Lappeenranta University of Technology
School of Business and Management
Degree Program in Strategic Finance and Business Analytics

Maria Polikarpova

FINANCIALIZATION IN THE US NATURAL GAS MARKET AND ITS INFLUENCE ON NATURAL GAS SPOT PRICE DYNAMICS

Examiners: Prof. Eero Pätäri
Dr. Elena John

Supervisors: Prof. Eero Pätäri

ABSTRACT

Lappeenranta University of Technology
School of Business and Management
Degree Program in Strategic Finance and Business Analytics

Maria Polikarpova

Financialization in the US natural gas market and its influence on natural gas spot price dynamics

Master's Thesis

60 pages, 8 figures, 14 tables, 4 appendices

Examiners: Professor Eero Pätäri
Dr. Elena John

Keywords: natural gas market, financialization, spot price, futures price, commodity

This thesis examines the influence of financialization of natural gas (NG) market or noncommercial traders on NG spot price in the US. As NG futures contract is one of the most popular instruments for speculators and it provides price discovery for NG spot price in the future, the dynamics of spot-futures prices are analyzed during the periods from 1997 to 2003 and from 2004 to 2016, respectively.

The descriptive statistics and the cointegration analyses demonstrated higher volatility of NG prices in the later period, as well as more persistent influence of shocks on the short- and long-term relationships between NG spot and futures prices. The seasonality analysis showed that summer period (in addition to winter period) has started to impact on NG spot price possibly due to wider application of NG as a fuel in increasing number of gas fired electrical power plants in the US.

The forecasting models of NG spot prices based on NG futures prices (or bases) and other explanatory variables did not show changes in patterns after 2003 and, therefore, the results demonstrate that noncommercial traders did not cause high fluctuations in NG prices. However, it was found that short positions of noncommercial traders had influenced NG spot price during the period from 1998 to 2010 when NG prices suffered from several high spikes and dips. At the same time, the estimate of maximum temperature anomaly was close to significant. These results can be attributed to special conditions of NG market at that time (weather disasters, inelastic demand, concentrated supply, unregulated NG price, and starting of shale gas extraction).

This thesis also suggests a trading strategy based on NG futures contracts. It shows that in calm time a negative basis (NG spot price net NG futures contract with 1-month maturity) should be a signal to long position in NG futures contract with 1-month maturity, whereas a positive basis should be a signal to short position in the same contract. However, the long and short positions for one-month NG futures need to be avoided or protected by call and put options during the period from November to January, as the dynamics of natural gas price is unpredictable in the conditions of weather anomalies and inelastic demand for NG.

ACKNOWLEDGEMENTS

I express my sincere gratitude to my supervisor Professor Eero Pätäri for helping me with this interesting research topic and guiding me through the process. I would like also to thank other professors and doctors which tought me in Lappenranta University of Technology and Norwegian School of Economics.

Special thanks go to my friends in Lappeeranta, Helsinki, Saint-Petersburg, Severodvinsk and Bergen who have supported my positive mood during studing process.

Maria Polikarpova (6.12.2016)

TABLE OF CONTENT

INTRODUCTION **4**

1.1 BACKGROUND 4

1.2 GOALS AND DELIMITATIONS 9

1.3 STRUCTURE OF THE THESIS 10

2 METHODOLOGY **12**

2.1 UNIT ROOT TESTS 12

2.2 COINTEGRATION ANALYSIS AND ERROR CORRECTION MODEL 14

2.3 FORECASTING MODEL BASED ON FUTURES PRICES 16

2.4 MARKOV-SWITCHING MODEL 18

2.5 TRADING STRATEGY 20

2.6 DATA 21

3. RESULTS **26**

3.1 TESTING ON UNIT ROOT 26

3.2 TESTING ON COINTEGRATION 29

3.3 TESTING ON SEASONALITY 32

3.4 FORECAST MODELS OF NG SPOT PRICE BASED ON FUTURES PRICE AND BASES 33

3.5 MARKOV-SWITCHING MODEL 38

3.6 TRADING STRATEGY AND ITS BACKTESTING 45

4 DISCUSSION AND CONCLUSIONS **49**

5. SUMMARY **53**

REFERENCES **55**

APPENDIX 1. Forecasts of NG spot price using futures and bases

APPENDIX 2. Forecasting models by Markow-Switching Model

APPENDIX 3. Trading strategy

APPENDIX 4. Code developed in R for simulation of tests and models

LIST OF SYMBOLS AND ABBREVIATIONS

ADF test	Augmented Dickey Fuller test
AIC	Akaikes Information Criteria
CME	Chicago Mercantile Exchange
DOLS	Dynamic Ordinary Least Squares
ECM	Error Correction Model
HACSE	Heteroscedastic and Autocorrelation Consistent Standard Errors
HH	Henry Hub
KPSS test	Kwiatkowski Phillips Schmidt and Shin test
LNG	Liquefied Natural Gas
MSM	Markov-Switching model
NG	Natural Gas
NOAA	National Center for Environmental Information
NYMEX	New York Mercantile Exchange
OLS	Ordinary Least Squares
VECM	Vector Error Correction Model
WTI	West Texas Intermediate
ZA test	Zitov Andrews test

A	coefficient matrix
c	coefficient
c	convenience yield (the CC model)
d	coefficient
d	difference
DU	indicator dummy variable for a mean shift at break point (the ZA test)
DT	indicator dummy variable for a trend shift at break point (the ZA test)
F	futures price
N	level of inventories
P	price
P	probability

r	interest rate
SP	matrix of NG spot and futures prices
S	spot price
s	storage costs
T	temperature
t	time
α	coefficient
β	coefficient
$\boldsymbol{\Gamma}$	coefficient matrix
σ	variance
τ	time trend
ε	error
u	error
ω	residuals
γ	coefficient
θ	coefficient
ρ	coefficient
δ	coefficient
μ	coefficient

INTRODUCTION

1.1 Background

In recent years the natural gas (NG) market has grown more sophisticated. This happened due to technical advances in storage and transport of NG resources, as well as due to the development of a more comprehensive market for NG and a corresponding futures market for NG hedging and trading. In the US, natural gas plays a crucial role in the heating market and electricity applications. The share of natural gas is still increasing thanks to its low price in recent years, technological advances in the extraction methods and favorable position of gas-fired electric power plants. (Nick and Thoenes 2014; API 2014)

NG price (Fig. 1) presented an upward trend until 2008–2009. The spikes of NG spot price were be linked to the weather shocks (abnormal cold winters and hurricanes) in 2000/2001, 2002/2003, 2005, and 2006 and seasonality in demand (Nick and Thoenes 2014). The second period (especially since 2005–2006) was associated with the expansion of shale gas extraction. The advancements in hydraulic fracturing and horizontal drilling allowed spurring natural gas domestic production (API 2014). The US extracted already about 48% of total dry NG production directly from shale and tight oil reserves in 2014, while in 2005 it accounted only for 5% of total dry NG production (API 2014; Mason and Wilmot 2014).

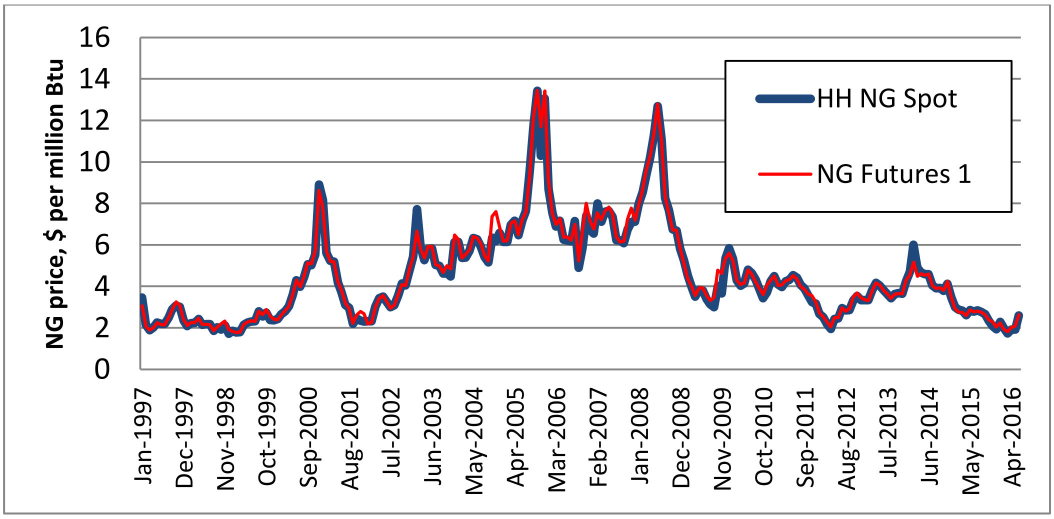
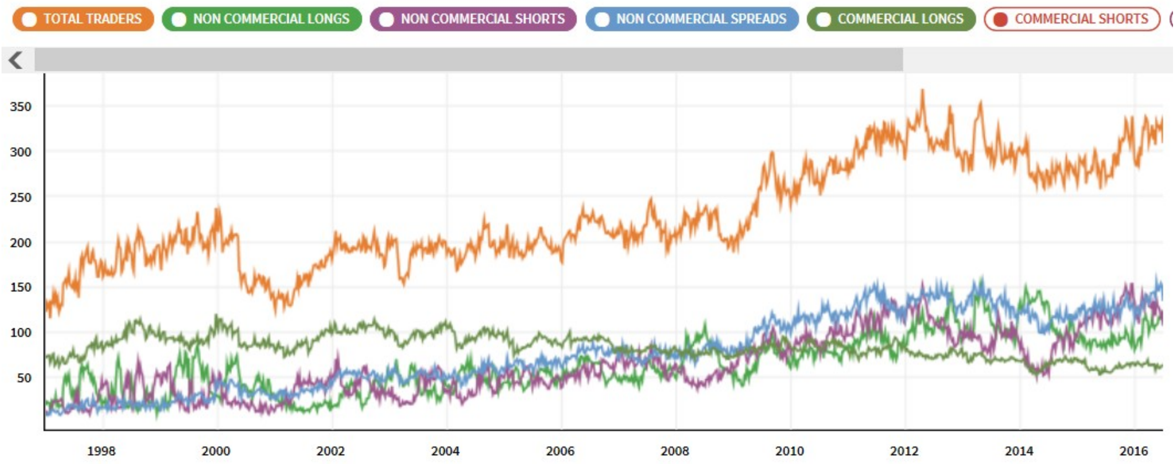


Fig. 1. NG spot and futures (with maturity in 1 month) prices. [EIA, 2016]

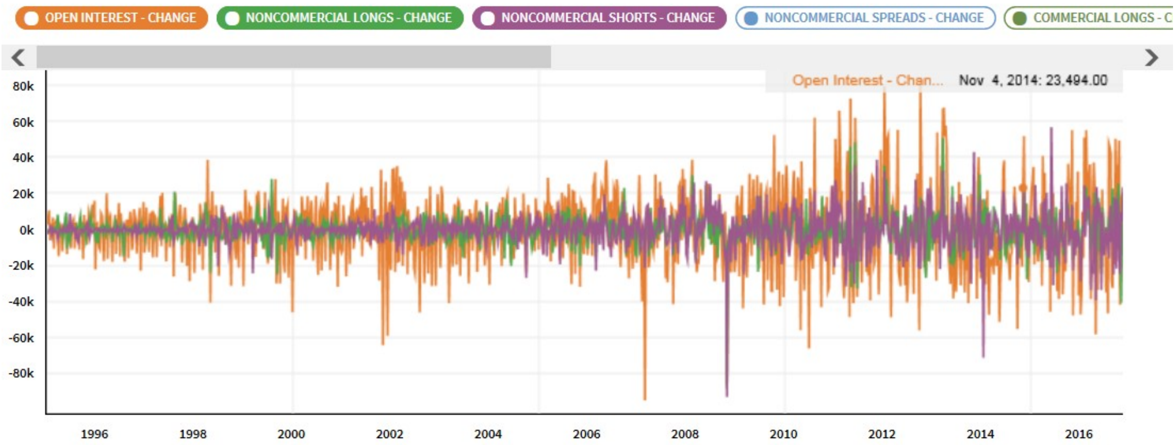
Several spikes between 2005 and 2010 were associated with fluctuations in tight demand – supply balance due the revolution in shale gas production, environmental policies, displacement of conventional foreign suppliers, and domination of relatively small number of large gas producers in the US (Sharipo and Palm 2006; API 2014). The large-scale emergence of shale gas and limited Liquefied Natural Gas (LNG) export capabilities resulted in downward pressure on US NG prices and, therefore, reduction of NG price range in the US since 2009 (see Fig. 1) (Ritz 2015).

Natural gas futures were first available on the New York Mercantile Exchange (NYMEX) in 1990-1993 (API 2014; EIA 2016). However, significant increase in energy futures from financial investor demand has been started after the US Commodity Futures Modernization Act in 2000. This Act introduced "more flexibility, allowing financial agents such as commodity index funds to enter them" (Lubnau and Todorova 2015, p.313). Financial innovations allowed for market participants easy and less expensive access to different financial instruments, such as options, futures, index funds (Fattough et al. 2012). Those might be some of the reasons why since then there has been a significant increase in both the volatility of the spot market, and in the volumes of natural gas futures traded which is inconsistent with previously observed trends. However, many researchers and economists point out that noncommercial traders (eg., hedge funds, investment banks) bring liquidity to the market and thereby allow commercial traders (NG producers and consumers) to hedge their risks at lower prices. (API 2014)

Fig. 2 (a) demonstrates the rise in positions of traders for natural gas in the NYMEX. As can be seen, the position of noncommercial traders accounts for a large part of the total open interest since 2004. Fig. 2 (b) illustrates a rise in change of noncommercial trader positions since 2004. Both figures indicate that noncommercial traders have become very active participants in NG financial market in the US.



(a)



(b)

Fig. 2. Commitment of traders for NG on the NYMEX (in 10000 mmbtu's) (a) and change in traders' positions (NG futures, NYMEX). (Quandle, 2016)

This dynamic has raised a question whether this increase in trading or "financialization" has had an influence on NG price dynamics in the US market. As natural gas takes a more prominent role in the emerging energy mix of the lower carbon economy, a greater understating of market dynamics will be beneficial for all – regulators, commercial traders, and non-commercial traders.

Several authors demonstrated that including of energy commodities in the portfolio offer statistically significant abnormal portfolio return and is often applied as a "low-cost diversification instrument" (Lubnau and Todorova 2015, p.313; Naryan and Liu 2015). Lubnau and Todorova (2015) had shown that the mean-reverting trading strategy using

Bollinger Bands for NG futures (2-, 3-, 4- and 5-month futures combined with the front-month futures) allows achieving the Sharpe ratios (quated on the annualyzed mean daily returns adjusted with the standard deviation) above 2 (1.69 the lowest result).

The researchers state that commodities emerged as a new asset class over the last fifteen years which supports the rise of commodity prices between 2002 and 2008 coincided with that period when money managers and institutional investors became active. (D'Ecclesia et al. 2014) Some reasons for this development include lower cost of investment in commodity markets, low risk aversion, low risk-free interest rate, low or too volatile returns on different financial assets, weakness of dollar, change in risk-aversion features, and excess liquidity in financial markets. (Fattough et al. 2012; Cheng and Xiong 2014) Kolodziej et al. (2014) demonstrate that in oil market the correlation between West Texas Intermediate (WTI) and S&P500 flips from negative to positive after 2008. The authors associate this change with significant reduction in risk-free interest rate that caused the incentive to hold WTI in their portfolios or in other words, to use it as financial asset. When interest rates declined, the holding of the crude oil as a financial asset became profitable due to positive capital gains and low convenience yield. (Kolodziej et al., 2014)

D'Ecclesis et al. (2014) apply Dynamic Ordinary Least Squares (DOLS) and Error Correction Model (ECM) approaches and demonstrate that the "hedging pressure" influences the real price of oil through quick reverting short-term deviations and the structural long-run equilibrium of the oil price. Alizadeh and Tamvakis (2016) show that trading volumes and returns are positively related only when the market is in backwardation and negatively related when the market is in contango, that is explained by the forward curve slope.

Over the years researchers have explored the factors responsible for movements in natural gas prices. Nick and Thoenes (2014) suggest that a series of factors including business cycle, international trade flows, demand and supply shocks/disruptions, export of LNG prices of energy substitutes, temperature or weather conditions and storage shocks play a part in determining a spot price, as well as its conditional mean and volatility. Several authors emphasize the impact of seasonal weather changes to both NG spot and future prices volatilities as the most important factor due to NG demand inelasticity (50% of heating needs and significant share of cooling needs (air conditioning) are meet with natural gas) (Brown and Mine 2008; Hartley et al. 2008; Mason and Wilmot 2014; Martinez and Torro 2015). The highest fluctuations of NG prices are associated with unexpected changes in temperatures and NG storage levels (API 2014). However, in the recent years the seasonality effect has diminished. The first reason is widespread shale gas extraction and because of it downward pressure on winter NG prices. The second reason is increased use of NG as a fuel for cooling (air conditioning) purposes and thereby upward pressure on summer NG prices (Martinez and Torro 2015). The gas-fired power plants became very competitive thanks to stricter environmental policies and low NG prices in the recent years. The producers of NG apply the underground storages in NG production areas and near the consumers (especially at low price during the off-peak periods) to meet the peak demand (API 2014).

Nick and Thoenes (2014) show that the supply disruptions and unexpected weather conditions have only transitory effect on NG prices while coal and crude oil prices influence on the long term development of NG prices. However, in the recent years the shale gas production and liquid spot markets could be the main reasons for the decoupling of oil and NG prices (Nick and Thoenes 2014). This factor has resulted in a considerable price volatility of NG compared with other fuels, such as crude oil and coal. (Mason and Wilmot 2014)

Several research papers have been conducted more specifically around the question of the impact of trade volumes on price volatility in the natural gas market. Early works, including Herbert (1995), established empirically a positive causal relationship between trade volumes and the volatility in the natural gas futures market (Fattouh 2016). Chevallier (2012) use high frequency data from oil and gas futures markets in the US to conclude that both trading volume and trading frequency have a statistically significant impact on various realized volatility measures. Alizadeh and Tamvakis (2016) examine the futures market and demonstrate that “trading volume decreases as maturity of futures contracts increases, while volatility increases as maturity decreases”.

In reviewing academic literature in relation to the mature global markets, a lack of integration of different geographical markets is identified. Siliverstovs et al. (2005) suggest that the regional natural gas markets of the US, Europe and Japan operate to a certain extent independently of each other. Among these regional natural gas markets the US and the UK market are the most developed, as well as the most efficient one (Wu, 2007). However, current widespread Liquified Natural Gas (LNG) is seen by many reseachers as "a driver of cross-continental market integration" (Wu 2007; Neumann et al. 2008; Ritz 2014, p.325).

1.2 Goals and delimitations

Given these findings, I limit the data and, thus, the scope of my analysis to the US market. The US NG market is the most liquid and competitive one where almost all NG is sold and bought in over 30 regional market hubs. The spot and futures markets are the main two distinct markets for NG trading. The futures contracts are traded on the New York Mercantile Exchange (NYMEX) with delivery at Henry Hub (Lousiana). The futures contract is one of the main instruments used by speculators in NG market.

The question of financialization impact on NG spot price is interesting. However, I have been unable to find any other existing quantitative research pertaining to the financialization impact on NG spot prices. Similar research has been conducted in relation to oil futures and spot prices, where the market is more developed and exhibits higher levels of maturity and liquidity than NG market. Shapiro and Pham (2006) studied significant price fluctuations in NG market in 2000–2006. The authors concluded that the concentrated market structure and unregulated price were the main reasons of high volatility of NG in that period. However, this study was just a qualitative analysis and the authors did not employ any empirical methods to prove their conclusions. The empirical approach adopted by Vansteenkiste (2011), where the author used a futures - spot spread model to discover an influence of noncommercial traders, is used for testing of the following hypothesis:

Financialization (non-commercial traders) has had an impact on the interaction of natural gas spot-futures prices

Econometric techniques are applied to test the possible influence of financialization on the interaction between NG spot price and NG futures price during the period from 1997 to 2016. Monthly data for natural gas spot price and futures prices for contracts with maturities in 1, 2, 3 and 4 months are employed in the analysis. The analysis has been executed using the software Excel (descriptive and correlation analyses, backtesting) and R software herein.

To my best knowledge, this type of analysis has not been done before. I investigate the most recent data (including the period of widespread nonconventional gas adoption and start of LNG trade). I analyze two separate periods (1997–2003 and 2004–2016), and thereby the dynamics of NG spot and futures prices independently in each period. The structural breaks found in NG spot and futures prices are included in the models to improve the reliability of the results. The application of Markov-Switching models allows recognizing and ranking the impact of fundamental factors, trader (commercial and noncommercial) positions and other exogenous factors. Testing of the above-described hypothesis in the interaction between NG spot and NG futures prices enables to recognize new patterns and, therefore, to provide some insights for new trading strategies for hedgers and speculators. However, the findings of this thesis are, of course, limited to the power of the applied tests and models, as well as the reliability of the applied data.

1.3 Structure of the thesis

In this thesis the main study objects are NG spot price, NG futures contract price (contracts with 1, 2, 3 and 4 months to maturity) and their bases. As futures contract presents one of the main trading instruments for speculators and it has a good forecasting power (ability of price discovery) on NG spot price, the interactions between NG spot and futures prices are examined to recognize the influence of traders on NG spot price. In order to provide more evidence, I study this interaction separately for the period from 1997 to 2003, and for the period from 2004 to 2016. This demarcation point between the subperiods (2003-2004) is chosen based on the analyses of other researchers and economists which point out that excessive trading of commodities by noncommercial traders started since 2003-2004 after introduction of the US Commodity Futures Modernization Act in 2000.

The work starts from literature review where the current dynamics of natural gas market of the US is described. I devote special attention to financialization phenomena of all commodity markets and give explanation of it from investment point of view. Section 2 describes the methodology with explanations for the chosen econometric techniques and their drawbacks. In Section 3 the applied time-series (NG spot price, NG futures contracts with maturities in 1, 2, 3 and 4 months and their bases) are tested on a unit root by the Augmented Dickey Fuller (ADF) test, the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test and the Zitov-Andrews (ZV) test to limit the spurious regression results in further analyses. Furthermore, the cointegration analysis is used to describe short-and long-term relationships of NG spot and futures prices. It also provides the evidence of the variable cointegration and, therefore, their forecasting power. In the next part, the forecasting models for NG spot price based on NG futures contract price (contracts with 1, 2, 3 and 4 months to maturity) and their bases are constructed by using linear regression models. The results of these forecasts and cointegration analysis enable to analyze the dynamics of interaction between NG spot and futures prices in two subsequent periods. To provide more evidence in the last part of Section 3, the forecasts of NG spot using Markov-Switching model with constant transition probabilities are applied. Besides the above described variables, the change of NG storage held in the underground storage, the changes in long and short futures positions of noncommercial and commercial traders, and maximum temperature anomality are applied as explanatory variables to forecast NG spot price. At the end of Section 3 a trading strategy is presented in details based on the results of the above-stated analyses. This trading strategy is tested using a backtest procedure.

