Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Multifractal analysis of petrol and diesel prices in the Czech Republic

Author: **Bc. Martin Baletka**

Supervisor: **PhDr. Ladislav Krištoufek**

Academic Year: **2012/2013**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 24, 2013

Signature

Acknowledgments

I am very grateful to my supervisor PhDr. Ladislav Krištoufek for his patience, priceless advice and constructive comments without which this thesis would not be possible.

Abstract

This thesis examines scaling properties of petrol and diesel prices in the Czech Republic and a crude oil price over the period from January 2004 to February 2013. Using generalised Hurst exponent and multifractal detrended fluctuation analysis techniques we find out that crude oil market is efficient, do not contain long memory and the returns exhibit monofractal behaviour. On the other hand, petrol and diesel markets in the Czech Republic are not efficient, because their returns contain long-range dependence in autocorrelations and exhibit multifractal behaviour caused mostly by fat-tailed distribution. Thus, fuels can be modelled by complex methods like Markov switching multifractal model.

JEL Classification	C15, C16, C46
Keywords	petrol, diesel, crude oil, long memory, multifractality, GHE, MF-DFA

Author's e-mail	`martin.baletka@ies-prague.org`
Supervisor's e-mail	`kristoufek@ies-prague.org`

Abstrakt

Tato práce zkoumá škálování cen benzínu a motorové nafty v České republice a ceny ropy na datech v období od ledna 2004 do února 2013. Použitím metod zobecněného Hurstova exponentu a multifraktální detrendované fluktuační analýzy jsme zjistili, že trh s ropou je efektivní, bez přítomnosti dlouhé paměti v autokorelacích a výnosy na trhu s ropou vykazují monofraktální škálování. Na druhou stranu český trh s pohonnými hmotami není efektivní, protože je ovlivněn dlouhou pamětí v autokorelacích výnosů benzínu a nafty, a vykazuje multifraktální škálování, které je způsobeno zejména distribucí výnosů s těžkými chvosty. Pro modelování pohonných hmot je tedy nutné použít složitější metody, jako například multifraktální Markov switching model.

Klasifikace JEL	C15, C16, C46
Klíčová slova	benzín, diesel, ropa, dlouhá paměť, multifraktalita, GHE, MF-DFA

E-mail autora	`martin.baletka@ies-prague.org`
E-mail vedoucího práce	`kristoufek@ies-prague.org`

Contents

List of Tables

List of Figures

Acronyms

MLE maximum likelihood estimator

ARFIMA autoregressive fractionally integrated moving average

R/S rescaled range

GHE generalised Hurst exponent

DMA detrended moving average

DFA detrended fluctuation analysis

MF-DFA multifractal detrended fluctuation analysis

VAT value added tax

CZK Czech koruna

USD U.S. dollar

RON research octane number

WTI West Texas Intermediate

CNB Czech National Bank

CACR Customs Administration of the Czech Republic

MITCR Ministry of Industry and Trade of the Czech Republic

MFCR Ministry of Finance of the Czech Republic

Master Thesis Proposal

Author	Bc. Martin Baletka
Supervisor	PhDr. Ladislav Krištoufek
Proposed topic	Multifractal analysis of petrol and diesel prices in the Czech Republic

Topic characteristics The thesis will focus on multifractal analysis of changes in price of petrol and diesel in the Czech Republic. Starting from raw time series of prices, which would need to be adjusted for taxes to be usable for our analysis, since both petrol and diesel are heavily taxed. The autocorrelation function would suggest strong significance with long lags. Using common sense only, one would expect them to show the same behaviour as crude oil, the core element of both fuels. However, crude oil market is much more efficient and does not show the same long memory as fuels. This suggests that both fuels are fractal processes. We will try to find a proper model for the time series using both fractal and multifractal analytics. To the author's best knowledge, this kind of analysis has not been performed with the fuel prices yet.

Hypotheses

1. Changes in petrol and diesel prices are strongly autocorrelated.
2. Changes in petrol and diesel prices have long memory and thick tailed distributions.
3. Changes in petrol and diesel prices are multifractal.

Methodology The time series cover long time data of petrol and diesel retail prices for the Czech Republic. To analyze them thoroughly adjustments for taxes will be required, because both excise tax and value added tax have been changed during the time window. Further, the time series will be examined via variance scaling, autoregressive fractionally integrated moving average (ARFIMA) and maximum likelihood estimator (MLE).

Outline

1. Introduction
2. Theoretical Background
3. Related Work
4. The Model
5. Empirical Verification
6. Conclusion

Core bibliography

1. Di Matteo, T., T. Aste & M. Dacarogna (2005): "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development." *Journal of Banking & Finance* **29(4)**: pp. 827–851.

2. Di Matteo, T. (2007): "Multi-scaling in finance." *Quantitative Finance* **7(1)**: pp. 21–36.

3. Kantelhardt, J. W. (2008): *"Fractal and Multifractal Time Series" Institute of Physics, Martin-Luther-University.*

4. Suleymanov, A. S., A. A. Abbasov & A. J. Ismaylov (2009): "Fractal analysis of time series in oil and gas production." *Chaos, Solitons & Fractals* **41(5)**: pp. 2474–2483.

5. Yamamoto, Y. & R. L. Hughson (1993): "Extracting fractal components from time series." *Physica D: Nonlinear Phenomena* **68(2)**: pp. 250–264.

6. Malamud, B. D. & D. L. Turcotte (1999): "Self-affine time series: measures of weak and strong persistence." *Journal of Statistical Planning and Inference* **80(1-2)**: pp. 173–196.

Author   Supervisor

Chapter 1

Introduction

Crude oil is a commodity of fundamental importance if not one the most important in the modern world. In 2011 humankind consumed 87 million barrels of crude oil per day. That is 10 million more than 10 years ago.[1] Over the past few years the price of crude oil has become much more volatile than ever before in recent history. The price[2] was virtually stable from 1987 to 1999 with the exception of the period of Gulf War (August 1990–February 1991), when the price peaked at 41.45 USD/barrel, with average of 29.08 USD/barrel. That is approximately double of daily averages of 17.26 USD/barrel during the pre-war period (1987-1990) and 17.67 USD/barrel during the post-war period (1991-1999). From 2004 the price gradually increased to its historical maximum of 143.95 USD/barrel in July 2008, then suddenly dropped to 33.73 USD/barrel in December 2008, followed by steady increase to 126.64 USD/barrel in May 2011. Since then it has been fluctuating around 110 USD/barrel. Overall the period from 2004–2013 exhibit extreme volatility in comparison to data from 1987–2003. This can be illustrated by the difference between minimum and maximum price in each period: *(i)* 1987–2003: 32.35 USD/barrel (min. 9.10, max. 41.45) compared to *(ii)* 2004–2013: 114.93 USD/barrel (min. 29.02, max. 143.95). The source of this fluctuations is often attributed to constantly growing demand for crude oil, as it is a vital component in many of our daily used products, and the uncertainty in supply in terms of both availability (political situation in most producing countries is often very unstable) and existence (costly discoveries and generally limited supply). Most importantly, crude oil is an essential ingredient in the production of petrol and diesel – fuels required for the operation of combustion engines used in cars and other self-propelled vehi-

[1] `http://www.indexmundi.com/energy.aspx?product=oil&graph=consumption`
[2] `http://tonto.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=D`

cles. There are more than 1 billion motor vehicles in the world today and the number is progressively rising. The total number of motor vehicles more than doubled during the past 10 years. International Organization of Motor Vehicle Manufacturers (OICA) reports that more than 60 million passenger cars and additional more than 20 million light commercial vehicles, heavy trucks, buses, coaches and other commercial vehicles were made in 2012 worldwide.[3] In the Czech Republic, the number of registered self-propelled vehicles increased from 4 844 019 at the end of 2003 to 6 142 719 at the end of 2011. Most of which consist of passenger cars where the number grew from 3 706 012 to 4 581 642 (Czech Statistical Office a;b). In other words the number of inhabitants per a passenger car dropped from 2.75 to 2.29. Regarding the Czech fuel market, it is characterised by a dense network of petrol stations, one of the highest in terms of per capita in Central Europe. Nevertheless, petrol and diesel prices do not exhibit the same fluctuations as crude oil price. Despite the high penetration of petrol stations, petrol and diesel market does not show the same market efficiency as crude oil market, which is classified as a commodity market. Petrol and diesel prices tend to be more sticky. Although a dramatic increase in price of crude oil is often used as an argument to rapid increase of petrol and diesel prices, dramatic decrease in price of crude oil of the same magnitude is unlikely to cause a rapid decrease in the price of petrol and diesel.

The thesis focuses on efficiency of the Czech fuel market by exploring the autocorrelation structure and performing the multifractal analysis of changes in price of petrol and diesel in the Czech Republic. Using common sense only, one would expect them to exhibit the same behaviour as crude oil, the core element of both fuels. However, we will show that crude oil market is much more efficient and does not contain the same long memory as fuels. Overall, petrol and diesel markets reveal fundamentally different properties in autocorrelations and multifractality than crude oil market. The objective of this thesis is to prove by employing generalised Hurst exponent (GHE) and multifractal detrended fluctuation analysis (MF-DFA) techniques that petrol and diesel markets are less efficient than crude oil market because they contain long-range dependence in autocorrelations, and are subject to multifractal behaviour, which arises mostly from the broad probability distribution function. To the best of our knowledge, this kind of analysis has not been performed with the fuel prices yet.

The thesis is structured as follows: Chapter 2 reviews relevant literature dealing with fractal analysis or fuel markets available up to date. Chapter 3

[3] `http://oica.net/category/production-statistics/`

explains the methodology used in this thesis. Chapter 4 describes the dataset used in the empirical part and gives insight into the Czech fuel market. Chapter 5 presents and discuss our empirical results. And finally Chapter 6 summarises our findings.

Chapter 2

Literature Review

This chapter summarises relevant literature regarding multifractal analysis up to date, that is particularly long-range dependence in autocorrelations also known as long memory, and multifractality per se.

Kantelhardt *et al.* (2002) develop MF-DFA, a method for the multifractal analysis of nonstationary time series based on a generalisation of the detrended fluctuation analysis (DFA). The MF-DFA consist of five steps, with the first three almost identical to the original DFA. By comparing their MF-DFA method to the standard DFA, they prove that both approaches are equivalent for stationary signals with compact support. Further, they show that the new method can reliably determine the multifractal scaling behaviour of time series. Moreover, comparing the MF-DFA results for original series with those for shuffled series it can be distinguished between multifractality due to long-range autocorrelations and multifractality due to a broad probability density function. Finally, they highlight that one of the reason for employing DFA or MF-DFA method is to avoid spurious detection of correlations that are a mere artefacts of nonstationarities in the time series.

Di Matteo *et al.* (2003) discuss scaling properties of four different stock market indices as a "stylised fact" and demonstrate that the $H(q)$ from GHE approach is a powerful instrument to characterise and differentiate the scaling structure of such markets. They find that very developed markets, represented by Nasdaq 100 (USA) and Nikkei 225 (Japan), have $H(2) \approx 0.46 < 0.5$, which contrasts with emerging markets, represented by WIG (Poland) and JSX (Indonesia), which have $H(2) \approx 0.58 > 0.5$. Additionally, the empirical analysis across a wide variety of stock markets confirms that the exponent $H(2)$ is sensitive to the degree of development of the market. They find out that developed stock markets show tendency to have $H(2) < 0.5$, while developing markets show tendency to have $H(2) > 0.5$. The robustness of the presented empirical approach is tested in several ways, including varying window sizes, varying maximum time-step, Jackknife method, and comparison with simulated Brownian motions. They verify that the observed differentiation among different degrees of market development is well above the numerical fluctuations and statistical errors. On the other hand, the analysis over different sub-periods reveals significant changes in the scaling behaviour of a given market in time. This indicates that the scaling structure of markets is an evolving quantity which is not only able to differentiate among markets at different development stage but can also catch the overall variability of the market conditions.

Di Matteo *et al.* (2005) extend the previous study and estimates long-term memories of developed and emerging markets using the scaling analysis to characterise their stage of development. They empirically analyse wide variety of 32 stock indices, 29 foreign exchange rates, treasury rates and eurodollar rates with various maturity dates to demonstrate the sensitivity of the generalised Hurst exponents $H(1)$ and $H(2)$ to the degree of development of the market. They find that for fixed income instruments, $H(2)$ is close to 0.5 while $H(1)$ is systematically larger than 0.5. On the other hand, they find that in stock markets generalised Hurst exponents show remarkable differences between mature and developing markets. Mature markets are characterised by low $H(1)$ and $H(2) \leq 0.5$ while developing markets have high $H(1)$ and $H(2) \geq 0.5$.

Di Matteo (2007) continues in the follow-up paper, where the multi-scaling methods are applied to financial data to discuss different tools used for estimating the scaling exponents, stressing their advantages and disadvantages. The study concludes that GHE approach is a suitable tool for describing the multiscaling properties in financial time series since the method is powerful robust and not biased, as other methods are. When estimations are performed in the frequency domain, there is some bias present; however using the generalised Hurst exponent method produce unbiased results even in this situation.

Alvarez-Ramirez *et al.* (2002) describe multifractal properties of crude oil price dynamics by using rescaled range (R/S) analysis, a technique originally from statistical physics. They use three types of crude oil, Brent (Europe), West Texas Intermediate (WTI) (USA), and Dubai (Persian Gulf). They find out that crude oil price is a persistent process including some long-run memory effects for daily fluctuations ($H_{R/S} \approx 0.58$). Moreover, they detected multifractal structures characterised by nonlinear dependence of the Hurst exponent

$H(q)$ and crossovers indicating several time scales in the evolution of crude oil price. Low time scale ranging from days to weeks with non-Gaussian process probably caused by market speculators, and larger one ranging from weeks to quarters. They conclude that the crude oil market dynamics is the consequence of different events acting at different time scales. Some of them with important influence for the future price in the crude oil market and others introduced by noisy speculators.

In the following study, Alvarez-Ramirez *et al.* (2008) discuss the Hurst exponent dynamics and affiliated long-term autocorrelations of the same international crude oil prices estimated this time with DFA for returns over the period from 1987 to 2007. Using this model-free approach developed in statistical physics reduces the effects of non-stationary market trends in the computation of the Hurst exponent, and focuses on the intrinsic autocorrelation structure of market fluctuations over different time horizons. To test for time-varying degree of autocorrelations, the DFA method is applied over subsample rolling windows with length from 30 to 300 business days. Their results indicate that over long horizons the crude oil market is consistent with the efficient market hypothesis. Nevertheless, for time horizons smaller than one month, meaningful autocorrelations cannot be excluded and the Hurst exponent shows cyclic non-periodic dynamics, well above 0.5, systematically in the 0.6–0.7 range. Additionally, a 1.85-year cycle in the time varying behaviour of the short-term Hurst exponent suggests that short-term inefficiencies can be exploited within this 1.85-year cycles for marginal price forecasting. Therefore, the market produce a time-varying short-term inefficient behaviour that becomes efficient in the long-term.

