nu = 20$ weeks. (d) Double logarithmic plots of the wavelet-based variograms for the log-returns sampled on different weekdays. The curves are shifted to make them all visible (Mondays through Sundays are sorted from bottom to top.) The dotted line is a power law corresponding $H = 0.4$. The bottom curves show the results of the wavelet-based variograms estimated from realizations of a OU process with parameters $1/\nu = 20$ weeks. For each time scale a we have plotted the mean (solid curve) and the 1/8 (upper and lower) quantiles (dashed curves).

models (Bollerslev, 1986), which are well-suited for efficient parameter estimation and forecasting. In parallel to to this approach, several authors have proposed to describe prices using continuous-time models with scaling properties. This idea dates back to the work

of Bachelier (1900), who modeled asset prices using Brownian motions. To deal with level effects, Bachelier's model has been modified, so that it is the *logarithmic* prices, rather than the prices themselves, that are described by Brownian motions. The resulting model can be written on the form

$$\log P(t) = \mu\, t + \sigma\, B(t)\,. \tag{6.9}$$

In 1963, Mandelbrot suggested to replace the Brownian motion $B(t)$ in equation (6.9) with a non-Gaussian self-affine process $Z(t)$. He deduced that in order to maintain uncorrelated increments, the process $Z(t)$ must be an α-stable Lévy process with $\alpha < 2$ (also called a Lévy flight). An interesting modification of the stable Lévy processes are the truncated Lévy flights (Mantegna and Stanley, 1994), which behave as heavy-tailed Lévy flights on short time scales, but converge to Brownian motions on long time scales.

The disadvantage of these self-affine (or truncated self-affine) processes is that increments are independent, so that the models do not describe volatility clustering. The dependence between log-returns can be introduced by explicitly modeling the volatility $\sigma = \sigma(t)$ as a stochastic process independent of $B(t)$ (Taylor, 1982). These processes are often called stochastic volatility (SV) models, and if the volatility $\sigma(t)$ is assumed to follow a OU process, then the process is referred to as a basic SV model. In the basic SV models, the logarithmic prices do not have the scaling properties of Lévy flights and truncated Lévy flights, and they can therefore not be characterized using scaling exponents or scaling functions. However, as is shown by e.g. Calvet and Fisher (2001, 2004) and Bacry et al. (2001, 2008), the volatility $\sigma(t)$ can be chosen so that the resulting model is multifractal.

The class of multifractal processes has its origin in statistical modeling of fully developed fluid turbulence, and the underlying ideas can be traced back to the works of Kolmogorov (1962) and Obukhov (1962). The analogies between turbulent fluids and financial markets were first discussed in the 1990s, by e.g. Ghashghaie et al. (1996), Mantegna and Stanley (1996), and Mandelbrot et al. (1997). Since then, several authors have proposed to model the fluctuation of asset returns using multifractal stochastic processes, which are cascade-type constructions inspired by the energy cascading in turbulence. See e.g. Lux (2012).

In this section we consider a particular multifractal model, the so-called MRW model, and discuss how this model can be modified in order to include anti-correlations. We begin by giving a brief description of the standard (uncorrelated) MRW model in section 6.3.1.

6.3.1 Stochastic volatility and MRW processes

Consider a discretization of the logarithmic price $X(t)$, i.e. $y_k = X(k\Delta t)$. This signal is modeled as

$$y_k = y_{k-1} + \sigma\,\sqrt{M_k}\,\varepsilon_k + \mu\,\Delta t\,, \tag{6.10}$$

where ε_k is a standard Gaussian white noise. The stochastic volatility term is defined as $M_k = c\,\exp(h_k)$, where h_k is a centered Gaussian process with co-variances

$$\mathrm{Cov}(h_k, h_l) = \lambda^2 \log^+ \frac{T}{(|k-l|+1)\Delta t}\,, \tag{6.11}$$

where $\log^+ a \stackrel{\text{def}}{=} \max\{0, a\}$. The constant c is chosen so that $1/c = \mathbb{E}[\exp(h_k)]$. Note that for $\lambda = 0$ the logarithmic price is a Brownian motion with drift, i.e. the price is modeled as a geometric Brownian motion.

From equations (6.10) and (6.11) one can derive equation (6.1), which shows that the MRW model describes algebraically decaying volatility dependence. This type of long-range dependence is common in financial time series, and the MRW model has shown to provide good descriptions of the fluctuations of stock prices and currency exchange rates (Bacry et al., 2008). However, the (drift-compensated) log-returns $x_k = y_k - y_{k-1} - \mu\,\Delta t$ are uncorrelated in this model, and therefore the model needs to be modified to make it capture the anti-correlations of spot-price data.

6.3.2 A dampled MRW model

The discrete-time analog of OU processes are auto-regressive models of order one (AR(1) processes). These can be written on the form

$$y_k = \phi y_{k-1} + \sigma\,\varepsilon_k + \mu\,\Delta t\,, \tag{6.12}$$

where $\phi = 1 - \nu\,\Delta t$. This model can now be generalized to include multifractal volatility by replacing the constant σ with the process $\sigma\sqrt{M_k}$, where $M_k = c\,\exp(h_k)$ is as defined above. The resulting model is given by the following equation

$$y_k = \phi y_{k-1} + \sigma\sqrt{M_k}\,\varepsilon_k\,. \tag{6.13}$$

We will refer to this process as a damped MRW model. We have here assumed that $\mu = 0$.[3] We note that the damped MRW model is not multifractal, since its structure functions are not power laws. However, the volatility dependence is the same as in the standard MRW model.

Keeping in mind that M_k and ε_k are independent processes, and that $\sqrt{M_k}$ is normalized to have unit variance, we can use equation (6.13) to derive some simple properties of the damped MRW model. The main observation is that the stochastic volatility term $\sqrt{M_k}$ does not effect the second order statistics. This means that the auto-correlation functions, variograms and power spectra are the same for the damped MRW process as for the corresponding AR(1) process. For instance, for $k > 0$, the auto-correlation function is

$$\mathbb{E}[y_0 y_k] = \frac{\sigma^2}{1-\phi^2}\,\phi^k \stackrel{\Delta t \ll 1/\nu}{\approx} \frac{\sigma^2}{1-\phi^2}\,e^{-\nu\,t}\,,$$

where $t = k\Delta t$. The auto-correlation function for the log-returns $x_k = y_k - y_{k-1}$ is then

$$\mathbb{E}[x_0 x_k] \sim -\Delta t^2 \frac{d^2}{dt^2}\frac{\sigma^2}{1-\phi^2}\,e^{-\nu\,t} = -\frac{\Delta t^2 \sigma^2 \nu^2}{1-\phi^2}\,e^{-\nu\,t}\,, \tag{6.14}$$

[3] This can be done without loss of generality since the parameter μ is easily estimated from data. One can then replace the logarithmic prices $X(t)$ with the drift-compensated signal $X(t) - \mu t$.

and (up to discreteness effects) the power spectrum of y_k is a Lorentzian:

$$S(f) \propto \frac{\nu}{\nu^2 + (2\pi f)^2}\,.$$

For general references on AR(1) processes see e.g. Percival and Walden (1993) and Brockwell and Davis (1991).

Although the damped MRW models and OU processes (or AR(1) processes) share the same correlations, there are essential differences between the models. As explained in the introduction and in section 6.3.1, the factor $\sqrt{M_k}$ introduces volatility clustering and "fat tailed" return distributions, features that are not contained in standard OU processes. This can be seen from figure 6.3. In figures 6.3(a) and 6.3(b) we show a realization of an OU process $X(t)$ and its increments $X(t+\Delta t)-X(t)$ respectively, and in figures 6.3(e) and 6.3(f) we show the corresponding plots for a damped MRW process. We clearly see that the increments of the damped MRW model has volatility clustering and spikiness that is not present in the OU process. If we compare with the Nord Pool data, which are shown in figures 6.3(i) and 6.3(j), we observe that these features are essential for accurate modeling of spot prices.

6.3.3 A fractional MRW model

The second class of models that we introduce are fractional MRW models. Here the term "fractional" refers to the replacement of the white Gaussian noise ε_k with a fractional Gaussian noise

$$\varepsilon_k^{(H)} = B_H(k+1) - B_H(k)\,.$$

Here $B_H(\cdot)$ is a fractional Brownian motion with self-affinity exponent H. This gives processes on the form

$$y_k = y_{k-1} + \sqrt{M_k}\,\varepsilon_k^{(H)}\,, \tag{6.15}$$

where M_k is as described in section 6.3.1.

The parameter ϕ is absent from this model since the anti-correlations of returns are described via a Hurst exponent $H < 1/2$. In fact, for $H \neq 1/2$, the correlations of $x_k = y_k - y_{k-1}$ are given by the expression

$$\mathbb{E}[x_0 x_k] \sim 2H(2H-1)\,k^{2H-2-\lambda^2/4}\,. \tag{6.16}$$

If we compare equation (6.16) with equation (6.14) we see that a distinguishing feature for the two models is that the auto-correlation function for log-returns decays exponentially for the damped MRW model, whereas it decays algebraically for the fractional MRW model.

As for the damped MRW model, the factor $\sqrt{M_k}$ introduces volatility clustering and "fat tailed" return distributions. In figures 6.3(c) and 6.3(d) we show a realization of a fractional Brownian motion $X(t)$ with $H = 0.4$ and its increments $X(t+\Delta t) - X(t)$ respectively. In figures 6.3(g) and 6.3(h) we show the corresponding plots for a fractional MRW process with $H = 0.4$. We see that the increments of the fractional MRW model has volatility clustering and spikiness that is not present in the fractional Brownian motion.