Section 4 presents main discussion and conclusions concerning the stated hypothesis based on the results from several analyses provided in Section 3. In addition to this discussion, several insights about current interactions between NG spot and futures prices are presented. These conclusions can be applied for hedgers and speculators in their trading strategies to utilize the current trends in NG spot-futures prices. At the end of the thesis (Section 5), the summary based on the analysis and results are outlined.

2 METHODOLOGY

2.1 Unit root tests

Traditionally, the relationship between two prices is simulated through Ordinary Least Squares (OLS) regression by using a level-level model. However, since 1980 it has has been proven inappropriate to adopt this approach when the time series of the analyzed prices are non-stationary. (Asche et al. 2003) The application of OLS regression to a time series with a unit root can generate spurious results which would be undesirable. So, before constructing the forecast of NG spot price based on NG futures prices and bases, I will test whether these time series are stationary and whether the linear combinations of these variables have stationary residuals (cointegation analysis in 2.2).

The Augmented Dickey Fuller (ADF) test is applied to explore the existence of a unit root in the analyzed time series. The existence of a unit root in a time series shows that it is non-stationary. This in turn means that a shock in price will persist indefinitely and, therefore, the price in the previous period is the best forecast for the current period. If the time series contains a unit root or is non-stationary, the differentiation of this time-series is required to produce a stationary time series. The first step is testing for a unit root. The null hypothesis of a unit root is tested against no unit root (stationarity) using the following model representing the ADF-test

$$\Delta P_t = \alpha_0 + \theta P_{t-1} + \sum_{i=1}^{k} \rho_i \, \Delta P_{i,t-1} + \varepsilon_t \tag{1}$$

H_0: θ=0, so a unit root

H_1: θ.>0, so no unit root

where $\boldsymbol{\Delta} P_t$ is the difference between the prices at time (t-1) and t, P_{t-1} is the price in the previous period (t-1), $\boldsymbol{\Delta} P_{i,t-1}$ is the lagged term for the difference in the price and ε_{t+1} is the forecast error, θ and ρ are the coefficients. (Nielsen, 2005) The number of lags is chosen based on the minimizing of Akaike Information Criteria (AIC) which thereby minimizes the information lost. If the time series demonstrate non-stationarity the differences of them can be applied to achieve stationarity in all of the time series. However, the ADF test has several drawbacks, such as low power in case of near unit root time series, especially when the time series consists of 30 or less years of observation. The ADF test results are also very sensitiv to structural breaks and trend component. To avoid these drawbacks the Kwiatkowski Phillips Schmidt and Shin (KPSS) test and the Zitov-Andrews (ZV) tests are applied as well.

The main advantage of the KPSS tests is its ability to test stationarity for time series which are near unit root and have a long trend. The setup of the test is represented by the following equation

$$P_t = \mu + \delta\tau + \rho \quad \sum_{i=1}^{t} \xi_i + u_t$$

$$\xi_i = \text{White noise } (0, \sigma^2) \qquad (2)$$

H_0: $\rho = 0$, so trend stationarity

H_1: process is integrated or level stationary

where τ is the time trend, P_t is the price in the previous period t, μ is the intercept, δ is the coefficient, ρ is the coefficients, and u_t and ε_t are the error terms.(Kwistkowski et al. 2012)

The ZV test allows solving problem of detecting a unit root in a stationary time-series when structural breaks are presented in the intercept or trend. The structural breaks are usually associated with global world economic events. The analysis of structural breaks is important, as their presence affects the stationarity and cointegration relationship that will be studied later. (Smyth and Narayan, 2015) The ZV test is based on the following regressions equations correspondingly for a structural break only in constant, in time trend, and in both (constant and time trend)

$$\Delta P_t = c + \alpha P_{t-1} + \beta t + \sum_{i=1}^{k} d_i \, \Delta P_{t-i} + \gamma DU_t + \varepsilon_t$$

$$\Delta P_t = c + \alpha P_{t-1} + \beta t + \sum_{i=1}^{k} d_i \, \Delta P_{t-i} + \theta DT_t + \varepsilon_t$$

$$\Delta P_t = c + \alpha P_{t-1} + \beta t + \sum_{i=1}^{k} d_i \, \Delta P_{t-i} + \gamma DU_t + \theta DT_t + \varepsilon_t \quad (3)$$

H_0: α=0, so a unit root

H_1: no unit root

where $\boldsymbol{\Delta} P_t$ is the difference between the prices at time (t-1) and t, P_{t-1} is the price in the previous period (t-1), $\boldsymbol{\Delta} P_{t-i}$ is the lagged term for the difference in the price, ε_t is the forecast error, DU_t is the indicator dummy variable for a mean shift at break point, DT_t is the indicator dummy variable for a trend shift at break point, θ, α, β, c, d and γ are the coefficients, ε_t are the error term. (Waheed et al. 2006)

In this thesis all three above-presented tests are applied. It is advisable to apply several tests to match their results and minimizing thereby the probability of the spurious regression. All unit root tests suffer from low power and their results are easily distorted if the size of the analyzed time series is too low.

2.2 Cointegration analysis and Error Correction model

The second step before the construction of the forecast is a cointegration analysis. If the linear combination of two non-stationary variables is stationary, then these variables are assumed as cointegrated, or they share a long run equilibrium (or the same stochastic trend) (Nielsen, 2005). So, these variables can be applied for the construction of the forecast. This procedure is necessary before constructing the forecast if the variables contain a unit root.

The cointegration analysis presents a two-step approach: collecting the residuals from an OLS regression (Eq.(7 – 8) below) between two variables and further testing the residuals on stationarity by the ADF-test. (Nielsen, 2005) If the residuals are stationary, then the applied variables in the OLS regressions are cointegrated. However, in case of structural breaks other tests are more appropriate. In this thesis the Johansen procedure (Eq. 4) is applied (Luetkepohl et al. 2004; Perron, 2005; Ghoddusi, 2016). This test is developed to detect a cointegration between several time series with a level shift at an unknown time. It applies a Vector Error Correction Model (VECM) to test the null hypothesis of *no cointegration between time series.*

$$\Delta SP_{\mathrm{t}} = \mu + A \cdot SP_{t-1} + \Gamma_1 \Delta SP_{t-1} + \cdots + \Gamma_{\mathrm{p}} \Delta SP_{\mathrm{t-p}} + w_t \quad (4)$$

where μ is the intercept and ΔSP_t is the differencing operator or matrix with NG spot and futures prices/bases, w_t is the residuals, A is the coefficient matrix for the first lag, Γ_p is the coefficient matrix for each differenced lag. (Luetkepohl et al. 2004)

The Error Correction Model (ECM) can be applied for modeling the short-term relationship between the cointegrating variables. Eq.(4) presents ECM, for example, for modeling of NG spot price and NG futures contract relationship

$$\Delta P_t^{S,NG} = \alpha + \rho_t e_{t-1}^{S,F} + \beta_1 \Delta P_{t-1}^{F,NG} + \beta_2 \Delta P_{t-1}^{S,NG} + u_t \quad (5)$$

where α is the intercept and $\Delta P_t^{S,NG}$ and $\Delta P_{t-1}^{S,NG}$ is the differences in NG spot price in periods t and (t-1) respectively, $e_{t-1}^{S,F}$ is the residuals with lag 1 from the regression of the NG spot price and the NG futures price, ΔP_{t-1}^{F} is the difference in the NG futures price of period (t-1), ρ_t is the adjustment parameter, β_1 and β_2 are slope coefficients and u_t is error term. All variables applied in this model should be cointegrated to minimize spurious regression.

In this thesis the shorted form of the ECM is applied to define the short-term relationships between NG spot price and NG futures prices and bases. The shorted form of the ECM is as follows (Peilong and Rui, 2010):

$$\Delta P_t^{S,NG} = \alpha + \rho_t e_{t-1}^{S,F} + u_t \tag{6}$$

$\boldsymbol{\rho}$ approaching 1: the price will be adjusted immediately after shock

$\boldsymbol{\rho}$ approaching 0: it takes a long time while the price will be adjusted after shock

$\boldsymbol{\rho}$ below 0: there are some conditions between variables that prompt them to back in equilibrium.

2.3 Forecasting Model Based on Futures Prices

In order to indentify the influence of investors on the natural gas market I apply several econometric techniques to construct the forecasted performance over 1-4 months horizons of natural gas futures during the periods from 1997 to 2003 and from 2004 to 2016. As it is known, futures are "a vehicle" with which producers and consumers may secure a future price on a given resource. So, the future price can be observed as a sum of the expected spot price and a risk premium. Both producers and consumers may be willing to pay a risk premium to obtain the benefit to secure a future price.

The Cost of Carry model for natural gas futures price can be written as follows

$$F_{t,T} = S_t e^{(r+s-c)(T-t)} \tag{7}$$

where S_t is the spot price at time t, F_T is the futures price at time t with delivery at time T, r is the risk-free interest rate while carrying NG futures contract, s is the storage costs of NG and c is the convenience yield. Two cases are usually presented in the market – contango and backwardation. If the convenience yield net storage costs is positive, then the futures curve is negatively sloped (backwardation), or NG spot price is above NG futures price. Contango is seen when NG spot price is below NG futures price as convenience yield net storage costs is negative. (Lombardini and Robays, 2011)

Before construction of the forecast the causality relationship between NG spot and futures prices should be defined. The Granger causality test can be applied to design the causal relationship between NG spot and futures prices. The test is based on detecting the relationship though a correlation between the current value of one variable and the past values of another variable. (Asche et al., 2013) Based on the previous studies, it is known that the futures contract Granger cause almost all physical prices due to its more liquidity and higher information content. (Asche et al., 2013) However, in the case of natural gas this relationship is opposite according to the results of many recent research papers. (Georgia 2012; Nicolau et al. 2013) Nicolau et al. (2013) demonstrated that the NG futures price Granger causes the NG spot price and, therefore, the NG futures price can be a predictor of the NG spot price. Based on these factors, it is assumed that NG futures price Granger causes NG spot price in the studied periods.

The forecasting model of NG spot price based on NG futures prices can be written also as follows

$$S_{t+1} = F_{t,t+1} + \varepsilon_{t+1} \tag{8}$$

where S_{t+1} is the spot price at time (t+1), $F_{t,\ t+1}$ is the futures price at time t with delivery at time (t+1) and ε_{t+1} is the forecast error. (Reichsfeld and Roache, 2011)

The forecasting model of the spot price based on the futures price with a risk can be written as follows

$$S_{t+1} = \alpha + \beta \cdot F_{t,t+1} + \varepsilon_{t+1} \tag{9}$$

where α is the intercept (risk), β is the slope coefficient of the futures price. (Reichsfeld and Roache, 2011)

The forecasting model of the spot price based on the futures price with a risk premium can be transformed in the next form to include a basis

$$\Delta S_{t+1} = \alpha + \beta(S_t - F_{t,t+1}) + \varepsilon_{t+1} \quad (10)$$

where $\boldsymbol{\Delta} S_{t+1}$ is the difference between the spot price at time of the futures maturity (t+1) and the spot price at time t, $F_{t,\, t+1}$ is the futures price at period t for the delivery at the period (t+1), ε_{t+1} is the forecast error, $\boldsymbol{\Delta} S_{t+1}$ is difference between the spot price at period t and the spot price at period (t+1), α is the intercept and β is the slope coefficient of the basis. (Reichsfeld and Roache 2011)

As NG spot and futures prices are prone to structural breaks due to economic, political and technological events, these breaks need to be assumed in the forecasting models of NG spot price. In the research community there are a lot of discussions about adequate methodology for that. To my best knowledge, there is no good way to include structural break in the forecasts of NG spot price using NG futures prices. So, herein the methodology suggested by Pesaran and Timmermann (2002) is applied, where the optimal window for the forecasts includes the observations after the structural breaks and before it. The inclusion of the observations before the breaks could decrease variance, but it may results in some bias of the coefficients. (Hansen 2012)

In this work Eq. (9) and (10) are applied to construct forecasts of NG spot prices based on the prices of NG futures contracts with maturities of 1, 2, 3 and 4 months and the bases defined by them. The forecasts are simulated for two periods (1997-2003 and 2004-2016) to analyze their dynamics and get some insights of financialization's influence on NG spot price.

2.4 Markov-Switching Model

In the commodity market, traders are distinguished in two groups: commercial and non-commercial. Commercial traders supply and consume natural gas, and utilize the futures market to hedge their exposure to fluctuations of gas price. It means that these agents have "rational expectations on risk and returns costlessly". (Vansteenkiste 2014, p.) Noncommercial traders try to achieve exposure to the gas price dynamics for speculative and diversification purposes and because of it they intervene in the gas futures market. Their trading strategies are based mainly on the previously observed historical patterns due to imperfect knowledge of the gas market determinants or its fundamentals. This fact results in additional gas demand unrelated to its real demand. It can be presented by a noise that is common for all non-commercial traders. (Vansteenkiste, 2014, p.8)

The number of non-commercial traders may increase volatility of natural gas spot and futures price and spikes' deepness. To test this assumption a 2-regime Markov-Switching model (MSM) with constant transition probabilities can be applied. The switching in the market regime can be associated with new policy, demand and supply shocks, macroeconomic events and financial crises. The MSM can derive explanatory power for time series nonlinear behavior, as it captures potential nonlinearities or asymmetries and define the adjustments for them. (Zeitlberger and Brauneis 2016) It is especially useful when the regime switching is driven by exogenous factors. (Basher et al. 2016)

The general idea behind the MSM is that there are two or more persistent and unobservable states S_t for L-dimensional time series process. The first-order Markov chain with switching probabilities defines the switch between these regimes through

$$\text{Probability}\ (S_t = j/S_{t-1} = i) \quad = p_{ij} \tag{11}$$

where S_t and S_{t-1} are the regimes, p_{ij} is the probabilities of switching (4 for two regimes and 9 for three regimes). (Zeitlberger and Brauneis 2016) Markov-switching model allows determination in which regime the market is traded by a probability value.

The fundamental model of the gas spot price can be presented by

$$S_t = \beta_0 + \beta_1 \cdot F_{t,T} + \beta_2 \cdot N_t + \beta_3 \cdot T_t + \varepsilon_t \tag{12}$$

where β_0 is the intercept, Nt is the level of inventories, T_{t+1} is the normalized temperature anomality, β_1- β_3 are the slope coefficient of the variables. Besides the fundamental variables a number of commercial and noncommercial trader long and short positions are added to the model.

2.5 Trading strategy

A trading strategy is suggested using an idea or belief that historical dynamics may repeat in the future. This idea refers to technical analysis methods which suggest examining past and present market activity to predict future patterns (Lissandrin 2015). As in this thesis the dynamics of NG spot-futures prices is analyzed, the trading strategy is developed to take long or short positions in NG futures contracts to get profit further from its adjustment to the forecasts.

The backtesting is applied herein to judge whether the suggested trading strategy is profitable over the past data and, therefore, can be implemented with some degree of confidence in the current period. This test is employed to simulate the results of the developed trading strategy (based on NG spot price forecasts' results), and modify or adjust them if the realized profit is negative. As one of the main rules of the backtesting is long enough sample period to include varying market conditions, the simulated backtest is employed over several current years (2010-2016). The current period is chosen to test the trading strategy on the most adequate market conditions which presents all current tendencies (impact of different seasons, widespread shale gas extraction, low interest rate) in the US NG market.

The second important requirement of a backtest is its maximum closeness to reality or including all the possible trading costs associated with the trading strategy. For this reason, several trading costs are included in the conducted backtest. The first of them is execution (transaction) cost charged by the broker for opening or closing the position. The second cost is holding costs that present deposit margin which buyer and seller of the contract deposit with the clearing house to guarantee their obligations. The third cost is exchange or clearing fee, which is charged by the clearing house for the services. The fourth cost is associated with mandatory fee to National Futures Association for all traders of futures contracts and data fee charged by the CME for providing data about traded contracts.

2.6 Data

Five variables were collected as inputs for the analysis – Henry Hub natural gas spot price, the NYMEX natural gas futures prices for the maturities of 1, 2, 3 and 4 months from the database of Energy Information Agency. (EIA 2016) The Henry Hub (HH) natural gas price is a benchmark for North American natural gas, as it is located in the central part of the US and presents interconnection point for 13 pipelines. (API 2014) Thanks to its central position Henry Hub (Louisiana) is a delivery point for NG futures traded in the New York Mercantile Stock Exchange (NYMEX). The underlying asset for one futures contract is 10,000 million British thermal units (MMBtu) of natural gas. The monthly time series data for these variables are extracted from the IET database for the period between January 1997 and May 2016. (EIA, 2016) The analysis is presented for two periods – from 1997 to 2003 and from 2004 to 2016. Two periods were selected due to the fact that several researchers have found evidence of the financialization phenomena in commodities market after 2003. (Turner et al. 2011; D'Ecclesia et al. 2014) All price time series have been converted into differences by taking the difference in the gas price over two consecutive periods

$$d_t = p_t - p_{t-1} \tag{13}$$

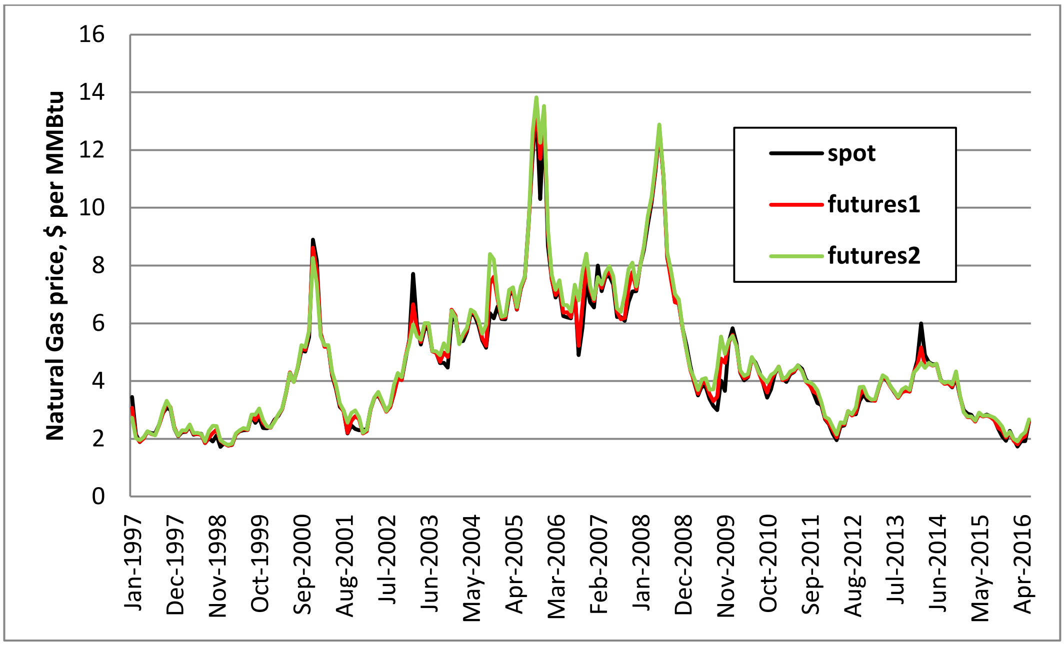
The differencing allowed stabilizing the mean of the time series being analyzed by eliminating trend and seasonality to some extent. This technique is very useful in case of natural gas prices where the seasonality has strong nature and trend is present (Fig. 1).

Fig. 3 illustrates a graphical analysis of NG spot and futures prices. Table 1 lists the descriptive statistics for natural gas spot price and natural gas futures prices for contracts with 1, 2, 3 and 4 months of maturity. All time series have positive and non-zero skewness, which provides evidence of asymmetry in their distributions. The increase of the futures price skewness (far from 0) in the period 2004–2016 shows that the distributions of NG spot and futures prices became more far from symmetrical in this period (Table 1). In both periods the skewness is positive, so the functions are less concave on the upside and, therefore, "tend to crash up". (Ashton 2011, p.1) It distinguishes the commodity prices from the stock prices, as the demand for the former is very inelastic, especially during the critical seasons, such as winter and summer for natural gas. The kurtosis is also positive, except for the futures contracts with the maturities of 3 and 4 months over the period from 1997 to 2003. The positive kurtosis demonstrates an existence of the fat tails in NG spot and futures prices distributions and, therefore, high possibility of unpredictable crashes. All in all, skewness and kurtosis values demonstrate that NG spot and futures prices do not follow the normal distribution.

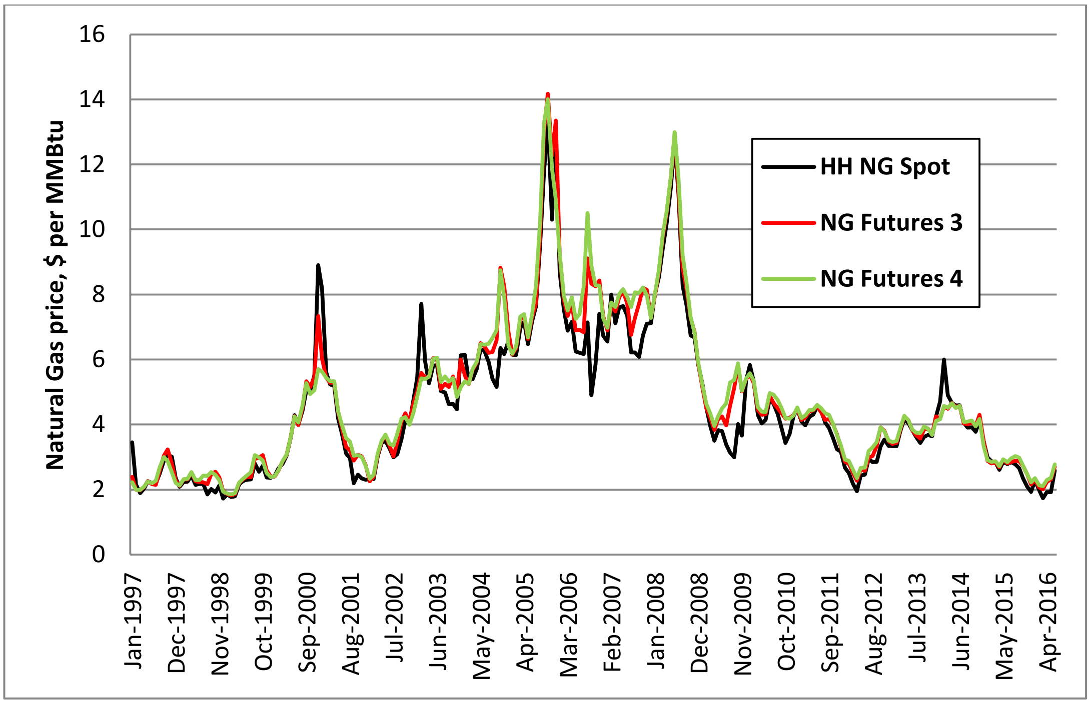
Table 1. Statistical characteristics of spot price and futures for natural gas during period from 1997 to 2016.