Further results to studies of efficiency of crude oil markets are provided in Alvarez-Ramirez *et al.* (2010). They estimate WTI crude oil spot price data for 1986–2009 with DFA with lagged autocorrelations. According to their results, the multiscaling pattern is not continuous. Instead, two discontinuities at one-quarter and one-year scales are found, indicating different sources of price fluctuations from speculative effects to fundamental supply and demand shocks. In contrast to previous results, the crude oil market seems to present important deviations from efficiency. Their results indicate positive or negative autocorrelations that might be masked by delay effects.

Barunik & Kristoufek (2010) study the sampling properties of the Hurst exponent estimation methods under heavy-tailed underlying processes. The authors present more realistic settings and practicable implications for not normally distributed returns of financial markets than the majority of existing studies that focus on estimation of expected values and confidence intervals based on simulations of standard normal process. They run extensive Monte Carlo simulation of α-stable distributed random variables with different lengths ranging from 2^9 to 2^{16} observations. The results show that R/S together with GHE are robust to heavy tail in the underlying process, however, detrended moving average (DMA) together with DFA and its multifractal generalisation MF-DFA deteriorate with increasing heavy tail in the underlying distributions. On normal data with $\alpha = 2$, all the methods hold the expected value for all time series lengths and therefore they seem to be better for H estimation than R/S. However, on non-normal simulations, the situation changes dramatically. The heavier tales of the underlying data are, the wider the confidence intervals of the estimates are. MF-DFA$(q = 1)$ tends to underestimate the expected $1/\alpha$ value. They conclude that MF-DFA methods as well as DMA are not appropriate for data with heavier tails and small sample size. On the other hand, both GHE tested methods proved to be very useful as they show the best properties.

Barunik *et al.* (2012) use GHE to examine multiscaling behaviour of financial time series, namely a collection of stock exchange indices, foreign exchange rates, and US treasury rates with different maturity. They provide evidence by comparing empirical and simulated data that GHE is robust and powerful tool in detecting various types of multiscaling. Nevertheless, they are faced with puzzling phenomenon, when the shuffled series express higher degree of multifractality than the original series. They presuppose that the puzzle is cause by short memory time-correlations in the data. Overall, they reason that the source of multifractality in financial time series comes mainly from the fat-tailed distribution of returns and time-correlations have the effect to reduce the measured multifractality.

Gu *et al.* (2010) provide empirical evidence of multifractality in the daily returns of WTI and Brent crude oil markets. The authors estimate generalised Hurst exponents with R/S and MF-DFA for data from 1987 to 2008. Apart from using the whole sample, they chose to split the dataset into three periods in order to offset the influence of two Gulf Wars to oil prices. The results indicate that the two crude oil markets become more and more efficient for long-term period, but have no such trend for short-period. The results also suggest that the multifractal structure of WTI and Brent markets are not only mainly attributed to the broad fat-tail distributions and persistence but also affected by some other factors, e.g. deregulating crude oil markets. The findings are in line with conclusions of Tabak & Cajueiro (2007), who employed R/S analysis on both Brent and WTI crude oil prices from 1983 to 2004 to analyse the the impacts of the deregulation of the crude oil market that took place in the 1980s and demonstrate that the crude oil markets became more efficient over time. Indeed, this is apparent especially in comparison of period of regulated prices in 1980s with period after deregulation in 1990s. Such findings are consistent with the previous results of Serletis & Andreadis (2004) who concluded that crude oil prices possess long-range dependence. Although, the degree of long-range dependence has decreased over time in terms of both mean and volatility returns.

He & Chen (2010) investigate multifractal features of the most important oil pricing benchmarks globally WTI and Brent mixtures of crude oil on a data from 1987 to 2009. Using MF-DFA and multifractal singular spectrum analysis, the authors find out that both markets exhibit multifractal properties influenced mainly by a nonlinear temporal correlations instead of a non-Gaussian distribution.

Wang *et al.* (2011) explore autocorrelations and cross correlations of WTI crude oil spot and futures return series from 1990 to 2010 using MF-DFA and detrended cross-correlation analysis. They conclude that both autocorrelated and cross-correlated behaviours are persistent for time scales smaller than a month, while for larger scales they are neither autocorrelated nor cross-correlated. Therefore efficient behaviour is indicated in the long term. Furthermore, the degrees of short-term cross-correlations are higher than those of autocorrelations, signalising that prediction of the oil spot or futures prices through analysing the history of both series together is easier and more accurate than analysing each series separately. By using MF-DFA, they find that the short-term correlations are strongly multifractal while the long-term correlations are nearly monofractal. Empirical evidence shows that the complexity of crude oil markets in the short term was much higher than that in the long term. Shuffling the original series proved that both long-range correlations and fat-tail distributions make important contributions to the multifractality.

Matia *et al.* (2003) research properties of 29 commodities and 2449 stocks over a period of approximately 15 years with MF-DFA to find out that price fluctuations for commodities have a significantly broader multifractal spectrum than for stocks. Moreover, they suggest that the multifractality is mainly caused by the broad distribution function.

Zunino *et al.* (2008) present evidence that multifractality degree is associated with the stage of market development by analysing 32 equity index returns for different countries with MF-DFA. The authors develop a model to test the relationship between the two and find robust evidence that higher multifractality degree corresponds with less developed market, implying that an inefficiency ranking can be derived from multifractal analysis techniques.

Yuan *et al.* (2009) provide analogous results for Shanghai stock price index daily returns from 1990 to 2008. Moreover, the authors divide the sample into three periods according to two financial reforms that were introduced during the main period to asses impact of the returns to the financial market risk.

Engelen *et al.* (2011) examine spot rate dynamics in the liquid petroleum gas shipping market over period from 1992 to 2009 with MF-DFA and R/S techniques. The authors study the effect of varying data-frequencies from daily, and weekly to monthly, and time scales between short and long term. The results indicate that the selected data-frequency affects the presence of multifractality. Weekly returns are multifractal only because of their fat tails while monthly returns are monofractal. On the other hand, daily returns are multifractal because of both fat tails and temporal autocorrelations, making it the richest and most valuable dataset for further modelling of the market dynamics. Moreover, MF-DFA showed that the daily returns are multifractal, persistent and include long-range dependence, which undermines market efficiency, therefore it can be modelled.

Chapter 3

Methodology

This chapter presents basic definitions and equations required for understanding of the concepts of long memory and multifractality in time series, based on Samorodnitsky (2007), Calvet & Fisher (2008), Di Matteo (2007) and Kantelhardt (2008).

A British hydrologist Harold Edwin Hurst introduced the R/S analysis in 1951 in his study of a water levels and the flow of water in the river Nile. Hurst was interested in dam design, he examined more than 600 years long (622–1281) data set of the water levels in the Nile river using a particular statistic approach R/S to solve the riddle of the Nile's great floods and to predict how much the Nile flooded from year to year. The R/S statistic is one of the most popular scaling method to estimate power-law correlation exponents from time series, however, is highly influenced by outliers and returns a biased estimation of the Hurst exponent. Peng *et al.* (1994) introduced the Detrended Fluctuation Analysis (DFA) during their studies of the correlation of molecular chains in deoxyribonucleic acid (DNA). This method has become widely used technique for determination of particularly monofractal scaling properties, because it is capable to avoid the spurious detection of apparent long-range correlations. Kantelhardt *et al.* (2002) introduced the Multifractal Detrended Fluctuation Analysis (MF-DFA) as a generalization of DFA. MF-DFA can be used for a global detection of multifractal behaviour, surprisingly the algorithm is not much difficult than the former DFA.

3.1 Long memory

The concept of long memory has useful implications for economists as it describes a time series with slow hyperbolic decay in autocorrelations, that is existence of some long-term relationship in values (e.g. prices, returns). In such case it should be possible to use previous values (i.e. in finance typically prices) to predict future values, which however, according to the Effective market hypothesis (Fama 1970) voids one of the basic features of efficient markets.

Long memory (also known as *long-range dependence*) can be defined in many different approaches – Guegan (2005) mentioned 11 different definitions. Therefore, it is no surprise that definitions of long memory vary from author to author. According to Samorodnitsky (2007), the most popular approach to long memory is through a slow (hyperbolic) decay of autocorrelations.

Definition 3.1 (Long memory). A stochastic process $X_t = (X_1, X_2, \ldots)$ with a finite variance $\mathrm{E}X_1^2 = \sigma^2 \in (0, \infty)$, covariances $R_n = \mathrm{Cov}\,(X_1, X_{n+1})$ and correlations $\rho_n = R_n/\sigma^2, n = 0, 1, \ldots$ has long memory if the so-called long-range correlations ρ_n declines as a power-law

$$\rho_n \propto n^{-\gamma} \tag{3.1}$$

for $n \to \infty$, where $0 < \gamma < 1$.

That is the autocorrelations are asymptotically proportional to the sample size n scaled to the power of $-\gamma$. This means that the autocorrelations are slowly decaying in a hyperbolic manner.

Equivalently, Definition 3.1 can be easily reorganised into the form used in Beran (1994).

Definition 3.2 (Long memory). A stochastic stationary process X_t is called a stationary process with long memory if there exists a real number $\gamma \in (0, 1)$ and a constant $c_\rho > 0$ such that

$$\lim_{n\to\infty} \frac{\rho_n}{c_\rho n^{-\gamma}} = 1. \tag{3.2}$$

Again this states that the autocorrelations ρ_n are asymptotically proportional to the scaled sample size n. Since $\gamma \in (0, 1)$ it also implies that autocorrelations are slowly decaying.

Some properties of long memory processes can be further demonstrated from either Definition 3.1 or Definition 3.2. Start with the stochastic process X_t with finite variance from Definition 3.1 and compute the variance of partial sum $S_n = X_1 + \ldots + X_n$.

$$\mathrm{Var} S_n = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \tag{3.3}$$

$$= \sigma^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \rho \left| i - j \right| = \sigma^2 \left(n + 2 \sum_{i=1}^{n-1} (n-i) \rho_i \right)$$

Here, the last sum closely relates to the pace of decay in the autocorrelations of the process X_t.

Often the pace of increase of the variance of partial sums itself is used to distinguish between a short memory process and a long memory process. Process has a short memory if

$$\lim_{n \to \infty} \frac{\mathrm{Var} S_n}{n} < \infty, \tag{3.4}$$

whereas a process with long memory can be identified by the diverging limit in (3.4), therefore

$$\lim_{n \to \infty} \frac{\mathrm{Var} S_n}{n} = \infty. \tag{3.5}$$

Another property of a process with long memory is that the autocorrelations of such process are not summable, which is a result of the slowly (hyperbolically) decaying autocorrelation function (e.g. Samorodnitsky 2007).

$$\sum_{n=0}^{\infty} |\rho_n| = \infty \tag{3.6}$$

It can be further shown that variance of partial sums of a process with long memory are proportional to power law scaled number of observations (e.g. Samorodnitsky 2007).

$$\mathrm{Var} S_n \propto n^{2-\gamma} \tag{3.7}$$

for $n \to \infty$ and $\mathrm{Var} S_n \to \infty$, where $0 < \gamma < 1$.

These properties are further used for construction of estimators of long-term memory parameters such as Hurst exponent.

3.2 Hurst exponent

Concepts of long memory and multifractality are interconnected through a constant called *the Hurst exponent*, which value determines whether or not the increments of a self-similar process with stationary increments has long memory (Samorodnitsky 2007).

The Hurst exponent arises from the rescaled range statistics developed by Hurst (1951). The construction of the R/S starts with a division of the time series (i.e. in finance typically returns) of length T into N adjacent sub-periods with length ν, so that $N \times \nu = T$. Then the rescaled range of cumulative deviations from the mean is calculated as R_i/S_i, where R_i is a range, i.e. the difference between maximum and minimum values, of the corresponding cumulative deviations from the mean, and S_i is a standard deviation of the corresponding returns. The procedure is repeated over each sub-period i of selected length ν. Eventually an average rescaled range $(R/S)_\nu$ is computed. Hurst (1951) noticed a peculiar scaling behaviour of such rescaled ranges as they all scale as

$$(R/S)_\nu \approx c\nu^H \tag{3.8}$$

with varying ν where c is a finite constant independent of ν.

The scaling law can be easily uncovered by performing an OLS regression on logarithms applied on each side of the equation. The H in (3.8) is known as the Hurst exponent and it relates to the concept of long memory through the exponent γ from (3.7) and it holds that

$$H = 1 - \frac{\gamma}{2}. \tag{3.9}$$

Therefore, the Hurst exponent describes the dynamics of a time series. Three scenarios can be distinguished. *(i)* The Hurst exponent $H = 0.5$ represents a self-determining process which current value is not correlated with its past values. *(ii)* Hurst exponent $0 < H < 0.5$ represents anti-persistent process, i.e. the series express mean-reverting behaviour and an increase in values is most likely to be followed by a decrease and vice versa. The closer the Hurst exponent is to 0, the stronger tendency to revert back to its mean value the series exhibit. And finally, *(iii)* Hurst exponent $0.5 < H < 1$ represents a persistent process, i.e. an increase in values is most likely to be followed by another increase or the other way around. The closer the Hurst exponent to 1, the stronger is the trend in a given time series. Thus in this case, the values rise and fall in a broader range than by pure random walk (Mitra 2012). Combining (3.9) and (3.7) reveals that the third scenario, Hurst exponent $0.5 < H < 1$, also indicate a long memory process.

In case the whole dynamics of a given process cannot be described by a single Hurst exponent, the multifractal (multiscaling) measures need to be applied.

3.3 Multifractality

Multifractality in its essence describes non-linear relationships at some parts of probability distribution function of a time series. In our case this translates particularly into different autocorrelation structure for certain parts of the corresponding distribution. Put differently, autocorrelations of small changes are not the same as autocorrelations of large changes for any multifractal time series. In other words, multifractality is a property of a time series that as a result of non-linear relationships require more than one Hurst exponent to describe its behaviour. It is a generalisation of (mono)fractal time series which evolution is fully described by a single Hurst exponent. In general, the underlying principle of fractal theory is that a simple process when repeated over infinitely many iterations becomes a very complex process. For example in finance, multifractality may be created as a result of different investment horizons. One can imagine a market with a large set of different agents, each with his own individual investment horizon. Each agent perceive information differently and reacts according to his own individual investment horizon and other preferences. Combining such iterations create multifractality in the market. Therefore the notion behind fractals is to uncover and describe the simple underlying process that is inside the complex process we normally observe in the markets.

The idea of multifractality has been formally described by Benoît Mandelbrot, a French mathematician, who recommended to adopt fat-tailed "fractal" processes as an alternative modelling approach built on the concept of scale invariance – meaning that the shape of the distribution of returns should be invariant with the change of time scale. Mandelbrot formalised this concept in his publication in 1964, where he introduced the self-similar process.

Definition 3.3 (Self-similar process). A random process $\{X(t)\}$ that satisfies

$$\{X(ct_1), \ldots, X(ct_k)\} \stackrel{d}{=} \left\{c^H X(t_1), \ldots, c^H X(t_k)\right\} \tag{3.10}$$

for some $H > 0$ and all $c, k, t_1, \ldots, t_k$ is called self-similar or self-affine. The number H is the self-similarity index, also known as scaling exponent or Hurst exponent, of the process $\{X(t)\}$.