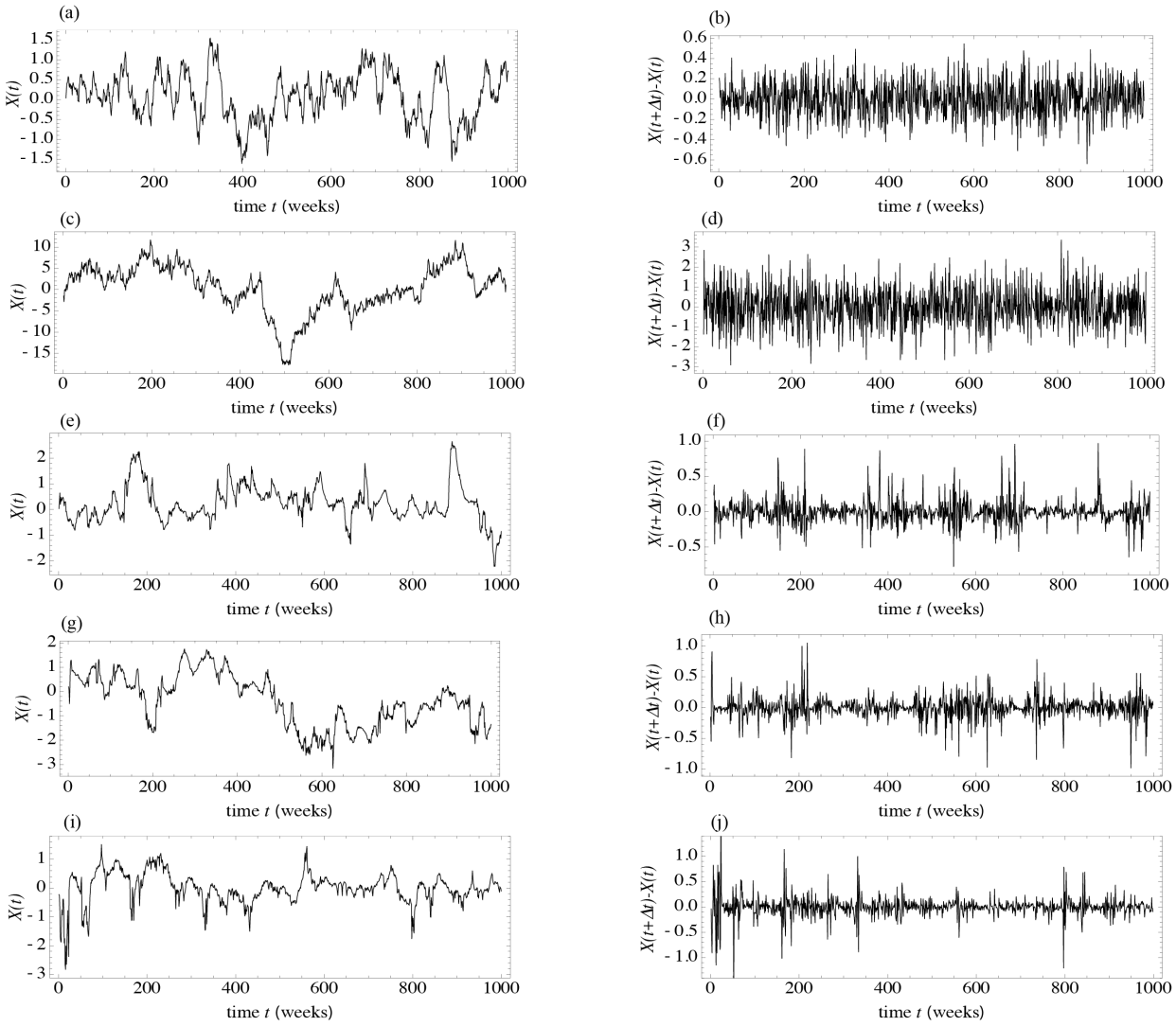


Figure 6.3: (a): A realization of an OU process with $1/\nu = 20$ weeks. (b): The increments of the signal in (a). (c): A realization of a fractional Brownian motion with $H = 0.4$. (d): The increments of the signal in (c). (e): A realization of a damped MRW model with $1/\nu = 20$ weeks and $\lambda = 0.7$. (f): The increments of the signal in (e). (g): A realization of a fractional MRW model with $H = 0.4$ and $\lambda = 0.7$. (h): The increments of the signal in (g). (i): The daily mean (logarithmic) spot price sampled every Monday. (j): The increments of the signal in (i).

6.4 Maximum likelihood estimators

In the following we present ML estimators for damped and the fractional MRW models. ML techniques for multifractal models have so-far been restricted to the so-called Markov-Switching Multifractal (MSM) (Calvet and Fisher, 2001, 2004), but were recently obtained

for the standard MRW model (Løvsletten and Rypdal, 2012a). The generalization to fractional and damped MRW models are new results.

6.4.1 Computation of the likelihood function for the standard MRW model

The standard MRW model can be written as $x_k = \sigma\sqrt{M_k}\,\varepsilon_k$, where the processes $M_k = c\exp(h_k)$ and ε_k are as described in section 6.3.1. Given a time series of n observations $z = (z_1, \ldots, z_n)$ (which we want to model with the process x_k) the ML estimator seeks the parameter vector $\theta = (\lambda, \sigma, T)$ that maximizes the likelihood function $\mathcal{L}_x(\theta|z)$, i.e.

$$\hat{\theta} = \mathrm{argmax}_\theta\, \mathcal{L}_x(\theta|z)\,,$$

where $\mathcal{L}_x(\theta|z) = p_x(z|\theta)$ is the n-dimensional probability density function (PDF) for the random vector $x = (x_1, \ldots, x_n)$, evaluated at the point z for a fixed parameter vector θ.

The challenge is to efficiently compute the PDFs $p_x(x)$. Such a method is presented in detail in Løvsletten and Rypdal (2012a), and here we will only explain the main ideas. The first step is to denote $h = (h_1, \ldots, h_n)$ and to write

$$p_x(x) = \int_{\mathbb{R}^n} p_{x,h}(x,h)\,dh = \int_{\mathbb{R}^n} p_{x|h}(x|h)p_h(h)dh\,, \tag{6.17}$$

where $p_{x,h}$ is the joint PDF for the pair $(x,h) \in \mathbb{R}^n \times \mathbb{R}^n$. Here $p_{x|h}$ is the conditional PDF of x given h, and p_h is the marginal PDF of h. The first term is easily computed by noting that $x|h$ is a Gaussian vector with independent entries. This gives

$$p_{x|h}(x|h) = \prod_{k=1}^{n} p_{x_k|h_k}(x_k|h_k)\,, \tag{6.18}$$

where

$$p_{x_k|h_k}(x_k|h_k) = \frac{1}{\sqrt{2\pi\, c\exp(h_k)}\,\sigma}\exp\Big(-\frac{x_t^2}{2\sigma^2\, c\exp(h_k)}\Big)\,. \tag{6.19}$$

By definition the vector h is centered and Gaussian with specified co-variances $\mathrm{Cov}(h_k, h_l) = \gamma(|k-l|)$, and so the second factor $p_h(h)$ is calculated by using standard techniques for Gaussian vectors: For each $k = 1, \ldots, n$ we define the regression coefficients $\varphi_i^{(k)}$ by the equations

$$\sum_{j=1}^{k} \varphi_j^{(k)}\gamma(|i-j|) = \gamma(i) \quad \text{for} \quad i = 1, \ldots, k\,. \tag{6.20}$$

Then it holds that

$$h_k = \varphi_1^{(k-1)}h_{k-1} + \cdots + \varphi_{k-1}^{(k-1)}h_1 + w_k\,, \tag{6.21}$$

where w_k are independent and centered Gaussian variables with variances

$$s_k^2 = \gamma(0) - \sum_{i=1}^{k-1}\varphi_i^{(k-1)}\gamma(i)\,.$$

We can now make an approximation by fixing a truncation parameter $K \in \mathbb{N}$, and for $k > K$ replacing the expression in equation (6.21) with

$$h_k = \varphi_1^{(K)} h_{k-1} + \cdots + \varphi_K^{(K)} h_{k-K} + w_k^{(K)} \,,$$

where $w_k^{(K)}$ are independent and centered Gaussian variables with variances s_{K+1}^2. With this approximation we obtain the following expression for $p_h(h)$:

$$\begin{aligned} \log p_h(h) &= -n \log \sqrt{2\pi} - \sum_{k=1}^{K} \log s_K - (n-K) \log s_{K+1} \\ &\quad - \sum_{k=1}^{K} \frac{\left(h_k - \varphi_1^{(k-1)} h_{k-1} - \cdots - \varphi_{k-1}^{(k-1)} h_1\right)^2}{2 s_k^2} \\ &\quad - \sum_{k=K+1}^{n} \frac{\left(h_k - \varphi_1^{(K)} h_{k-1} - \cdots - \varphi_K^{(K)} h_{k-K}\right)^2}{2 s_{K+1}^2} \,. \end{aligned}$$

By combining this expression with equations (6.18) and (6.19) we have an expression for the full likelihood $p_{x,h}(x,h)$, and what remains is to calculate the integral in equation (6.17). This integral can be accurately approximated using Laplace's method (Laplace, 1986). This method entails writing $p_{x,h}(x,h) = \exp(n f_x(h))$ and assuming that the function $f_x(h)$ has a global maximum h^* as a function of h. For large n the contribution to the integral in equation (6.17) is concentrated around h^*, and the hence we can approximate it by making a second order Taylor expansion of $f_x(h)$ about h^*. The result is the approximation

$$\begin{aligned} p_x(x) &\approx \exp(f_x(h^*)) \int_{\mathbb{R}^n} \exp\left(\frac{1}{2}(h-h^*)\Omega_x(h-h^*)^T\right) dh \\ &= (2\pi)^{n/2} |\det \Omega_x|^{-1/2} \, p_{x,h}(x,h^*) \,, \end{aligned}$$

where Ω_x is the Hessian matrix of $f_x(h)$ evaluated at the point h^*.