	NG Spot	*NG Futures 1*	*NG Futures 2*	*NG Futures 3*	*NG Futures 4*
1997-2003					
Mean	3.425	3.448	3.483	3.474	3.449
St.Deviation	1.555	1.509	1.456	1.355	1.267
Kurtosis	1.615	0.933	0.376	-0.576	-1.091
Skewness	1.327	1.173	1.023	0.763	0.591
2004-2016					
Mean	5.133	5.223	5.368	5.499	5.591
St.Deviation	2.384	2.243	2.503	2.555	2.548
Kurtosis	1.541	1.499	1.344	1.092	0.626
Skewness	1.179	1.191	1.157	1.104	0.986

The results of Table 1 and Fig.3 show higher volatility of the NG prices (higher standard deviations) after 2004, which can be related to the financial crisis in 2008 and other events, such as extension of shale gas extraction since 2004-2005 and monopoly of 20 large gas producing companies (60% of all gas production in the US) . (Mason and Wilmot, 2014) This can be easily recognized in Fig. 3, where the NG spot and futures prices demonstrate several spikes during the period from 2003 to 2009. Figure 3 illustrates that there is a greater difference between spot price and futures contracts for the longer maturities. It can be identified as a negative tendency in later period, meaning that NG spot prices were often lower than the respective future contracts. This fact implies that the market participants supposed that the value of NG prices should rise over time. All in all, in the first period the prices of NG futures contracts demonstrates low (negative in two cases) kurtosis and low skewness compared to the period from 2004 to 2016.



(a)



(b)

Fig. 3. Spot and future prices of natural gas (in $/MM) during period from 1997 to 2016: a) Natural gas spot and futures prices for contracts with 1 and 2 months of maturity;
b) Natural gas spot and futures prices for contracts with 3 and 4 months of maturity.

The correlation analysis presented in Table 2 tells to what degree the variables are in relation with each other. In this case the correlation analysis indicates a high (in range of 0.921-0.994) correlation between all analyzed variables. The correlations between variables are slightly higher during the period from 2004 to 2016, which supports the idea that the futures price can be good predictors of the spot price at the corresponding period.

Table 2. Correlation Analysis for NG spot and futures prices during period from 2004 to 2016.

	NG Spot	*NG Futures 1*	*NG Futures 2*	*NG Futures 3*	*NG Futures 4*
1997-2003					
NG Spot	1	0.994	0.981	0.957	0.921
2004-2016					
NG Spot	1	0.994	0.983	0.972	0.958

Table 3 describes the bases behavior of the applied variables, where the basis was defined by subtracting NG futures prices from NG spot price. Due to the cost of carry and market predictions for the future, the basis tends to be negative for the most of commodities. This phenomenon is seen in Table 3 as well.

As seen in Table 3, the bases are slightly negative in the period before 2004 (except basis 1) and deeply negative during the period from 2004 to 2016. This seems to indicate that the market is in contango in the second period. This insight is supported by other studies concerning energy commodities (Kemp, 2010) that make the claim that from 2004-2005 the NYMEX contracts started to be traded more often in contango than in backwardation. This effect and higher volatility of the basis in the second period may be caused by a shale-induced supply, an increase in trade volumes, market structure specifics and higher market liquidity, as it is shown in Fig. 2. All the bases display negative skewness and excess kurtosis (except bases 3 and during the period from 1997 to 2003). This shows that their distributions have fatter tails with longer left tails than a normal distribution.

Table 3. Statistical characteristics of natural gas bases during period from 1997 to 2016.

	Basis 1	***Basis 2***	***Basis 3***	***Basis 4***
1997-2003				
Mean	0.002	-0.033	-0.024	0.001
St.Deviation	1.509	1.456	1.355	1.267
Kurtosius	0.933	0.377	-0.577	-1.091
Skewness	-1.173	-1.023	-0.763	-0.591
2004-2016				
Mean	-1.773	-1.918	-2.049	-2.141
St.Deviation	2.443	2.503	2.555	2.548
Kurtosius	1.499	1.345	1.093	0.626
Skewness	-1.191	-1.157	-1.104	-0.986
**Basis= Spot Price - Futures Price*				

For Markov-Switching model several variables are collected. Firstly, maximum temperature anomalies (contiguous in the US) are handled from the database of National Center for Environmental Information (NOAA). (NOAA, 2016) Secondly, the changes in noncommercial and commercial long and short positions are collected from the Quandle database. (Quandle, 2016) Thirdly, the changes in natural gas storage level are extracted from Quandle database as well. (Quandle, 2016)

3. RESULTS

3.1 Testing on unit root

Initially all time series were tested based on the presence of a unit root using the ADF-test. Due to the influence of seasonality on the commodity markets, several lags were tested to identify the correct number of lags. The initial number of lags was set to 12 for the monthly time series. The inclusion of a trend was also tested with the null hypothesis that the coefficient of trend would be zero. All the time series demonstrated a unit root or non-stationarity at the 5 % significance level. So, the differencing was applied to avoid a unit root and thereby to minimize the spurious regression results in the following forecasts (Table 4).

The results of the ADF-test and the KPSS-test demonstrate stationarity at the 5% significance level for NG spot price and NG futures prices for contracts with maturities of 1, 2, 3 and 4 months in both periods (1997–2003 and 2004 – 2016). The results of the KPSS-test show that the analyzed time-series are stationary around the trend and around the constant mean, as the null hypotheses of the trend and mean stationarity cannot be rejected at the 5% level (Table 4).

Next the ZA is used to detect structural breaks and test the time series on stationary in intercept, trend, and both intercept and trend. The results show that the prices of NG spot and NG futures contracts are trend non-stationary during the period from 1997 to 2003, whereas the prices of NG futures contract with maturity in 3 and 4 months are also non-stationary in the intercept (Table 4). In all these cases the t-statistics are below the critical t-value at the 5% level. In the later period (2004-2016), NG spot prices and NG futures price for the contract with the maturity in 1 month demonstrate stationarity in both intercept and trend, as the null hypotheses of a unit root are rejected at the 5% level. However, the null hypotheses of a unit root are rejected only at the 10% level for NG futures prices with maturities in 2, 3 and 4 months. It demonstrates that these time series are trend non-stationary at the 5% level. The ZA tests also demonstrate the structural breaks in the intercept in November–December 2000 and October–November 2005 and structural breaks in the trend in October–November 2001 and January–March 2006 (Table

4). This result coincides with the structural breaks recognized in the work of Ghoddusi (2016), where these breaks were associated with the take-off of nonconventional gas production. (Ghoddusi, 2016; Nick and Thoenes, 2014) These results demonstrate that the structural breaks can be also associated with temperature anomaly and inelastic demand, as the most of the breaks have happened in the late autumns and winter periods.

Table 4. ADF test, KPSS and ZA for testing of unit root in natural gas spot price and futures prices for contracts with 1, 2, 3 and 4 months of maturity.

	NG spot price	***Futures 1***	***Futures 2***	***Futures 3***	***Futures 4***
1997-2003					
t-ADF (N_{lags})	-7.408*** (1)	-7.033*** (1)	-6.577 *** (1)	-8.283*** (1)	-6.596*** (1)
KPSS-test, trend	0.046*	0.045*	0.047*	0.053*	0.058*
KPSS-test	0.066*	0.063*	0.081*	0.059*	0.075*
t-ZA, intercept (break point)	-4.857. (48, 2000)	-4.969. (48, 2000)	-4.835. (48, 2000)	-4.399 (47, 2000)	-3.603 (49, 2000)
t-ZA, trend (break point)	-3.131 (58, 2001)	-3.159 (58, 2001)	-3.132 (58, 2001)	-2.943 (59, 2001)	-2.623 (58, 2001)
t-ZA, intercept and trend (break point)	-4.824. (48, 2000)	-4.956. (48, 2000)	-4.784 (48, 2000)	-4.391 (47, 2000)	-3.559 (49, 2000)
2004-2016					
t-ADF (N_{lags})	-5.173*** (8)	-5.802*** (8)	-5.961*** (8)	-5.817*** (8)	-4.478* (9)
KPSS-test, trend	0.035*	0.035*	0.037*	0.039*	0.040*
KPSS-test	0.045*	0.046*	0.051*	0.056*	0.071*
t-ZA, intercept (break point)	-5.679** (21, 2005)	-5.392** (22, 2006)	-5.251* (22, 2006)	-5.270* (22, 2006)	-5.026* (20, 2005)
t-ZA, trend (break point)	-5.017** (23, 2006)	-4.524* (24, 2006)	-4.371. (23, 2006)	-4.462. (24, 2006)	-4.483. (24, 2006)
t-ZA, intercept and trend (break point)	-6.139** (19, 2005)	-5.921** (22, 2006)	-5.780** (22, 2006)	-5.720* (20, 2005)	-5.741** (20, 2005)
*.Significance at 10% significance level, *Significance at 5% significance level ** Significance at 1% significance level, ***Significance at 0.1% significance level*					

Furthermore, the time series of the bases in two periods are tested on the stationary using the ADF-test (Table 5). The results from the ADF-tests demonstrate stationarity for bases at 1% significance level in both periods, as the null-hypothesis of a unit root is rejected based on the t-statistics. However, all time series require differentiation to achieve stationary. The KPSS-tests also demonstrate stationarity around trend and constant under the null hypotheses against the alternative of non-stationarity based on the statistics listed in Table 5.

Table 5. ADF test for testing of unit root in bases.

	Basis 1	***Basis 2***	***Basis 3***	***Basis 4***
1997-2003				
t-ADF (N_{lags})	-3.479*** (12)	-2.568* (12)	-6.166** (3)	-7.832** (1)
KPSS-test, trend	0.026*	0.030*	0.029*	0.026*
KPSS-test	0.046*	0.058*	0.051*	0.043*
t-ZA, intercept (break point)	-3.312 (78, 2003)	-3.159 (79, 2003)	-3.804 (48, 2000)	-4.232 (48, 2000)
t-ZA, trend (break point)	-3.431 (73, 2003)	-3.156 (73, 2003)	-3.234 (73, 2003)	-3.382 (84, 2003)
t-ZA, intercept and trend (break point)	-4.391 (68, 2002)	-3.917 (68, 2002)	-3.832 (48, 2001)	-4.237 (48, 2001)
2004-2016				
t-ADF (N_{lags})	-9.346*** (10)	-6.039*** (12)	-4.279** (12)	-4.454** (11)
KPSS-test, trend	0.008*	0.012*	0.054*	0.014*
KPSS-test	0.003*	0.014*	1.246	0.018*
t-ZA, intercept (break point)	-5.497** (70, 2009)	-5.053* (32, 2006)	-4.849* (32, 2006)	-4.502 (32, 2006)
t-ZA, trend (break point)	-5.291* (120, 2014)	-4.809* (120, 2014)	-4.361. (119, 2014)	-4.079 (119, 2014)
t-ZA, intercept and trend (break point)	-5.486* (70, 2009)	-5.759** (32, 2006)	-5.855** (32, 2006)	-5.612** (32, 2006)
*.Significance at 10% significance level, *Significance at 5% significance level ** Significance at 1% significance level, ***Significance at 0.1% significance level*				

The ZA tests tell that all NG bases are non-stationary in intercept and trend during the period from 1997 to 2003 in accordance with the t-statistics. During the period from 2004 to 2016 only NG bases 1 and 2 demonstrate stationarity in trend and intercept, as the null hypotheses of a unit root are rejected at the 5% level (Table 5). The ZA tests for NG bases 3 and 4 demonstrates that these time series are trend non-stationary at the 5% level. Several structural breaks in the intercept in December 2000, August 2002, June–July 2003, October 2006, December 2009 and in the trend in January 2003, November 2003 and

January–February 2014 are recognized by the ZA test. It may be associated with different factors which were already mentioned above (shale gas production, weather anomaly (cold temperatures and hurricanes), inelastic demand etc.).

These results demonstrate that NG spot and futures prices and bases cannot be assumed as stationary and a cointegration analysis needs to be applied before construction of the forecasts for the NG spot price based on the NG futures prices and bases. It can minimize spurious regression and meaningless coefficients.

3.2 Testing on cointegration

A cointegration analysis consists of testing long-term relationship and short-term relationship between NG spot prices and NG futures prices and bases.

Long-run relationship

The results of the applied cointegration analysis are shown in Table 6. The residuals resulting from OLS regressions of NG spot price and its futures prices for contracts with 1, 2, 3 and 4 months maturities demonstrate stationarity, as the null hypothesis of the presence of unit root are rejected at least at the 5% significance levels based on the t-statistics given in Table 6. Due to presence of structural breaks in the analyzed time series, the Johansen procedure for cointegration analysis is applied. It demonstrates that NG spot and futures prices are cointegrated in both periods (Table 6).

Further, the residuals of the regressions of NG spot prices and bases in two periods are tested on presence of a unit root using the ADF-test (Table 7). The results of the ADF-test demonstrate stationarity of the residuals in all cases, as the null hypothesis of a unit root is rejected at least at the 10% significance levels based on the t-statists listed in Table 7. The cointegration test by Johansen procedure supports the above results of cointegration between NG spot prices and bases in both periods, as the t-statistics are above its critical values at the 10% level.

Table 6. ADF test for residuals of regression of NG spot and futures prices and cointegration test by Johansen procedure for NG spot and futures prices (for contracts with 1, 2, 3 and 4 months of maturity)

	NG spot and futures 1	*NG spot and futures 2*	*NG spot and futures 3*	*NG spot and futures 4*
1997-2003				
t-ADF (N_{lags}), residuals	-5.477*** (1)	-4.948*** (1)	-4.803** (1)	-4.618** (1)
t- Johansen proc	15.58.	18.6*	22.80**	22.75**
2004-2016				
t-ADF (N_{lags}), residuals	-2.619* (11)	-3.579* (9)	-3.071* (9)	-3.744** (12)
t- Johansen proc	42.85**	42.29**	40.12**	37.92**
*.Significance at 10% significance level, *Significance at 5% significance level, ** Significance at 1% significance level, ***Significance at 0.1% significance level*				

Table 7. ADF test for residuals of regression of the natural gas spot price and bases and cointegration test by Johansen procedure for NG spot and futures prices.

	NG spot and basis 1	*NG spot and basis 2*	*NG spot and basis 3*	*NG spot and basis 4*
1997-2003				
t-ADF (N_{lags}), residuals	-2.874. (1)	-2.830. (2)	-2.830. (2)	-2.948** (4)
t- Johansen proc.	15.70.	15.38.	18.27*	20.96**
2004-2016				
t-ADF (N_{lags}), residuals	-3.505* (2)	-3.576** (2)	-3.680* (2)	-3.690** (2)
t- Johansen proc	29.18**	28.45**	27.53**	38.06**
*.Significance at 10% significance level, *Significance at 5% significance level, ** Significance at 1% significance level, ***Significance at 0.1% significance level*				

The presented results demonstrate the long-run relationship between the natural gas spot price and the futures prices for contracts with maturities in 1, 2, 3 and 4 months or the bases. This implies that these time series are cointegrated and, therefore, NG spot prices may be forecasted by NG futures prices or bases.

Short-run relationship (ECM)

The model presented in Eq. (6) can be applied to the short-term relationship between the cointegrated variables (NG spot and futures prices). The results demonstrate that the adjustment coefficient is significantly different from zero for NG futures contracts with maturities of 1 and 3 months towards their equilibrium with correspondingly NG spot prices , as the null hypothesis (*the adjustment coefficient is equal to zero*) is rejected at 5% significance level based on t-statistic (Table 8). As the adjustment coefficient is below zero, it implies that there are some conditions precluding NG spot price towards long-term equilibrium with NG futures prices. However, this state is supported only during the period from 1997 to 2003 and for NG futures contract with maturity of 1 month during the period from 2004 to 2016.

Table 8. Forecasting of NG spot price based on NG futures prices for contracts with 1, 2, 3 and 4 months of maturity.

Δ NG spot	*Coefficien (ρ)*	*t-stat (p-prob)*	*Partial R-sq*	*Coefficien (ρ)*	*t-stat (p-prob)*	*Partial R-sq*
	1997-2004			***2004-2016***		
Futures 1 =f($e_{S,F}$)	**-0.278****	-1.913 (0.059)	0.070	**-0.164*****	-1.799 (0.074)	0.080
Futures 2 =f($e_{S,F}$)	**-0.256***	-1.561 (0.122)	0.080	-0.021	-0.212 (0.832)	0.030
Futures 3 =f($e_{S,F}$)	**-0.377****	-1.979 (0.051)	0.130	**0.186****	1.843 (0.068)	0.050
Futures 4 =f($e_{S,F}$)	0.179	0.633 (0.528)	0.030	0.003	0.026 (0.979)	0.080
*.Significance at 5% significance level, *Significance at 5% significance level* *** Significance at 1% significance level, ***Significance at 0.1% significance level*						

The model presented in Table 8 shows that the adjustment coefficients for the current differences in NG spot prices are insignificantly different from zero for NG futures contracts with maturities of 2 months during the period from 2004 to 2016 and also for the contracts with maturity of 4 months for both periods, as the null hypothesis (*the adjustment coefficient is equal to zero*) cannot be rejected at 5% significance level. This implies that it takes a long time for NG spot price to adjust back to its long-term equilibrium with its futures price after a shock. In case of the futures contract with maturity of 3 month in the period from 2004 to 2016 the adjustment coefficient approaches 1, so the spot price of NG adjusts back fast to its long-term equilibrium with its futures price after a shock.

All in all, the results demonstrate a small inconsistency in price dynamics in two subperiods, especially in case of short-term relationship for NG spot and futures contracts with maturities of 2 and 3 months. The shocks have more persistent influence on NG price dynamics in the period before 2004. It means that NG variables attain a new equilibrium, or slowly return to the previous one after a shock. These shocks may include take-off of nonconventional gas, different political or economical incidents, reasons associated with concentrated market structure or news which causes changes in NG consumption and production.

3.3 Testing on seasonality

The natural gas as a fuel is applied for energy generation purposes (for heating and cooling or air conditioning) and, therefore, the price of the natural gas can be exposed to some seasonal fluctuations. In Table 9 the regression analysis of the natural gas spot price based on the season dummies is presented for two periods: 1997-2003 and 2004-2016. The four dummy (winter, spring, summer and autumn) are created to test the seasonality.

The regressions' estimates indicate that NG spot price changes are affected by winter and summer seasons during the period from 2004 to 2016, as the coefficients with the winter and summer dummies are significant at 10% significance level based on t-statistic (Table 9). However, the summer dummy is insignificant based on its t-statistic during the period from 1997 to 2003 while the winter dummy demonstrates its impact on NG spot price changes.

All in all, these results identify slightly different patterns of NG spot prices between the periods of 1997–2003 and 2004–2016. The significance of the summer period on NG spot prices may be explained by the fact that the consumption of natural gas increased by 4 000 million cubic meters since 2004 thanks to significant consumption of natural gas by electric power producers (gas-fired power plants) (API 2014; EIA 2016). The electric power producers generate electricity mainly for conditioning purposes. This idea is supported by the conclusion of Martinez and Torro (2015) that in the recent years seasonality effect has diminished due higher consumption of NG for cooling purposes and downward pressure on NG prices due widespread shale gas extraction. In any case these results show that the adjustments for seasonality or maximum temperature anomaly need be included in the future analyses.

Table 9. Seasonality of NG spot price using winter, spring, summer and autumn seasons.

Δ NG spot	*Coefficient*	*t-value (prob)*	*Coefficient*	*t-value (prob)*
	1997-2003		***2004-2016***	
Constant	**0.268***	**1.759 (0.082)**	0.157	1.109 (0.269)
D_winter	**-0.427****	**-2.027 (0.046)**	**-0.348.**	**-1.707 (0.090)**
D_spring	-0.227	-1.054 (0.295)	-0.023	-0.119 (0.905)
D_summer	-0.311	-1.441 (0.154)	**-0.355.**	**-1.770 (0.079)**
D_autumn	**-**	**-**	-	-
*.Significance at 10% significance level , *Significance at 5% significance level ** Significance at 1% significance level, ***Significance at 0.1% significance level*				

3.4 Forecast models of NG spot price based on futures price and bases

The description of the constructed forecast models of the natural gas spot prices based on its futures contracts with different maturities (Eq. 9) and bases (Eq. 10) are listed in Table 10 and Table 11. The analyses presented in the other research papers indicate that NG futures price should be a very good forecaster to provide price discovery for NG spot price, especially for maturities shorter than 2 years (as these futures contracts are the most liquid). (FedResBSF 2005; IMF 2011) As structural breaks are detected in the applied time series of NG spot and NG futures and bases, the methodology suggested by Pesaran and Timmermann (2002) will be applied herein.

Forecast of NG spot prices based NG futures prices

During the period from 1997 to 2003 the ZA test detected several structural breaks in the intercept and trend components. To minimize the bias in the forecasts' estimates, the observations of the structural breaks 47 (November 2000), 48 (December 2000), 58 (October 2001) and 59 (November 2001) are excluded from the analysis. This approach is applied in accordance with the methodology suggested by Pesaran and Timmermann (2002) where only observations before and after the structural breaks are included in the optimal window for the forecasts.

First, all the models are tested on heteroskedasticity and autocorrelation of the residuals using the Breusch-Pagan test and the Durbey-Watson test, correspondingly. Heteroskedasticity presents a case when the variance of the unobserved error, conditional on independent variables, is not constant. The autocorrelation of the error term means the correlation of the error terms at several different time periods. Both these cases result in invalid OLS t-statistics and confidence intervals that may lead to inaccurate results of the hypothesis testing. To avoid this problem, Heteroscedastic and Autocorrelation Consistent Standard Errors (HACSE) is applied herein. This method for standard errors calculation allows for the correction of heteroscedasticity and autocorrelation without changing the coefficients. (Eiker 1967; White 1980; Huber 1967)

Testing the null hypothesis "*slope coefficients equal to zero*" shows, that it cannot be rejected at the 90% confidence level based on t-statistic for NG futures contracts with maturity of 1 month in both periods and for NG futures contracts with maturity of 4 month in the period from 1997 to 2003 (Table 10).