This signify that self-similar process is invariant in distribution under suitable scaling of time. Which in plain English means that self-similar process behaves identically in a statistical sense when observed from close or far away, that is under different scales. A typical example of self-similar process in nature is a fern, which self-similar structure is repeated from the whole blade (leaf), through individual leaflets to their individual subleaflets and so forth.

Multifractals originated from natural sciences, where multifractal measures have proven to be useful in many applications in numerous fields including astrology, biology and medicine, network traffic modelling, seismology, and many others. As already mentioned, fractals in their essence are sets that can be constructed by interating a simple transformation. Multifractal measures are similarly to fractal sets built by iterating a simple transformation. For example, Calvet & Fisher (2008) states the binomial measure on $[0, 1]$, derived as the limit of a multiplicative cascade. Consider two positive numbers m_0 and m_1, that $m_0 + m_1 = 1$, and the uniform probability measure μ_0 on the interval $[0, 1]$. In the first step of the cascade, a measure μ_1 is defined by uniformly spreading the mass m_0 on the left subinterval $[0, \frac{1}{2}]$ and mass m_1 on the right subinterval $[\frac{1}{2}, 1]$. In the second step, each subinterval from previous step is split in half to create sub-subintervals. And again the mass m_0 of the subinterval $\mu_1[0, \frac{1}{2}]$ is allocated to the left sub-subinterval $[0, \frac{1}{4}]$ and mass m_1 of the subinterval $\mu_1[0, \frac{1}{2}]$ on the right sub-subinterval $[\frac{1}{4}, \frac{1}{2}]$. Analogically, the mass m_0 of the subinterval $\mu_1[\frac{1}{2}, 1]$ is allocated to the left sub-subinterval $[\frac{1}{2}, \frac{3}{4}]$ and mass m_1 on the right sub-subinterval $[\frac{3}{4}, 1]$. This way we obtain measure μ_2, that is described as follows.

$$\mu_2[0, \frac{1}{4}] = m_0 m_0, \qquad \mu_2[\frac{1}{4}, \frac{1}{2}] = m_0 m_1, \tag{3.11}$$

$$\mu_2[\frac{1}{2}, \frac{1}{4}] = m_1 m_0, \qquad \mu_2[\frac{3}{4}, 1] = m_1 m_1. \tag{3.12}$$

Infinite sequence of measures μ_k is then generated by iterations of the procedure above. Generalising this construction in order to allow intervals to be uniformly split into arbitrary number (b$\geq$ 2) of cells and randomising the allocation of mass into subintervals leads us to multifractal measures.

Definition 3.4 (Multifractal measure). A random measure μ defined on $[0,1]$ is called multifractal if it satisfies for all $q \in Q$:

$$\mathrm{E}\left([t, t+\Delta t]^q\right) \sim c(q)(\Delta t)^{\tau(q)+1} \text{ as } \Delta t \to 0, \tag{3.13}$$

where Q is an interval containing $[0,1]$, and $\tau(q)$ and $c(q)$ are deterministic functions defined on Q.

Rich local properties characterising multifractal measures are described by the Local Hölder exponent.

Definition 3.5 (Local Hölder exponent). Let g be a function defined on the neighbourhood of a given date t. The number

$$\alpha(t) = Sup\left\{\beta \geq 0 : |g(t+\Delta t) - g(t)| = O\left(|\Delta t|^\beta\right) \text{ as } \Delta t \to 0\right\} \tag{3.14}$$

is called the local Hölder exponent or local scale of g at t.

In other words, the Hölder exponent describes the local changes (or variability) of the function at a point of time. The infinitesimal variations of the function can be expressed as being of order $|dg| \approx (dt)^{\alpha(t)}$ around instant t. Here, more abrupt variations coincide with lower values of $\alpha(t)$. For function g that is bounded around t, the Hölder exponent $\alpha(t)$ is non-negative.

Each typical continuous process used in finance (i.e. monofractal series) have a unique Hölder exponent. This contrasts with multifractal measures, which contain continuum of local exponents. For illustration, $\alpha(t) = 0$ at points of discontinuity, $\alpha(t) = 1$ at non-singular differentiable points, and $\alpha(t) = 1/2$ for a Brownian motion. Fractional Brownian motion is characterised by a single unique exponent $\alpha(t) = H$.

Mandelbrot (1974; 1989) suggested multifractal spectrum as a convenient representation for the distribution of Hölder exponents in multifractals. To estimate the distribution of the local Hölder exponent $\alpha(t)$ at a random point, one divides the unit interval $[0,1]$ into b^k subintervals $[t_i, t_i + \Delta t]$ of length $\Delta t = b^{-k}$, and for each subinterval computes the *coarse Hölder exponent*

$$\alpha_k(t_i) \equiv \ln |g(t_i, \Delta t)| \,/ \ln \Delta t.$$

The number of coarse Hölder exponents contained between α and $\alpha + \Delta\alpha$ is denoted by $N_k(\alpha)$.

Definition 3.6 (Multifractal spectrum). The limit

$$f(\alpha) \equiv \lim \left(\frac{\ln N_k(\alpha)}{\ln b^k} \right) \text{ as } k \to \infty \tag{3.15}$$

represents a renormalised probability distribution of local Hölder exponents, and is called the multifractal spectrum.

The advantage of multifractal spectrum in comparison with a plain histogram is that the spectrum makes it easier to detect the events that occur many times during the construction but at vanishing frequency. The quantity $f(\alpha)$ corresponds with the fractal dimension of the set of instants with local Hölder exponent α.

The relation between multifractal spectrum and scaling function, which is the link towards the explicit formulae for the multifractal spectrum $f(\alpha)$ based on the scaling function $\tau(q)$, is derived as follows (see e.g. Calvet & Fisher 2008).

Proposition 3.1 (Multifractal spectrum and scaling function). *The multifractal spectrum $f(\alpha)$ is the Legendre transformation*

$$f(\alpha) = \inf_q \left[\alpha q - \tau(q)\right] \tag{3.16}$$

Direct application of this proposal results into the explicit formulae for the multifractal spectrum. For example, consider a multiplier M with lognormal distribution $-\ log_b M \sim N(\lambda, \sigma^2)$, then Proposition 3.1 implies the multifractal spectrum as a quadratic function

$$f(\alpha) = 1 - (\alpha q - \lambda)^2 / [4(\lambda - 1)], \tag{3.17}$$

parametrised by the unique real number $\lambda > 1$.

Corresponding with multifractal measures, we can define multifractal process by its moment-scaling properties.

Definition 3.7 (Multifractal process). A stochastic process $\{X(t)\}$ is called multifractal if it has stationary increments and satisfies the moment scaling rule

$$\mathrm{E}\left(|X(t+\Delta t) - X(t)|^q\right) \sim c_X(q)(\Delta t)^{\tau_X(q)+1} \tag{3.18}$$

as Δt converges to zero.

The function $\tau_X(q)$ is known as the scaling function. By setting the parameter $q = 0$ in the equation (3.18) it can be demonstrated that all scaling functions have the same intercept $\tau_X(0) = -1$. It has been verified that the scaling function $\tau_X(q)$ is weakly concave (e.g. Calvet & Fisher 2008).

A self-similar process has a linear scaling function $\tau_X(q)$. Direct implication of the invariance condition $X(t) \stackrel{d}{=} t^H X(1)$ is $\mathrm{E}\left(|X(t)|^q\right) = t^{Hq}\mathrm{E}\left(|X(1)|^q\right)$, and therefore the scaling rule (3.18) holds with

$$\tau_X(q) = Hq - 1. \tag{3.19}$$

Because the intercept $\tau_X(0) = -1$ is fixed, a linear scaling function is fully determined by its slope H. That is the reason why self-similar processes are often called *uniscaling*, *unifractal*, or *monofractal*. Self-similar processes do not capture the changes in distribution of returns at different time horizons, so typical for most financial data.

Generally, we can distinguish between two types of processes: *(i)* a process with $H(q) = H$, constant independent on q; and *(ii)* a process with $H(q)$ not constant (i.e. dependent on q).[1] A process of the first type is called monofractal (or equivalently uni-fractal, or uni-scaling) because its scaling behaviour is determined from a single constant H which is the original Hurst exponent H (also known as self-affine index). This exponent H is equal to 0.5 for a Brownian motion, and is a constant different from 0.5 for fractional Brownian motion. A process of the second type is called multifractal (or equivalently multi-scaling) (Feder & Bak 1989) because multiple different exponents defines the scaling behaviour of different q^{th}-moments of the distribution.

3.4 Sources of multifractality

There is a wide consensus in the literature about the sources of the multifractality. One cause is the autocorrelation structure (long memory and different long-range correlations for small and large fluctuations) and the other cause is the distribution of data (fat-tailed distribution and extreme events) (e.g. Matia *et al.* 2003; Jia *et al.* 2012).

Generally, in order to identify the contributions of the two sources to the multifractality of a time series, two other time series can be introduced (Jia

[1] Note that we use H without parentheses as the classical Hurst exponent for self-similar process and $H(q)$ as the generalised Hurst exponent

et al. 2012) – *(i) a reshuffled series*, which perfectly destroys all long-range correlations and at the same time keeps the fluctuation distributions obtained by shuffling the original series, and *(ii) a surrogate series* obtained from the Fourier transform, which eliminate the fat-tailed distribution and preserves the long memory of the original series.

In this thesis, we implement the procedures as follows. *(i)* During shuffling, each observation in a vector of data is assigned an index from 1 to N, where N is the length of the series. Then a new vector of indices is created as a random permutation of the original vector of indices. Finally, the shuffled series is created by reorganising the original observations according to the new vector of indices. Such shuffled series has the same distribution as the original one, however the autocorrelation structure is destroyed. *(ii)* Apart from the shuffling method, we create a surrogate series as an additional evidence of the true source of multifractal behaviour of our data. The procedure creates discrete Fourier transformation of the original series (signal), randomise (shuffles) the positions of the frequencies included in the signal. Finally, the surrogate series is created by inverse discrete Fourier transformation of the shuffled frequencies. Such surrogate series has the same autocorrelation structure as the original one, but the distribution of observations is approximately Gaussian.

Therefore, for series with a multifractal behaviour caused purely by different long-range correlations for small and large fluctuations, the shuffled series will indicate simple random behaviour $H_{shuff}(q) = 0.5$ (i.e. non-multifractal scaling), whereas for the surrogate series the original $H(q)$ dependence will not change, $H_{surr}(q) = H(q)$. On the other hand for series with a multifractal behaviour caused purely by fat-tailed distribution, shuffled series will report the original $H(q)$ dependence unchanged, $H_{shuff}(q) = H(q)$, whereas the surrogate series will indicate simple random behaviour $H_{surr}(q) = 0.5$ (i.e. non-multifractal scaling). If both sources of multifractality contributes to the overall multifractal behaviour of the series, then the shuffled and surrogate series will exhibit weaker multifractality, i.e. multifractality degree of the original series $\Delta H = \Delta H_{max} - \Delta H_{min}$ will be larger than the ΔH_{shuff} and ΔH_{surr}.

In order to estimate the scaling properties of crude oil, petrol, and diesel time series, we utilise two independent methods: Generalised Hurst exponent, and multifractal detrended fluctuation analysis as described in the next sections.

3.5 Generalised Hurst exponent

We employ the generalised Hurst exponent (GHE) method based on Di Matteo *et al.* (2003). In a nutshell, GHE is a tool to directly examine the statistical properties of data by analysing the q-order moments of the distribution of the increments (Mandelbrot 1997) and (Barabási & Vicsek 1991). During the initialisation process, all variables are detrended by eliminating linear drift (if present) as suggested by Di Matteo *et al.* (2005). This serves as a precaution against the known fact that financial time series exhibit variations in their statistical properties with time. Otherwise we could detect false variations and dependencies on time window T.

The statistical evolution of stochastic variable $X(t)$ is defined as

$$K_q(\tau) = \frac{\langle |X(t+\tau) - X(t)|^q \rangle}{\langle |X(t)|^q \rangle}, \tag{3.20}$$

where the time interval τ can evolve between ν and τ_{max}.

The generalised Hurst exponent $H(q)$is defined from the scaling behaviour of $K_q(\tau)$ if it follows the relation (Barabási & Vicsek 1991)

$$K_q(\tau) = \left(\frac{\tau}{\nu}\right)^{qH(q)}. \tag{3.21}$$

It is worth mentioning, that some values of q are associated with special features. E.g. for $q = 2$ the $K_q(\tau)$ is proportional to the autocorrelation function and $H(2)$ therefore describes the scaling behaviour of the autocorrelation function. Implying that values of $H(2) \in (0.5, 1)$ indicate presence of long-range dependence, the stronger the closer the exponent is to 1. $H(2) = 0.5$ indicates self-determining random walk process, thus no presence of long-range dependence. And finally, values of $H(2) \in (0, 0.5)$ indicate negative autocorrelations, therefore violent fluctuations of the variable. Additionally, for $q = 1$, $H(1)$ describes the scaling of the absolute values of the increments, which is in fact closely related to the classical Hurst exponent H, that is indeed associated with the scaling of the absolute spread in the increments.

One can plot values of $qH(q)$ against q to visually judge multifractality. As already mentioned in Section 3.3, a monofractal series is described by a single Hurst exponent H, $H(q) = H$ for all q. Therefore $qH(q)$ of monofractal process will be a straight line. Whereas $H(q)$ of a monofractal series is dependent on q, therefore the corresponding $qH(q)$ will translate into a non-linear curve. The more bent the curve is, the stronger the multifractal behaviour of the given variable is.

3.6 Multifractal detrended fluctuation analysis

We apply the multifractal detrended fluctuation analysis (MF-DFA) method as described in Kantelhardt *et al.* (2002). In essence, the method focuses on fluctuations around trend in the data. The procedure itself can be described in five steps, where the first three steps are essentially the same as in the conventional DFA procedure (see e.g. Peng *et al.* 1994).

Assume a series X_t of length N and of compact support. The support is defined as the set of the indices t with non zero values X_t, and it is compact if $X_t = 0$ for an insignificant fraction of the series only. The value of $X_t = 0$ is interpreted as having no value at this n.

- *Step 1:* Calculate the cumulative deviations from mean ("profile")

$$Y(i) \equiv \sum_{t=1}^{i} \left[X_t - \langle X \rangle \right], \, i = 1, \ldots, N. \tag{3.22}$$

- *Step 2:* Split the profile $Y(i)$ into $N_s \equiv (N/s)$ non-overlapping segments of equal length s. Usually a short block of the series is left at the end, simply because the length of the series N is scarcely a multiple of the time scale s. Not to ignore the information from this leftover part, we repeat this step from the end of the series to the beginning. Therefore, we obtain $2N_s$ segments in total.

- *Step 3:* Calculate the polynomial fit (local trend) for each of the $2N_s$ segments, and the variance

$$F^2(\nu, s) \equiv \frac{1}{s} \sum_{i=1}^{s} \left\{ Y\left[(\nu - 1)\, s + i \right] - y_\nu\,(i) \right\}^2 \tag{3.23}$$

for each segment ν, $\nu = 1, \ldots, N_s$ and

$$F^2(v, s) \equiv \frac{1}{s} \sum_{i=1}^{s} \left\{ Y\left[N - (\nu - N_s)\, s + i \right] - y_\nu\,(i) \right\}^2 \tag{3.24}$$

for $\nu = N_s+1, \ldots, 2N_s$. Where, $y_\nu(i)$ is the fitting polynomial in segment ν. Polynomial fit can be of arbitrary order, usually linear, or quadratic is used. In this thesis we employ linear detrending.