6.4.2 Computation of the likelihood function for the damped MRW model

Let y_k be the damped MRW model defined by equation (6.13) and x_k be the standard MRW model. We can write

$$y_k = \phi y_{k-1} + x_k \,,$$

and then, for $y = (y_1, \ldots, y_n)$ and an initial condition y_0, we have

$$p_y(y) = \int p_{y_0}(y_0) p_x(y_1 - \phi y_0, y_2 - \phi y_1, \ldots, y_n - \phi y_{n-1}) \, dy_0 \,.$$

For large n, the ML method is insensitive to the initial condition y_0, and therefore we choose $y_0 = 0$, which gives

$$p_y(y) = p_x(y_1, y_2 - \phi y_1, \ldots, y_n - \phi y_{n-1}) \,.$$

We then have a simple relationship between the likelihood functions of the process y_k and the the likelihood functions for the process x_k:

$$\mathcal{L}_y\big(z_1,\dots,z_n \,|\, (\lambda,\sigma,T,\phi)\big) = \mathcal{L}_x\big(z_1, z_2-\phi z_1,\dots,z_n-\phi z_{n-1} \,|\, (\lambda,\sigma,T,\phi)\big)\,.$$

6.4.3 Computation of the likelihood function for the fractional MRW model

In the fractional MRW model the white noise process ϵ_k is replaced by a fractional Gaussian noise $\varepsilon_k^{(H)}$ with Hurst exponent H. By definition this is a centered Gaussian process with co-variance

$$\beta(|k-l|) \stackrel{\text{def}}{=} \mathrm{Cov}(\varepsilon_k^{(H)},\varepsilon_l^{(H)}) = \frac{1}{2}\Big\{(|k-l|+1)^{2H} - 2|k-l|^{2H} + (|k-l|-1)^{2H}\Big\}\,.$$

In the same way as for process h_k in section 6.4.1, we can write

$$\varepsilon_k^{(H)} = \xi_1^{(k-1)}\varepsilon_{k-1}^{(H)} + \cdots + \xi_{k-1}^{(k-1)}\varepsilon_1^{(H)} + w_k\,, \tag{6.22}$$

where the regression coefficients are defined via the equations

$$\sum_{j=1}^{k}\xi_j^{(k)}\beta(|i-j|) = \beta(i)\,,\quad i=1,\dots,k\,, \tag{6.23}$$

and w_k are independent and centered Gaussian variables with variances

$$r_k^2 = \beta(0) - \sum_{i=1}^{k-1}\xi_i^{(k-1)}\beta(i)\,.$$

Again we fix a truncation parameter $K\in\mathbb{N}$, and for $k>K$ we replace the expression in equation (6.22) with the expression

$$\varepsilon_k^{(H)} = \xi_1^{(K)}\varepsilon_{k-1}^{(H)} + \cdots + \xi_K^{(K)}\varepsilon_{k-K}^{(H)} + w_k^{(K)}\,,$$

where $w_k^{(K)}$ are independent and centered Gaussian variables with variances r_{K+1}^2. For $k>K$ it follows that the conditional PDF of x_k, given both $h_1,\dots,h_k$ and $x_1,\dots,x_{k-1}$, satisfies

$$\begin{aligned}\log p_{x_k|h_1,\dots,h_k,x_1,\dots,x_{k-1}}(x_k|h_1,\dots,h_k,x_1,\dots,x_{k-1}) &= -\log(\sigma\sqrt{2\pi c}) - \frac{h_k}{2}\\ -\log r_{K+1} - \frac{1}{2\sigma^2 c r_{K+1}^2}&\left(x_k\exp(-h_k) - \sum_{i=1}^{K}\xi_i^{(K)}x_{k-i}\exp(-h_{k-i})\right)^2.\end{aligned}$$

From this we obtain an expression for the conditional PDF of x given h:

$$p_{x|h}(x|h) = \sum_{k=1}^{n} p_{x_k|h_1,\dots,h_k,x_1,\dots,x_{k-1}}(x_k|h_1,\dots,h_k,x_1,\dots,x_{k-1})\,. \tag{6.24}$$

Equation (6.24) is substituted into equation (6.17). The factor $p_h(h)$ and the integral in equation (6.17) are computed as explained in section 6.4.1.

time series	$1/\nu$ (weeks)	λ	T (weeks)	σ
daily means	29	0.68	112	0.13
daily maxima	22	0.86	14	0.17
weekly means	40	0.55	132	0.14
weekday 1	22	0.70	26	0.18
weekday 2	20	0.57	99	0.16
weekday 3	24	0.59	99	0.16
weekday 4	26	0.64	71	0.18
weekday 5	26	0.65	25	0.16
weekday 6	39	0.72	32	0.17
weekday 7	41	0.78	88	0.25

Table 6.2: ML estimates for the damped MRW model using the method described in section 6.4.2. The first data set consists of the daily mean logarithmic spot price after having subtracted a linear trend. In the second signal we consider daily maxima of the logarithmic price rather than the daily mean price. The third signal is obtained from the first signal my taking seven-day means. The seven remaining signals are obtained from the first signal by sampling every seventh day.

6.4.4 Implementation of the ML estimators

The methods described above are implemented as packages in the R programming language. Equations (6.20) and (6.23) are efficiently solved using the Durbin-Levinson algorithm (Trench, 1964; McLeod et al., 2007), so the most intensive computations are determining the maxima h^* in the Laplace approximation. This is done using analytic expressions for the derivatives of the functions $f_x(h)$, which roots are found numerically using the derivative-free SANE algorithm (La Cruz et al., 2006).

The estimates $\hat{\theta}$ are found by numerically optimizing the likelihood functions.

6.5 Results

The ML estimators for the damped MRW model and the fractional MRW model are applied to the data from the Nord Pool market and the results are shown in tables 6.2 and 6.3. As discussed in section 6.2 we use daily averaged data sampled once a week. Hence we have one time series for each day of the week, and these are referred to as "weekday 1" to "weekday 7" in tables 6.2 and 6.3. For comparison we have also included two time series that are sampled daily. These are the daily means and the daily maxima of the logarithmic spot prices. In addition we have looked at the time series consisting of weekly means of the logarithmic prices.

For the weekly sampled time series ("weekday 1" to "weekday 7") the estimates for the damped MRW model give characteristic time scales $1/\nu$ that vary between 20 and 40

time series	H	λ	T (weeks)	σ
daily means	0.47	0.68	17	0.10
daily maxima	0.40	0.83	16	0.16
weekly means	0.59	0.58	97	0.14
weekday 1	0.45	0.67	101	0.20
weekday 2	0.45	0.57	146	0.17
weekday 3	0.45	0.60	99	0.17
weekday 4	0.44	0.64	97	0.18
weekday 5	0.44	0.63	101	0.18
weekday 6	0.51	0.71	73	0.19
weekday 7	0.50	0.80	97	0.26

Table 6.3: ML estimates for the fractional MRW model using the method described in section 6.4.3. The data sets are as explained in the caption of table 6.2.

weeks. The mean value is 28 weeks and the standard deviation is 8.2 weeks. These results are consistent with the preliminary analysis discussed in section 6.2, where we showed that the wavelet-based variograms fit well with the wavelet-based variogram of an OU process with $1/\nu = 20$ weeks. Also, the power spectrum of the spot prices fits with a Lorentzian spectrum with this characteristic time scale.

The highest estimate of $1/\nu$ are found for the time series "weekday 7". This indicates that the anti-correlations are weaker if we consider only the spot prices for Sundays. This is also seen from the estimates for the fractional MRW model. Table 6.3 shows that the estimates of the Hurst exponent is 0.44-0.45 for all the weekdays, but around 0.50 for Saturdays and Sundays. If these differences between weekends and weekdays are significant, then it suggests that the anti-persistence in electricity spot prices is, at least partly, caused by macroeconomic variations, and not solely by the mean-reverting mechanisms of natural factors such as weather and water levels.

Relative to what is observed in stock markets we find large values of the intermittency parameter λ, both for the damped MRW model and for the fractional MRW. For the fractional MRW model the estimates of λ vary between the 0.57 and 0.80 for the different weekdays, with an average of 0.66 and a standard deviation of 0.08. For the damped MRW model the estimates vary from 0.57 to 0.77, with a mean of 0.66 and a standard deviation of 0.07. In both cases we observe higher estimates of λ for Saturdays and Sundays than for the rest of the week. This could mean that the volatility clustering is stronger for weekend prices than for the rest of the week, but it may also be a simple level-effect. In commodity prices one always observes that the fluctuation level increases with the price itself. This means that when the price is very low, the increments have smaller amplitudes, and this prevents the prices from becoming negative. If the conditional standard deviation of the fluctuation $\delta P_\tau(t) = P(t+\tau) - P(t)$ is proportional to the price level $P(t)$, i.e.

$$\text{stdev}[\delta P_\tau(t) \,|\, P(t) = P] \propto P\,, \tag{6.25}$$

where stdev[·] denotes standard deviation, then one can normalize for the level effect by taking the logarithm of the price. This is the case for geometric Brownian motion. If the fluctuation level is proportional to the price itself, as in a geometric Brownian motion, then one can normalize for the level effect by taking the logarithm of the price. Although this proportionality hypothesis is a good approximation for most price levels, it might be inaccurate for very low values of the spot price. If this is the case, the fluctuations during times with low prices may be amplified by the logarithmic transformation. This can cause biased estimates of the intermittency parameter and a spurious difference between weekend prices and the rest of the week.

Remark 1 *In this paper we analyze daily mean spot prices. In other situations, it is relevant to consider hourly prices, which can be negative. We note that a first step towards dealing with negative prices would be to model the low-level fluctuations by*

$$\text{stdev}[\delta P_\tau(t) \,|\, P(t) = P] \propto (P - P_0)^\kappa \,,$$

and to estimate the parameters κ and P_0. See e.g. Rypdal and Løvsletten (2011). Here κ is called the constant elasticity variance (CEV) parameter, and P_0 represents a base line for the price. For $\kappa = 1$, the appropriate transformation is $X(t) = \log(P(t) - P_0)$. If the estimated P_0 is lower than the lowest price in the time series, then this transformation also allows negative prices.

We finally remark that the estimates of T are known to be very unstable, and it is difficult to draw any meaning from these results.