In all other cases the results demonstrate that the null hypothesis cannot be rejected based on t-statistic and t-HACSE statistic shown in Table 10. This means that these contracts can be applied as a forecaster of NG spot prices. The intercepts are also insignificant for all forecasts, as based on t-statistic or t-HACSE statistic the null hypothesis is not rejected at the 90% confidence interval in any of the cases.

Table 10. Forecasting of NG spot prices based on NG futures prices for contracts with 1, 2, 3 and 4 months of maturity.

NG spot	*α (t-value)*	*β (t-value)*	*R-sq*	*DW[1] autocorr. (p-value)*	*W-test[2] normality (p-value)*	*BP-test[3] heterosc. (p-value)*
			1997-2003			
NG Futures 1	0.040 (0.767)	**0.189* (2.258)**	0.039	2.252 (0.852)	**0.881 (0.000)**	6.653 (0.010)
NG Futures 2	0.041 (0.661)	0.098 (0.717)	0.008	1.953 (0.412)	**0.905 (0.000)**	3.953 (0.071)
NG Futures 3	0.044 (0.686)	-0.082 (-0.468)	0.003	1.995 (0.342)	**0.909 (0.000)**	0.058 (0.810)
NG Futures 4	0.026 (0.444)	**0.311. (1.860)**	0.037	2.061 (0.582)	**0.882 (0.000)**	0.006 (0.940)
			2004-2015			
NG Futures 1	-0.013 (-0.234)	**0.171. (1.961)**	0.028	2.095 (0.708)	**0.936 (0.000)**	0.146 (0.702)
NG Futures 2	-0.022 (-0.375)	0.058 (0.419)	0.003	1.829 (0.152)	**0.933 (0.000)**	**5.801 (0.016)**
NG Futures 3	-0.016 (-0.256)	0.042 (0.289)	0.001	1.796 (0.114)	**0.934 (0.000)**	**8.132 (0.004)**
NG Futures 4	-0.014 (-0.252)	0.012 (0.124)	0.001	1.793 (0.112)	**0.937 (0.000)**	2.230 (0.135)
*.Significance at 10% significance level, *Significance at 5% significance level ** Significance at 1% significance level, ***Significance at 0.1% significance level*						

[1]H0- hypothesis of the DW test: no autocorrelation in the residuals
[2]H0- hypothesis of the W test: normal distribution of the residuals
[3]H0- hypothesis of the BP test: homoskedasticity or constant variance of the unobserved error

As is shown in Table 10, only NG futures prices with the maturity of 1 month for both periods demonstrate their forecasting ability for NG spot prices. The slope coefficients are close in the periods 1997–2003 and 2004–2016. The residuals of these forecast models have homoscedastic variance, are normally distributed and without autocorrelation in

accordance with the p-values of the Breusch-Pagan test, the Shapiro-Wilk test and the Durbin-Watson test, correspondingly. However, these models demonstrate low R-sq values (about 0.03), so they cannot be applied for the forecasting purposes.
Fig. 1.1 (Appendix 1) illustrates the fitted line of the forecasting models of Table 10. The forecasting errors are high in all models. It is clearly seen that the forecasting graph are biased in all the periods.

Forecast of NG spot prices based bases

The following forecasting models of NG spot price changes based on the bases (Table 11) defined by NG futures price with maturities in 1, 2, 3 and 4 months (Eq.10) demonstrate better forecasting performance (higher R-sq) than the models based on NG futures prices (Eq. 9).

The analysis indicates that in both periods the null hypotheses "*slope coefficients equal to zero*" are rejected for all the models and, therefore, the slope coefficients are statistically different from zero at the 95 % confidence level based on their t-statistics. For all models, the intercepts are statistically insignificant from zero, as the null hypothesis cannot be rejected on their t-statistic at the 95% confidence level. However, only the residuals of models for basis 1 have homoscedastic variance, are normally distributed and without autocorrelation based the p-values of the Breusch-Pagan test, the Shapiro-Wilk test, and the Durbin-Watson test, correspondingly. In all other cases the tests demonstrate that the residuals do not have homoscedastic variance, are not normally distributed, and are with autocorrelation. It means that these models can not be applied for the forecasting purposes.

All in all, we see slightly different dynamics for the forecasting models in two periods, except NG basis 1. Although NG spot price and NG basis have negative relationship in all models, in the period from 2004 to 2016 this dynamics became less obvious. The graphs presented in Fig. 1.2 (Appendix 1) illustrate high bias of the fitted and real paths and high errors in cases of all the models discussed above. It supports the above-stated conclusion that the bases defined by NG futures contracts with maturities from 1 to 4 months forecast poorly the spot price changes during the period from 1997 and 2016. It is worth noticing from Fig. 1.1 and 1.2 (Appendix 1), that the volatilities are significantly higher in the period after 2004 (not only during the financial crisis in 2008). As already mentioned, the reasons can be extension of shale gas extraction, seasonality (cold winters and hurricanes) and problems associated with NG market structure (high concentration).

Table 11. Forecasting of NG spot price based on basis defined by NG futures prices for contracts in 1, 2, 3 and 4 months of maturity.

NG spot	*α (t-value)*	*β (t-value)*	*R-sq*	*DW[1] autocorr. (p-value)*	*W-test[2] normality (p-value)*	*BP-test[3] heterosc. (p-value)*
			1997-2003			
Basis 1	-0.010 (-0.116)	**-2.502*** (-4.684)**	0.236	2.348 (0.937)	**0.936 (0.001)**	0.971 (0.324)
Basis 2	0.006 (0.052)	**-1.937*** (-3.912)**	0.390	2.083 (0.653)	**0.810 (0.024)**	**7.680 (0.006)**
Basis 3	-0.017 (-0.232)	**-1.134*** (-6.855)**	0.560	**1.516 (0.021)**	0.986 (0.602)	**7.849 (0.006)**
Basis 4	-0.017 (-0.227)	**-1.242*** (-4.951)**	0.545	1.785 (0.187)	**0.930 (0.001)**	**9.431 (0.002)**
			2004-2015			
Basis 1	0.009 (0.109)	**-2.337*** (-7.483)**	0.282	2.601 (0.999)	**0.876 (0.000)**	2.618 (0.106)
Basis 2	0.010 (0.107)	**-0.484* (-2.181)**	0.033	1.946 (0.374)	**0.867 (0.000)**	1.207 (0.272)
Basis 3	0.003 (0.028)	**-0.998*** (-5.038)**	0.157	1.846 (0.180)	**0.818 (0.000)**	0.390 (0.532)
Basis 4	0.002 (0.015)	**-0.706*** (-4.913)**	0.144	**1.673 (0.028)**	**0.822 (0.000)**	1.221 (0.269)

*.Significance at 10% significance level, *Significance at 5% significance level*
*** Significance at 1% significance level, ***Significance at 0.1% significance level*

[1]H0- hypothesis of the DW test: no autocorrelation in the residuals
[2]H0- hypothesis of the W test: normal distribution of the residuals
[3]H0- hypothesis of the BP test: homoskedasticity or constant variance of the unobserved error

3.5 Markov-Switching model

As the previous analysis does not demonstrate the obvious changes in NG spot prices in two studied periods, a Markov-Switching model with constant transition probabilities is applied in this part to test the influence of fundamental factors (changes in natural gas storage levels and maximum temperature anomaly), commercial and noncommercial traders' positions, and possible endogenous factors. All time series are tested on a unit root to avoid a spurious regression by application of the MSM.

In the period from 1997 to 2016

Eq. (12) with the changes in short and long positions of commercial and noncommercial traders is applied. Table 12 lists the estimates of 2 regimes simulated by using Markov–Switching model. Based on the previous results (Section 3.4) NG futures contract with maturity in one month and basis (NG spot price net NG futures contract with maturity in one month) are applied as explanatory variables.

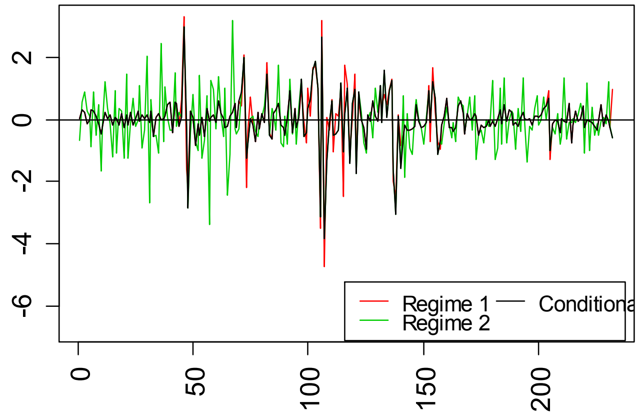
Based on the estimated coefficients and t-statistics (Table 12), the null hypotheses "*slope coefficient equal to zero*" are rejected at the 5% level for the estimate of NG futures contracts with maturity of 1 month in the first regime (spikes and dips of NG spot price) and the short position of commercial traders in the second regime (calm time), and for the estimate of NG basis in the second regimes. The estimate of NG basis cannot reject the null hypothesis at the 5% level in the first regime although its t-statistic is close to it. It means that all other variables do not demonstrate significant influence on NG spot price. The defined regimes and their probabilities are illustrated in Fig.4 and Fig. 2.1 in Appendix 2. It can be seen that the most of the spikes and dips (regime 1) in Fig. 4 coincide with the structural breaks in NG spot price, NG futures price, and basis, which were identified using the ZA test in points 48, 58-59, 68, 73, 84, 107–110, 118, 156 and 206–207 (Tables 4 and 5).

Table 12. Forecasting of NG spot price based on NG futures prices for contracts with 1 month of maturity (or basis) and other variables during the period from 1997 to 2016.

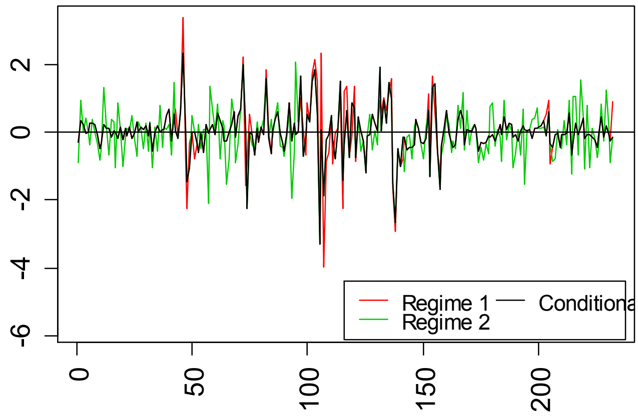
	based on futures 1		based on basis 1	
NG spot	Regime1: *β (t-value)*	Regime2: *β (t-value)*	Regime1: *β(t-value)*	Regime2: *β(t-value)*
Intercept	-0.004 (-0.144)	0.050 (0.274)	-0.011 (-0.391)	0.047 (0.343)
Futures 1/ Basis 1	**0.248*** (4.140)**	0.019 (0.097)	**-0.339. (-1.757)**	**-2.230*** (-4.638)**
Noncommercial position (long)	-	-	-	-
Noncommercial position (short)	-	-	-	-
Commercial position (long)	-	-	-	-
Commercial position (short)	-	**-0.0001*** (9.908)**	-	-
Storage	-	-	-	-
Max temperature anomaly	0.013 (0.484)	0.115 (0.922)	-0.002 (-0.070)	-0.121 (-1.164)
R-sq	0.260	0.123	0.209	0.342
RSE	0.314	1.339	0.312	1.091

*.Significance at 10% significance level, *Significance at 5% significance level*
*** Significance at 1% significance level, ***Significance at 0.1% significance level*

The null hypotheses "*slope coefficient equal to zero*" for maximum temperature anomaly is not rejected at the 5% level, although the previous analysis demonstrates the influence of winter and summer seasons on NG spot price. This result can be caused by the fact that monthly time series of NG spot price is applied, so the influence of temperature anomaly (daily or weekly) is minimized. However, in calm times the t-statistic of maximum temperature anomaly is close to significant. The short position of commercial traders demonstrates its impact on NG spot price in calm time that supports the idea of Shapiro and Pham (2006) that NG market of the US is exposed to decisions of 20 gas producers who concentrate 60% of NG supply.



(a)



(b)

Fig. 4. Fitted lines defined for NG spot price based on futures 1(a) and basis (b) by MSM in 1997–2016

The interesting finding of this analysis is the fact that NG futures price with maturity of one month has positive relationship with NG spot price in the regime 1, where NG spot price demonstrates significant spikes and dips (Fig. 4). At the same time the basis (NG spot price net NG futures contract with maturity in one month) shows a negative relationship with NG spot price in "calm" regime 2 (Fig. 4). These facts can be applied for trading strategies. To test these findings over the most current data, the MSM is applied to the period from 2004 to 2016. The most current data is chosen so, as this period is the most useful for construction of a trading strategy, as this environment is the most close to the current one and thereby it contains more close market, economical, and political features and trends.

In the period from 2004 to 2016

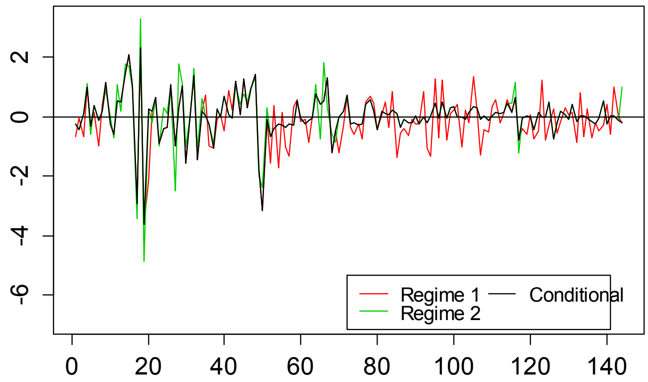
The same tendencies are seen during the period from 2004 to 2016 as during the period 1997-2016 in Table 13. The defined regimes and their probabilities are illustrated in Fig.5 and Fig. 2.2 in Appendix 2 correspondingly. NG futures price have positive and significant relationship with NG spot price in the period of high spikes and dips (regime 2 in Fig. 5), while in calm times the basis demonstrates negative and significant relationship with NG spot price (regime 1 in Fig. 5).

Table 13. Forecasting of NG spot price based on NG futures prices for contracts with 1 month of maturity (or basis) and other variables during the period from 2004 to 2016.

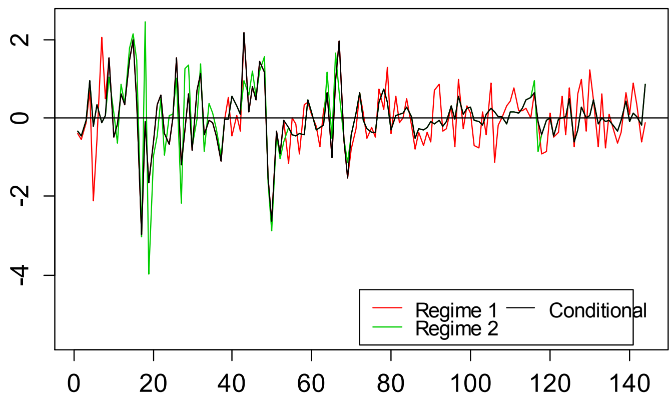
	based on futures 1		based on basis 1	
NG spot	Regime1: *β (t-value)*	Regime2: *β (t-value)*	Regime1: *β(t-value)*	Regime2: *β(t-value)*
Intercept	0.078 (0.348)	-0.052 (-1.374)	0.047 (-0.861)	**-0.041. (-1.797)**
Futures 1/ Basis 1	-0.020 (-0.071)	**0.300*** (3.332)**	**-2.266*** (-4.175)**	**-0.272** (-2.795)**
Noncommercial position (long)	-	-	-	-
Noncommercial position (short)	-	-	-	-
Commercial position (long)	-	-	-	-
Commercial position (short)	-	-	-	-
Storage	-	-	-	-
Max temperature anomaly	0.265 (1.451)	0.003 (0.089)	-0.005 (-0.042)	0.018 (0.537)
R-sq	1.244	0.316	0.367	0.187
RSE	0.145	0.286	1.71	0.332

*.Significance at 10% significance level, *Significance at 5% significance level*
*** Significance at 1% significance level, ***Significance at 0.1% significance level*

It should be pointed out that the estimate of maximum temperature anomaly becomes more close to significant in regime 1 or in calm times. This fact can be associated with increasing influence of weather factor on NG spot price (in winter and summer seasons) due to wider use of NG as a fuel for electricity generation. As was mentioned above, the regime of high spikes and dips coincides with the structural breaks identified by the ZA test in points 19–24, 32, 70, 118–120 (Tables 4 and 5).



(a)



(b)

Fig. 5. Fitted lines defined for NG spot price based on futures1 (a) and basis1 (b) by MSM in 2004–2016

However, in this case the estimate of the short position of commercial traders does not show its significance at the 5% level in any of the regimes (Table 13). In order to find out more about the influence of the short position of commercial traders on NG spot price I apply the MSM to the period between 1998 and 2009 where the most of NG spot spikes and dips occur based on Fig. 5.

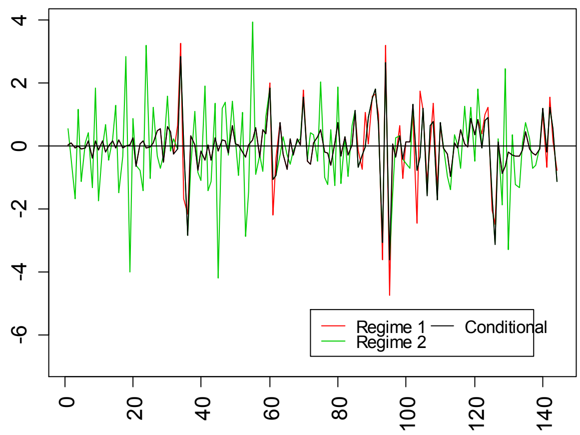
In the period from 1998 to 2009

The period from 1998 to 2009 demonstrates the highest volatility of NG spot price based on Fig. 4 and 5. The defined regimes and their probabilities are illustrated in Fig.5 and Fig. 2.2 in Appendix 2, correspondingly.

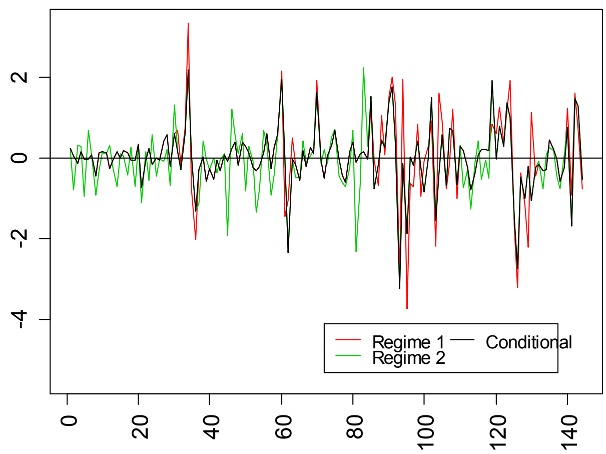
Table 14. Forecasting of NG spot price based on NG futures prices for contracts with 1 month of maturity (or basis) and other variables during the period from 1998 to 2009.

	based on futures 1		based on basis 1	
NG spot	Regime1: *β (t-value)*	Regime2: *β (t-value)*	Regime1: *β(t-value)*	Regime2: *β(t-value)*
Intercept	0.018 (0.535)	0.037 (0.203)	-0.009 (-0.230)	0.092 (0.598)
Futures 1/ Basis 1	**0.283*** (3.568**	0.054 (0.299)	**-0.579** (-2.824)**	**-2.397*** (-4.688)**
Noncommercial position (long)	-	-	-	-
Noncommercial position (short)	-	-	-	-
Commercial position (long)	-	-	-	-
Commercial position (short)	-	**-0.0001*** (11.546)**	-	-
Storage	-	-	-	-
Max temperature anomaly	0.019 (0.585)	0.090 (0.647)	0.011 (0.334)	-0.160 (-1.379)
R-sq	0.414	0.096	0.402	0.342
RSE	0.304	1.364	0.313	1.091
*.Significance at 10% significance level, *Significance at 5% significance level ** Significance at 1% significance level, ***Significance at 0.1% significance level*				

In this case both regimes are characterized by big number of spikes and dips of NG spot price (Fig. 6). The results shown in Table 14 indicate that the short positions of commercial traders cause significant impact on NG spot price in the period from 1998 to 2010 (regime 2) where NG spot and futures prices fluctuate the most in last 20 years.



(a)



(b)

Fig. 6. Fitted lines defined for NG spot price based on futures1 (a) and basis1 (b) by MSM in 1998–2010

However, the estimate of commercial traders' short position is significant at the 5 % level only in the model with NG futures price and it is insignificant in another model (with NG basis). The estimate of maximum temperature anomaly is again close to significant based on its t-statistic in regime 2. These findings support partly the hypothesis of Shapiro and

Pham (2006) that the combination of industry concentration, temperature fluctuations and inelastic demand could result in significant NG price fluctuations. However, several spikes and dips (November–December 2000, October–November 2005 and January–March 2006) can be caused by other exogenous factors, as these fluctuations are in regime 1 where the estimate of commercial traders' short position does not show its significant at the 5 % level.

All in all, this analysis shows that the changes in positions of noncommercial traders (financialization) do not play a significance role in the fluctuations of NG spot price. It seems that other exogenous factors (macroeconomic cycles, shale gas production, temperature fluctuations, hurricanes) and commercial traders have higher power over NG spot price and its huge fluctuations, especially in conditions of inelastic demand for NG and NG market supply concentration.

3.6 Trading strategy and its backtesting

The analysis in the previous sections demonstrates that NG spot price has negative and significant relationship with NG basis in calm times (Section 3.4–3.5). This observation can be applied to develop a trading strategy.

As the difference in NG spot price has a negative and significant relationship with NG basis (NG spot price net NG futures price with maturity in one month), the trading strategy can be formulated using the following long and short entry signals:

1. If NG basis is negative at this month, then the NG futures contract (contracts) with one-month maturity should be bought (long position). In the next month, this NG futures contract (contracts) should be closed just before expiration.
2. If NG basis is positive at this month, then the NG futures contract (contracts) with one-month maturity should be sold (short position). In next month this NG futures contract (contracts) should be closed just before expiration, or NG should be bought at NG spot price and be delivered.

To test this trading strategy a backtest is employed using NG spot prices and NG futures contracts prices during the period from January 2010 to October 2016. This period is chosen in accordance with the results in Table 13 and Fig. 5, where NG spot price has negative and significant relationship with NG basis only in calm time that corresponds mainly to the period from 2010 to 2016. This period is chosen also based on the fact that it represents the most current market conditions and trends and, therefore, it can provide higher reliability of the backtest results to the live trading. Due to the limited number of observations, the backtest and study periods coincide, except the period from June 2016 to October 2016.