- *Step 4:* Compute the q^{th} order fluctuation function by averaging over all segments

$$F_q(s) \equiv \left\{ \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} \left[F^2(\nu, s) \right]^{q/2} \right\}^{1/q}, \tag{3.25}$$

  where q is in general any real number. However $q = 0$ requires special treatment as $\frac{1}{0} \to \infty$, see the next step. Steps 2 to 4 are usually repeated several times to find out the relation between fluctuation function $F_q(s)$ and time scale s for different weights q.

- *Step 5:* Analyse log–log plots $F_q(s)$ versus s for each q to visualise the scaling behaviour of the fluctuation functions. For long-range power-law correlated series X_t, the fluctuation function $F_q(s)$ increases for large values of s as a power-law

$$F_q(s) \approx s^{H(q)}. \tag{3.26}$$

For stationary series, the $H(2)$ is equal to the Hurst exponent, therefore, the function $H(q)$ is called the generalised Hurst exponent.

As already mentioned in previous step, we cannot determine the value of $H(0)$ directly because of the diverging exponent in (3.25) for $q = 0$. Thus, the 'zero-order' fluctuation function is defined as

$$F_0(s) \equiv \exp\left\{ \frac{1}{4N_s} \sum_{\nu=1}^{2N_s} \ln \left[F^2(\nu, s) \right] \right\} \sim s^{h(0)}. \tag{3.27}$$

In case of monofractal time series, the scaling behaviour of the variances $F^2(\nu, s)$ is constant across all segments ν and the averaging procedure in step 4 will return the same scaling behaviour, thus $H(q)$ is independent of q. For multifractal time series, the scaling behaviour of the variances $F^2(\nu, s)$ changes across segments ν as a result of different scaling of small and large fluctuations (i.e. deviations from the polynomial fit within a given segment), which affects the average $F_q(s)$, therefore the generalised Hurst exponent $H(q)$ is dependent on q. Positive values of q emphasise, periods with large fluctuations as a direct implication of (3.25). For this reason, typically, in case of multifractal series, large fluctuations have lower $H(q)$ than small fluctuations.

The accuracy of $H(q)$ determined by MF-DFA depends on the length of the series analysed. Kantelhardt (2008) states that for $q \in \langle -10, 10 \rangle$ and length of time series $N = 10\ 000$, the systematic and statical error (i.e. 'standard deviation') can be expected up to $\Delta H(q) = \pm 0.1$.

Additionally, Kantelhardt *et al.* (2002) provides the relation between scaling function $\tau(q)$ and generalised Hurst exponent $H(q)$ as

$$\tau(q) = qH(q) - 1. \tag{3.28}$$

Thus multifractal spectrum $f(\alpha)$ related with $\tau(q)$ via a Legendre transformation is (Peitgen *et al.* 2004)

$$f(\alpha) = \alpha q - \tau(q), \tag{3.29}$$

where $\alpha = \tau'(q)$ is the Hölder exponent and $f(\alpha)$ denotes the dimension of the subset of the series characterised by α.

Combining (3.28) and (3.29) we obtain direct relations amongst $f(\alpha)$, α, and $H(q)$ as

$$\alpha = H(q) + qH'(q) \qquad \text{and} \qquad f(\alpha) = q[\alpha - H(q)] + 1. \tag{3.30}$$

The width of multifractal spectrum $f(\alpha)$ is an indicator of degree of multifractality $\Delta\alpha$

$$\Delta\alpha = \alpha_{max} - \alpha_{min}. \tag{3.31}$$

The wider the spectrum the stronger multifractal properties are present in the time series.

Chapter 4

Data

The data sample used in this thesis is daily price of crude oil and daily retail prices of petrol (Natural 95)[1] and diesel in the Czech Republic. The analysis is performed on data within time window of 1st January 2004 – 28th February 2013. That is 3 020 observations in case of petrol price series, 3 019 observations in diesel price series (value for 9th April 2004 is missing in the original data for unknown reasons), and 2 317 observations in crude oil price series.[2] Different length of time series is not to pose any problem for the methodology, because we are comparing behaviour/statistics of separate time series rather than comparing time series as such against each other.

Daily price of crude oil in the Czech Republic is calculated from the Brent spot price FOB in the European market.[3] We chose to use Brent because Czech National Bank (CNB) uses Brent as the benchmark for Czech price of crude oil. For our analysis and easier comparison with fuel prices, which are traditionally stated in litres, the volumes of crude oil is converted from barrels into litres (1 barrel = 158.987 litres), and from USD into CZK using CNB spot exchange rates for each date of our time series.[4] Because the crude oil price is provided by a foreign entity and banking holidays differ across countries, we use only the dates for which a crude oil price is stated. If such a day is a banking holiday in the Czech Republic, therefore an exchange rate is not updated by the CNB, we use the previous business day exchange rate for conversion into CZK.

[1] regular unleaded petrol 95 RON

[2] petrol and diesel prices are available online at `http://www.finance.cz/makrodata-eu/pohonne-hmoty/`

[3] crude oil price time series are available online at `http://www.eia.gov/dnav/pet/hist/leafhandler.ashx?n=pet&s=rbrte&f=d`

[4] CNB exchange rates are available online at `http://www.cnb.cz/cs/financni_trhy/devizovy_trh/kurzy_devizoveho_trhu/vybrane_form.jsp`

Petrol and diesel prices are nation wide averages collected and calculated by CCS, which is the main entity for issuance of so called fuel cards, accepted at most of the petrol stations in the Czech Republic. Approximately 5 000 petrol stations in the Czech Republic and Slovakia accepts CCS fuel cards as of 31$^{\text{st}}$ December 2012; As of April 2007, the cards were accepted at about 2 300 petrol station in the Czech Republic. The averages are calculated as a simple average of unit price of petrol/diesel at every petrol station where at least one transaction occurs that given day (every petrol station produce just one unit price of petrol/diesel, no weights nor adjustments for the volume sold nor for number of customers).

Retail fuel prices are reported including taxes – in case of the Czech Republic it is a value added tax (VAT) and excise tax. Both fuels are heavily taxed and the degree of taxation changed during the observed time window, therefore raw price series need to be adjusted for taxes to obtain a tax-free prices comparable across time. Taxation changed three times during our time frame (2004 – 2013), VAT was 22% until the end of April 2004, 19% from May 2004 to the end of the year 2009, 20% from January 2010 to December 2012, and the final change so far came in the begining of 2013, when the VAT was raised to 21%. Changes in VAT are summarised in Table 4.1.

Table 4.1: Changes of VAT in the Czech Republic

from	to	VAT
1. 1. 1993	31. 12. 1994	23%
1. 1. 1995	30. 4. 2004	22%
1. 5. 2004	31. 12. 2009	19%
1. 1. 2010	31. 12. 2012	20%
1. 1. 2013	present	21%

Source: Ministry of Finance of the Czech Republic

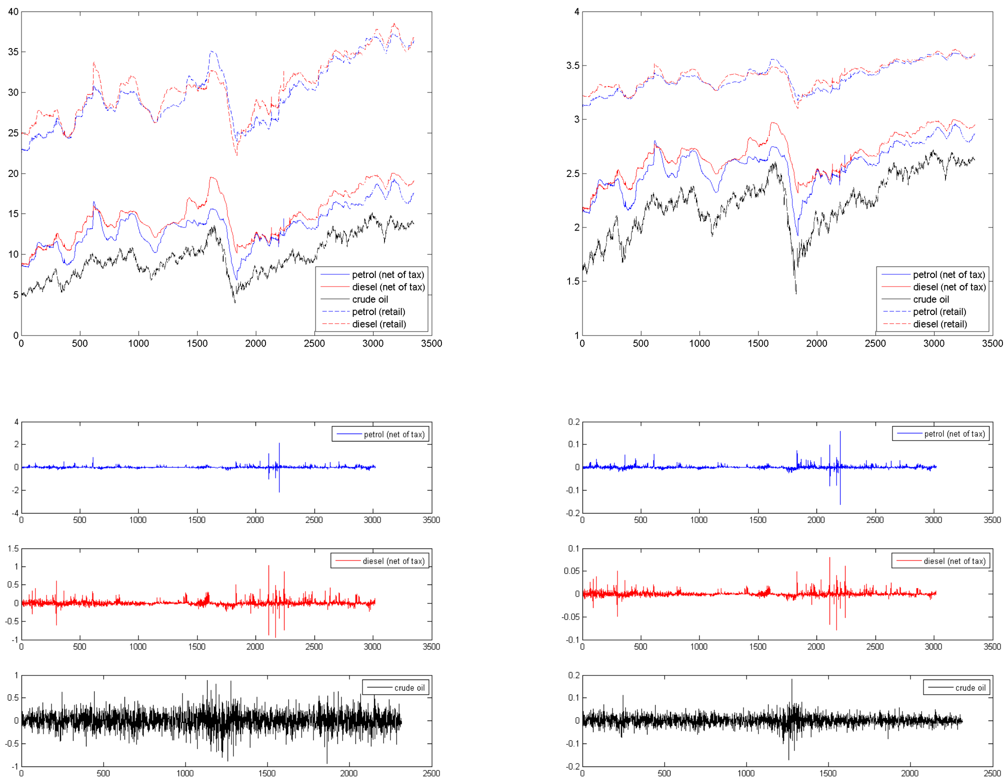
According to the Customs Administration of the Czech Republic (2010), excise taxes were increased on 1$^{\text{st}}$ January 2010 from 11 840 CZK per 1 000 litres to 12 840 CZK per 1 000 litres and from 9 950 CZK per 1 000 litres to 10 950 CZK per 1 000 litres, for petrol and diesel respectively.[5]

According to our calculations, tax component of petrol ranged from 50% to 69.3%; Tax component of diesel ranged from 44.3% to 64.4%. Historical

[5] Excise tax is considered as a part of the VAT tax base, therefore VAT is calculated from the *net of tax price + excise tax*

average calculated from our sample is 56.9%, and 51.4% taxation of petrol, and diesel, respectively. Crude oil price is reported net of any tax. Plot of net of tax prices in comparison to crude oil price in the upper panel of Figure 4.1 reveals similar positive trends in prices for all variables disrupted by a sudden drop during the recent crisis in the second half of 2008. The lower panel shows volatility clusters in petrol and diesel returns occurring particularly in the later stages of the crisis from late 2009 to early 2010.

Figure 4.1: Petrol, diesel and crude oil simple-prices and simple-returns (left), and log-prices and log-returns (right)



4.1 Czech Fuel Market

The fuel market in the Czech Republic is characterised by a dense network of petrol stations. In terms of petrol stations per capita, it has one of the highest penetrations in Central Europe. Demand for fuels in 2004 was 4.92 million litres, equally split into petrol and diesel. Fuel demand reached its maximum in 2006 with total demand of 6.41 million litres, this was caused by increase in demand for diesel to 3.77 million litres. From 2007 the total demand gradually decreased to 5.33 million litres in 2010. Demand for gasoline was rather stable, from 2.30 million litres in 2004 to its maximum of 2.44 million litres in 2006 and then steadily decreased to 1.83 million litres in 2012. Demand for diesel grew from 2.50 million litres in 2005 to 3.77 million litres in 2006 and then slowly decreased to 3.48 million litres in 2012. (MITCR)

According to the Ministry of Industry and Trade of the Czech Republic (MITCR) (Dušek & Purnoch 2013), there are 6 790 petrol stations in total in the Czech Republic (as of 31st December 2012), which can be further divided into three types (see Table 4.2).

(i) Public, where anyone can buy fuel. Typical public petrol station sell multiple kinds of fuels (2 745 out of 3 728).

(ii) "Petrol stations with restricted access and sales" (Restricted), which are typically used only by their respective owners and/or operators. These are usually operated in business areas (e.g. quarries, sawmills, farms, transport organisations, construction yards) and the sale of fuel is carried out according to a specific contract. The vast majority of these petrol stations sell only one kind of fuel (typically only diesel) (411 out of 472).

(iii) Private, which are used only by their owners within the company. These are located exclusively in business areas (e.g. farms, freight and passenger transport companies, quarries, sand pits, garages of municipal services, construction yards, etc.). An overwhelming 97.3% of these stations "sell" only a single kind of fuel, typically diesel (2 519 out of 2 590).

Table 4.2: Types of petrol stations in the Czech Republic as of 31st December 2012

	petrol stations	share
Public	3 728	54.9%
Restricted	472	7.0%
Private	2 590	38.1%
Total	6 790	100%

Source: Ministry of Industry and Trade of the Czech Republic

Dušek & Purnoch (2013) highlights that the legislative changes introduced in the beginning of 2011 were the main driver of the apparent growth in 2011 and 2012 in the category of Restricted petrol stations in Table 4.3, as some

Table 4.3: Number of registered petrol stations 2006–2012

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Public	NA	NA	3 649	3 610	3 578	3 615	3 672	3 717	3 728
Restricted	NA	NA	39	92	208	251	293	397	472
Private	NA	NA	2 670	2 658	2 638	2 633	2 626	2 576	2 590
Total	NA	NA	6 358	6 360	6 424	6 499	6 591	6 690	6 790

Source: Ministry of Industry and Trade of the Czech Republic

private petrol stations had to change their status to Restricted to confirm with the new law.

No reliable comparable data for annual statistics of registered petrol stations for years 2004 and 2005 were available at the time of writing this thesis. Total number of petrol stations in the Czech republic for this years is expected to be around 2200 in 2004, and 3600 in 2005, according to estimations of MITCR (Dušek 2004; 2005). However, because of different legislation in force until 2006, not all petrol stations were subject to the data collection process. Additionally, petrol stations were divided into categories according to size of each operator (in terms of number of petrol stations network) unlike the three categories (i.e. Public, Restricted, Private) as from 2006. Therefore, we assume that the data collected in 2004 and 2005 are rather comparable to the Public petrol stations only of the sum of Public and Restricted petrol stations. This assumption is in line with the actual number of Private and Restricted petrol stations reported in 2006.

The Czech Republic has very limited crude oil resources in its territory, therefore most of the crude oil is imported (98%). Most of the imports come from Russia (61.40%), Azerbaijan (28.62%), Kazakhstan (9.94%), and Algeria (0.04%). In the first two quarters of 2012 a total of 3 337 700 tonnes of crude oil were imported, see Table 4.4 (Dušek 2012).

Judging by the level of market penetration, one would expect the Czech fuel market to be highly competitive, therefore changes in the price of petrol and diesel should not be autocorrelated (Fama 1970). However, it is not the case, and the autocorrelation structure is more complicated and contains long memory, as we are going to present in the next section.

Table 4.4: Crude oil imports in the first two quarters of 2012

country of origin	imports in thousands of tonnes	share
Algeria	1.3	0.04%
Azerbaijan	955.1	28.62%
Kazakhstan	331.9	9.94%
Russia	2 049.4	61.40%
Total	3 337.7	100%

Source: Ministry of Industry and Trade of the Czech Republic

4.2 Descriptive statistics

A simple log-transformation is applied to make the interpretation of the results easier (e.g. change in log-return can be interpreted as a percentage change in price). Additionally, from now on in order to simplify the following text, we will refer to log-price as price, and log-return as return.