6.6 Comparing the models

As discussed in sections 6.1 and 6.2, the two models presented in this paper, the damped MRW model and the fractional MRW model, represent two different ways of describing the anti-correlations in electricity spot prices. In the damped MRW model the correlations decay exponentially according to equation (6.3), whereas the correlation decay is a power law, as in equation (6.5), for the fractional model. Our preliminary analysis has revealed that the wavelet-based variograms has a scaling regime up to a time scale of about $20 - 50$ weeks, and flatten for time scales longer than this. In the same way, the power spectrum flattens for low frequencies. These results can be seen as indications that the damped MRW model is more appropriate than the fractional MRW model for describing spot prices. See figure 6.2(d). On the other hand, the features mentioned above could be effects of the yearly oscillation, since it is known that the existence of periodicities often produce characteristic "S-shapes" in variograms. We are therefore considering two competing hypotheses. The first hypothesis is that there exists a characteristic scale $1/\nu$ which is not related to the annual periodicity in the signal. More precisely that the correlation decays exponentially with a rate $1/\nu$, and that the non-periodic component of the spot prices can be well described by the damped MRW model. The second hypothesis is that the non-periodic

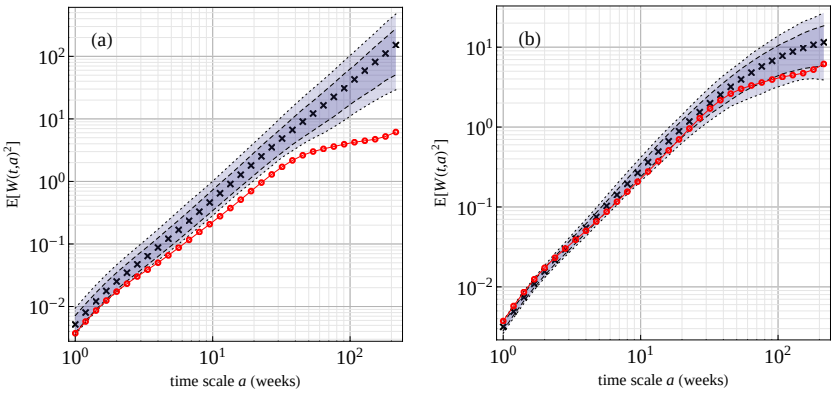


Figure 6.4: (a): Wavelet-based variograms estimated from realizations of the fractional MRW model with parameters $H = 0.45$, $\lambda = 0.67$, $T = 101$ weeks and $\sigma = 0.20$. For each time scale a we have plotted the mean (crosses) and the 1/8 (dashed) and 1/40 (dotted) upper and lower quantiles. Prior to the analysis we have added a sinusoidal oscillation with a one-year period and an amplitude 0.25 to each realization. The curve with circles is the wavelet-based variogram for the daily mean (logarithmic) spot price sampled every Monday. (b): The same as (a), but in this case we have simulated the damped MRW model with parameters $1/\nu = 21.9$ weeks, $\lambda = 0.70$, $T = 26$ weeks and $\sigma = 0.18$.

component of the spot price has a algebraic correlation decay on time scales from a day to several years, and that the characteristic scale which is observed is an effect of the annual periodicity. In this case the fractional MRW model is a good description of the spot price data.

To test the hypotheses we have simulated an ensemble consisting of 500 realizations of the fractional MRW model. The parameters are chosen equal to those estimated for the time series "weekday 1", i.e. $H = 0.45$, $\lambda = 0.67$, $T = 101$ weeks and $\sigma = 0.20$. For each of the realizations we have estimated the wavelet-based variogram $V(a) = \mathbb{E}[|W(t,a)|^2]$, where $W(t,a)$ is the wavelet transform defined in equation (6.6). For each time scale a the mean is calculated together with the 1/8 and 1/40 (upper and lower) quantiles. These curves are plotted in figure 6.4(a). To simulate the effect of the annual periodicity we have added a sinusoidal oscillation with a one-year period and an amplitude 0.25. This amplitude is chosen according to the estimate shown in figure 6.1(d). The analysis shows that the periodicity is not strong enough to produce the "break" in the variograms. Hence this "break" represents a characteristic time scale that should be included in the model. In figure 6.4(b) we show the same analysis as in 6.4(a), but here we have simulated the damped MRW model with parameters $1/\nu = 22$ weeks, $\lambda = 0.70$, $T = 26$ weeks and $\sigma = 0.18$. The result shows that wavelet-based variogram of the spot prices is much better reproduced by the damped model than the fractional model.

6.7 Concluding remarks

The main point of this work is to present stochastic models for electricity spot prices that capture both the slowly decaying volatility dependence and the anti-correlated returns. We also present ML methods that efficiently and accurately estimate parameters for these models. For the data from the Nord Pool market the ML estimates show that the intermittency parameter λ is significantly higher than what is observed in stock markets, confirming the exceptional nature of electricity spot markets.

Another important result is that the damped MRW model performs better than the fractional MRW model, and based on this we conclude that there *is* a characteristic scale for the correlation decay of returns (which is not related to the annual oscillations). Estimates show that this time scale is 20-30 weeks. The use of Hurst exponents to characterize correlations is incompatible with the existence of such a characteristic scale, and hence the results of this paper indicate that Hurst-type analysis is inappropriate for electricity spot prices. This conclusion is supported by the fact that the ML estimates of H actually are quite close to 0.5. If we keep the one-year oscillation in mind, and take into account that estimates of H tend to be slightly reduced in the presence of a periodic component, then it seems unlikely that $H < 0.5$ with any certainty based on the ML estimates. Since the anti-correlations are consistently captured by the damped MRW model, our interpretation is that the fractional model struggles to produce clear evidence for anti-correlations because it imposes a scale free correlation function which is not compatible with the data.

Another result of this paper becomes evident by considering data sampled on different weekdays. From table 6.2 it appears that the anti-correlations are stronger for weekdays than for weekends, which suggests that anti-correlations are partly caused by human activity.

We finally remark that the damped MRW model (toghether with a one-year oscillation) is suitable for forecasting spot prices, and this is the topic of ongoing research. For this purpose, the main advantage of multifractal models is that they efficiently exploit the memory effects in the volatility for forecasts of future prices.

7
Paper 4

Abstract. We perform significance testing of linear trends in regional temperature records for the time period year 1900–2014. Instrumental data records on $5^\circ \times 5^\circ$, $2^\circ \times 2^\circ$, and equal-area grids, are examined, and it is demonstrated that the stochastic properties of the temperature fluctuations vary substantially with geographic location. Many temperature records are consistent with error models for internal variability that exhibit long-range memory (LRM), whereas the temperature fluctuations of the tropical oceans are strongly influenced by the El Niño Southern Oscillation (ENSO), and therefore seemingly more consistent with random processes of short-range memory (SRM) type. We demonstrate that the choice of null model in the ENSO regions strongly affects the significance of the warming trends. By using a null model with LRD we find significant trends (on the 0.05-level) in only about 65% of grid points, and if we use a null model of SRM type this number is approximately 95%. However, if we in each grid point use a likelihood ratio test to choose the null model most consistent with the data, then we obtain significant trends in just over 80% of grid points. We also discuss how variations in the scaling properties on decadal time scales could influence the results of trend significance tests. We demonstrate that if there actually is LRM on time scales longer than a decade in the regions of Earth where the model selection tests prefers a SRM model, then this can change our conclusions about the significance of the warming trends in these regions.

7.1 Introduction

To evaluate the rate of global warming it is useful to model temperature variations on decadal time scales as superpositions of stochastic processes and trend signals. Of course, such decompositions are not canonical since they depend crucially on both the trend models and the stochastic components (the errors), and it is therefore essential to construct models that consistently describe the so-called "stochastic signatures" of the observations. In particular, in order to be meaningful with respect to trend-significance testing, the models should correctly reproduce the relations between the fluctuation levels on different time scales. For the global mean surface temperature there is strong evidence of long-range dependence (LRD) (Lennartz and Bunde, 2009; Rypdal et al., 2013), and it is therefore reasonable to choose error models that exhibit scaling and slowly decaying auto-correlation functions (ACFs). Examples of such models are the fractional Gaussian noises (fGn), which are stationary processes characterized by dimensionless Hurst exponents $H \in (0, 1)$. The Hurst exponent determines the asymptotic correlation decay via the asymptotic formula $\sim 2H(2H-1)\tau^{2H-2}$ for the auto-correlation function. Alternative models include fractional differenced noise and its generalizations, which are typically defined via the parameter $d = H - 1/2$. Due to the strong persistence, standard estimators of variability and uncertainty are typically unsuitable in climatic time series (Koutsoyiannis, 2003).

The issue of the reality of the global warming signal is best addressed by considering globally averaged signals. Østvand et al. (2013) demonstrated that the strongest statistical significance is obtained by considering the global land record, but that significance on the 95% significance level also is demonstrated in the global ocean record under an fGn null hypothesis (error model). Nevertheless, number of authors have also focused on the significance of trends in local temperature records (e.g. Lennartz and Bunde, 2009; Franzke, 2012; Bunde et al., 2014). For such records the statistical significance is lower because of a much higher level of climate noise, but the significance issue is still interesting because it is intimately related to the ratio between the estimated warming signal and the natural variation on the time scale on which the trend is estimated. This ratio, which could be perceived as a local/regional warming impact factor, could be an important measure of the load of forced climate change on ecosystems and human communities. It is a more relevant measure than the nominal warming in a given location, since the capacity for adaptation to anthropogenic warming will depend on the adaptive capacity already stablished to the natural variability. This serves as the main motivation for assessing trend significance and warming impact factor for local-regional temperature records in a global grid. We will study the geographical distribution of these quantities, and the dependence on the selection of model for the background climate noise (error model). We shall also attempt to find the most reasonable error model, which will not be the same in all geographical locations.

The remainder of this paper is organized as follows: In section 7.2 we discuss various trend models and error models that are relevant for surface temperature data records. In section 7.3 we outline the data analysis methods employed in the paper. This includes parameter estimation, model selection and trend significance testing on several data sets.

Detailed descriptions of the statistical methods are included as supplementary material. The main results are presented and discussed in section 7.4.

7.2 Selection of trend model and error model

In Østvand et al. (2013) it is demonstrated that the instrumental record for global temperature (monthly data for the period 1880-today) has significant linear trends under an fGn-assumption on the errors. However, it is also shown that a linear-trend model

$$Y_t = a_0 + a_1 t + X_t \tag{7.1}$$

actually is inconsistent with observations, and that an oscillatory correction (with a 70yr period) is significant. An even more refined model for the trend can be constructed by exploiting our knowledge of the radiative forcing of the climate system, for instance using the dataset constructed by Hansen et al. (2005, 2011). In Rypdal and Rypdal (2014) it is shown that the global temperature is well described by a stochastic linear-response model on the form[1]

$$Y_t = \int_{-\infty}^{t} (t-s)^{H-3/2}\big(F(s)ds + \sigma dB_s\big) = \int_{-\infty}^{t} (t-s)^{H-3/2}F(s)ds + X_t\,. \tag{7.2}$$

As in the linear-trend model (7.1), the errors X_t in model (7.2) are given by a fGn with Hurst exponent $H \in (0,1)$. However, the trend in (7.2) is a convolution of the forcing data $F(t)$ with a power-law kernel, rather than a simple linear expression.