The trading of NG futures contracts is accompanied by several fees and costs. This analysis assumes brokerage commission of 2.5$ per transaction (CME 2016). The holding costs is assumed 2200$ for initial margin and 2420$ (110% of initial margin for non-member) for maintenance margin based on the current policy of the Chicago Mercantile Exchange (CME) for CME NYMEX Natural Gas Futures (Quandl 2016). This analysis assumes clearing fee of 1.45$ per contract for non-member (CME 2016). The mandatory fee to National Futures Association is assumed 0.01$ per contract (CME 2016). The firth cost is data fee that is charged by the CME for providing data about traded contracts. The data fee to the CME is assumed 15$ per month for non-professional traders (CME 2016). The above-stated costs are current, while the backtest is conducted over the period from 2010 to 2016. This fact somewhat limits the reliability of the backtest results.

For backtesting, the loan interest rate and deposit interest rate are assumed equal to 2.5% and 1% above Libor 0.5%, correspondingly. These rates are assumed constant over the whole tested period (2010–2016). In this strategy it is assumed also that the initial and maintenance margins are taken as loans at 3% per annum before these amounts are earned by profits from the trades of NG futures contracts. The next assumption of this strategy is that all the profits above the initial and maintenance margins and the other above-listed costs can be deposited to earn 3% per annum. Although the maintenance and initial margins are usually earning an interest of 3-month Libor (assumed 0.5%) (CME, 2016), in this analysis only the maintenance margin earns an interest, as it is assumed the initial margin is subject to daily adjustments.

At the first attempt the backtesting demonstrates that the developed strategy generates a loss of about 29000$ due to fail of trading strategy in the period close to winter each year (Table 7). It is associated with high dependence of NG spot price on the weather conditions due to inelastic demand for natural gas. The previous analyses show also several structural breaks or a lot of dips and spikes (high volatility) in these periods.

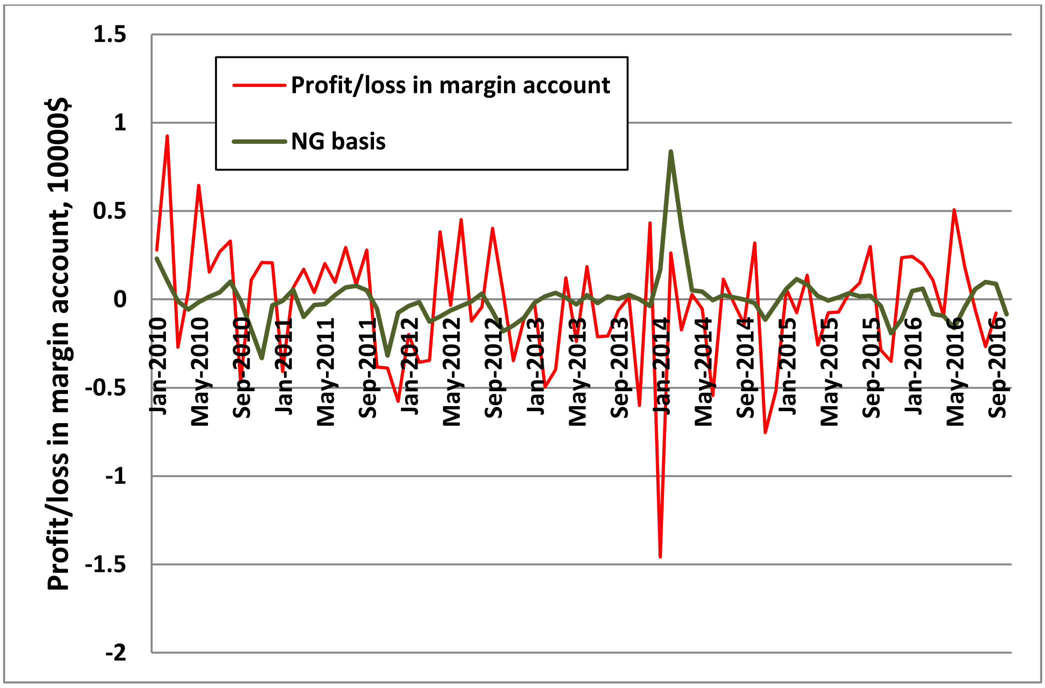


Fig. 7. Backtesting for trading strategy using NG futures in 2010-2016

In order to avoid this drawback of the trading strategy need to be adjusted and the third rule is suggested:

3. In November, December and January no long or short positions in NG futures contract should be taken or each long and short positions in NG futures contracts should be protected by put option and call option correspondingly.

The trading strategy based on the three above listed rules (no position in NG futures in November, December, and January) allows generating a profit of about 23600$ without any initial investments. This strategy is described in details in Table 3 (Appendix 3) and illustrated in Fig. 8. This positive result of a backtest does not guarantee that the suggested strategy allow generating a positive return in the future, but it provides some level of confidence that it can be used for trading in a real time.

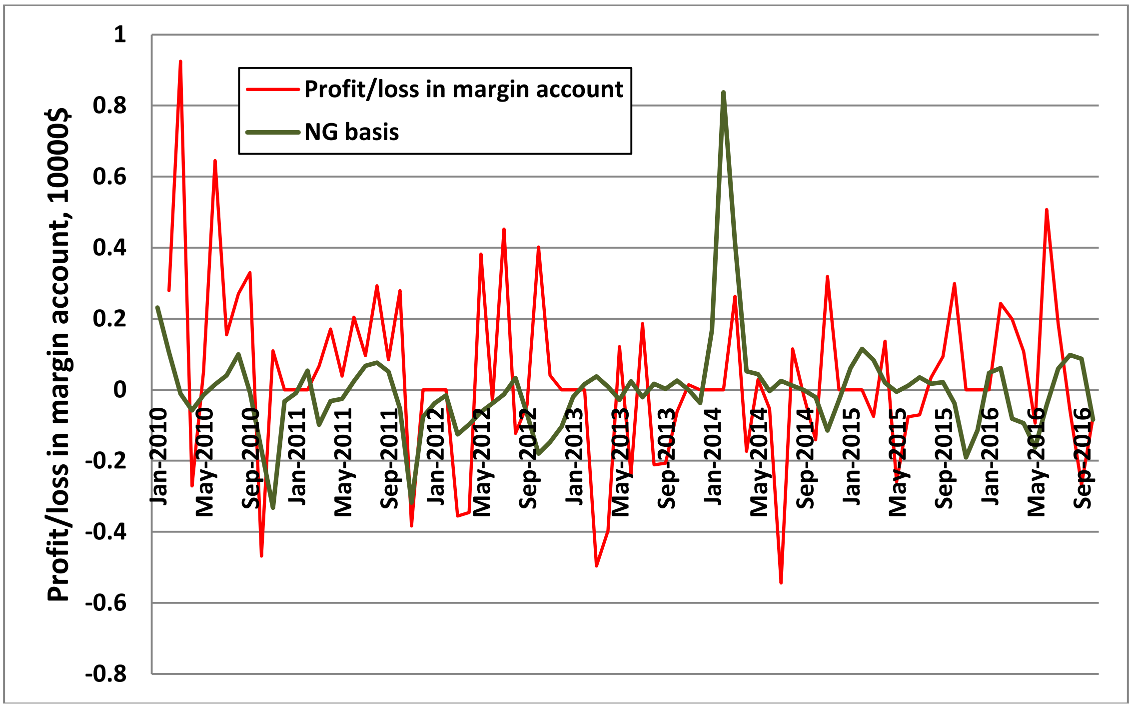


Fig. 8. Backtesting for trading strategy using NG futures in 2010-2016 (without positions in November, December and January)

.

4 DISCUSSION AND CONCLUSIONS

Several econometric techniques are employed in this thesis to recognize the influence of financialization on NG spot prices in the US. As many researchers and economists have pointed out that NG futures contract provides the best forecast for NG spot price, and that it is one of the popular contracts among the speculators, the interaction between NG spot and NG futures with maturities of 1, 2, 3, 4 months is the main focus of the analysis. For this purpose, the monthly time series of NG spot price and futures contracts are collected from the EIA database for the period between January 1997 (the earliest available information) to May 2016 (the upper limit is determined by a desire to test the developed trading strategy over several most recent observations).

First, the descriptive statistics analysis of NG spot and futures prices is done. It demonstrates that NG spot and futures prices become more volatile and are tended to be more unpredictable in the period after 2004 compared to the period from 1997 to 2003. Higher volatility of NG spot and futures prices can be attributed to different macroeconomic shocks, weather disasters, temperature shocks, concentrated market structure, financial crises, large-scale emergence of shale gas, and limited LNG export capabilities.

The ADF test and the KPSS test show that all time series being examined do not contain a unit root or they become stationary after one differentiation. However, the ZA test identifies that most of the utilized time series are trend non-stationary in the period before 2004. This test also demonstrates several structural breaks for intercept and trend in both periods. The structural breaks of NG spot and futures prices happened in 2000-2001 and 2005-2006 and NG bases - in 2001, 2003, 2006, 2009 and 2014. This result coincides with the structural breaks detected by Ghoddusi (2016), and can be attributed to the shale gas take-off since 2001. The second reason of the structural breaks can be unpredictable weather anomalies in the conditions of inelastic demand for NG, as the most of structural breaks occur during the period from October to March.

As non-stationary is detected from the unit root tests, the cointegration analysis is performed next. The results confirm the existence of a cointegration relationship among the variables under analysis. To support this conclusion, the Johansen procedure was applied to provide the cointegration test in presence of the structural breaks. It also confirms that the applied time series are cointegrated (in long-term relationship) at the 5% significance level, hence, they can be used for forecasting in both periods. However, the short-term analysis by Error Correction model shows that the shocks have usually persistent influence on NG spot-futures long-term relationship in both periods (except NG spot- NG futures with maturity in 3 months during the period from 2004 to 2016). This mostly results in attaining of new equilibrium of NG spot-futures interaction. It supports the previous finding of several structural breaks in the studied time series.

As the next step, the seasonality analysis is done to analyze the influence of temperature critical seasons on NG spot price. The regression with season dummies demonstrated that the estimates of winter and summer dummies became significantly different from zero at the 10% level in the period after 2004 compared with the 1997–2004 period, when only the estimate of winter season dummy shows its significance at the 5% level. This result confirms the current trend in the US energy structure, that is the appearance of many gas-fired electrical power units to meet the demand in the cooling or conditioning especially during the summer seasons.

In addition, the forecast models of NG spot price based on NG futures contracts with different maturities and bases are simulated to take account of several structural breaks in their long-term relationships. This analysis demonstrates the forecasting power of NG futures with maturity of one month and bases (NG spot price net NG futures contracts with maturity of 1, 2, 3 and 4 months) in both periods. However, only NG futures contract with one-month maturity and the related basis demonstrate the highest R-sq rates and no problems in residuals.

To test the stated hypothesis of influence of noncommercial traders on NG spot-futures price dynamics, the forecasting analysis using a Markov-Switching model (MSM) is implemented next. The application of fundamental factors (changes in natural gas storage level and maximum temperature anomaly) and commercial and noncommercial traders' positions in addition to NG futures contract with maturity of one month and the related basis based as explanatory variables supports previously detected dynamics. NG spot price has a positive and significant relationship with NG futures contract in maturity of one month and negative and significant relationship with basis (NG spot price net NG futures contract in maturity in one month). All other applied explanatory variables (except short position of commercial traders) do not demonstrate significant impact on NG spot price during the period from 1997 to 2016. The regimes identified by the MSW for NG spot price regimes shows that the regime of high spikes and dips coincides the structural breaks resulted in the analysis where the ZA test is employed.

The analysis using MSM shows that the short position of commercial traders has negative and significant impact on NG spot price, and in addition, that maximum temperature anomaly has close to significant influence on NG spot price in relatively calm regime. To support these findings, two separate analyses of NG spot price using the same variables are presented for the periods of 1998–2010 (time of high spikes and dips in NG spot price) and 2004–2016 (relatively calm years). The overlap of the periods is associated with the limited number of observations and the limit for minimum number observations in order to run the MSW. The results of the first test (1998–2010) show that NG spot price is exposed to a negative influence of the short position of commercial traders and maximum temperature anomaly (close to significant). However, this tendency does not exist in the period from 2004 to 2016 when NG spot price faces only partly (from 2004 to 2010) such spikes and dips that occur during the period 1998–2010. The influence of commercial traders on NG spot price during the period 1998–2010 can be related to the hypothesis suggested previously by Shapiro and Pham (2006) that at that time NG spot price was defined in conditions of inelastic demand, several weather disasters, concentrated supply (20 gas producers controlled 60% of NG market supply) and unregulated price. However, this hypothesis was supported only by qualitative analysis of Shapiro and Pham (2006), whereas in this thesis, it is quantitatively supported to same extent. The analysis demonstrated that there are several spikes and dips of NG spot price in the period from 1998 to 2010 that cannot be attributed to the applied variables or factors.

All in all, *in accordance with the presented results, the above-stated hypothesis of the influence of noncommercial traders on NG spot price cannot be accepted.* At the same several tendencies of NG spot price dynamics are identified. NG spot price can be forecasted by price of NG futures contact with one-month maturity in the periods of high spikes and dips. NG basis (NG spot price net NG futures contract in maturity in one month) can be successfully applied as a forecaster during calm periods. These findings can be used by market participants to benefit from this dynamics.

The trading strategy based on NG one-month futures contract is suggested. It is based on two entry signals (long position in futures contract in case of negative NG basis, and short position in futures contract in case of positive basis) and one rule (no positions in November, December, and January) identified during the backtesting of the trading strategy. The backtest showed a profit of about 23600$ if this trading strategy is employed during the period from 2010 to 2016.

5. SUMMARY

The significant increase of noncommercial traders in different commodity markets raises a question whether it has impacted on commodity prices and whether it has been one of the reasons for several spikes and dips seen in 2000s. In this thesis, I decided to examine this phenomena in the U.S. natural gas market, which is one of the most liquid natural gas markets. As natural gas futures prices are assumed to be a reliable forecaster of natural gas spot price, and the related contracts are widely used by market participants to get exposure to NG price risk, the dynamics of natural gas spot-futures prices during the period from January 1997 to May 2016 was the main subject of interest in this thesis.

Several econometric techniques were applied to test the hypothesis that financialization influences on the interaction between natural gas spot and futures prices. Each of the chosen analyses were applied separately for two subperiods (1997–2003 and 2004–2016) to identify the changes in trends and dynamics. The descriptive statistics shows that NG spot and futures prices became more volatile in the later period, and their dynamics were subjected to several structural breaks. These facts may result in tendency that the shocks have had a persistent influence on NG spot-futures long-term relationship since 2000. The seasonality analysis identified that summer period (in addition to winter period) became to play a crucial role in NG spot price, which could be associated with an increase of NG consumption by electricity generators in the US.

It was impossible to accept or confirm the above-stated hypothesis that financialization influenced on NG spot-futures price mechanism and, therefore, possibly on NG spot price. The linear forecasting models demonstrated that the relationship between natural gas spot and prices of futures contracts with maturities of 1, 2, 3 and 4 months are similar in both studied periods being examined. However, the Markov-Switching model with constant transition probabilities showed an interesting tendency. NG spot price was partly exposed to the short position of commercial traders during the period from 1998 to 2010 when NG price faced its highest spikes and dips. Both forecasting methods also allowed identifying some interesting patterns and constructing a trading strategy based on them for the current time. In calm times, the natural gas basis (NG spot price net price of NG futures with one-month maturity) can be used as a good forecaster, while in the periods of high spikes and dips, the natural gas futures with one-month maturity are more preferable. However, the trading strategy need to be changed for the winter period due to unpredictable dynamic of NG spot price associated with inelastic demand for natural gas.

The results presented in this thesis can be used by policy makers, different groups of traders (hedgers, speculators etc), economists and researchers, as they demonstrate new insights about tendencies in natural gas market of the US. As natural gas as a fuel is playing crucial role in the world economy, the new knowledge about natural gas trading pattern and its exposure to different factors is valuable. However, several limitations of this study should be pointed out. The first of them is associated with the power and reliability of the applied tests. As it is known, some of them have low power in the presence of structural breaks and non-stationarity. The second limitation is associated with the assumptions applied in this thesis (NG futures price Granger cause NG spot price, constant costs and interest rates in the backtesting). Some of these assumptions can cause errors and, therefore, lead to wrong conclusions. The third limitation is associated with the sampling of the chosen data range, as monthly data can be a poor source in identification of trader influence on NG spot price (it may be the case, that the trader influence can be captured only by using more frequent sample data). The fourth limitation is related to the reliability and completeness of the applied parameters. Several analyses demonstrate that the chosen variables are not able to predict natural gas spot price, which raises a question whether some important parameters (macroeconomic trends, influence of different political and economical shocks) are possibly omitted or whether the prices are predictable at all.

The above-stated limitations invoke new topics for further research: it could be interesting to apply ARCH and GARCH to identify the patterns in natural gas market since 2001-2003. The interaction between natural gas spot and futures prices can be examined by using daily or weekly time series to check the robustness of the results presented in this thesis. As LNG trade is developing and the US is an important supplier of it, it could interesting to make an analysis of the impact of this LNG trade on natural gas market of the US.

REFERENCES

Alizadeh, A.H., Tamvakis, M. Market conditions, trader types and price-volatility relation in energy futures markets. *Energy Economics*, vol. 56, 2016, pp. 134–149.

American Petroleum Institute (API). Understanding Natural Gas Market, 2014. From: http://www.api.org/~/media/files/oil-and-natural-gas/natural-gas-primer/understanding-natural-gas-markets-primer-high.pdf, retrieved: 3.04.2016

Asche, F., Gjolberg, O., and T. Volker, 2003. Price relationship in the petroleum market: an analysis of crude oil and refined product prices. *Energy Economics*, vol. 25, 2003, pp. 289–301.

Asche, F., Misund, B., Sikveland, M. The relationship between spot and contract gas prices in Europe. *Energy Economics*, vol. 38, 2013, pp. 212–217.

Ashton, M. Skewness and kurtosis: what commodities investors need to know. Seeking Alpha, 2011. From: http://seekingalpha.com/article/269053-skewness-and-kurtosis-what-commodities-investors-need-to-know, retrieved: 19.11.2016.

Basher, S.A., Haug, A.A., Sadorsky, P. The impact of oil shocks on exchange rates: a Markov-switching approach. *Energy Economics*, vol. 54, 2016, pp. 11–23.

Brown, S., Mine, Y. What Drives Natural Gas Prices? *The Energy Journal*, vol. 29, no.2, 2008, pp. 45–60.

Cheng, J.-H., Xiong, W. Why do hedgers trade so much? *Journal of Legal Studies*, vol. 43, 2014, pp. 183–201.

Chevallier, J. The impact of nonlinearities for carbon markets analysis. *International Economics*, vol. 2, 2011, pp.131–150.

CME, 2016. . From: http://www.cmegroup.com, retrieved: 3.04.2016.

Coppola, A. Forecasting oil price movements: exploiting the information in the futures market, *Journal of Futures Market*, 2007, vol.28, no.1. From: http://onlinelibrary.wiley.com/doi/10.1002/fut.20277/pdf, retrieved: 3.04.2016.

D’Ecclesia, R.L., Magrini, E., Montalbano, P. and Triulzi, U. Understanding recent oil price dynamics: a novel empirical approach. *Energy Economics*, vol. 46, 2014, pp. 11–17.

EIA, 2016. From: http://www.eia.gov/dnav/ng/ng_pri_fut_s1_m.htm, retrieved: 19.07.2016

Eiker, F. Limit theorems for regression with unequal and dependent errors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* 1967, pp. 59–82.

Fattouh, B., Kilian, L. and Mahadeva, L. The role of speculation in oil markets: What have we learned so far? *CERP Discussion Paper*, no. DP8916, 2012.

Federal Reserve Bank of San Francisco. Do oil futures help predict future oil prices, 2005, NTNU. From: http://www.frbsf.org/economic-research/files/el2005-38.pdf, retrieved: 19.11.2016.

Georgia, C. The nexus between natural gas spot and futures prices. Evidence from the NYMEX market. Masters' thesis. International Hellenic University, 2012

Hamouz, M., Kittler, J., Kamarainen, J.-K., Paalanen, P., Kälviäinen, H., Matas, J., Feature-based Affine-invariant Localization of Faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 27, No. 9, September 2005, pp. 1490–1495.

Hartley, P.R., Medlock, K.B., Rosthal, J.E. The Relationship of Natural Gas to Oil Prices. *The Energy Journal*, vol. 29, no.3, 2008, pp. 47–65.

Herber, J.H. Trading volume, maturity and natural gas futures price volatility. *Energy Economics*, vol. 17, no.4, 1995, pp. 293–299.

Huber, P.J. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,*1967*,* pp. 221–233.

Ghoddusi, H. Integration of physical and futures prices in the US natural gas market. *Energy Economics*, vol. 56, 2016, pp. 229–238.

Gonzalez, R.C., Woods, R.E., Digital Image Processing, 3rd edition, Prentice-Hall, Inc., USA, 2002.

Jang, J.-S. R., Sun C.-T., Mizutani E.: Neuro-Fuzzy and Soft Computing, 1st edition, Prentice-Hall, Inc., USA, 1997.

Kemp, J. Don't bank on return of backwardation, Reuturs, 2010. From: http://blogs.reuters.com/great-debate/2010/05/20/dont-bank-on-return-of-backwardation/, retrieved: 3.04.2016.

Kolodziej, M., Kaufmann, R.K., Kulatilaka, N., Bicchetti, D., Maystre, N. Crude Oil: commodity or financial asset? *Energy Economics*, vol. 46, 2014, pp. 216–223.

Liu, Y., Handroos, H., Alkkiomäki, O., Kyrki, V., Kälviäinen, H., Development of a Hybrid Position/Force Controlled Hydraulic Parallel Robot for Impact Treatment, *Proceedings of the 7th International Conference on Machine Automation (ICMA2008)*, September 24-26, Awaji, Japan, 2008, pp. 41–47.

Lissandrin, M., Daly, D. and Sornette, D. Statistical testing of DeMark technical indicators on commodity futures, *Energy Economics*, vol. 51, 2015, pp. 30–49.

Lubnau, T. and Todorova, N. Trading on mean-reversion in energy futures markets, *International Journal of Forecasting*, vol. 51, 2015, pp. 312–319.

Luetkepohl, H., Saikkonen, P. and Trenkler, C. Testing for the cointegration rank of a VAR process with level shift at unknown time. *Econometrica*, vol. 72(2), 2004, pp. 647–662.