First of all, we check the distribution of our data by Jarque-Bera test. The null hypothesis that the data come from normal distribution is strongly rejected for each series. Histograms in Figure 4.3 summarise the distribution of each series. All the price series seem to be of a multi-modal distributions with truncated right end. The crude oil price histogram indicate tri-modal distribution, the petrol and diesel price histograms suggests multi-modal distribution. Multimodality can be attributed to different stages of the local market at a given time, as the prices tend to gently fluctuate around a certain price for a few weeks or even months before any sudden jump that typically comes at the beginning of year or summer season occur. Between mid-2005 and mid-2008, the net of tax price of petrol and diesel fluctuated around 15 CZK/l, then suddenly dropped in the end of 2008 to 10 CZK/l (petrol) and 6 CZK/l (diesel). Ever since the prices evolve around a positive trend with repeated periods of seeming price stability that usually lasts for a few weeks. If we add back tax components and compare the retail prices over the period of 2004–2013 across European countries, we will see that the Czech petrol and diesel retail prices have grown from relatively low to one of the highest in Europe. However, that is partly a result of changes in taxation regimes in the Czech republic, therefore this information is provided only to give the big picture, although it cannot be applied directly to the net of tax prices we use in the analysis. Histograms of returns show unimodal leptokurtic distributions with mean close to zero and fat tails. Especially, the histograms of petrol and diesel returns exhibit very large excess kurtosis and extreme fat tails. Narrowness of the distributions suggests that prices tend to evolve in a rather small steps (i.e. returns are close to zero), however, fat tails indicate that some extreme events occur from time to time.

To find out whether the series are stationary, we employ Augmented Dickey-Fuller (ADF) test, which tests for the presence of unit root in a process, together with Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for trend stationarity. Combining the two tests is necessary to distinguish among series that appear to be stationary, series that appear to have unit root, and series that are not informative enough and one cannot be sure about their stationarity or integration. In our case, ADF test (with constant – to capture non-zero mean) fails to reject the presence of unit root in the crude oil price and diesel price series with p-values 0.18. ADF test on petrol price returns statistics with p-value close to 0.05, therefore we should reject the unit root hypothesis at 5% significance level, on the other hand the null hypothesis cannot be rejected at 1% significance level. As for the returns, ADF test strongly rejects unit root

Table 4.5: Descriptive statistics

	crude oil price	petrol price	diesel price	crude oil return	petrol return	diesel return
observations	2317	3020	3019	2316	3019	3018
mean	2.2242	2.5546	2.6524	4.57e-04	2.56e-04	3.03e-04
st.dev.	0.2743	0.2054	0.1989	0.0218	0.0077	0.0059
mode	1.3827	2.4827	2.3933	-0.1702	0	0
median	2.2346	2.5848	2.6516	0.0011	0	0
min	1.3827	1.9214	2.0345	-0.1702	-0.1629	-0.0790
max	2.7199	2.9598	2.9999	0.1813	0.1575	0.0970
1st q.	2.0542	2.4338	2.5125	-0.0118	-0.0015	-0.0013
3rd q.	2.4409	2.6933	2.8056	0.0126	0.0015	0.0013
skewness	-0.3088	-0.3960	-0.1508	-0.0330	0.9006	2.1563
kurtosis	2.4532	2.7721	2.3647	8.4410	150.4155	69.7746
Jarque-Bera	65.7018	85.4847	62.2136	2857.25	2.73e+06	563038
J-B p-val.	5.41e-15	2.74e-19	3.09e-14	<0.01	<0.01	<0.01
ADF stats	-2.2716	-2.8880	-2.2881	-8.9968	-7.1750	-8.1415
ADF p-val.	0.1813	0.0468	0.1759	4.89e-16	1.11e-10	1.92e-13
KPSS stats	12.2002	10.4815	12.0477	0.0664	0.1295	0.3540

hypothesis in each series. Therefore the return series do not have a unit root. KPSS test rejects the null hypothesis of trend-stationarity for each of the price series at 5% significance level (critical value 0.461), and fails to reject the null hypothesis for the returns at the same 5% significance level. Therefore, according to the KPSS test, returns are stationary around deterministic trend. Combination of the ADF and KPSS test hence implies that the price series are not stationary, whereas the return series are stationary. Therefore, we proceed with analysis with return series only. Descriptive statistics are summarised in Table 4.5.

Plots of the autocorrelation function (ACF) of returns (see Figure 4.2) reveal more interesting patterns. The ACF of crude oil returns is almost clear of any autocorrelation, suggesting effective crude oil market (e.g. Fama 1970). Ljung-Box Q test confirms that within a 5 days long trading week there is no autocorrelation in crude oil returns, the Q-statistics at lag 5 is 1.90 with p-value 0.86. On the other hand, Ljung-Box Q test for petrol and diesel series indicate that returns are autocorrelated, the Q-statistics at lag 5 is 224.03 and 175.87, respectively (p-values <0.01). Thus, judging by the (P)ACF plots and Ljung-Box Q tests, both petrol and diesel returns seems to contain long memory, therefore it makes sense to study their properties by methods presented in the previous chapter.

Figure 4.2: ACF (left) and PACF (right) plots of crude oil, petrol, and diesel returns

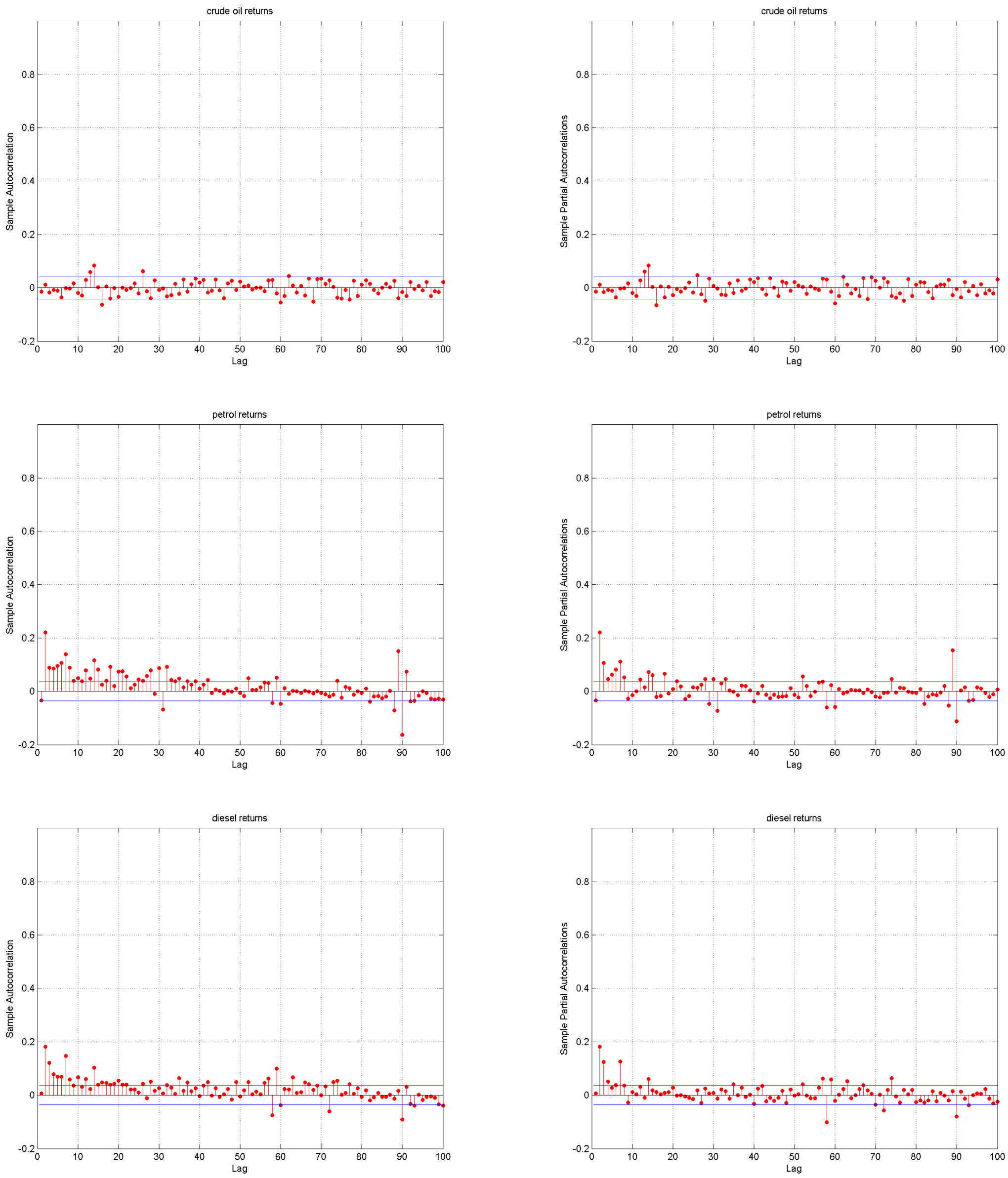
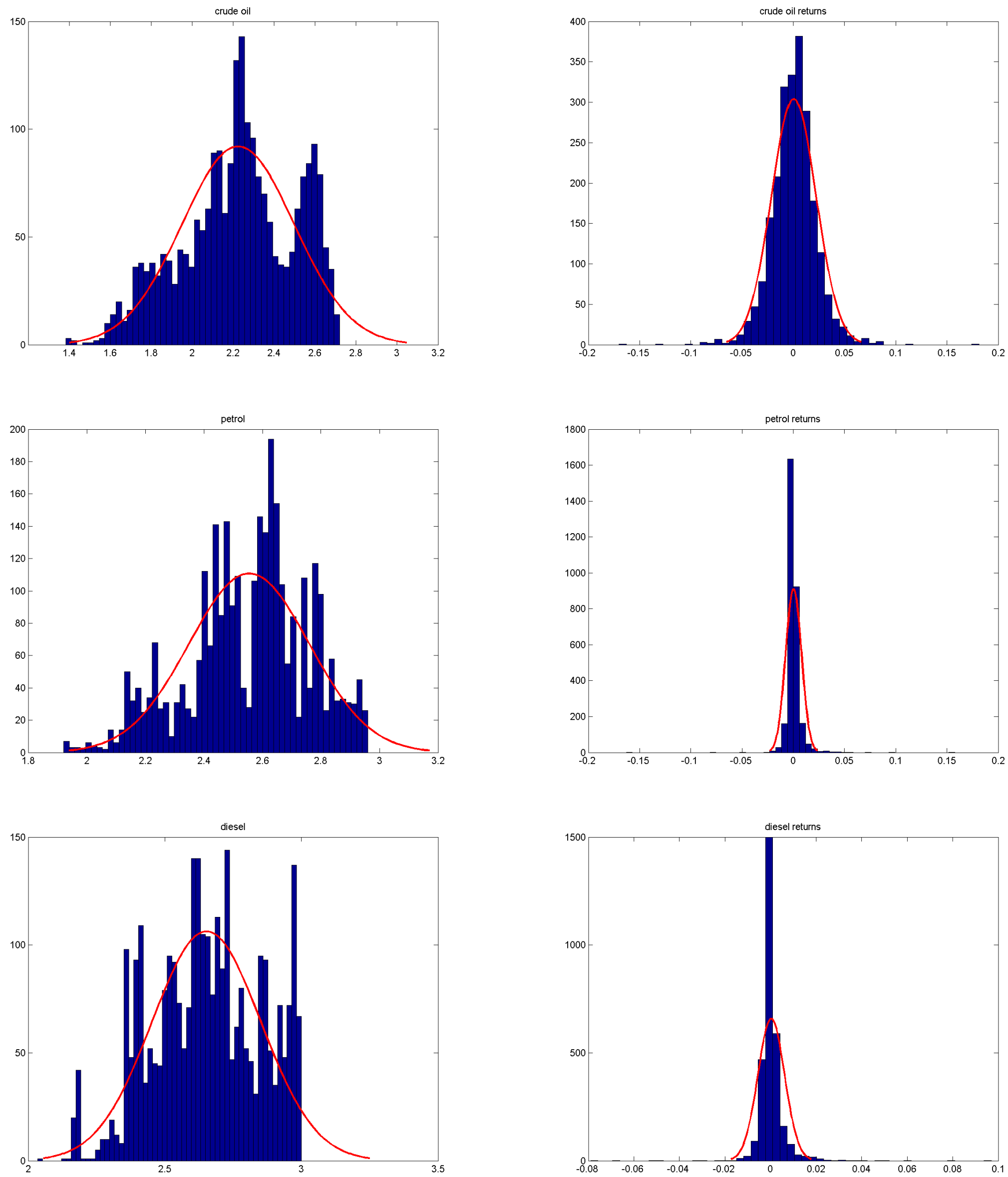


Figure 4.3: Histograms of crude oil, petrol, and diesel prices (left) and their returns (right)



Chapter 5

Empirical Results

In this section, we report the empirical findings of our analysis. We start with generalised Hurst exponent (GHE) method to estimate presence of long memory and multifractality in the fuel market in the Czech Republic. Then, we continue with multifractal analysis of the Czech fuel market by employing multifractal detrended fluctuation analysis (MF-DFA). Description of the data including descriptive statistics and insight into local Czech market is to be found in Chapter 4.

5.1 Generalised Hurst exponent

We start our analysis by employing GHE method to examine the fuel (logarithmic) returns, particularly for presence of long memory in the Czech fuel market. The scaling behaviour of crude oil, petrol and diesel returns is depicted in Figure 5.1. Each curve represents scaling behaviour for different fixed value of q^{th} moments of $K_q(\tau)$ defined in (3.20) ranging from $q = 0.1$ to $q = 4$, where τ describes time scale ranging from 1 to 19 days as suggested by Di Matteo *et al.* (2003). Scaling at integer values of q are highlighted for better clarity. Scaling behaviour for $q = 1$ and $q = 2$ is particularly of our interest as it conveys an intuitive interpretation. Behaviour at $q = 1$ describes scaling of absolute returns, which can be also perceived as the degree of tail heaviness. Scaling behaviour for $q = 2$ can be interpreted as scaling of autocorrelations of returns. As the Figure 5.1 illustrates, scaling behaviour given by (3.21) holds for all our time series. Scaling is almost perfectly linear for values of $q \leq 2$. Higher values around $q = 3$ scales almost linearly, particularly in case of petrol and diesel. Scaling for larger values of q is more complicated (non-linear), especially in case of crude oil returns. However, as we are interested particularly in scaling behaviour for $q = 2$, i.e. scaling of autocorrelations, this non-linear scaling for higher q is not a problem for our research. In fact, the perfect linearity of scaling does not hold for higher values of q for empirical data in general.