In principle we could apply linear-response models on the form (7.2) to analyze local and regional temperatures as well. However, since the fluctuation levels in local temperatures are several orders of magnitude larger than in the global mean, it is unreasonable to expect that we can extract detailed information from the global signal by analyzing local data. This is illustrated in figure 7.1, where we have plotted monthly de-seasonalized temperature data for the city of Moscow together with the global mean temperature anomaly. Due to the low trend-to-noise ratio in local data, it reasonable to revert to simple linear trend models on the form (7.1).

The aim of this paper is to discuss detection of global warming in local and regional temperature records by analyzing linear trend models on decadal time scales. In this task we are faced with a number of practical and theoretical obstacles. The first, and perhaps most important problem, is to choose the error models. The reason that this is challenging, is that the temporal characteristics of the temperature field vary considerably with the geographic location. A an illustration of this, in figure 7.1(b) we have plotted a comparison of the spectral power densities of the temperature signals in the inland city of Moscow and a reconstructed temperature signal for a $5^\circ \times 5^\circ$ grid box in the tropical

[1]This stochastic integral is divergent and should only be regarded as a formal expression. A well-defined formulation of the model is given in (Rypdal and Rypdal, 2014).

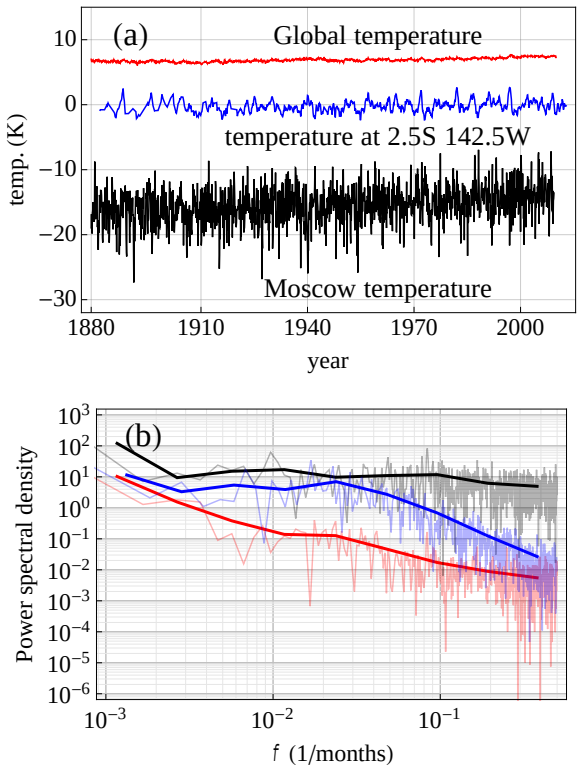


Figure 7.1: (a): The black curve is the monthly temperature data for Moscow. The blue curve is the monthly reconstructed temperature for the $5° \times 5°$ grid centered at 2.5S, 142.5W. The red curve is the global mean temperature anomaly plotted with monthly resolution. The time series are shifted vertically to make them visible. (b): The power spectral densities of the the three time series in (a). The smooth curves are obtained by averaging over logarithmic bins. The colors of the power spectral densities are the same as used for the signal in (a).

Pacific ocean (at 2.5S, 142.5W). As we see from this figure, the temperature in Moscow has a more or less flat spectrum, consistent with a white noise process, whereas the data from the tropical pacific has something closer to a red-noise spectrum. We also see that neither of these signals have spectra similar to the global temperature, which spectrum is shown as the red curve in figure 7.1(b). Significance testing of trends on decadal time scales depend crucially on the power of the error models at frequencies $f \sim (100 \text{ yrs})^{-1}$, and how these fluctuation levels relate to the fluctuation levels on the shorter time scales is substantially different for Moscow than for the tropical Pacific ocean. Consequently we have to use different error models for these two geographical locations.

7.2.1 Two classes of null models

As a first step in this direction we can attempt to model all error processes using one class of parametric models. For instance staying within the scope of LRD, we can use fGns and allowing Hurst exponents to vary with longitude and latitude. The estimation of the Hurst exponent H is actually a non-trivial obstacle for monthly temperature records. For ocean temperatures we typically observe imperfect scaling, which implies that different Hurst exponents can be found using estimators that emphasize on different time scales. The maximum likelihood (ML) estimator of H is sensitive to scaling properties at the shortest time scales, and not the decadal scales which are those relevant in trend significance testing. An alternative is to use a non-parametric estimator H, for instance using detrended fluctuation analysis (DFA) (Peng et al., 1994) or simple variograms. The advantage of these estimators is that one can easily choose the range of time scales that are used to determine H.

An alternative to the approaches described above is to allow different error models in the geographic locations where there are substantial deviations from scaling. This is most evident in tropical Pacific where the El Niño oscillation increases the power on a range of frequencies corresponding to time scales roughly from a year to a decade. Due to the unpredictability of ENSO episodes we will consider simple stochastic error models, and from a purely empirical point of view we suggest to use the Ornstein-Uhlenbeck (OU) class of models, i.e. models defined by the stochastic differential equations (SDEs)

$$dX_t = -\frac{1}{\tau} X_t \, dt + \sigma \, dB_t. \tag{7.3}$$

The power spectral density of an OU process is Lorentzian, with $S(f) \sim f^{-2}$ for $f \gg 1/\tau$ and $S(f) \sim f^0$ for $f \ll 1/\tau$. Hence we have two scaling regimes, one corresponding to Brownian motion (i.e. $H = 3/2$) on short time scales, and one corresponding to white noise (i.e. $H = 1/2$) on long time scales. The transition between these time scales is given by the characteristic time τ, which is also the e-folding time for the ACF. In discrete time an OU process becomes an autoregressive process of order one (an AR(1) process), and we observe that the AR(1) models actually describe the correlation structure of the temperature signals better than fGns in regions of the world strongly dominated by ocean oscillations. To test this hypothesis we perform a model selection test based on likelihood ratios, and the results of these tests show that OU processes are preferred over fGns in the tropical ocean regions. Over land the fGn class of models is typically preferred, except in some inland regions where the estimated Hurst exponent is close to 0.5 (note that $H = 1/2$ corresponds to white noise which is the intersection of the two model classes).

7.3 Data analysis methods

Four data sets are analyzed in this project. These are the HADCRUT4 surface temperature anomalies (Morice et al., 2012), which combines the land temperatures CRUTEM4 (Jones et al., 2012) and sea surface temperatures HADSST3 (Kennedy et al., 2011). We

also use the NOAA MLOST V3.5.3 data developed by Smith and Reynolds (2005). In both these datasets the mean temperature in $5^\circ \times 5^\circ$ grids are provided with monthly time resolution. In addition to these we use Berkeley Earth's equal-area data set (with 15984 time series), and the GISS Surface Temperature Analysis (GISSTEMP) (Hansen et al., 2010) which is given on $2^\circ \times 2^\circ$ grids. The latter data sets are spatially complete, which makes it more convenient to work with. The time period analyzed is January 1900 to April 2013. The main conclusions of our analysis are very similar for all the four data sets, and for convenience we only present the figures for the GISSTEMP results in this paper. The corresponding figures for the other three data sets are included in the supplementary material.

Remark: *The supplementary material is not included in this thesis, but the statistical methods are described in chapter 3. The corresponding figures for the other three data sets are available upon request.*

The data analysis consists of three main parts:

1. Parameter estimation with respect to OU and fGn processes for each grid point.
2. Model selection and trend significance testing based on the preferred model for each grid point.
3. The robustness of the results are tested by exploring the effect of a different scaling regime on the longest time scales. This point is discussed in section 7.4.

The accurate estimation of the parameters (σ, H) and (σ, τ) for the fGn and OU models respectively are crucial for the conclusions drawn in points (ii) and (iii). For the Hurst exponent we observe that direct application of the ML estimator tends to give higher estimates of H compared with the variogram technique or the de-trended fluctuation analysis (DFA). In the latter we have control over which time scales that contribute to the estimate. We also observe that the discrepancy between the two methods disappears if the signals are coarse grained over four-month windows prior to the ML estimation, i.e., if a new, coarser time series is produced by dividing the series into 4-month segments and averaging the data points within each segment. Subsequently, the estimates are corrected for bias based on a Monte Carlo study of the estimator. The details of this procedure are described in the supplementary material. The fluctuation level is estimated using the standard ML technique.

Since likelihoods are computed as a part of the parameter estimation procedure, we can use a straight forward likelihood ratio test to determine if a OU process or a fGn is best supported by the data in each grid point. For each regional signal we estimate the trend $\hat{a}_1$ over the time period from 1900 to 2014, and based on the which model is preferred in each region we compute the p-value for the trend estimate based on a null model which is either of OU- or fGn-type. In regions where the OU process is preferred over the fGn, the computation of the p-value is straight forward using standard techniques

for AR(1) processes. In regions where the fGn model is preferred we have to preform trend significance with respect to a LRM model, which is more challenging. We adapt the method proposed by Ko et al. (2008). This suggest that the $(1-\alpha)$-confidence interval for the trend estimate $\hat{a}_1$ is well approximated by

$$\hat{a}_1 \pm \hat{\sigma} c(\hat{H})^{1/2} t_{\hat{n}_e - 2, 1-\alpha/2}. \tag{7.4}$$

Here $\hat{\sigma}$ is the estimated fluctuation level (standard deviation) of the time series, $\hat{H}$ is the estimated Hurst exponent and $t_{n,q}$ is the lower q-quantile of the student-t distribution with n degrees of freedom. The function $c(\hat{H})$ is defined in the supplementary material, and describes how the uncertainty in the trend estimate varies with the Hurst exponent. In fact the factor $\hat{\sigma} c(\hat{H})^{1/2}$ is the standard deviation of estimated psuedo-trends under the assumption that the data is generated by a fGn with scale parameter $\sigma = \hat{\sigma}$ and Hurst exponent $H = \hat{H}$. We note that

$$c(H)^{1/2} \sim n^{2-H}. \tag{7.5}$$

The number $\hat{n}_e = n_e(\hat{H})$ is called the estimated effective sample size, and is defined in the supplementary material. The actual sample size n of the time series is replaced with the effective sample size to compensate for the reduced statistics that results from a strongly correlated time series. The p-value approximation corresponding to confidence intervals in equation (7.4) are

$$p = 2\left(1 - F\Big(\Big|\frac{\hat{a}_1}{\hat{\sigma} c(\hat{H})^{1/2}}\Big|\Big)\right),$$

where F denotes the (cumulative) distribution function for the student-t distribution with $\hat{n}_e - 2$ degrees of freedom.