Lombardi, M.J., van Robays, I. Do financial investors destabilize the oil price? Working Paper Series 1346. ECB, 2011. From: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1346.pdf?eaea02dc58fdd3aba2f9a0263a7d534a, retrieved: 19.11.2016.

Madaleno, M., Pinho, C. Wavelet dynamics for oil-stock world interactions. *Energy Economics*, vol. 45, 2014, pp. 120–133.

Martinez, B., Torro, H. European natural gas seasonal effects on futures hedging. *Swiss Finance Institute Research* Paper no.15-56, ETH Zurich.

Mason, C.F. and Wilmot, N.A. Jump processes in natural gas markets. *Energy Economics*, vol. 46, 2014, pp. 569–579.

Mensi, W., Hammoudeh, S., Seong-Min, Y. How do OPEC news and structural breaks impact returns and volatility in crude oil markets? Further evidence from a long memory process. *Energy Economics*, vol. 42, 2014, pp. 343–354.

Naryan, P.K.., Liu, R. A unit root model for trending time-series energy variables. *Energy Economics*, vol. 50, 2015, pp. 391–402.

National Centers for Environmental Information, 2016. From: https://www.ncdc.noaa.gov/, retrieved: 25.11.2016.

Neumann, A. Linking Natural Gas Markets – Is LNG doing its job? *DIW Berlin Discussion Paper*, n.822, 2008.

Nick, S. and Thoenes, S. What drives natural gas prices? — A structural VAR approach. *Energy Economics*, vol. 45, 2014, pp. 517–527.

Nicolau, M., Palomba, G. And Traini, I. Are futures prices influenced by spot prices or vice-versa? An analysis of crude oil, natural gas and gold markets. Working Paper, Universita Politecnica delle Marche, 2013. From: http://econpapers.repec.org/paper/ancwpaper/394.htm, retrieved: 19.11.2016.

Nielsen, H.B. Non-stationary time series, cointegration and spurious regression, Lectures of Econometrics 2. 2005. From: http://www.econ.ku.dk/metrics/econometrics2, retrieved: 3.04.2016.

Peilong, S., Rui, G. Empirical research on the price of oil and refined oil based on the cointegration relationship. Shanxi University of Finance and Economics, Chine. 2010. From: http://www.seiofbluemountain.com, retrieved: 3.04.2016.

Pindyck, R.S. Long-run evolution of energy prices, Massachusetts Institute of Technology, 1998.

Perron, P. Dealing with structural breaks. Draft for Handbook of Econometrics. Boston University, 2005. From: http://people.bu.edu/perron/papers/dealing.pdf, retrieved: 19.11.2016.

Pesaran, M.H., and Timmermann, A. Market timing and return prediction under model instability. *Journal of Empirical Finance*, vol. 9, 2002, pp. 495–510.

Quandl, 2016. Available from: https://www.quandl.com. retrieved: 28.11.2016

Reichsfeld, D.A., Roache, S.K. (2011). Do commodity futures help forecast spot prices? *IMF* Working Paper. From: http://core.ac.uk/download/pdf/6431324.pdf, retrieved: 3.04.2016.

Ritz, R.A. Price discrimination and limits to arbitrage: an analysis of global LNG markets. *Energy Economics*, vol. 45, 2015, pp. 324–332.

Reichsfeld, D.A. and S.K. Roache. Do commodity futures help forecast spot prices? *IMF Working Paper*, 2011. From: https://www.imf.org/external/pubs/ft/wp/2011/wp11254.pdf, retrieved: 19.11.2016.

Shapiro, R. and Palm, N. An analysis of spot and futures price for natural gas: the role of economic fundamentals, market structure, speculation and manipulation. SONECOM, 2006. From: http://www.citizensutilityboard.org/pdfs/blog/NatLegalPolicyCenterGasManip.pdf, retrieved: 15.10.2016

Siliverstovs, B., L'Hégaret, G., Neumann, A., Hirschhausen, C. International market integration for natural gas? A cointegration analysis of prices in Europe, North America and Japan. *Energy Economics*, vol. 27, no.4, 2005, pp. 603–615.

Smyth, R., Narayan, P.K. Applied econometrics and implications for energy economic research. *Energy Economics*, vol. 50, 2015, pp. 351–358.

Timmermann, A. and M.H. Pesaran. Market timing and return prediction under model instability. FMG Discussion Paper 412, Financial Market Group, 2002. From: https://ideas.repec.org/p/fmg/fmgdps/dp412.html, retrieved: 19.11.2016.

Turner, A., Farrimond, J., Hill, J. The oil trading markets. 2003-2010. Analysis of market behavior and possible policy responses. The Oxford Institute for Energy Studies WPM 42. 2011.

Vansteenkiste, I. What is driving oil future prices? Fundamental versus speculation. *EU Central Bank*. From: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1371.pdf, retrieved: 3.04.2016

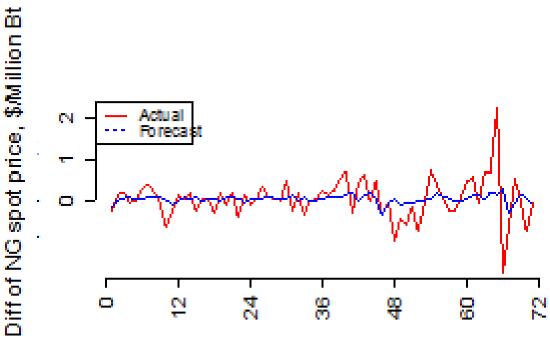
Waheed, M., Alam, T. and S. Ghauri. Structural breaks and unit root: evidence from Pakistani macroeconomic time series, MPRA Paper, State Bank of Pakistan, 2006. From: https://mpra.ub.uni-muenchen.de/1797/1/MPRA_paper_1797.pdf, retrieved: 19.11.2016.

White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, vol. 48, no.4, 1980, pp. 817–838.

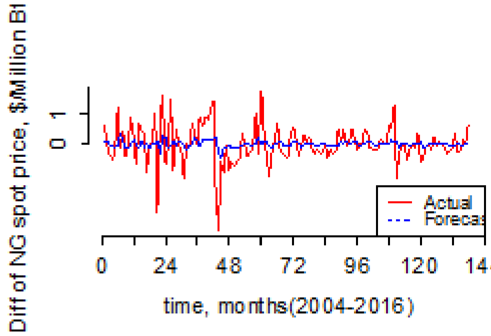
Wu, X. Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, vol. 29, 2007, pp. 46–63.

Zeitlberger, A. and Brauneis, A. Modelling carbon spot and futures price returns with GARCH and Markov switching GARCH model. *Central European Journal of Operations Research*, vol. 24(1), 2016, pp. 149–176.

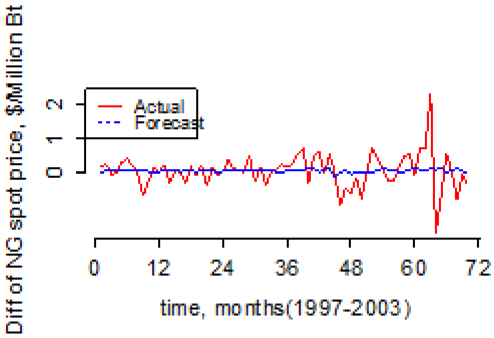
APPENDIX 1. Forecasts of NG spot price using futures and bases



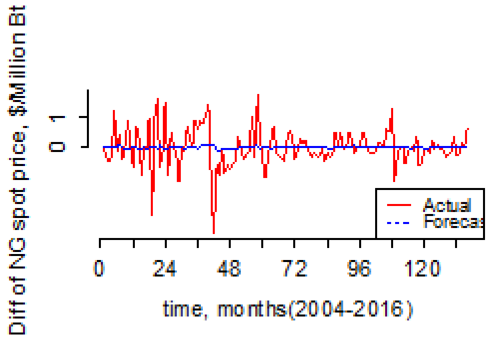
Model for NG spot price based on NG futures contract 1 month (1997-2003)



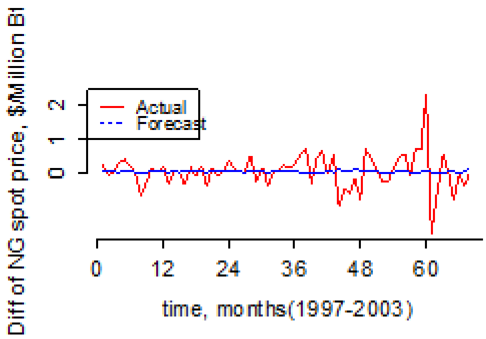
Model for the WIT spot price based on NG futures contract 1 month (2004-2015)



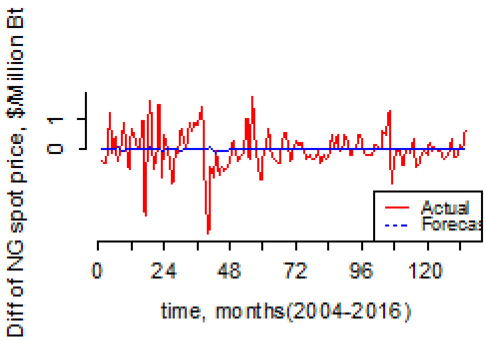
Model for the NG spot price based on NG futures contract 2 months (1997-2003)



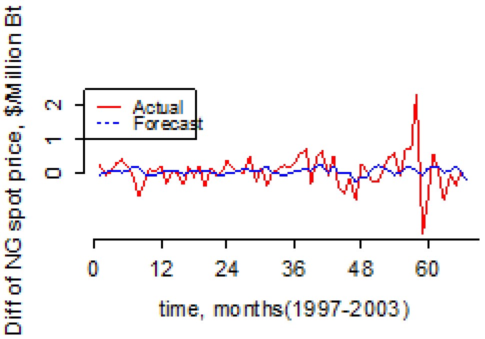
Model for the WIT spot price based on NG futures contract 2 months (2004-2015)



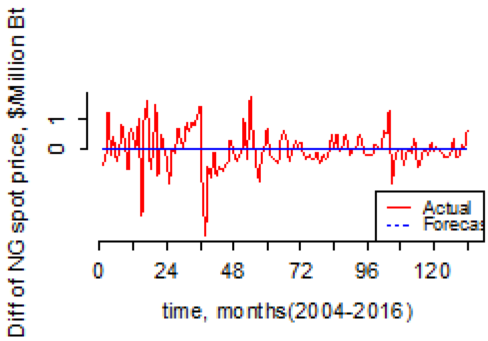
Model for the NG spot price based on NG futures contract 3 months (1997-2003)



Model for the WIT spot price based on NG futures contract 3 months (2004-2015)



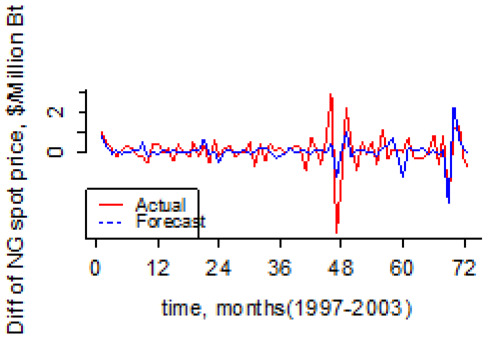
Model for the NG spot price based on NG futures contract 4 months (1997-2003)



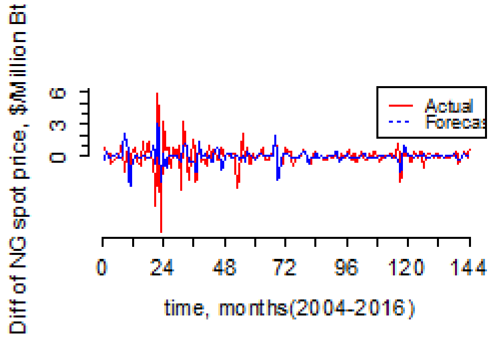
Model for the WIT spot price based on NG futures contract 4 months (2004-2015)

Fig. 1.1. Fitted lines defined by NG futures prices for contracts with 1, 2, 3 and 4 months of maturity.

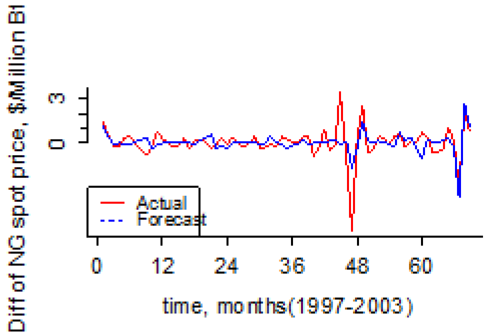
APPENDIX 1. (continues)



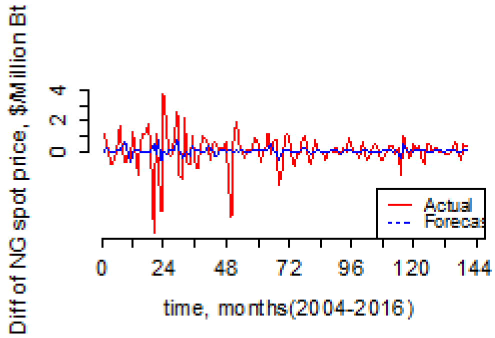
Model for NG spot price based on basis for the NG futures contract 1 month (1997-2003)



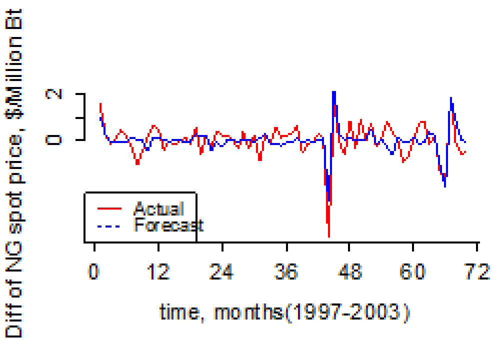
Model for NG spot price based on basis for the NG futures contract 1 month (2004-2015)



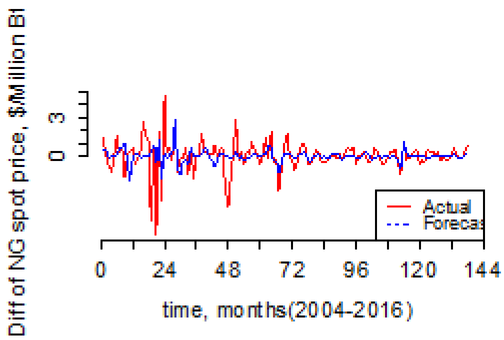
Model for WIT spot price based on basis for the NG futures contract 2 months (1997-2003)



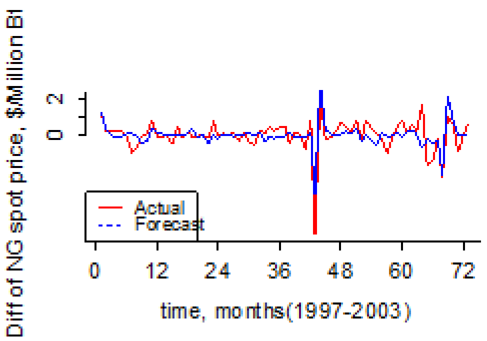
Model for NG spot price based on basis for the NG futures contract 2 months (2004-2015)



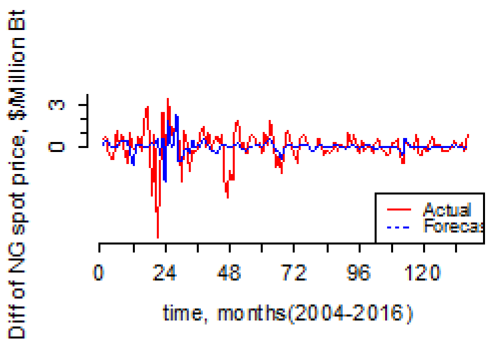
Model for WIT spot price based on basis for the NG futures contract 3 months (1997-2003)



Model for WIT spot price based on basis for the NG futures contract 3 months (2004-2015)



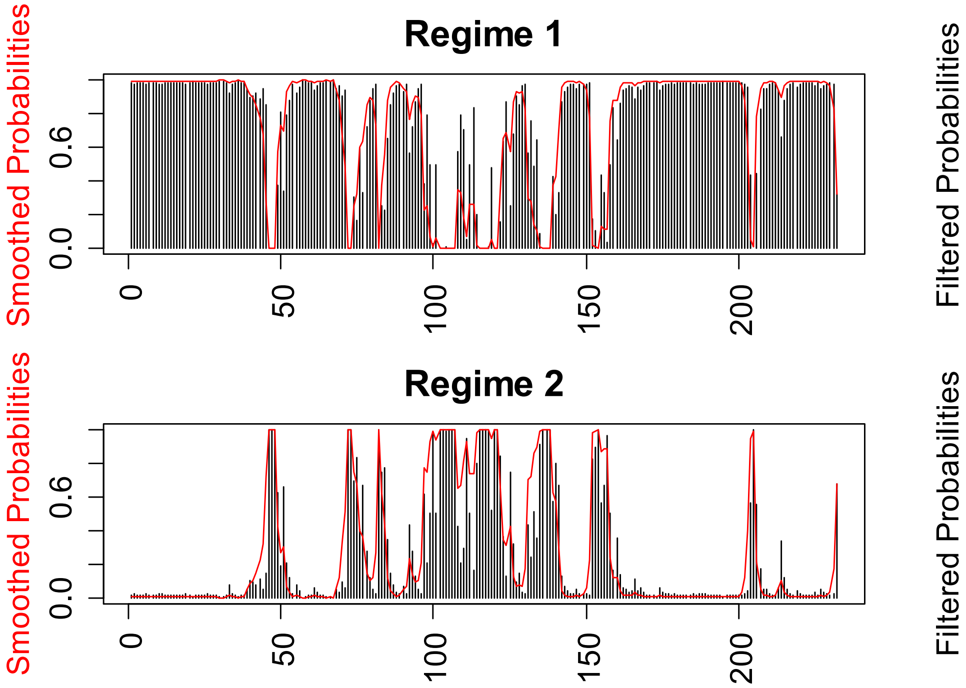
Model for WIT spot price based on basis for the NG futures contract 4 months (1997-2003)



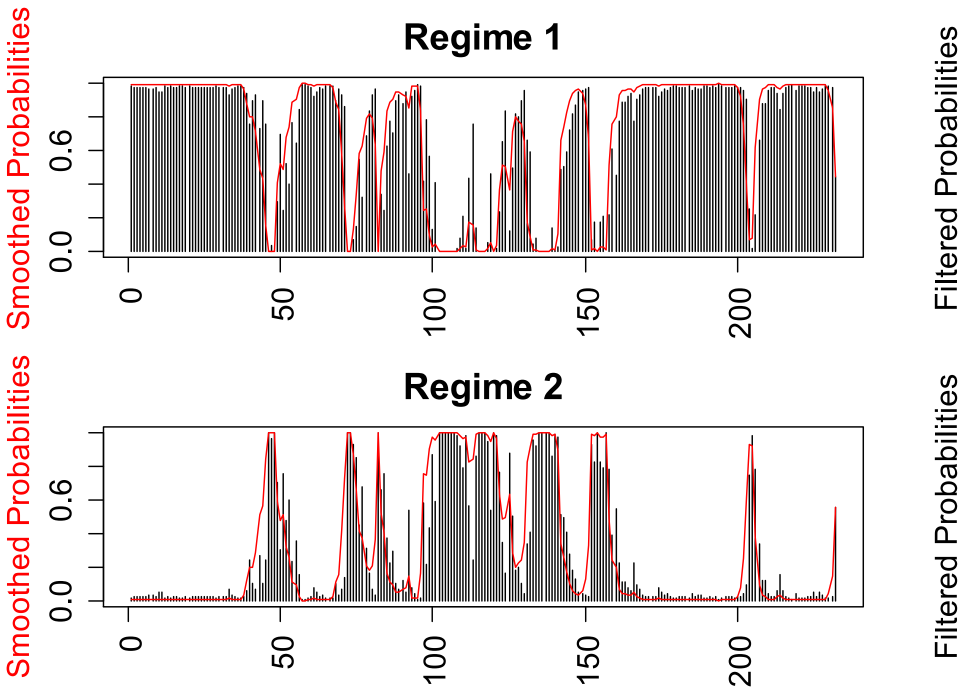
Model for WIT spot price based on basis for the NG futures contract 4 months (2004-2015)

Fig. 1.2. Fitted lines defined by NG bases for futures contracts with 1, 2, 3 and 4 months of maturity.

APPENDIX 2. Forecasting models by Markow-Switching Model



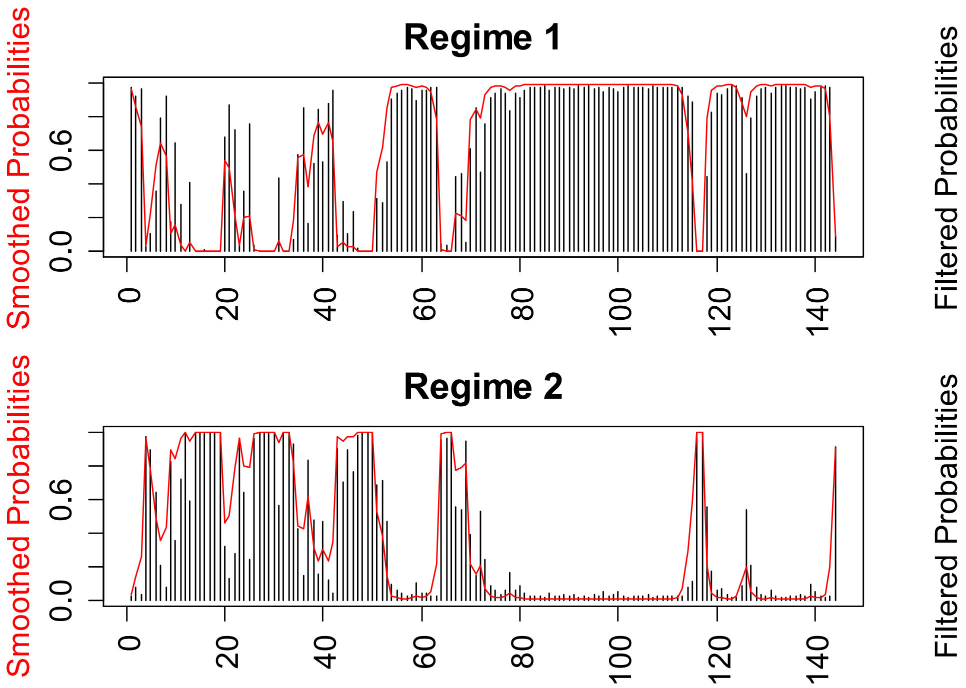
(a)



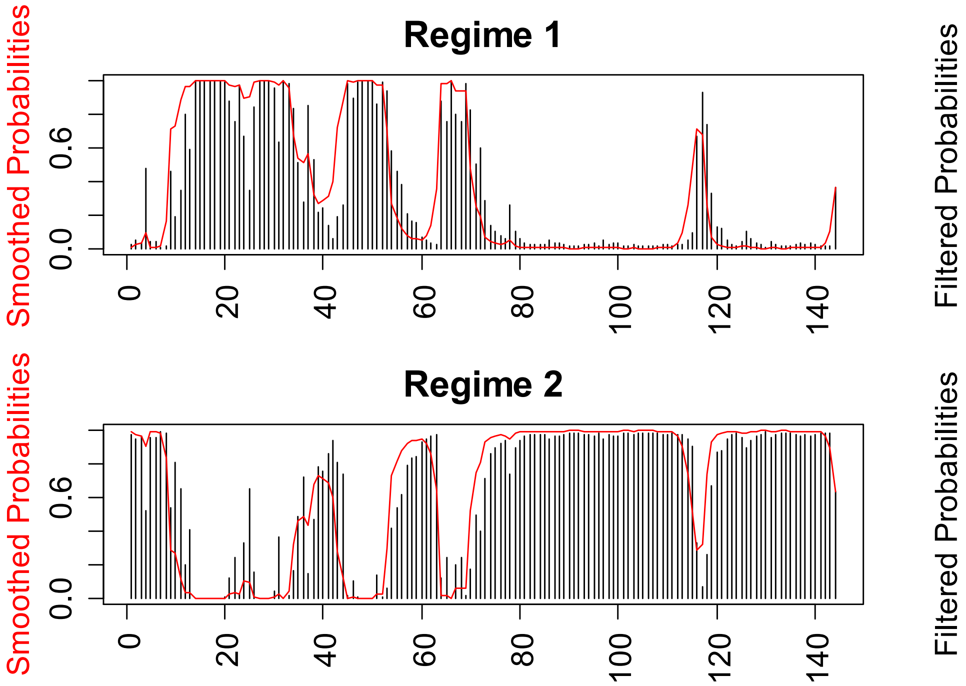
(b)

Fig. 2.1. Filtered probabilities for NG spot price in two regimes (using NG futures price (a)/NG basis (b) and other time series as explanatory variables) in 1997-2016.