Figure 5.2 depicting $qH(q)$ as a function of q provides evidence of multiscaling (multifractal) behaviour of petrol and diesel returns, and monoscaling (monofractal) behaviour of crude oil returns. That is because monoscaling series should display linear relationship between $qH(q)$ and q, as a single Hurst exponent is capable to describe the whole dynamics of the underlying process. Hence, the generalised Hurst exponent $H(q)$ is independent on q in case of monofractal data. On the other hand, multifractal data exhibit non-linear relationship in $qH(q)$ against q plot, which points out multiscaling behaviour. In this case, one Hurst exponent is not enough to describe the whole dynamics of the underlying process, therefore the generalised Hurst exponent $H(q)$ depends on q. Such information can be interpreted in terms of complexity of relationships among autocorrelations. Autocorrelations of petrol and diesel returns differ in certain parts of the corresponding probability distribution function. In other words, autocorrelations of small changes in petrol or diesel returns are not equivalent to autocorrelations of large changes of corresponding fuel return series and they exhibit complex non-linear relationships. Whereas crude oil returns demonstrate linear relationship among autocorrelations, that is autocorrelations for both small and large changes are identical. As an implication of our findings, petrol and diesel returns cannot be successfully modelled by traditional processes used in finance. Instead, a model with built-in multifractal scaling is required for proper modelling of petrol and diesel returns in the Czech fuel market. The specific type of model depends also on presence of long memory, which we examine in the next step.

Presence of long memory in our time series can be inferred from the actual values of generalised Hurst exponents. Values of generalised Hurst exponents $H(q)$ for integer values of q are summarised in Table 5.1. There, particularly exponents for $q = 1$ and $q = 2$ are to be pointed out as they are associated with special features. $H(1)$ describes scaling behaviour of absolute values of increments (i.e. absolute returns), which is closely related to the classical Hurst exponent H for self-similar processes. Thus a series with $H(1) = 0.5$ indicates a self-determining random walk process which current value is not dependent on its past values. A series with $H(1) \in (0, 0.5)$ indicates anti-persistent process, i.e. the series express mean-reverting behaviour and an increase in values is

Figure 5.1: Scaling behaviour of crude oil, petrol, and diesel returns, highlighted lines for $q=1(\diamond), q=2(\square), q=3(\circ), q=4(\triangle)$

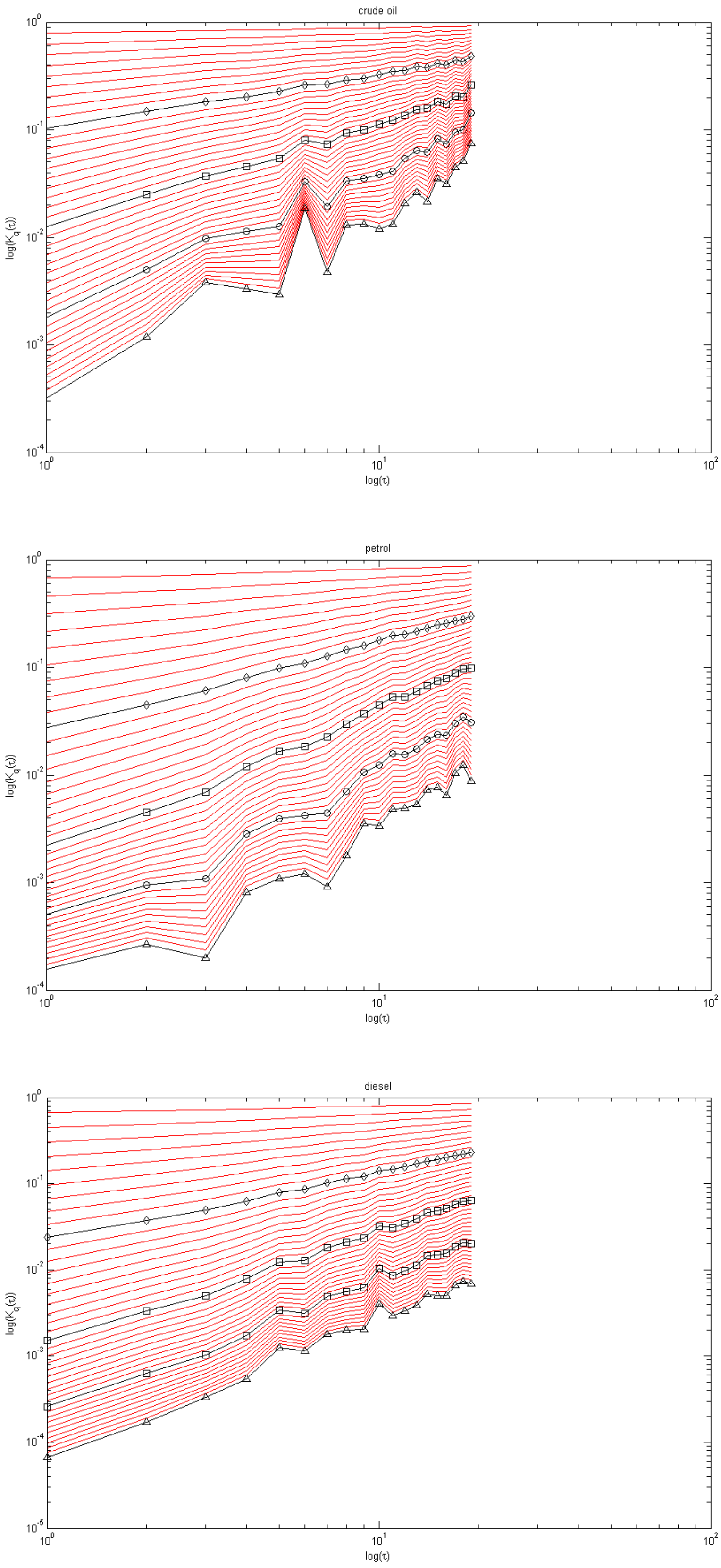
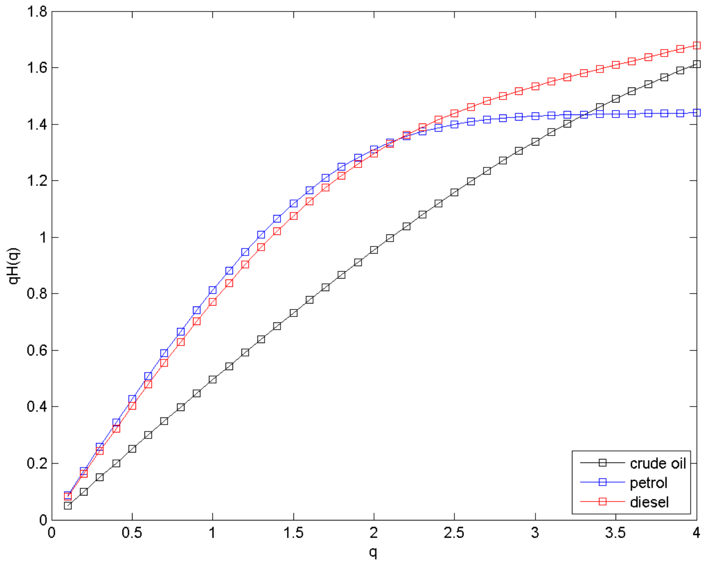


Figure 5.2: $qH(q)$ as a function of q



most likely to be followed by a decrease and vice versa. The closer the Hurst exponent is to 0, the stronger tendency to revert back to its mean value the series exhibit. And finally, a series with $H(1) \in (0.5, 1)$ indicates a persistent process, i.e. an increase in values is most likely to be followed by another increase or the other way around. The closer the Hurst exponent to 1, the stronger is the trend in a given time series. On the other hand $H(2)$ is an indicator of long-range dependence of autocorrelations (i.e. long memory). A series with $H(2) = 0.5$ indicates absence of long-range dependence. A series with $H(1) \in (0.5, 1)$ corresponds to anti-persistence and indicates strong negative autocorrelations that decay after certain lag or decay exponentially. The closer $H(2)$ is to 0, the more violently the process fluctuate. And lastly, a series with $H(2) \in (0.5, 1)$ corresponds to persistence and indicates positive autocorrelations that decay as a power law (i.e. slower than exponentially). The closer $H(2)$ is to 1, the stronger the long-range dependence is. There is no straightforward economic interpretation for $H(q \geq 3)$, however running the analysis for wider range of parameters q allows us to read the scaling behaviour from $qH(q)$ plot.

Our empirical results show that both $H(1)$ and $H(2)$ of crude oil returns are very close to 0.5 (0.50 and 0.48, respectively), demonstrating negligible anti-persistence in autocorrelations and random walk like behaviour. However $H(1)$ and $H(2)$ of petrol and diesel is significantly larger than 0.5. Values of

Table 5.1: Generalised Hurst exponent

	crude oil	petrol	diesel
$H(1)$	0.4960	0.8136	0.7714
std.dev.	(0.0057)	(0.0180)	(0.0217)
$H(2)$	0.4776	0.6554	0.6482
std.dev.	(0.0086)	(0.0249)	(0.0131)
$H(3)$	0.4462	0.4761	0.5115
std.dev.	(0.0127)	(0.0406)	(0.0073)
$H(4)$	0.4029	0.3602	0.4029
std.dev.	(0.0209)	(0.0462)	(0.0209)

$H(1)_{petrol} = 0.81$ and $H(1)_{diesel} = 0.77$ suggest strong persistence of both series and points out clusters of high volatility. Moreover, values of $H(2)_{petrol} = 0.66$ and $H(2)_{diesel} = 0.65$ reveal significant long-range dependence in autocorrelations which is an evidence against market efficiency (Fama 1970).

We can summarise our results from GHE stating that petrol and diesel markets are not efficient, as their returns are autocorrelated and contain long-range dependence in autocorrelations. Additionally, they are subject to multifractal behaviour. Therefore, in order to model petrol and diesel returns, one has to implement a process including both long memory, and multifractality. An example of such model is Markov switching multifractal model (Calvet & Fisher 2008). On the other hand, no indication of long-range dependence is found in monofractal crude oil returns. It is therefore enough to use traditional models in case of crude oil, however, because of random walk like behaviour there is not much to be modelled in its returns. Such finding about crude oil are in line with our expectations and the fact that crude oil is traded in commodity markets. In order to provide further evidence, we perform MF-DFA in the next part of this chapter.

5.2 Multifractal detrended fluctuation analysis

Following GHE method, we perform MF-DFA[1] over the Czech crude oil, petrol and diesel (logarithmic) returns to provide further evidence of long memory and multifractal behaviour in the Czech fuel market.

[1] our MATLAB code for MF-DFA is based on Multifractal Toolbox by Ihlen (2012), available online at `http://www.ntnu.edu/inm/geri/software`

According to the algorithm of MF-DFA, we calculate the fluctuations $F_q(s)$ as defined in (3.25) for scales s ranging from 8 to $T/4$ observations, where T is the length of a given time series, as generally recommended (e.g. Kantelhardt *et al.* 2002). Fluctuations $F_q(s)$ as a function of s for $q \in [-4, 4]$ are illustrated in Figure 5.3, where the lower curves of each series correspond to lower values of q. Fluctuations for $q = 2$ (i.e. second moments) are emphasised in the plot as they represent the scaling behaviour of autocorrelations of the returns. Linear relationship in this scaling behaviour is at least reasonably good for all values of q, although it becomes weaker particularly for very high or very low (very negative) q for some fluctuations typically in the larger range of scales s. Overall the scaling is very reasonable, most importantly, scaling at $q = 2$ is almost linear, therefore we can state that power-law scaling between $F_q(s)$ and s defined in (3.26) holds (although it weakens for q on the edge of our selected range) and we can perform MF-DFA on the whole sample.

Presence of multifractality in our time series can be detected from the plot of generalised Hurst exponent $H(q)$ as a function of q displayed in the left panel of Figure 5.4 . It shows that crude oil return series exhibit monofractal behaviour, while petrol and diesel returns exhibit multifractal behaviour. The reasoning behind this statement is following. In case of monofractal series, a single Hurst exponent is able to fully describe dynamics of a given series, therefore, the

Figure 5.3: Fluctuation function for crude oil (black), petrol (blue) and diesel (red), with highlighted lines for $q = 2$

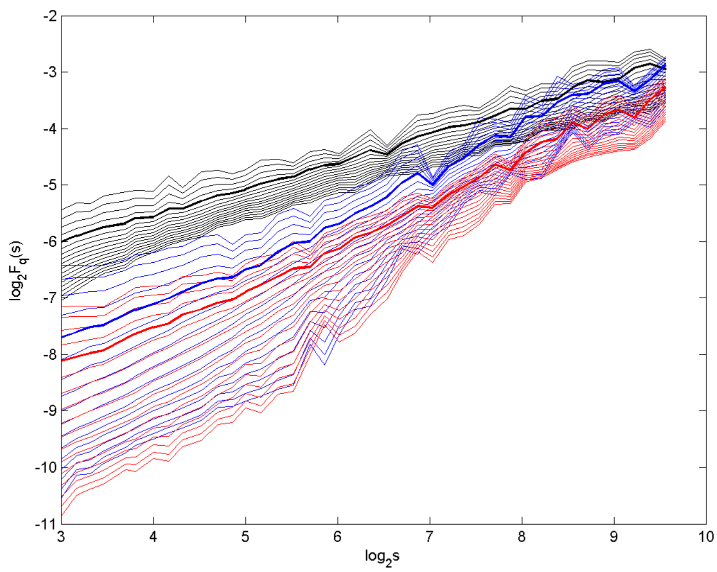
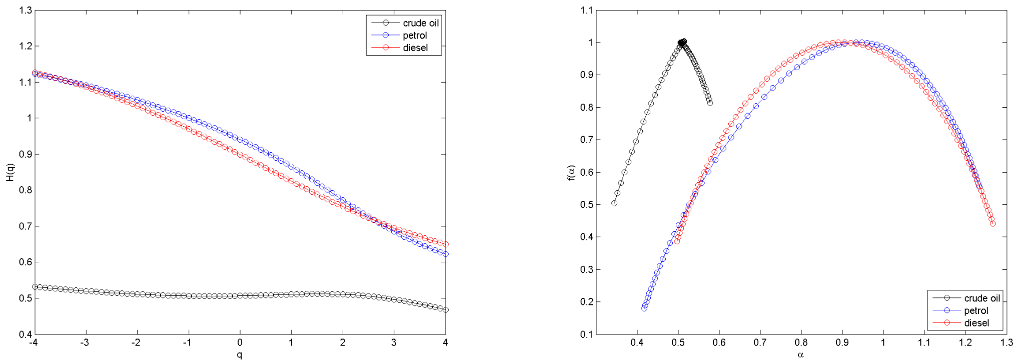


Figure 5.4: Generalised Hurst exponent (left) and multifractal spectrum (right)



generalised Hurst exponent $H(q)$ is independent on q. On the other hand, in case of multifractal series, one Hurst exponent is not capable to fully describe dynamics of a given series, therefore, $H(q)$ vary with different values of q. The later behaviour is apparent from the plot for petrol and diesel returns as $H(q)$ decrease from values close to 1.1 to approximately 0.6 suggesting that small-scope fluctuations (described by negative q) have stronger persistence than larger fluctuations (described by positive q).