We remark that regression techniques for linear trends under an AR(1) hypothesis follow along the same lines as the fGn case, and we refer the reader to Lee and Lund (2004). Further details about the data analysis techniques are found in the supplementary material.

7.4 Results and discussion

Fig. 7.2(a) shows the estimated trends for the period 1900-2014 in the GISSTEMP data set. We observe warming over all of Earth's surface, except for a small region in the North-Atlantic. The warming trends are generally weaker in the ocean surface compared to over land, and in particular we observe weaker trend over the Pacific ocean. Fig. 7.2(b) shows the estimated fluctuation level (sample standard deviation) of the temperature signal after subtracting the estimated linear trend.

Very large fluctuation levels are observed over land compared with the oceans, and hence it is unclear which regions have the strongest trends relative to the natural fluctuations. The estimated Hurst exponents are shown in Fig. 7.2(c), and we observe stronger in ocean temperatures than in land temperatures. In North-America and in Eurasia the estimated

model is close to a white-noise process, i.e. $H \approx 0.5$, where as we apparently have strong LRM in the oceans, in particular in the tropical Pacific. A similar picture is seen in Fig. 7.2(d). Here we have plotted the estimated correlation length in an OU process. We see that the estimated correlation time varies from a few months over much of Earth's land areas, to a couple of years in the tropical Pacific and tropical Atlantic oceans.

Based on the parameter estimates presented in Fig. 7.2 we can compute the p-values for the estimated trends. As illustrated in Fig. 7.3(a-b), these p-values depend crucially on the chosen null model. In Fig. 7.3(a) we have shown a map of the p-values computed with respect to the fGn model, and in 7.3(b) we have show the corresponding p-values computed with respect to the OU model. A striking feature in these plots are that the temperature trends for grid points in the Pacific ocean are determined as significant with respect to a SRM model, but cannot be concluded as significant if we apply a LRM model. Hence, our interpretation of the significance of the local warming trends in the Pacific ocean depend on which model is best suited to describe the correlation structure in these data. In 7.3(c) we have plotted the results of the likelihood ratio model selection test. It shows that an OU process is preferred over a fGn in much of the Pacific ocean, and in Fig. 7.3(d) we have combined Figs. 7.3(a) and 7.3(b) so that the p-value for the preferred model is plotted in each grid point. When combining the two models we have more grid points with significant warming than what is predicted by only using the fGn model, but less than predicted by just using the OU process.

Remark. *The p-values have been adjusted for multiple testing using the False Discovery Rate (FDR) method (Benjamini and Hochberg, 1995).*

7.4.1 Exploring the effect of a second scaling regime

It is important to realize that in a 100 year-long time series the statistics is very poor on decadal time scales. They are sufficient to estimate a linear trend with reasonably small errors, but there is very little information about the fluctuation function $\Sigma(\Delta t)$ on these time scales. Here $\Sigma(\Delta t)$ denotes standard deviation of the signal obtained after coarse graining the monthly temperature record over time intervals of length Δt. Trend testing is based on the assumption that the model choice, which is justified by analysis of the stochastic properties on the shorter time scales, also is valid on the longer time scales. For instance, if we find that a OU process with correlation time τ is a good description of short-scale (up to a few years) the temperature fluctuations in the ENSO regions, then we hypothesize that the OU process correctly will prescribe the fluctuation levels on the longer time scales via the relation

$$\Sigma(\Delta t) = C\Delta t^{-1/2} \tag{7.6}$$

for $\Delta t \gg \tau$, with $C = \sigma/\sqrt{2\tau}$. In the same way, if a regional temperature signal is modeled as LRM process with Hurst exponent H, then this implies a fluctuation function on the form

$$\Sigma(\Delta t) = \sigma\Delta t^{1-H}\,. \tag{7.7}$$

Due to the poor statistics we cannot verify the correctness of these relations on the longest time scales, and it is therefore conceivable that the fluctuation function scales differently with Δt on these scales. We can for instance imagine a fluctuation function with a sharp scaling break at a time scale s:

$$\Sigma(\Delta t) = \begin{cases} \sigma_1 \Delta t^{H_1 - 1} & \Delta t < s \\ \sigma_2 \Delta t^{H_2 - 1} & \Delta t > s \end{cases} . \tag{7.8}$$

Continuity of the fluctuation function implies that $\sigma_2 = s^{H_1 - H_2}\sigma_1$. If T denotes the longest time scale relevant for the trend, then the second scaling regime changes the fluctuation at this scale by a factor

$$k = \left(\frac{T}{s}\right)^{H_2 - H_1} . \tag{7.9}$$

From Eq. 7.5 we see that the standard deviation of the pseudo trends is proportional to $T\Sigma(T)$:

$$\sigma c(H)^{1/2} \propto \sigma \left(\frac{1}{T}\right)^{2-H} = T\Sigma(T) . \tag{7.10}$$

Thus, if the inclusion of a second scaling regime yields the change $\Sigma(T) \to k\Sigma(T)$, then for the purpose of testing the significance of a linear trend, this has the same effect as scaling the estimated trend according to $\hat{a} \to \hat{a}/k$.

If the second scaling regime is given by the exponent $H_2 \approx 0.8$, consistent with the scaling exponents observed for global temperature fluctuations on decadal time scales, then we have $H_2 - H_1 \approx 0.3$ in in regions where $\Sigma(\Delta t) \sim \Delta t^{-1/2}$. From inspection of the estimated power spectral densities of the temperature time series (see Fig. 7.1(b)), we can be relatively certain that if there is a second scaling regime, then it does not dominate on time scales $\lesssim$ 20 years, but if we let $s = 20$ years and $T = 114$ years, we obtain $k = 1.7$. In Fig. 7.4(a) and (b) we have plotted the p-values obtained after adjusting the test observable by a factor k in the regions where the model selection test did not (with statistical significance) favor a LRM model. The k-values are $k = 2$ and $k = 3$ in the two figures respectively. These figures should be compared with Fig. 7.3(d), and we observe that the existence of LRM on time scales longer than 20 years would substantially changes our conclusions about the significance of the regional warming trends. The most striking differences are in the Pacific ocean, the Southern Oceans and in the northern parts of Eurasia.

Then main result of this paper is that our conclusions about trend significance for the last century on gridded temperature data are highly model dependent, and that a direct application of a single class of models will give misleading results. We also show that trend significance is sensitive to the persistence on time scales that are relatively long compared to what can be analyzed from the instrumental data. In future work we will explore the possibility of strengthening the results presented in this paper by using long data sets from various climate models to discern the nature of the regional temperature variability on time scales from a few decades to a few centuries.

Table 7.1: Percentage significant trends at the 1% ($p < 0.01$) and 5% ($p < 0.05$) significance level assuming a fGn null hypothesis (2. column), AR(1) null hypothesis (3. column). In the last column (prefered model) the trend signficance are tested against the model selected by the likelihood-ratio criteria. Note that the different data sets have have different coverages of Earth's surface. In particular, the HadCrut4 data covers less surface than the other data sets (see the suplementary material).

	fGn		AR(1)		Prefered model	
	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$
GISSTEMP	47%	68%	87%	93%	68%	82%
Berkely Earth Data	52%	72%	90%	95%	71%	83%
HadCrut4	35%	54%	77%	85%	55%	69%
NOAA MLOST	45%	65%	88%	94%	70%	83%

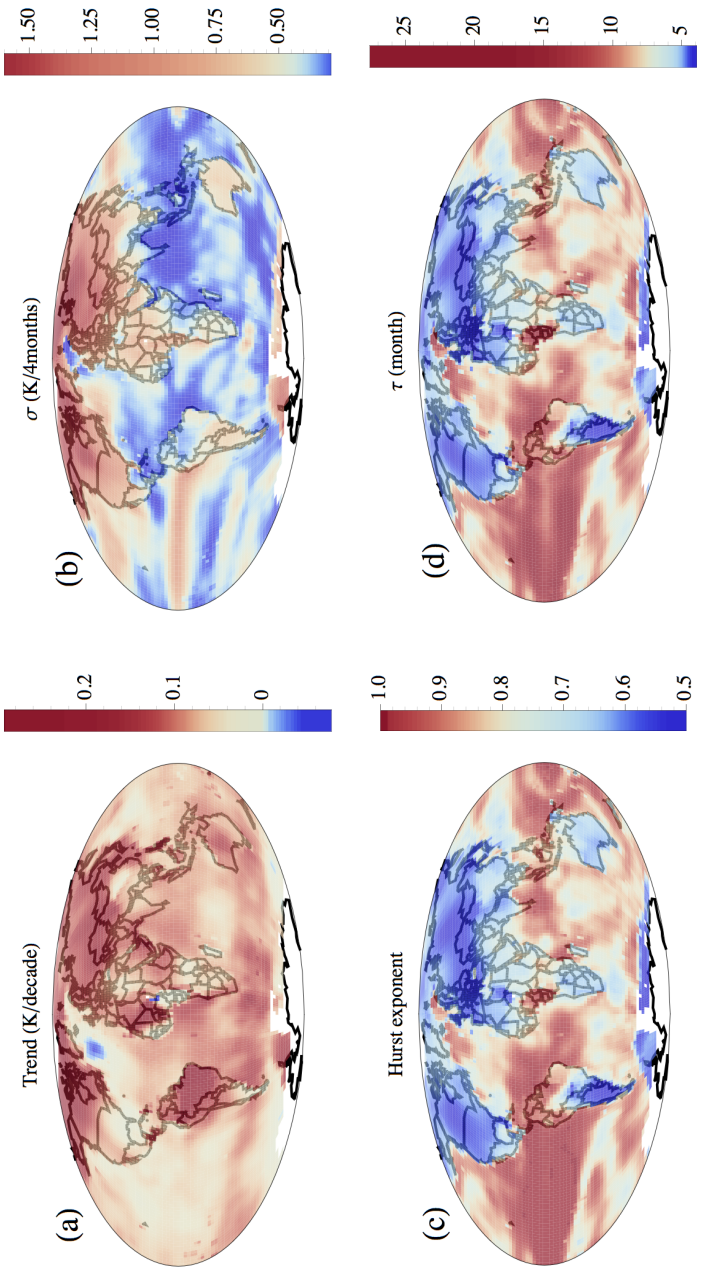


Figure 7.2: (a): Estimeted linear trend for the period 1900-2014 in each $2° \times 2°$ grid of the GISSTEMP data set. (b): Estimated fluctuation level σ. (c) Hurst exponents estimated using MLE. (d) Correlation time τ in the OU process (AR(1) process) estimated using MLE. All estimates are preformed subsequent to a 4-month coarse graining.