APPENDIX 2. (continues)



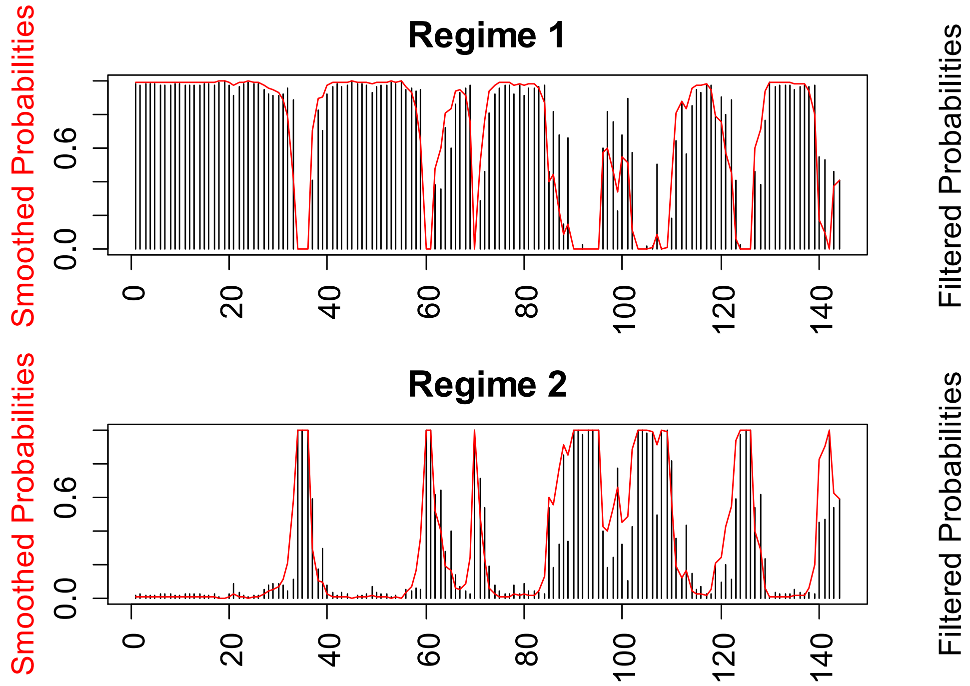
(a)



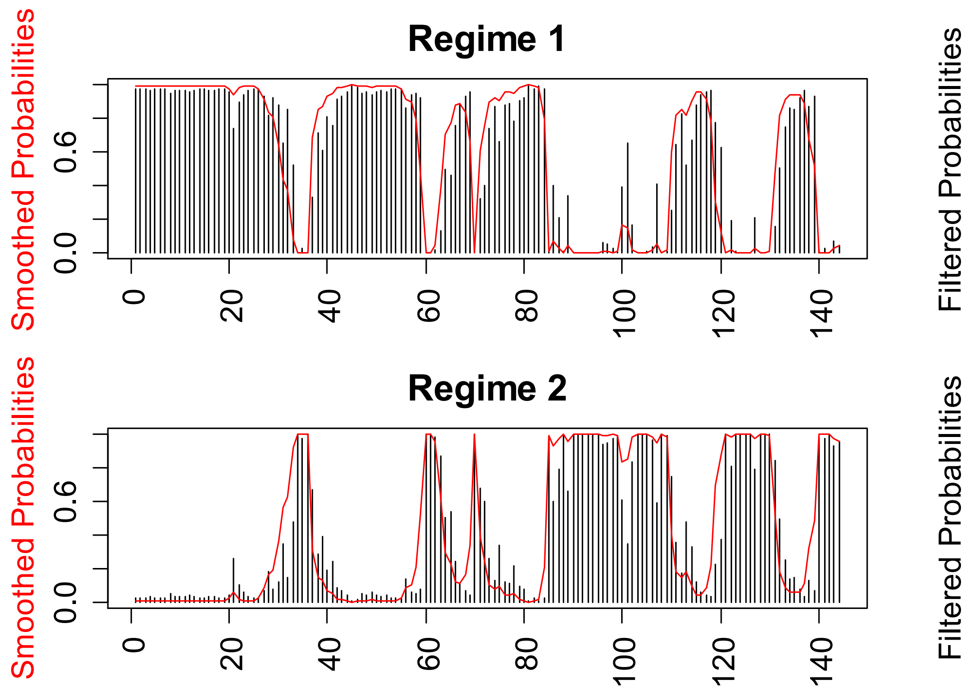
(b)

Fig. 2.2. Filtered probabilities for NG spot price in two regimes (using NG basis and other time series as explanatory variables) in 2004-2016.

APPENDIX 2. (continues)



(a)



(b)

Fig. 2.3. Filtered probabilities for NG spot price in two regimes (using NG futures price (a)/NG basis (b) and other time series as explanatory variables) in 1998-2010.

APPENDIX 3. Trading strategy

Table 3. Backtesting for trading strategy in 2010-2016.

Date	NG spot $ per MBtu	NG Futures, $ per MBtu	Basis, $ per MBtu	Long/ short position in futures (10000 MBtu each)	Cash Flow, $	Margin account, $	Initial and Maintenance margin, $	Loan/ Deposit to support Margin account, $	Cumulative costs/ profit for loan/ deposit, $	Other costs, $
01-2010	5.83	5.599	0.231	short	0	0	4620	-4620	0	-19
02-2010	5.32	5.215	0.105	short	+55990 -53200	2790	4620	-4030	-10.54	-41.9
03-2010	4.29	4.301	-0.011	long	+52150 -42900	9250	4620	5220	-19.61	-64.8
04-2010	4.03	4.088	-0.058	long	-43010 +40300	-2710	4620	2510	-12.08	-87.72
05-2010	4.14	4.155	-0.015	long	-40880 +41400	520	4620	3030	-7.93	-110.6
06-2010	4.8	4.785	0.015	short	-41550 +48000	6450	4620	9480	-3.13	-133.6
07-2010	4.63	4.59	0.04	short	+47850 -46300	1550	4620	11030	9.72	-156.5
08-2010	4.32	4.22	0.1	short	+45900 -43200	2700	4620	13730	24.52	-179.4
09-2010	3.89	3.898	-0.008	short	+42200 -38900	3300	4620	17030	42.69	-202.3
10-2010	3.43	3.6	-0.17	short	+36000 +34300	-4680	4620	12350	64.99	-225.2
11-2010	3.71	4.042	-0.332	0	+37100	1100	4620	13450	81.43	-244.2
12-2010	4.25	4.283	-0.033	0	0	0	4620	13450	99.25	-244.2
01-2011	4.49	4.499	-0.009	0	0	0	4620	13450	117.08	-244.2
02-2011	4.09	4.036	0.054	short	0	0	4620	13450	134.90	-263.2
03-2011	3.97	4.069	-0.099	long	+40360 -39700	660	4620	14110	152.72	-286.1
04-2011	4.24	4.272	-0.032	long	-40690 +42400	1710	4620	15820	171.36	-309
05-2011	4.31	4.336	-0.026	long	-42720 +43100	380	4620	16200	192.15	-331.9
06-2011	4.54	4.516	0.024	short	-43360 +45400	2040	4620	18240	213.40	-354.8
07-2011	4.42	4.353	0.067	short	+45160 -44200	960	4620	19200	237.21	-377.8
08-2011	4.06	3.984	0.076	short	+43530 -40600	2930	4620	22130	262.22	-4007
09-2011	3.9	3.849	0.051	short	+39840 -39000	840	4620	22970	290.89	-423.6
10-2011	3.57	3.624	-0.054	long	+38490 -35700	2790	4620	25760	320.61	-446.5
11-2011	3.24	3.558	-0.318	0	-36240 +32400	-3840	4620	21920	353.82	-465.5
12-2011	3.17	3.246	-0.076	0	0	0	4620	21920	382.23	-465.5
01-2012	2.67	2.708	-0.038	0	0	0	4620	21920	410.64	-465.5
02-2012	2.51	2.526	-0.016	long	0	0	4620	21920	439.05	-484.4
03-2012	2.17	2.296	-0.126	long	-25260 +21700	-3560	4620	18360	467.45	-507.4
04-2012	1.95	2.048	-0.098	long	-22960 +19500	-3460	4620	14900	491.41	-530.3
05-2012	2.43	2.493	-0.063	long	-20480 +24300	3820	4620	18720	511.05	-553.2
06-2012	2.46	2.498	-0.038	long	-24930 +24600	-330	4620	18390	535.45	-576.1

Date	NG spot $ per MBtu	NG Futures, $ per MBtu	Basis, $ per MBtu	Long/ short position in futures (10000 MBtu each)	Cash Flow, $	Margin account, $	Initial and Maintenance margin, $	Loan/ Deposit to support Margin account, $	Cumulative costs/ profit for loan/deposit, $	Other costs, $
07-2012	2.95	2.963	-0.013	long	-24980 +29500	4520	4620	22910	559.45	-599
08-2012	2.84	2.807	0.033	short	-29630 +28400	-1230	4620	21680	589.10	-622
09-2012	2.85	2.918	-0.068	long	+28070 -28500	-430	4620	21250	617.20	-645
10-2012	3.32	3.5	-0.18	long	-29180 +33200	4020	4620	25270	644.77	-668
11-2012	3.54	3.687	-0.147	0	-35000 +35400	400	4620	25670	677.37	-686.8
12-2012	3.34	3.444	-0.104	0	0	0	4620	25670	710.47	-686.8
01-2013	3.33	3.35	-0.02	0	0	0	4620	25670	743.56	-686.8
02-2013	3.33	3.314	0.016	short	0	0	4620	25670	776.66	-705.7
03-2013	3.81	3.773	0.037	short	+33140 -38100	-4960	4620	20710	809.75	-724.7
04-2013	4.17	4.161	0.009	short	+37730 -41700	-3970	4620	16740	836.65	-747.6
05-2013	4.04	4.068	-0.028	long	+41610 -40400	1210	4620	17950	858.58	-770.5
06-2013	3.83	3.806	0.024	short	-40680 +38300	-2380	4620	15570	882.03	-793.4
07-2013	3.62	3.641	-0.021	long	+38060 -36200	1860	4620	17430	902.50	-816.4
08-2013	3.43	3.413	0.017	short	-36410 +34300	-2110	4620	15320	925.30	-839.3
09-2013	3.62	3.618	0.002	short	+34130 -36200	-2070	4620	13250	945.45	-862.2
10-2013	3.68	3.654	0.026	short	+36180 -36800	-620	4620	12630	963.02	-885.1
11-2013	3.64	3.64	0	0	+36540 -36400	140	4620	12770	979.82	-904.1
12-2013	4.24	4.277	-0.037	0	0	0	4620	12770	996.79	-904.1
01-2014	4.71	4.542	0.168	0	0	0	4620	12770	1013.76	-904.1
02-2014	6	5.163	0.837	short	0	0	4620	12770	1030.73	-923.1
03-2014	4.9	4.486	0.414	short	+51630 -49000	2630	4620	15400	1047.70	-946
04-2014	4.66	4.608	0.052	short	+44860 -46600	-1740	4620	13660	1067.96	-968.9
05-2014	4.58	4.536	0.044	short	+46080 -45800	280	4620	13940	1086.05	-991.8
06-2014	4.59	4.594	-0.004	long	+45360 -45900	-540	4620	13400	1104.48	-1015
07-2014	4.05	4.025	0.025	short	-45940 +40500	-5440	4620	7960	1122.24	-1038
08-2014	3.91	3.899	0.011	short	+40250 -39100	1150	4620	9110	1133.20	-1061
09-2014	3.92	3.921	-1E-03	long	+38990 -39200	-210	4620	8900	1145.59	-1084
10-2014	3.78	3.801	-0.021	long	-39210 +37800	-1410	4620	7490	1157.73	-1106
11-2014	4.12	4.235	-0.115	0	-38010 +41200	3190	4620	10680	1168.10	-1125
12-2014	3.48	3.509	-0.029	0	0	0	4620	10680	1182.45	-1125
01-2015	2.99	2.929	0.061	0	0	0	4620	10680	1196.81	-1125
02-2015	2.87	2.755	0.115	long	0	0	4620	10680	1211.17	-1144

Date	NG spot $ per MBtu	NG Futures, $ per MBtu	Basis, $ per MBtu	Long/ short position in futures (10000 MBtu each)	Cash Flow, $	Margin account, $	Initial and Maintenance margin, $	Loan/ Deposit to support Margin account, $	Cumulative costs/ profit for loan/deposit, $	Other costs, $
03-2015	2.83	2.747	0.083	long	+27550 -28300	-750	4620	9930	1225.53	-1167
04-2015	2.61	2.591	0.019	long	+27470 -26100	1370	4620	11300	1238.95	-1190
05-2015	2.85	2.856	-0.006	short	+25910 -28500	-2590	4620	8710	1254.08	-1213
06-2015	2.78	2.769	0.011	long	-28560 +27800	-760	4620	7950	1265.98	-1236
07-2015	2.84	2.805	0.035	long	+27690 -28400	-710	4620	7240	1276.93	-1259
08-2015	2.77	2.753	0.017	long	+28050 -27700	350	4620	7590	1286.98	-1282
09-2015	2.66	2.639	0.021	long	+27530 -26600	930	4620	8520	1297.48	-1305
10-2015	2.34	2.378	-0.038	short	+26390 -23400	2990	4620	11510	1309.14	-1328
11-2015	2.09	2.281	-0.191	0	-23780 +20900	-2880	4620	11510	1324.53	-1347
12-2015	1.93	2.044	-0.114	0	0	0	4620	11510	1339.93	-1347
01-2016	2.28	2.233	0.047	0	0	0	4620	11510	1355.33	-1347
02-2016	1.99	1.929	0.061	long	0	0	4620	13940	1370.72	-1366
03-2016	1.73	1.812	-0.082	short	+19290 -17300	1990	4620	15930	1389.15	-1389
04-2016	1.92	2.014	-0.094	short	-18120 +19200	1080	4620	17010	1410.08	-1411
05-2016	1.92	2.083	-0.163	short	-20140 +19200	-940	4620	16070	1432.35	-1434
06-2016	2.59	2.634	-0.044	short	-20830 +25900	5070	4620	21140	1453.44	-1457
07-2016	2.82	2.761	0.059	long	-26340 +28200	1860	4620	23000	1480.88	-1480
08-2016	2.82	2.722	0.098	long	+27610 -28200	-590	4620	22410	1510.63	-1503
09-2016	2.99	2.903	0.087	long	+27220 -29900	-2680	4620	19730	1539.65	-1526
10-2016	2.98	3.064	-0.084	0	+29030 -29800	-770	4620	18960	1565.33	-1549
							4620	18960	1590.03	-1572
Totally	23598 $									