The ΔH defined as $H(q)_{max} - H(q)_{min}$ is very large (0.50 for petrol and 0.48 for diesel), therefore, petrol and diesel returns seems to be multifractal. On the contrary, crude oil returns seems to be monofractal as its plot depicts almost stable linear relationship between $H(q)$ and q. It only starts to slightly bend between $q = 2.5$ and $q = 4$. Source of this bending behaviour can be attributed to not-perfectly linear scaling of fluctuations $F_q(s)$ for higher values of q. This behaviour is likely to be caused by multifractal noise and would be probably eliminated if the time series is longer (Grech & Pamula 2011). Additionally, ΔH of crude oil is 0.06 which is within the confidence bands for monofractal series with finite sample of similar length as proved by simulating multifractal behaviour of monofractal series of finite sample in Grech & Pamula (2011).

In the right panel of Figure 5.4 displaying the multifractal spectrum we can clearly see that the width of multifractal spectrum (i.e. degree of multifractaly) $\Delta\alpha = \alpha_{max} - \alpha_{min}$ is almost the same, and large enough for both petrol (0.82) and diesel returns (0.77) to confirm multifractality. $\Delta\alpha$ for crude oil return is 0.23 which is larger than expected, however, the same reasoning as in previous paragraph applies, i.e. the multifractal-like behaviour is likely to be caused by multifractal noise and would be probably eliminated if the time series is longer (Grech & Pamula 2011). Additionally, it has been statistically proved (e.g. Grech & Pamula 2011) that length of series affect the ability of accurate estimation of $\Delta\alpha$ and that monofractal time series with relatively short sample (4096 observations) can exhibit $\Delta\alpha$ of approximately 0.23 despite being monofractal. Therefore, in our case with 2317 observations of crude oil prices, $\Delta\alpha = 0.23$ is within the confidence bands, and we cannot interpret this result as evidence of multifractality in crude oil returns. Multifractal spectrum can be interpreted as a typical distribution of the measure (Mandelbrot 1990), thus we can say that $H(q)$ of crude oil tends to be around 0.5 (i.e. monofractal), while $H(q)$ of petrol and diesel tend to be spread out over many values (i.e. multifractal).

Such findings can be interpreted in the same way as our GHE results. That is, there are complex non-linear relationships in autocorrelations across the probability distribution function of petrol and diesel returns in the Czech market. Such non-linearities imply that autocorrelations of small changes in petrol or diesel returns are different from autocorrelations of large changes of corresponding fuel return series. On the other hand, crude oil returns exhibit linear relationships along the whole probability distribution function, in other words autocorrelations for both small and large changes are the same. Presence of multifractality in time series requires specific models, because traditional models used in finance (e.g. ARIMA-(G)ARCH family) are not able to cope with multifractal behaviour. Nevertheless, the actual type of model depends on both multifractality and presence of long memory, thus we need to continue and analyse the generalised Hurst exponents estimated by MF-DFA to make any conclusions regarding long memory.

Generalised Hurst exponents $H(q)$ for integer values of q together with ΔH and $\Delta\alpha$ are summarised in Table 5.2. The bigger ΔH (or $\Delta\alpha$) the more complicated the underlying process, and therefore scaling behaviour is. Thus, more complexity is to be found in the given market. Analogically to GHE case, particularly values of $H(1)$ representing scaling behaviour of absolute returns, and $H(2)$ indicating presence of long-range dependence in autocorrelations should be emphasised and discussed.

Our results indicate that the crude oil market is very efficient as its returns have $H(1) \approx H(2) = 0.51 \approx 0.5$, suggesting negligible persistence in autocorrelation and random walk like behaviour. However, petrol and diesel markets seems to have long-range dependent autocorrelations as their $H(2)$ is equal to 0.77 and 0.75, respectively. Moreover, values of $H(1)$ equal to 0.87 and 0.82,

Table 5.2: MF-DFA generalised Hurst exponent

	crude oil	petrol	diesel
$H(-4)$	0.5313	1.1224	1.1265
$H(-3)$	0.5200	1.0905	1.0858
$H(-2)$	0.5112	1.0500	1.0334
$H(-1)$	0.5065	1.0001	0.9696
$H(0)$	0.5069	0.9400	0.8980
$H(1)$	0.5107	0.8650	0.8243
$H(2)$	0.5107	0.7723	0.7544
$H(3)$	0.4968	0.6855	0.6953
$H(4)$	0.4686	0.6225	0.6501
ΔH	0.0627	0.4999	0.4763
$\Delta\alpha$	0.2334	0.8165	0.7691

respectively, suggesting strong persistence of both petrol and diesel returns, therefore high (low) returns of petrol and diesel are likely to be followed by high (low) returns, eventually pointing out clusters of volatility. Therefore, not only multifractality but also long memory is present in both petrol and diesel markets in the Czech Republic. In order to successfully model their returns, one needs to implement a more complex model capable to cope with both multifractality and long memory. Calvet & Fisher (2008) suggest to use Markov switching multifractal model for such conditions in the market. On the other hand, crude oil market is monofractal and free of long-range dependence. Therefore, it seems to be efficient and there is not much to be modelled. This is in line with our expectations and results from GHE method in Section 5.1.

To identify the source of multifractality in our data, we introduce two additional types of series *(a) shuffled series*, and *(b) surrogate series*. Both procedures are thoroughly described in Section 3.4. The generating process of such series from the original dataset and following estimation by MF-DFA was repeated 500 times each to ensure statistical significance of our results.

Let us first discuss the results of the simulation visually by description of Figure 5.5 depicting one of the 500 realisations (the axis are fixed at the same scale for all variables to ensure easier comparison). Left panels of the figure show generalised Hurst exponents for each variable and right panels display multifractal spectra. It is apparent that shuffling the series have very small effect on ΔH as $H(q)$ still varies in q. The curve depicting generalised Hurst exponent for shuffled crude oil returns is almost identical to the original one.

The curves depicting $H(q)$ for shuffled petrol and shuffled diesel returns are shifted down and particularly in case of diesel it is slightly straightened up. However, they both still exhibit strong multifractal behaviour. The spectra of petrol and diesel is shifted to left and only negligibly narrower, while the spectrum of crude oil is shortened to less than a half of its original length. On the other hand, all surrogate series exhibit constant $H(q)$ across all q and therefore multifractal spectra are concentrated in a tiny narrow arc of points and multifractal degree $\Delta\alpha$ is close to zero. Thus multifractality seems not to be present in surrogate series. Corresponding histograms of ΔH and $\Delta\alpha$ of shuffled and surrogate series are visualised in Figure 5.6 . The histograms also suggests that surrogate series are probably free of multifractality as the mean value of ΔH and $\Delta\alpha$ for each variable is close to zero. On the other hand, mean value of ΔH and $\Delta\alpha$ for each shuffled series is significantly different from zero, therefore, multifractality seems to be present.

To support these observations with statistically significant test, we use the simulated data to calculate 0.025 and 0.975 quantiles and thus form a 95% confidence intervals for ΔH and $\Delta\alpha$ for all shuffled and surrogate series. Finally, the original values from MF-DFA results are compared with the corresponding critical intervals to asses the true source of multifractality. Therefore, the corresponding null hypotheses are $\mathrm{H}_{0,shuffled}$: Multifractality completely due to distributional properties, and $\mathrm{H}_{0,surrogate}$: Multifractality completely due to (linear) autocorrelations, respectively. The confidence intervals summarised in Table 5.3 imply that both petrol and diesel returns are multifractal (large ΔH and $\Delta\alpha$). The proximity of original values to the confidence interval of shuffled series indicate that the major source of multifractality of both fuels arise from distributional properties (i.e. fat-tailed distribution of returns), whereas

Table 5.3: Original values and 95% confidence intervals of simulation results

		crude oil	petrol	diesel
ΔH	original series	0.0627	0.4999	0.4763
	shuffled series	$[0.0511, 0.1511]$	$[0.2811, 0.4418]$	$[0.2417, 0.3733]$
	surrogate series	$[0.0088, 0.0841]$	$[0.0059, 0.0747]$	$[0.0053, 0.0617]$
$\Delta\alpha$	original series	0.2334	0.8165	0.7691
	shuffled series	$[0.1270, 0.3151]$	$[0.5525, 0.8071]$	$[0.4623, 0.6864]$
	surrogate series	$[0.0262, 0.1749]$	$[0.0172, 0.1223]$	$[0.0177, 0.1194]$

autocorrelations of returns as such have some small contribution to the overall multifractality. The contribution of autocorrelations is slightly bigger in case of diesel returns than in case of petrol returns, as the original values are relatively further away from the confidence interval of shuffled series. In case of crude oil returns, the simulation suggest that we cannot reject the hypothesis that multifractality (or multifractal noise) is completely due to distributional properties, however, as already mentioned, the actual values of ΔH and $\Delta\alpha$ are too small to indicate presence of multifractality and are most likely to be caused by a multifractal noise.

Figure 5.5: Results of shuffling and surrogate procedures

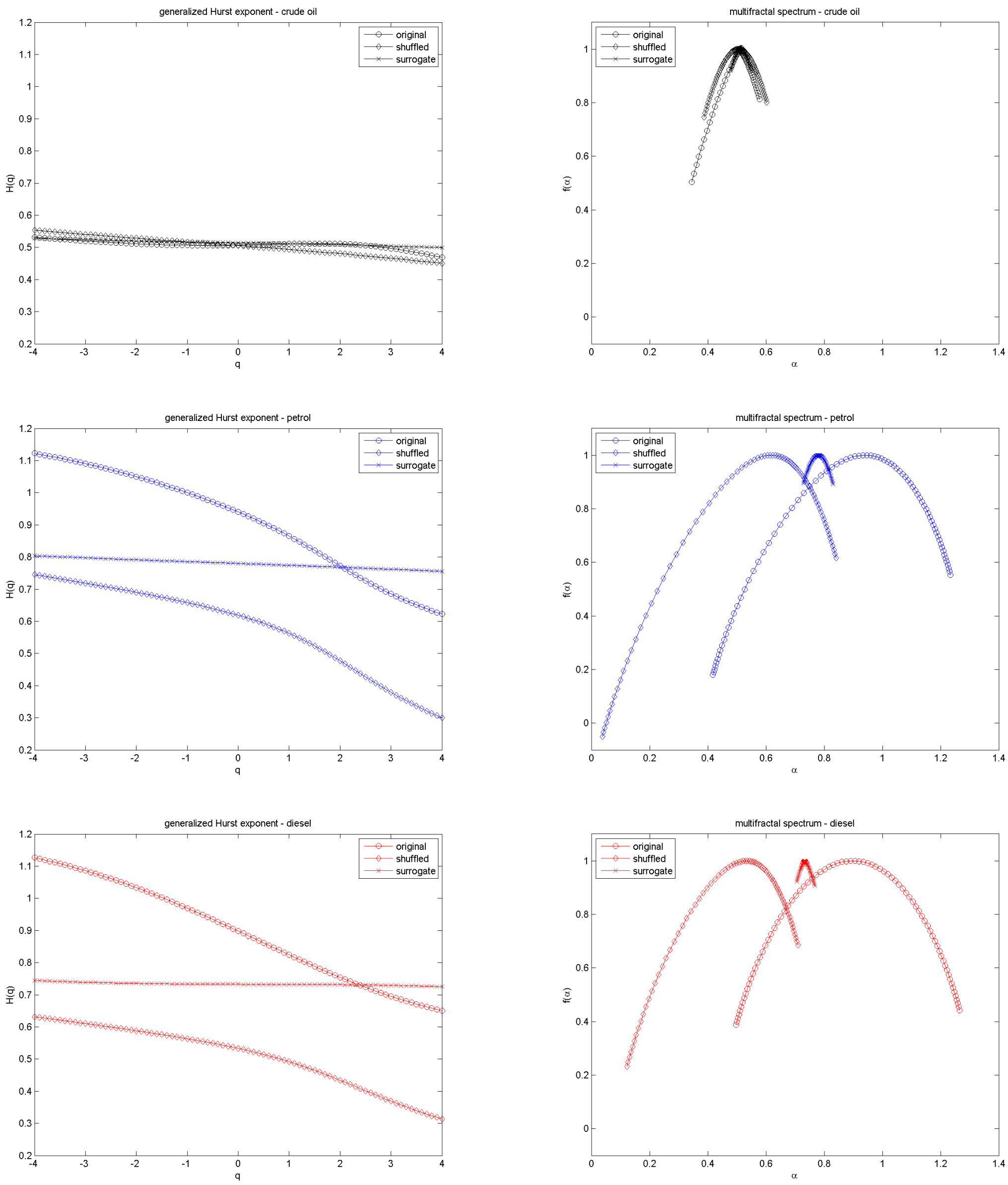
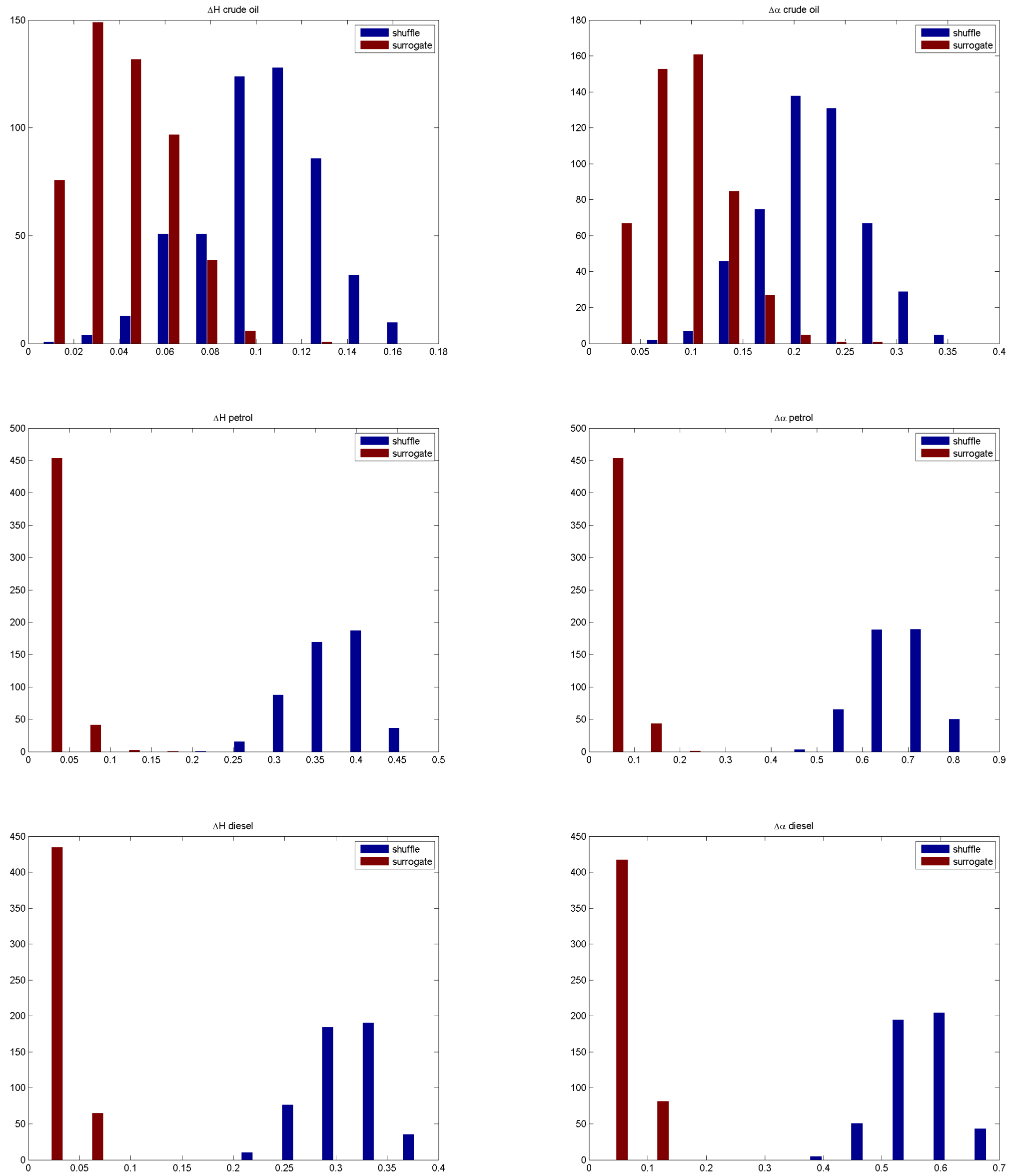