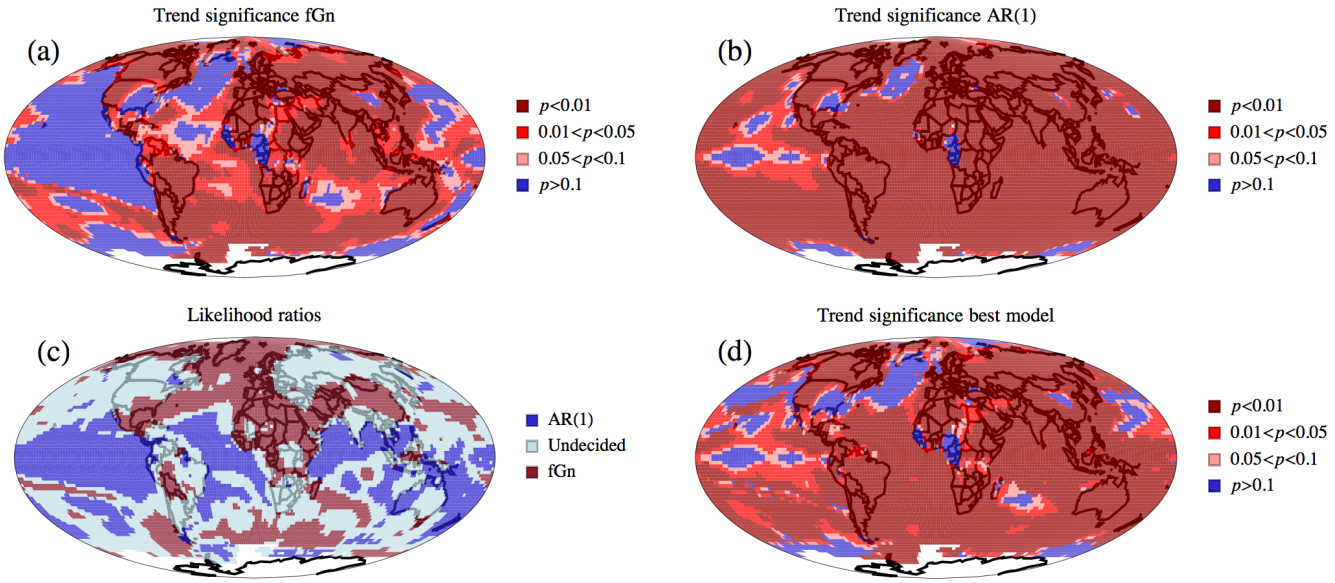


Figure 7.3: (a): The distribution of p-values based on a fGn null model. (b): The distribution of p-values based on an AR(1) null model. (c): Shows the results of the likelihood ratio model selection test. In the grid points marked as red the data is more consistent with a fGn error model, and in the grid points marked as blue the data is more consistent with an AR(1) error model. In the grid points marked as light blue, one model is not significantly preferred over the other. (d): Shows the distribution of p-values when the most model with the highest likelihood is chosen as the null model in each grid point.

Figure 7.4: (a): The figure is a modification of Fig. 7.3(d), where the test observable is divided by the factor $k = 2$ in the grid points where the model selection test is undecided or prefers the SRM model. (b): As in (a) but with $k = 3$.

8 Bibliography

Andersen, T. G. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77*(2), 343–377.

Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'É.N.S 17*, 21–86.

Bacry, E., J. Delour, and J. Muzy (2001). Multifractal random walk. *Physical Review E 64*(2), 026103.

Bacry, E., A. Kozhemyak, and J. F. Muzy (2008). Continuous cascade models for asset returns. *Journal of Economic Dynamics and Control 32*(1), 156–199.

Bacry, E. and J. F. Muzy (2003). Log-Infinitely Divisible Multifractal Processes. *Communications in Mathematical Physics 236*(3), 449–475.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics 73*(1), 5–59.

Baillie, R. T. and S.-K. Chung (2002). Modeling and forecasting from trend-stationary long memory models with applications to climatology. *International Journal of Forecasting 18*(2), 215–226.

Bali, T. G. and L. Wu (2006). A comprehensive analysis of the short-term interest-rate dynamics. *Journal of Banking and Finance 30*(4), 1269–1290.

Bates, D. and M. Maechler (2014). *Matrix: Sparse and Dense Matrix Classes and Methods*.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B ....*

Benth, F. E., J. Kallsen, and T. Meyer-Brandis (2007). A Non-Gaussian Ornstein–Uhlenbeck Process for Electricity Spot Price Modeling and Derivatives Pricing. *Applied Mathematical Finance 14*, 153–169.

Beran, J., Y. Feng, S. Ghosh, and R. Kulik (2013). *Long-Memory Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bølviken, E. (2014). *Computation and Modelling in Insurance and Finance*. Cambridge University Press.

Bos, M. S., R. Fernandes, S. Williams, and L. Bastos (2012). Fast error analysis of continuous GNSS observations with missing data. *Journal of Geodesy 87*(4), 1–10.

Brenner, R. J., R. H. Harjes, and K. F. Kroner (1996). Another Look at Models of the Short-Term Interest Rate. *Journal of Financial and Quantitative Analysis 31*, 85–107.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2 ed.). Springer.

Bunde, A., J. Ludescher, C. L. E. Franzke, and U. Büntgen (2014). How significant is West Antarctic warming? *Nature Geoscience 7*(4), 246–247.

Bye, T. and E. Hope (2005). Deregulation of Electricity Markets: The Norwegian Experience. *Economic and Political Weekly 40*, 5269–5278.

Calvet, L. and A. Fisher (2001, November). Forecasting multifractal volatility. *Journal of Econometrics 105*(1), 27–58.

Calvet, L. E. and A. Fisher (2004). How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Processes. *Journal of Financial Econometrics 2*(1), 49–83.

Casella, G. and R. L. Berger (2002). *Statistical Inference*. Duxbury.

Chapman, D. A. and N. D. Pearson (2001). Recent Advances in Estimating Term-Structure Models. *Financial Analysts Journal 57*, 77–95.

Chapman, S. C., B. Hnat, G. Rowlands, and N. W. Watkins (2005). Scaling collapse and structure functions: identifying self-affinity in finite length time series. *Nonlinear Processes in Geophysics 12*(6), 767–774.

Cohn, T. A. and H. F. Lins (2005). Nature's style: Naturally trendy. *Geophysical Research Letters 32*(23), L23402.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). A Theory of the Term Structure of Interest Rates. *Econometrica 53*, 385–407.

Craigmile, P. and D. Mondal (2013). Estimation of long range dependence in gappy gaussian time series.

Dahlhaus, R. (1989). Efficient Parameter Estimation for Self-Similar Processes. *The Annals of Statistics 17*(4), 1749–1766.

Dahlhaus, R. (2006). Correction: Efficient Parameter Estimation for Self-Similar Processes. *The Annals of Statistics 34*(2), 1045–1047.

Das, S. R. (2002). The surprise element: jumps in interest rates. *Journal of Econometrics 106*(1), 27–65.

Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative Finance 7*, 21–36.

Durham, G. B. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics 70*(3), 463–487.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik 322*(8), 549–560.

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

Erlwein, C., F. E. Benth, and R. Mamon (2010). HMM filtering and parameter estimation of an electricity spot price model. *Energy Economics 32*, 1034–1043.

Erzgräber, H., F. Strozzi, J. M. Zaldívar, H. Touchette, E. Gutiérrez, and D. K. Arrowsmith (2008). Time series analysis and long range correlations of Nordic spot electricity market data. *Physica A: Statistical Mechanics and its Applications 387*(26), 8–6574.

Franzke, C. (2012). On the statistical significance of surface air temperature trends in the Eurasian Arctic region. *Geophysical Research Letters 39*, L23705.

Ghashghaie, S., W. Breymann, J. Peinke, P. Talkner, and Y. Dodge (1996). Turbulent cascades in foreign exchange markets. *NATURE 381*(6585), 767–770.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010). GLOBAL SURFACE TEMPERATURE CHANGE. *Reviews of Geophysics 48*(4), RG4004.

Hansen, J., M. Sato, P. Kharecha, and K. von Scuckmann (2011). Earth's energy imbalance and implications. *Atmos. Chem. Phys. 11*, 13421–13449.

Hansen, J., M. Sato, R. Ruedy, L. Nazarenko, A. Lacis, G. A. Schmidt, G. Russell, I. Aleinov, M. Bauer, S. Bauer, N. Bell, B. Cairns, V. Canuto, M. Chandler, Y. Cheng, A. Del Genio, G. Faluvegi, E. Fleming, A. Friend, T. Hall, C. Jackman, M. Kelley, N. Kiang, D. Koch, J. Lean, J. Lerner, K. Lo, S. Menon, R. Miller, P. Minnis, T. Novakov, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, D. Shindell, P. Stone, S. Sun, N. Tausnev, D. Thresher, B. Wielicki, T. Wong, M. Yao, and S. Zhang (2005). Efficacy of climate forcings. *Journal of Geophysical Research 110*(D18), D18104.

Haslett, J. and A. E. Raftery (1989). Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. *Applied Statistics 38*(1), 1.

Hong, Y., H. Li, and F. Zhao (2004). Out-of-Sample Performance of Discrete-Time Spot Interest Rate Models. *Journal of Business and Economic Statistics 22*(4), 457–473.

Hurst, H. E. (1957). A suggested statistical model of some time series which occur in nature. *NATURE 180*(4584), 494

Huybers, P. and W. Curry (2006). Links between annual, Milankovitch and continuum temperature variability. *NATURE 441*(7091), 329–332.

Johannes, M. (2004). The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *The Journal of Finance 59*(1), 227–260.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research 117*(D5), D05127.

Kahane, J. P. (1985). Sur le chaos multiplicatif. *Ann. Sci. Math. Quebec 9*(2), 105–150.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research 116*(D14), D14104.

Ko, K., J. Lee, and R. Lund (2008). Confidence intervals for long memory regressions. *Statistics & Probability Letters 78*(13), 1894–1902.