APPENDIX 4. Code developed in R for simulation of tests and models

```
#Unit root tests
#1997-2003, spot
library("CADFtest")
ADFt = CADFtest(ds[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(ds[1:85], null = "Trend")
kpss.test(ds[1:85])
#summary(ur.kpss(rs1))
library("fUnitRoots")
za.ds = ur.za(ds[1:85], model="intercept", lag=12)
summary(za.ds)
za.ds = ur.za(ds[1:85], model="trend", lag=12)
summary(za.ds)
za.ds = ur.za(ds[1:85], model="both", lag=12)
summary(za.ds)
#ZA test: t < -4.82, so stationary
#the ADF test provides p-values smaller than 0.01 in both periods, so stationarity
#KPSS test: p-value greater than 0.05 then stationary

#2004-2016, spot
ADFt = CADFtest(ds[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(ds[86:234], null = "Trend")
kpss.test(ds[86:234])
za.ds = ur.za(ds[86:234], model="intercept", lag=12)
summary(za.ds)
za.ds = ur.za(ds[86:234], model="trend", lag=12)
summary(za.ds)
za.ds = ur.za(ds[86:234], model="both", lag=12)
summary(za.ds)

#1997-2003, 1 month futures
ADFt = CADFtest(df1[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df1[1:85], null = "Trend")
kpss.test(df1[1:85])
za.df1 = ur.za(df1[1:85], model="intercept", lag=12)
summary(za.df1)
za.df1 = ur.za(df1[1:85], model="trend", lag=12)
summary(za.df1)
za.df1 = ur.za(df1[1:85], model="both", lag=12)
summary(za.df1)
```

```
#2004-2016, 1 month futures
ADFt = CADFtest(df1[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df1[86:234], null = "Trend")
kpss.test(df1[86:234])
za.df1 = ur.za(df1[86:234], model="intercept", lag=12)
summary(za.df1)
za.df1 = ur.za(df1[86:234], model="trend", lag=12)
summary(za.df1)
za.df1 = ur.za(df1[86:234], model="both", lag=12)
summary(za.df1)

#1997-2003, 2 month futures
ADFt = CADFtest(df2[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df2[1:85], null = "Trend")
kpss.test(df2[1:85])
za.df2 = ur.za(df2[1:85], model="intercept", lag=12)
summary(za.df2)
za.df2 = ur.za(df2[1:85], model="trend", lag=12)
summary(za.df2)
za.df2 = ur.za(df2[1:85], model="both", lag=12)
summary(za.df2)

#2004-2016, 2 month futures
ADFt = CADFtest(df2[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df2[86:234], null = "Trend")
kpss.test(df2[86:234])
za.df2 = ur.za(df2[86:234], model="intercept", lag=12)
summary(za.df2)
za.df2 = ur.za(df2[86:234], model="trend", lag=12)
summary(za.df2)
za.df2 = ur.za(df2[86:234], model="both", lag=12)
summary(za.df2)

#1997-2003, 3 month futures
ADFt = CADFtest(df3[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df3[1:85], null = "Trend")
kpss.test(df3[1:85])
za.df3 = ur.za(df3[1:85], model="intercept", lag=12)
summary(za.df3)
za.df3 = ur.za(df3[1:85], model="trend", lag=12)
summary(za.df3)
```

```
za.df3 = ur.za(df3[1:85], model="both", lag=12)
summary(za.df3)

#2004-2016, 3 month futures
ADFt = CADFtest(df3[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df3[86:234], null = "Trend")
kpss.test(df3[86:234])
za.df3 = ur.za(df3[86:234], model="intercept", lag=12)
summary(za.df3)
za.df3 = ur.za(df3[86:234], model="trend", lag=12)
summary(za.df3)
za.df3 = ur.za(df3[86:234], model="both", lag=12)
summary(za.df3)

#1997-2003, 4 month futures
ADFt = CADFtest(df4[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df4[1:85], null = "Trend")
kpss.test(df4[1:85])
za.df4 = ur.za(df4[1:85], model="intercept", lag=12)
summary(za.df4)
za.df4 = ur.za(df4[1:85], model="trend", lag=12)
summary(za.df4)
za.df4 = ur.za(df4[1:85], model="both", lag=12)
summary(za.df4)

#2004-2016, 4 month futures
ADFt = CADFtest(df4[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(df4[86:234], null = "Trend")
kpss.test(df4[83:234])
za.df4 = ur.za(df4[86:234], model="intercept", lag=12)
summary(za.df4)
za.df4 = ur.za(df4[86:234], model="trend", lag=12)
summary(za.df4)
za.df4 = ur.za(df4[86:234], model="both", lag=12)
summary(za.df4)
#1997-2003, basis 1
ADFt = CADFtest(db1[1:85], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db1[1:85], null = "Trend")
kpss.test(db1[1:85])
za.db1 = ur.za(db1[1:85], model="intercept", lag=12)
summary(za.db1)
```

```
za.db1 = ur.za(db1[1:85], model="trend", lag=12)
summary(za.db1)
za.db1 = ur.za(db1[1:85], model="both", lag=12)
summary(za.db1)

#2004-2016, basis 1
ADFt = CADFtest(db1[86:234], max.lag.y = 12, criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db1[86:234], null = "Trend")
kpss.test(db1[86:234])
za.db1 = ur.za(db1[86:234], model="intercept", lag=12)
summary(za.db1)
za.db1 = ur.za(db1[86:234], model="trend", lag=12)
summary(za.db1)
za.db1 = ur.za(db1[86:234], model="both", lag=12)
summary(za.db1)

#1997-2003, basis 2
ADFt = CADFtest(db2[1:85], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db2[1:85], null = "Trend")
kpss.test(db2[1:85])
za.db2 = ur.za(db2[1:85], model="intercept", lag=12)
summary(za.db2)
za.db2 = ur.za(db2[1:85], model="trend", lag=12)
summary(za.db2)
za.db2 = ur.za(db2[1:85], model="both", lag=12)
summary(za.db2)

#2004-2016, basis 2
ADFt = CADFtest(db2[8:234], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db2[86:234], null = "Trend")
kpss.test(db2[86:234])
za.db2 = ur.za(db2[86:234], model="intercept", lag=12)
summary(za.db2)
za.db2 = ur.za(db2[86:234], model="trend", lag=12)
summary(za.db2)
za.db2 = ur.za(db2[86:234], model="both", lag=12)
summary(za.db2)

#1997-2003, basis 3
ADFt = CADFtest(db3[1:85], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db3[1:85], null = "Trend")
```

```
kpss.test(db3[1:85])
za.db3 = ur.za(db3[1:85], model="intercept", lag=12)
summary(za.db3)
za.db3 = ur.za(db3[1:85], model="trend", lag=12)
summary(za.db3)
za.db3 = ur.za(db3[1:85], model="both", lag=12)
summary(za.db3)

#2004-2016, basis 3
ADFt = CADFtest(db3[86:234], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(b3[86:234], null = "Trend")
kpss.test(b3[86:234])
za.db3 = ur.za(db3[86:234], model="intercept", lag=12)
summary(za.db3)
za.db3 = ur.za(db3[86:234], model="trend", lag=12)
summary(za.db3)
za.db3 = ur.za(db3[86:234], model="both", lag=12)
summary(za.db3)

#1997-2003, basis 4
ADFt = CADFtest(db4[1:85], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db4[1:85], null = "Trend")
kpss.test(db4[1:85])
za.db4 = ur.za(db4[1:85], model="intercept", lag=12)
summary(za.db4)
za.db4 = ur.za(db4[1:85], model="trend", lag=12)
summary(za.db4)
za.db4 = ur.za(db4[1:85], model="both", lag=12)
summary(za.db4)

#2004-2016, basis 4
ADFt = CADFtest(db4[86:234], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
require("tseries")
kpss.test(db4[86:234], null = "Trend")
kpss.test(db4[86:234])
za.db4 = ur.za(db4[86:234], model="intercept", lag=12)
summary(za.db4)
za.db4 = ur.za(db4[86:234], model="trend", lag=12)
summary(za.db4)
za.db4 = ur.za(db4[86:234], model="both", lag=12)
summary(za.db4)
```

```
#COINTEGRATION ANALYSIS
library("urca")
library("CADFtest")
#NG spot and futures1
Engle=lm(s~f1)
summary(Engle)
res1=resid(Engle)
ADFt = CADFtest(res1[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(res1[87:235], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
f = f1[1:86]
sf = cbind(s1,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)

s2 = s[87:235]
f = f1[87:235]
sf = cbind(s2,f)
cosf = cajolst(sf, trend=TRUE, K=12, season=12)
summary(cosf)#r = 0 t >tcrit, so cointegration

#NG spot and futures2
Engle=lm(s~f2)
summary(Engle)
res2=resid(Engle)
ADFt = CADFtest(res2[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(res2[87:235], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
f = f2[1:86]
sf = cbind(s1,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)
s2 = s[87:235]
f = f2[87:235]
sf = cbind(s2,f)
cosf = cajolst(sf, trend=TRUE, K=12, season=12)
summary(cosf)


#NG spot and futures3
Engle=lm(s~f3)
summary(Engle)
res3=resid(Engle)
```

```
ADFt = CADFtest(res3[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(res3[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
f = f3[1:86]
sf = cbind(s1,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)
s2 = s[87:235]
f = f3[87:235]
sf = cbind(s2,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)

#NG spot and futures4
Engle=lm(s~f4)
summary(Engle)
res4=resid(Engle)
ADFt = CADFtest(res4[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(res4[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
f = f4[1:86]
sf = cbind(s1,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)
s2 = s[87:235]
f = f4[87:235]
sf = cbind(s2,f)
cosf = cajolst(sf, trend=TRUE, K=5, season=12)
summary(cosf)

#COINTEGRATION ANALYSIS FOR BASES AND SPOT PRICES
#NG spot and basis 1
Engle=lm(s~b1)
summary(Engle)
resb1=resid(Engle)
ADFt = CADFtest(resb1[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(resb1[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
b = b1[1:86]
sb = cbind(s1,b)
```

```
cosb = cajolst(sb, trend=TRUE, K=5, season=12)
summary(cosb)
s2 = s[87:235]
b = b1[87:235]
sb = cbind(s2,b)
cosb = cajolst(sb, trend=TRUE, K=12, season=12)
summary(cosb)

#NG spot and basis 2
Engle=lm(s~b2)
summary(Engle)
resb2=resid(Engle)
ADFt = CADFtest(resb2[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(resb2[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
b = b2[1:86]
sb = cbind(s1,b)
cosb = cajolst(sb, trend=TRUE, K=6, season=12)
summary(cosb)
s2 = s[87:235]
b = b2[87:235]
sb = cbind(s2,b)
cosb = cajolst(sb, trend=TRUE, K=12, season=12)
summary(cosb)

#NG spot and basis 3
Engle=lm(s~b3)
summary(Engle)
resb3=resid(Engle)
ADFt = CADFtest(resb3[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(resb3[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
b = b3[1:86]
sb = cbind(s1,b)
cosb = cajolst(sb, trend=TRUE, K=6, season=12)
summary(cosb)
s2 = s[87:235]
b = b3[87:235]
sb = cbind(s2,b)
cosb = cajolst(sb, trend=TRUE, K=12, season=12)
summary(cosb)
```

```
#NG spot and basis 3
Engle=lm(s~b4)
summary(Engle)
resb4=resid(Engle)
ADFt = CADFtest(resb4[1:86], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(resb4[87:226], max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

s1 = s[1:86]
b = b4[1:86]
sb = cbind(s1,b)
cosb = cajolst(sb, trend=TRUE, K=6, season=12)
summary(cosb)
s2 = s[87:235]
b = b4[7:235]
sb = cbind(s2,b)
cosb = cajolst(sb, trend=TRUE, K=7, season=12)
summary(cosb)

#ERROR_CORRECTION MODEL
#install.packages("apt")
library("apt")
#NG spot and futures1
df = ts(s[1:86])
dff = ts(res1[1:86])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

df = ts(s[87:235])
dff = ts(res1[87:235])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

#NG spot and futures2
df = ts(s[1:86])
dff = ts(res2[1:86])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

df = ts(s[87:235])
dff = ts(res2[87:235])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)
```

```
#NG spot and futures3
df = ts(s[1:86])
dff = ts(res3[1:86])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

df = ts(s[87:235])
dff = ts(res3[87:235])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

#NG spot and futures4
df = ts(s[1:86])
dff = ts(res4[1:86])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)

df = ts(s[87:235])
dff = ts(res4[87:235])
fit = ecmSymFit(df, dff, lag=1)
summary(fit)
ecmDiag(fit)


#Seasonality Testing

k1 = c(1,1,0,0,0,0,0,0,0,0,0,1)
k2 = c(0,0,1,1,1,0,0,0,0,0,0,0)
k3 = c(0,0,0,0,0,1,1,1,0,0,0,0)
k4 = c(0,0,0,0,0,0,0,0,1,1,1,0)

d1 = rep(k1, 19)
d2 = rep(k2, 19)
d3 = rep(k3, 19)
d4 = rep(k4, 19)


fit1 = lm(ds[1:86] ~ d1[1:86]+d2[1:86]+d3[1:86]+d4[1:86])
summary(fit1)

fit2 = lm(ds[87:234] ~ d1[87:234]+d2[87:234]+d3[87:234]+d4[87:234])
summary(fit2)
```

```
#Forecasting models of NG spot by futures
#1997-2003

#1 month (1997-2003)
ds1 = c(s[3:47],s[51:57],s[62:80]) - c(s[2:46],s[50:56],s[61:79])
ADFt = CADFtest(ds1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df1 = c(f1[2:46],f1[50:56],f1[61:79]) - c(f1[1:45],f1[49:55],s[60:78])
ADFt = CADFtest(df1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod1 = lm(ds1 ~ df1)
summary(mod1)
res1 = residuals(mod1)
shapiro.test(res1) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod1) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod1) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(mod1)
coeftest(mod1, y)

mod11 = predict(mod1)
plot.ts(ds1, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(mod11, col="blue")
xlabel = seq(0, 80, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("topleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)


#2 months (1997-2003)
ds2 = c(s[4:47],s[52:57],s[62:81]) - c(s[3:46],s[51:56],s[61:80])
ADFt = CADFtest(ds2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df2 = c(f2[2:45],f2[50:55],f2[60:79]) - c(f2[1:44],f2[49:54],f2[59:78])
ADFt = CADFtest(df2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod2 = lm(ds2 ~ df2)
summary(mod2)
res2 = residuals(mod2)
shapiro.test(res2) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod2) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
```

```
dwtest(mod2) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

mod22 = predict(mod2)
plot.ts(ds2, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(mod22, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("topleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)

#3 months (1997-2003)
ds3 = c(s[5:46],s[54:58],s[64:82]) - c(s[4:45],s[53:57],s[63:81])
ADFt = CADFtest(ds3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df3 = c(f3[2:43],f3[50:54],f3[61:79]) - c(f3[1:42],f3[49:53],f3[60:78])
ADFt = CADFtest(df3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod3 = lm(ds3 ~ df3)
summary(mod3)
res3 = residuals(mod3)
shapiro.test(res3) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod3) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod3) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

mod33 = predict(mod3)
plot.ts(ds3, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(mod33, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("topleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)

#4 months (1997-2003)
ds4 = c(s[5:47],s[54:58],s[65:83]) - c(s[4:46],s[53:57],s[64:82])
ADFt = CADFtest(ds2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df4 = c(f4[2:44],f4[50:54],f4[61:79]) - c(f4[1:43],f4[49:53],f4[60:78])
ADFt = CADFtest(df2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod4 = lm(ds4 ~ df4)
summary(mod4)
res44 = residuals(mod4)
shapiro.test(res44) #p-value <0.05, so non-normal distribution (H0 - normal distribution)
```

```
library("lmtest")
bptest(mod4) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod4) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(mod4)
coeftest(mod4, y)

mod44 = predict(mod4)
plot.ts(ds4, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(mod44, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("topleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)


#2003-2016
#1 month, excl. 104-109
ds12 = c(s[89:103],s[112:235]) - c(s[88:102],s[111:234])
ADFt = CADFtest(ds12, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df12 = c(f1[88:102],f1[111:234]) - c(f1[87:101],f1[110:233])
ADFt = CADFtest(ds12, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod12 = lm(ds12 ~ df12)
summary(mod12)
res12 = residuals(mod12)
shapiro.test(res12) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod12) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod12) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

mod122 = predict(mod12)
plot.ts(ds12, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(mod122, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
   lty=1:2, cex=0.8)
```

```
#2 months
ds22 = c(s[90:103],s[113:235]) - c(s[89:102],s[112:234])
ADFt = CADFtest(ds22, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df22 = c(f2[88:101],f2[111:233]) - c(f2[87:100],f2[110:232])
ADFt = CADFtest(ds22, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod22 = lm(ds22 ~ df22)
summary(mod22)
res22 = residuals(mod22)
shapiro.test(res22) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod22) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod22) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(mod22)
coeftest(mod22, y)

mod222 = predict(mod22)
plot.ts(ds22, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
       axes = FALSE,col="red")
lines(mod222, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
      lty=1:2, cex=0.8)


#3 months
ds33 = c(s[91:103],s[114:235]) - c(s[90:102],s[113:234])
ADFt = CADFtest(ds33, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df33 = c(f3[88:100],f3[111:232]) - c(f3[87:99],f3[110:231])
ADFt = CADFtest(ds33, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod33 = lm(ds33 ~ df33)
summary(mod33)
res33 = residuals(mod33)
shapiro.test(res33) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod33) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod33) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)
```

```
library("sandwich")
library("lmtest")
y = vcovHAC(mod33)
coeftest(mod33, y)

mod333 = predict(mod33)
plot.ts(ds33, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
       axes = FALSE,col="red")
lines(mod333, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
      lty=1:2, cex=0.8)

#4 months
ds44 = c(s[92:103],s[115:235]) - c(s[91:102],s[114:234])
ADFt = CADFtest(ds44, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
df44 = c(f4[88:99],f4[111:231]) - c(f4[87:98],f4[110:230])
ADFt = CADFtest(ds44, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
mod44 = lm(ds44 ~ df44)
summary(mod44)
res44 = residuals(mod44)
shapiro.test(res44) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(mod44) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(mod44) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

mod444 = predict(mod44)
plot.ts(ds44, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
       axes = FALSE,col="red")
lines(mod444, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
      lty=1:2, cex=0.8)
```

```
#Forecast models using bases

#1997-2003
library("CADFtest")
#1 month
ds1 = c(s[2:67],s[70:72],s[75:78]) - c(s[1:66],s[69:71],s[74:77])
ADFt = CADFtest(ds1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ds1 = diff(ds1)
db1 = c(s[1:66],s[69:71],s[74:77]) - c(f1[1:66],f1[69:71],f1[74:77])
db1 = diff(db1)
ADFt = CADFtest(db1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb1 = lm(ds1 ~ db1)
summary(modb1)
resb1 = residuals(modb1)
shapiro.test(resb1) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb1) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb1) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

modb11 = predict(modb1)
plot.ts(ds1, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(modb11, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("bottomleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)

#2 months
library("CADFtest")
ds2 = c(s[3:67],s[71:72],s[76:78]) - c(s[1:65],s[69:70],s[74:76])
ds2 = diff(ds2)
ADFt = CADFtest(ds2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db2 = c(s[1:65],s[69:70],s[74:76]) - c(f2[1:65],f2[69:70],f2[74:76])
db2 = diff(db2)
ADFt = CADFtest(db2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb2 = lm(ds2 ~ db2)
summary(modb2)
resb2 = residuals(modb2)
shapiro.test(resb2) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb2) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
```

```
dwtest(modb2) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(modb2)
coeftest(modb2, y)

modb22 = predict(modb2)
plot.ts(ds2, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
       axes = FALSE,col="red")
lines(modb22, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("bottomleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)

#3 months
ds3 = c(s[4:47],s[52:72],s[76:81]) - c(s[1:44],s[49:69],s[73:78])
ds3 = diff(ds3)
ADFt = CADFtest(ds3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db3 = c(s[1:44],s[49:69],s[73:78]) - c(f3[1:44],f3[49:69],f3[73:78])
db3 = diff(db3)
ADFt = CADFtest(db3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb3 = lm(ds3 ~ db3)
summary(modb3)
resb3 = residuals(modb3)
shapiro.test(resb3) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb3) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb3) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(modb3)
coeftest(modb3, y)

modb33 = predict(modb3)
plot.ts(ds3, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
       axes = FALSE,col="red")
lines(modb33, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("bottomleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)
```

```
#4 months

library("CADFtest")
ds4 = c(s[5:47],s[53:83]) - c(s[1:43],s[49:79])
ds4 = diff(ds4)
ADFt = CADFtest(diff(ds4), max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db4 = c(s[1:43],s[49:79]) - c(f4[1:43],f4[49:79])
db4 = diff(db4)
ADFt = CADFtest(db4, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb4 = lm(ds4 ~ db4)
summary(modb4)
resb4 = residuals(modb4)
shapiro.test(resb4) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb4) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb4) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

library("sandwich")
library("lmtest")
y = vcovHAC(modb4)
coeftest(modb4, y)

modb44 = predict(modb4)
plot.ts(ds4, xlab="time, months(1997-2003)", ylab="Diff of NG spot price, $/Million Btu",
    axes = FALSE,col="red")
lines(modb44, col="blue")
xlabel = seq(0, 87, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 9, by = 1)
axis(2, at = ylabel)
legend("bottomleft",legend=c("Actual", "Forecast"),col=c("red", "blue"), lty=1:2, cex=0.8)


#2004-2016
#1 month
library("CADFtest")
dss1 = c(s[86:153],s[156:203],s[206:235]) - c(s[85:152],s[155:202],s[205:234])
dss1 = diff(dss1)
ADFt = CADFtest(dss1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db1 = c(s[85:152],s[155:202],s[205:234]) - c(f1[85:152],f1[155:202],f1[205:234])
db1 = diff(db1)
ADFt = CADFtest(db1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
```

```
modb12 = lm(dss1 ~ db1)
summary(modb12)
resb12 = residuals(modb12)
shapiro.test(resb12) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb12) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb12) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

modb122 = predict(modb12)
plot.ts(dss1, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
        axes = FALSE,col="red")
lines(modb122, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("topright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
       lty=1:2, cex=0.8)

#2 months
library("CADFtest")
dss2 = c(s[87:115],s[119:203],s[207:235]) - c(s[85:113],s[117:201],s[205:233])
dss2 = diff(dss2)
ADFt = CADFtest(dss2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db2 = c(s[85:113],s[117:201],s[205:233]) - c(f2[85:113],f2[117:201],f2[205:233])
db2 = diff(db2)
ADFt = CADFtest(db2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb22 = lm(dss2 ~ db2)
summary(modb22)
resb22 = residuals(modb22)
shapiro.test(resb22) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb22) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb22) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

modb222 = predict(modb22)
plot.ts(dss2, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
        axes = FALSE,col="red")
lines(modb222, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
       lty=1:2, cex=0.8)
```

```
#3 months
library("CADFtest")
dss3 = c(s[88:115],s[120:202],s[207:235]) - c(s[85:112],s[117:199],s[204:232])
dss3 = diff(dss3)
ADFt = CADFtest(dss3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db3 = c(s[85:112],s[117:199],s[204:232]) - c(f3[85:112],f3[117:199],f3[204:232])
db3 = diff(db3)
ADFt = CADFtest(db3, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb32 = lm(dss3 ~ db3)
summary(modb32)
resb32 = residuals(modb32)
shapiro.test(resb32) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb32) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb32) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)

modb322 = predict(modb32)
plot.ts(dss3, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
        axes = FALSE,col="red")
lines(modb322, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
       lty=1:2, cex=0.8)

#4 months
library("CADFtest")
dss4 = c(s[88:115],s[121:202],s[208:235]) - c(s[84:111],s[117:198],s[204:231])
dss4 = diff(dss4)
ADFt = CADFtest(dss4, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
db4 = c(s[84:111],s[117:198],s[204:231]) - c(f4[84:111],f4[117:198],f4[204:231])
db4 = diff(db4)
ADFt = CADFtest(db4, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
modb42 = lm(dss4 ~ db4)
summary(modb42)
resb42 = residuals(modb42)
shapiro.test(resb42) #p-value <0.05, so non-normal distribution (H0 - normal distribution)

library("lmtest")
bptest(modb42) #p-value <0.05, so heteroscedastic (H0 - homoskedastic)
dwtest(modb42) #p-value <0.05, so autocorrelation (H0 -no autocorrelation)
```

```
library("sandwich")
library("lmtest")
y = vcovHAC(modb42)
coeftest(modb42, y)

modb422 = predict(modb42)
plot.ts(dss4, xlab="time, months(2004-2016)", ylab="Diff of NG spot price, $/Million Btu",
        axes = FALSE,col="red")
lines(modb422, col="blue")
xlabel = seq(0, 144, by = 12)
axis(1, at = xlabel)
ylabel = seq(0, 16, by = 1)
axis(2, at = ylabel)
legend("bottomright",legend=c("Actual", "Forecast"),col=c("red", "blue"),
       lty=1:2, cex=0.8)

#Analysis by Markov-Switching model
install.packages("tseries")
install.packages("urca")
install.packages("CADFtest")

install.packages("forecast")
install.packages("timeDate")
library("forecast")

install.packages("sandwich")
install.packages("lmtest")
install.packages("zoo")
library("tseries")
library("zoo")
library("urca")
library("sandwich")
library("CADFtest")

#model definition
install.packages("MSwM")
install.packages("nlme")
install.packages("parallel")
library("nlme")
library("parallel")
library("MSwM")

data = read.csv("C:/Users/Maria/Desktop/Gas Market/R-thesis/Data1.csv", header = TRUE)
ls()
s=data[,"spot"]
f1=data[,"futures1"]
d1=data[,"long"]
d2=data[,"short"]
```

```
cd1=data[,"clong"]
cd2=data[,"cshort"]
stor=data[,"storage"]
mt=data[,"NMATA"]

ds=diff(s)
df1=diff(f1)
db1=diff(s-f1)
stor=diff(stor)
d1=diff(d1)
d2=diff(d2)
cd1=diff(cd1)
cd2=diff(cd2)
mt=diff(mt)

library("CADFtest")
ADFt = CADFtest(ds, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(df1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(db1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)
ADFt = CADFtest(stor, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ADFt = CADFtest(d1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ADFt = CADFtest(d2, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ADFt = CADFtest(cd1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ADFt = CADFtest(cd1, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ADFt = CADFtest(mt, max.lag.y = 12, type = "trend", criterion = "AIC")
summary(ADFt)

ds=ds[2:234]
df1=df1[1:233]
db1=db1[1:233]
d1=d1[1:233]
d2=d2[1:233]
cd1=cd1[1:233]
cd2=cd2[1:233]
mt=mt[1:233]
stor=stor[1:233]
```

```
ds=ds[90:234]
df1=df1[89:233]
db1=db1[89:233]
d1=d1[89:233]
d2=d2[89:233]
cd1=cd1[89:233]
cd2=cd2[89:233]
mt=mt[89:233]
stor=stor[89:233]

ds=ds[14:157]
df1=df1[13:156]
db1=db1[13:156]
d1=d1[13:156]
d2=d2[13:156]
cd1=cd1[13:156]
cd2=cd2[13:156]
mt=mt[13:156]
stor=stor[13:156]

#by basis1
mod = lm(ds ~ df1+d1+d2+cd1+cd2+mt+stor)
summary(mod)

#fit regime switching model
#fit regime switching model
m1=msmFit(mod,k=2,sw=c(T, T, T, T, T, T, T, T, T), control=list(parallel=F))

#plot the probabilities
par("mar")
par(mar=c(3,3,3,3))

plot.ts(m1)
plot(m1, which=2)

summary(m1)

plotProb(m1, which=1)
plotProb(m1, which=2)

plotDiag(m1, which=1)
plotDiag(m1, which=2)
plotDiag(m1, which=3)

#by basis1
mod = lm(ds ~ db1+d1+d2+cd1+cd2+mt+stor)
summary(mod)
```

```
#fit regime switching model
```
#fit regime switching model
```
m1=msmFit(mod,k=2,sw=c(T, T, T, T, T, T, T, T, T), control=list(parallel=F))

#plot the probabilities
par("mar")
par(mar=c(3,3,3,3))

plot.ts(m1)
plot(m1, which=2)

summary(m1)

plotProb(m1, which=1)
plotProb(m1, which=2)

plotDiag(m1, which=1)
plotDiag(m1, which=2)
plotDiag(m1, which=3)
```