Figure 5.6: Histograms ΔH (left) and $\Delta\alpha$ (right) of simulated shuffled and surrogate series



Using a similar reasoning as Calvet & Fisher (2008) about financial markets, where different agents apply variety of strategies and investment horizons, that through multiplicative cascade of individual decisions and simultaneous flow of information create multifractal market, and applying this to Czech fuel market could bring us an idea about the economic interpretation of the source of discovered multifractality in fuels. For simplicity, let us imagine two types of agents in the market. One is a private individual with passenger car, the other is a business entity with commercial vehicle or truck. In general, one can expect that a private individual would be more price sensitive and would substitute to other means of transport more easily than a business entity if fuel is too expensive, therefore he makes decisions within his relatively long "investment horizon". On the other hand, business entity cannot afford to stop doing its business just because of the price of fuel, therefore has relatively low price sensitivity and makes decisions within its very short "investment horizon". Each type of agents perceive information in the market differently and makes decisions according to its own investment horizon. Such effects, when combined together creates complicated flows of information. Extending this thought by allowing more diversity within each type of agents, who make simultaneous decisions within their own unique investment horizon forms the multiplicative cascade that in the end creates multifractality in the market. There is even more diversity to be found in the real world, therefore, we suggest that in practice, the source of multifractality in petrol and diesel markets in Czech Republic could be attributed to heterogeneous demand created by wide range of individual customers with different preferences and buying strategies (i.e. investment horizons).

Chapter 6

Conclusion

In this thesis, we investigated and compared the efficiency of Czech petrol and diesel markets together with European crude oil market on daily prices from January 2004 to February 2013 with generalised Hurst exponent (GHE) and multifractal detrended fluctuation analysis (MF-DFA) techniques. We adjusted the retail prices to net of tax prices prior to the analysis, because very large fraction (approximately 50%) of the daily retail price of petrol and diesel is a tax component and taxation regimes changed a few times during the period of interest. Crude oil price, reported in USD was converted into CZK using Czech National Bank spot exchange rates for corresponding date. For convenience and easier interpretation, the analysis was performed with logarithmic returns. Autocorrelation function of petrol and diesel returns combined with Ljung-Box Q test at lag 5, representing 5 days long business week, confirmed our hypothesis of strong autocorrelation (p-values <0.01), which is not to be found in crude oil returns (p-value 0.86). Furthermore, petrol and diesel returns come from heavily fat-tailed distribution, suggesting that the prices usually evolve in rather small steps, however, some extreme jumps occur from time to time.

Results from both GHE and MF-DFA supported our hypothesis about efficient crude oil market. Depending on the method, only negligible anti-persistence or persistence was detected in crude oil returns ($H(1)_{\text{GHE}} = 0.48$, $H(1)_{\text{MF-DFA}} = 0.51$) and no significant indication of long-range dependence was found ($H(2)_{\text{GHE}} = 0.50$, $H(2)_{\text{MF-DFA}} = 0.51$) suggesting random walk like behaviour and efficient market. Moreover, both methods suggested that crude oil returns are monofractal, with $\Delta H = 0.06$, and $\Delta\alpha = 0.23$ suggesting only some multifractal noise, which is likely to be due to relatively short length of the series. This, however, was not statistically proved because our null hypotheses were defined in order to uncover the source of multifractality instead. Therefore it is enough to use traditional models to analyse crude oil market, however, because of the random walk like behaviour there is not much to be modelled in its returns. Such finding about crude oil are in line with our expectations and the fact that crude oil is traded in commodity markets.

On the other hand, petrol market in the Czech Republic proved to be inefficient because its returns are strongly persistent ($H(1)_{\text{GHE}} = 0.81$, $H(1)_{\text{MF-DFA}} = 0.87$) and contain long-range dependence in autocorrelations ($H(2)_{\text{GHE}} = 0.66$, $H(2)_{\text{MF-DFA}} = 0.77$). Strong persistence indicate formation of clusters of high volatility and presence of long memory violates market efficiency. Additionally, GHE suggested presence of multifractality in petrol returns, which was subsequently confirmed by the width of multifractal spectrum in MF-DFA ($\Delta H = 0.50$, $\Delta\alpha = 0.82$). Presence of multifractality indicates non-linearities in scaling of autocorrelations as they differ across the probability distribution function. As a result, it implies that petrol market cannot be successfully modelled by traditional models used in finance. Instead, a more complex model capable to implement both long memory and multifractality is required. For instance, Markov switching multifractal model would be appropriate under such circumstances.

Similarly to petrol, diesel market in the Czech Republic proved to be inefficient because for the same reasons. Diesel returns are strongly persistent as well, nevertheless the degree of persistence is slightly smaller ($H(1)_{\text{GHE}} = 0.77$, $H(1)_{\text{MF-DFA}} = 0.82$). Long-range dependence in autocorrelations was again detected with negligibly smaller values than in petrol returns ($H(2)_{\text{GHE}} = 0.65$, $H(2)_{\text{MF-DFA}} = 0.75$). As expected, GHE suggested presence of multifractality in petrol returns, which was subsequently confirmed by the width of multifractal spectrum in MF-DFA ($\Delta H = 0.48$, $\Delta\alpha = 0.77$). Analysis of diesel returns revealed very similar results to analysis of petrol returns, therefore, all conclusion regarding petrol market from the previous paragraph apply to diesel market as well.

To identify the source of multifractality in petrol and diesel returns, we ran simulations for shuffled and surrogate series. We have statistically proved that the major source of multifractality in petrol and diesel markets in the Czech Republic is fat-tailed distribution of their returns. Linear autocorrelations of returns have only small contribution to the overall multifractality, although, their relative contribution is slightly bigger in case of diesel market than in petrol market.

In the end of the thesis, we presented a thought that in practice, multifractality in petrol and diesel markets in the Czech Republic could be caused by heterogeneous demand, created as a result of wide range of numerous interacting agents, each with different individual preferences and consuming strategies.

Finally, we suggest that future research should attempt to model data from petrol and diesel market in the Czech Republic with Markov switching multifractal model or similar methods implementing both long memory and multifractality.

Bibliography

Alvarez-Ramirez, J., J. Alvarez, & E. Rodriguez (2008): "Short-term predictability of crude oil markets: A detrended fluctuation analysis approach." *Energy Economics* **30(5)**: pp. 2645 – 2656.

Alvarez-Ramirez, J., J. Alvarez, & R. Solis (2010): "Crude oil market efficiency and modeling: Insights from the multiscaling autocorrelation pattern." *Energy Economics* **32(5)**: pp. 993 – 1000.

Alvarez-Ramirez, J., M. Cisneros, C. Ibarra-Valdez, & A. Soriano (2002): "Multifractal hurst analysis of crude oil prices." *Physica A: Statistical Mechanics and its Applications* **313(3–4)**: pp. 651 – 670.

Barabási, A.-L. & T. Vicsek (1991): "Multifractality of self-affine fractals." *Physical Review A* **44(4)**: p. 2730.

Barunik, J., T. Aste, T. D. Matteo, & R. Liu (2012): "Understanding the source of multifractality in financial markets." *Physica A: Statistical Mechanics and its Applications* **391(17)**: pp. 4234 – 4251.

Barunik, J. & L. Kristoufek (2010): "On hurst exponent estimation under heavy-tailed distributions." *Physica A: Statistical Mechanics and its Applications* **389(18)**: pp. 3844 – 3855.

Beran, J. (1994): *Statistics for long-memory processes*, volume 61. Chapman & Hall/CRC.

Calvet, L. E. & A. Fisher (2008): *Multifractal volatility: theory, forecasting, and pricing.* Academic Press.

Customs Administration of the Czech Republic (2010): "Změna sazeb spotřebních daní." [online] `http://www.celnisprava.cz/cz/crceskebudejovice/aktuality/Stranky/zmena-sazeb-spotrebnich-dani.aspx`.

Czech Statistical Office (2004a): "Statistical yearbook of the czech republic 2004." [online] http://www.czso.cz/csu/2004edicniplan.nsf/publ/10n1-04-_2004.

Czech Statistical Office (2012b): "Statistical yearbook of the czech republic 2012." [online] http://www.czso.cz/csu/2012edicniplan.nsf/publ/0001-12-r_2012.

Di Matteo, T. (2007): "Multi-scaling in finance." *Quantitative Finance* **7(1)**: pp. 21–36.

Di Matteo, T., T. Aste, & M. Dacorogna (2003): "Scaling behaviors in differently developed markets." *Physica A: Statistical Mechanics and its Applications* **324(1–2)**: pp. 183 – 188. Proceedings of the International Econophysics Conference.

Di Matteo, T., T. Aste, & M. M. Dacorogna (2005): "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development." *Journal of Banking & Finance* **29(4)**: pp. 827–851.

Dušek, L. (2004): "Zpráva o síti čerpacích stanic PHM v ČR za 1. pololetí 2004." *Ropa a ropné produkty 2004*, Ministry of Industry and Trade of the Czech Republic.

Dušek, L. (2005): "Zpráva o síti čerpacích stanic PHM v ČR za 1. pololetí 2005." *Ropa a ropné produkty 2005*, Ministry of Industry and Trade of the Czech Republic.

Dušek, L. (2012): "Ropa a ropné produkty za 1. pololetí roku 2012." *Ropa a ropné produkty 2012*, Ministry of Industry and Trade of the Czech Republic.

Dušek, L. & M. Purnoch (2013): "Zpráva o aktualizaci a stavu evidence čerpacích stanic pohonných hmot v Čr k 31. 12. 2012." *Evidence čerpacích stanic pohonných hmot 2012*, Ministry of Industry and Trade of the Czech Republic.

Engelen, S., P. Norouzzadeh, W. Dullaert, & B. Rahmani (2011): "Multifractal features of spot rates in the liquid petroleum gas shipping market." *Energy Economics* **33(1)**: pp. 88 – 98.

Fama, E. F. (1970): "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance* **25(2)**: pp. pp. 383–417.

Feder, J. & P. Bak (1989): "Fractals." *Physics Today* **42**: p. 90.

Grech, D. & G. Pamula (2011): "How much multifractality is included in monofractal signals?" *arXiv preprint arXiv:1108.1951* .

Gu, R., H. Chen, & Y. Wang (2010): "Multifractal analysis on international crude oil markets based on the multifractal detrended fluctuation analysis." *Physica A: Statistical Mechanics and its Applications* **389(14)**: pp. 2805 – 2815.

Guegan, D. (2005): "How can we define the concept of long memory? an econometric survey." *Econometric reviews* **24(2)**: pp. 113–149.

He, L.-Y. & S.-P. Chen (2010): "Are crude oil markets multifractal? evidence from mf-dfa and mf-ssa perspectives." *Physica A: Statistical Mechanics and its Applications* **389(16)**: pp. 3218 – 3229.

Hurst, H. E. (1951): "Long-term storage capacity of reservoirs." *Trans. Amer. Soc. Civil Eng.* **116**: pp. 770–808.

Ihlen, E. A. (2012): "Introduction to multifractal detrended fluctuation analysis in matlab." *Frontiers in Physiology* **3**.

Jia, Z., M. Cui, & H. Li (2012): "Research on the relationship between the multifractality and long memory of realized volatility in the sseci." *Physica A: Statistical Mechanics and its Applications* **391(3)**: pp. 740 – 749.

Kantelhardt, J. W. (2008): "Fractal and Multifractal Time Series." *Institute of Physics, Martin-Luther-University* **(4)**.

Kantelhardt, J. W., S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, & H. Stanley (2002): "Multifractal detrended fluctuation analysis of nonstationary time series." *Physica A: Statistical Mechanics and its Applications* **316(1–4)**: pp. 87 – 114.

Mandelbrot, B. B. (1964): *Self-similar Random Process and the Range.*

Mandelbrot, B. B. (1974): "Intermittent turbulence in self-similar cascades-divergence of high moments and dimension of the carrier." *Journal of Fluid Mechanics* **62(2)**: pp. 331–358.

Mandelbrot, B. B. (1989): "Multifractal measures, especially for the geophysicist." *Pure and Applied Geophysics* **131(1-2)**: pp. 5–42.

Mandelbrot, B. B. (1990): "Negative fractal dimensions and multifractals." *Physica A: Statistical Mechanics and its Applications* **163(1)**: pp. 306–315.

Mandelbrot, B. B. (1997): *Fractals and scaling in finance: Discontinuity and concentration.* Springer Verlag.

Matia, K., Y. Ashkenazy, & H. E. Stanley (2003): "Multifractal properties of price fluctuations of stocks and commodities." *EPL (Europhysics Letters)* **61(3)**: p. 422.

Mitra, S. (2012): "Is hurst exponent value useful in forecasting financial time series?" *Asian Social Science* **8(8)**: p. p111.

Peitgen, H.-O., H. Jürgens, & D. Saupe (2004): *Chaos and fractals: new frontiers of science.* Springer.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, & A. L. Goldberger (1994): "Mosaic organization of DNA nucleotides." *Phys. Rev. E* **49**: pp. 1685–1689.

Samorodnitsky, G. (2007): "Long range dependence." *Foundations and Trends in Stochastic Systems* **1(3)**: pp. 163–257.

Serletis, A. & I. Andreadis (2004): "Random fractal structures in north american energy markets." *Energy Economics* **26(3)**: pp. 389 – 399.

Tabak, B. M. & D. O. Cajueiro (2007): "Are the crude oil markets becoming weakly efficient over time? a test for time-varying long-range dependence in prices and volatility." *Energy Economics* **29(1)**: pp. 28 – 36.

Wang, Y., Y. Wei, & C. Wu (2011): "Detrended fluctuation analysis on spot and futures markets of west texas intermediate crude oil." *Physica A: Statistical Mechanics and its Applications* **390(5)**: pp. 864 – 875.

Yuan, Y., X. tian Zhuang, & X. Jin (2009): "Measuring multifractality of stock price fluctuation using multifractal detrended fluctuation analysis." *Physica A: Statistical Mechanics and its Applications* **388(11)**: pp. 2189 – 2197.

Zunino, L., B. Tabak, A. Figliola, D. Pérez, M. Garavaglia, & O. Rosso (2008): "A multifractal approach for stock market inefficiency." *Physica A: Statistical Mechanics and its Applications* **387(26)**: pp. 6558 – 6566.