Koedijk, K. G., F. G. J. A. Nissen, P. C. Schotman, and C. C. P. Wolff (1997). The Dynamics of Short-Term Interest Rate Volatility Reconsidered. *European Finance Review 1*(1), 105–130.

Kolmogorov, A. N. (1941). The local structure of turbulence in incoressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk. SSSR 31*(301-305).

Kolmogorov, A. N. (1962). A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics 13*, 82–85.

Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal 48*(1), 3–24.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Ann. Math. Stat. 22*, 79–86.

La Cruz, W., J. M. Martínez, and M. Raydan (2006, July). Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation 75*(255), 1429–1449.

Laplace, P. S. (1986). Memoir on the Probability of the Causes of Events. *Statistical Science 1*(3), 364–378.

Lee, J. and R. Lund (2004). Revisiting simple linear regression with autocorrelated errors. *Biometrika 91*(1), 240–245.

Lee, J. and R. Lund (2008). Equivalent sample sizes in time series regressions. *Journal of Statistical Computation and Simulation*.

Lennartz, S. and A. Bunde (2009). Trend evaluation in records with long-term memory: Application to global warming. *Geophysical Research Letters 36*(16), L16706.

Levinson, N. (1946). The Wiener rms (root mean square) error criterion in filter design and prediction. *J- Math. Phys. 25*, 261–278.

Lobato, I. N. and C. Velasco (2000). Long Memory in Stock-Market Trading Volume. *Journal of Business and Economic Statistics 18*(4), 410–427.

Longstaff, F. A., K. C. Chan, G. A. Karolyi, and A. B. Sanders (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate. *The Journal of Finance 47*(3), 1209–1227.

Løvsletten, O. (2010). Empirical analysis and stochastic modeling of temporal fluctuations in the Norwegian InterBank Offered Rate.

Løvsletten, O. and M. Rypdal (2012a). Approximated maximum likelihood estimation in multifractal random walks. *Physical Review E 85*, 046705.

Løvsletten, O. and M. Rypdal (2012b). R package mrw. `http://www.complexityandplasmas.net/nordforsk/Papers.html`.

Lux, T. (2003). The multi-fractal model of asset returns: Its estimation via GMM and its use for volatility forecasting. *Christian-Albrechts-Universität Kiel*.

Lux, T. (2006). The Markov-Switching Multifractal Model of asset returns : GMM estimation and linear forecasting of volatility. *Economics Working Papers 17*.

Lux, T. (2008). The Markov-Switching Multifractal Model of Asset Returns. *Journal of Business and Economic Statistics 26*(2), 194–210.

Lux, T. (2012). Turbulence in financial markets: the surprising explanatory power of simple cascade models, Quantitative Finance. *Quantitative Finance 1*, 632–640.

Malo, P. (2006). Multifractality in nordic electricity markets. *Helsinki School of Economics Working Papers*, 26.

Malo, P. (2009). Modeling electricity spot and futures price dependence: A multifrequency approach. *Physica A: Statistical Mechanics and its Applications 388*, 4763–4779.

Mandelbrot, B. B. (1963). The Variation of Certain Speculative Prices . *The Journal of Business 36*, 394–419.

Mandelbrot, B. B., L. Calvet, and A. Fisher (1997). A Multifractal Model of Asset Returns.

Mann, M. E. (2010). On long range dependence in global surface temperature series. *Climatic Change 107*(3-4), 267–276.

Mantegna, R. N. and H. E. Stanley (1994). Stochastic Process with Ultraslow Convergence to a Gaussian: The Truncated Lévy Flight. *Physical Review Letters 73*, 2946–2949.

Mantegna, R. N. and H. E. Stanley (1996). Turbulence and financial markets. *NATURE 383*, 587–588.

Mantegna, R. N. and H. E. Stanley (2000). *An Introduction to Econophysics*. Correlations and Complexity in Finance. Cambridge University Press.

Martino, S., K. Aas, O. Lindqvist, L. R. Neef, and H. Rue (2011). Estimating stochastic volatility models using integrated nested Laplace approximations. *The European Journal of Finance 17*(7), 487–503.

McLeod, I. A., H. Yu, and Z. L. Krougly (2007). Algorithms for Linear Time Series Analysis: With R Package. *Journal of Statistical Software 23*(5), 1–26.

Mitchell, W. C. (1915). The making and using of index numbers. *Bulletin of the United States Bureau of Labor Statistics 173*, 5–114.

Mondal, D. and D. B. Percival (2008). Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics 62*(5), 943–966.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research 117*(D8), D08101.

Muzy, J.-F. and E. Bacry (2002). Multifractal stationary random measures and multifractal random walks with log infinitely divisible scaling laws. *Physical Review E 66*(5), 056121.

Muzy, J. F., E. Bacry, and A. Arneodo (1991). Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Physical Review Letters 67*, 3515–3518.

Muzy, J.-F., R. Baïle, and E. Bacry (2013). Random cascade model in the limit of infinite integral scale as the exponential of a nonstationary 1/f noise: Application to volatility fluctuations in stock markets. *Physical Review E 87*(4), 042813.

Norouzzadeh, P., W. Dullaert, and B. Rahmani (2007). Anti-correlation and multifractal features of Spain electricity spot market. *Physica A: Statistical Mechanics and its Applications 380*, 333–342.

Obukhov, A. M. (1962). Some Specific Features of Atmospheric Turbulence. *Journal og Geophysical Research 67*, 3011–3014.

Østvand, L., K. Rypdal, and M. Rypdal (2013). Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years. *arxiv*.

Palma, W. (2007). *Long-memory time series: theory and methods*. John Wiley & Sons.

Pathirana, A. and S. Herath (2002). Multifractal modelling and simulation of rain fields exhibiting spatial heterogeneity. *Hydrology and Earth System Sciences 6*, 695–708.

Peng, C. K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994). Mosaic organization of DNA nucleotides. *Physical Review E 49*(2), 1685–1689.

Percival, D. B. and A. T. Walden (1993). *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge University Press.

Pesin, Y. (1997). *Dimension Theory in Dynamical Systems*. Chicago Lectures in Mathematics.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rea, W., M. Reale, and J. Brown (2011). Long memory in temperature reconstructions. *Climatic Change 107*(3-4), 247–265.

Riedi, R. H., M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk (1999). A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory 45*(3), 992–1018.

Robert, R. and V. Vargas (2010). Gaussian multiplicative chaos revisited. *Ann. Probab. 38*(2), 605–631.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(2), 319–392.

Rypdal, K., L. Østvand, and M. Rypdal (2013). Long-range memory in Earth's surface temperature on time scales from months to centuries. *Journal of Geophysical Research: Atmospheres 118*(13), 7046–7062.

Rypdal, M. and O. Løvsletten (2011). Multifractal modeling of short-term interest rates. *arXiv.org*.

Rypdal, M. and O. Løvsletten (2013, January). Modeling electricity spot prices using mean-reverting multifractal processes. *Physica A: Statistical Mechanics and its Applications 392*(1), 194–207.

Rypdal, M. and K. Rypdal (2010). Stochastic modeling of the AE index and its relation to fluctuations in Bz of the IMF on time scales shorter than substorm duration. *Journal of Geophysical Research 115*.

Rypdal, M. and K. Rypdal (2011). Discerning a linkage between solar wind turbulence and ionospheric dissipation by a method of confined multifractal motions. *Journal of Geophysical Research 116*, A02202.

Rypdal, M. and K. Rypdal (2012). Is there long-range memory in solar activity on time scales shorter than the sunspot period? *Journal of Geophysical Research 117*, A04103.

Rypdal, M. and K. Rypdal (2014). Long-memory effects in linear-response models of Earth's temperature and implications for future global warming. *Journal of Climate*.

Rypdal, M., E. Sirnes, O. Løvsletten, and K. Rypdal (2013). Assessing market uncertainty by means of a time-varying intermittency parameter for asset price fluctuations. *Physica A: Statistical ....*

Samuelson, P. A. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review 6*, 41–49.

Shumway, R. H. and D. S. Stoffer (2010). *Time series analysis and its applications: with R examples* . Springer.

Simonsen, I. (2003). Measuring anti-correlations in the nordic electricity spot market by wavelets. *Physica A: Statistical Mechanics and its Applications 322*(0), 10–606.

Simonsen, I. (2005). Volatility of power markets. *Physica A: Statistical Mechanics and its Applications 355*, 10–20.

Simonsen, I., A. Hansen, and O. M. Nes (1998). Determination of the Hurst exponent by use of wavelet transforms. *Physical Review E 58*, 2779–2787.

Skaug, H. and J. Yu (2009). Automated likelihood based inference for stochastic volatility models.

Smith, T. M. and R. W. Reynolds (2005). A Global Merged Land–Air–Sea Surface Temperature Reconstruction Based on Historical Observations (1880–1997). *J. Climate 18*(12), 2021–2036.

Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2013). Climate change 2013: The physical science basis. *Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge Univ Press, New York)*.

Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes—A Study of Daily Sugar Prices, 1961–79. In *Time Series Analysis, Theory and Practice*, pp. 203–226. North Holland.

Thiele, T. N. (1880). Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en 'systematisk'Karakter. *Det Kongelige Danske Videnskabernes Selskabs Skrifter-Naturvidenskabelig og Mathematisk Afdeling*, 381–408.

Trench, W. F. (1964). An Algorithm for the Inversion of Finite Toeplitz Matrices. *Journal of the Society for Industrial and Applied Mathematics 12*, 515–522.

Uhlenbeck, G. E. and L. S. Ornstein (1930). On the Theory of the Brownian Motion. *Physical Review 36*, 824–841.

Varadhan, R. and P. Gilbert (2009). BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software 32*(4), 1–26.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica 57*, 307–333.

Weron, R. (2000). Energy price risk management. *Physica A: Statistical Mechanics and its Applications 285*(1–2), 127–134.

Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices*. A Statistical Approach. Wiley.

Weron, R. and B. Przybylowicz (2000). Hurst analysis of electricity price dynamics. *Physica A: Statistical Mechanics and its Applications 283*(3-4), 7.

Weron, R., I. Simonsen, and P. Wilman (2003). Modeling highly volatile and seasonal markets: evidence from the Nord Pool electricity market. In *The Application of Econophysics: Proceedings of the Second Nikkei Econophysics Symposium*, pp. 10.

Zwiers, F. W. and H. von Storch (1995). Taking Serial Correlation into Account in Tests of the Mean. *Journal of Climate 8*(2), 336–351.