Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Efficiency, predictability and liquidity in the commodity futures markets

Author: **Vojtěch Čermák**

Supervisor: **PhDr. Ladislav Krištoufek, Ph.D.**

Academic Year: **2014/2015**

Declaration of Authorship

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

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

Prague, May 15, 2015

Signature

Acknowledgments

I am grateful to PhDr. Ladislav Krištoufek. I am extremly thankful and indebted to him for sharing expertise, and sincere and valuable guidance and encouragement extended to me.

Abstract

This thesis examines efficiency of several CME commodity futures and its relation to market liquidity over the ten years period. The goal is to find ARMA model that is better than white noise in terms of forecasting power and carry out analysis of market liquidity if we find such model. This is done by comparing selected ARMA models to white noise. In order to do that, we use Diebolt - Mariano test on forecast errors obtained by pseudo out - of - sample analysis using rolling window with re - estimation. Concern of furhter analysis are factors, that can influence the DM statistics. Main findings are, that we are able to find such ARMA model for small enough time period within the ten years period for almost all commodities. For most commodities, this sub period is not long enough to violate efficient market hypothesis. Only for palladium and lean hog futures this period is longer than one year. These two futures shows strong signs of inefficiency, as its predictability is not out - weighted by liquidity restrictions.

Keywords commodity futures,market efficiency,market liquidity, predictability

Author's e-mail cermak.vojtech@seznam.cz
Supervisor's e-mail kristoufek@icloud.com

Abstrakt

Tato práce zkoumá efektivitu několika CME komoditních futures a její vztah k likviditě během desetiletého období. Cílem je najít ARMA model, který je lepší než šum z pohledu kvality předpovědi a dále provést analýzu likvidity daného trhu v případě že se nám podaří takový model nalézt. Toto je provedeno porovnáváním vybraných ARMA modelů ku šumu. K tomu je použit Diebolt - Mariano test na chyby v předpovědi získané pseudo out - of - sample analýzou za použití klouzavého okna s opětovnou regresí. Další analýza se zabývá vybranými faktory které mohou ovlivnit velikost DM statistiky. Hlavním zjištěním je, že pro téměř každou komoditu jsme schopni nalézt dostatečně malé období, kde je ARMA model lepší než šum. Pro většinu komodit je toto období tak malé, že na nemá vliv na hypotézu efektivních trhů. Jen pro palladium a lean hog futures jsme schopni nalézt toto období větší než rok. Tyto dvě komodity ukazují silné známky neefektivity, která vychází z toho, že schopnost predikovat není vyvážena sníženou likviditou v daném období.

Klíčová slova komoditní futures, efektivní trh, likvidita trhu, prediktabilita

E-mail autora `cermak.vojtech@seznam.cz`
E-mail vedoucího práce `kristoufek@icloud.com`

Bibliographic note

ČERMÁK, Vojtěch. *Efficiency, predictability and liquidity in the commodity futures markets*. Prague 2015. 56 pages. Bachelor thesis. Charles University in Prague, Faculty of Social Sciences, Institute of Economical Studies. Supervisor: PhDr. Ladislav Krištoufek, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

DM Diebolt Mariano

MAE Mean absolute error

RMSE Rooted mean squared error

CME Chicago Mercantile Exchange

CBOT Chicago Board of Trade

NYMEX New York Mercantile Exchange

MA Moving average model

AR Autoregressive model

EHM Efficient Market Hypothesis

ARMA Autoregressive - Moving average model

ADF Augumented Dickey Fuhler test

Bachelor Thesis Proposal

Author	Vojtěch Čermák
Supervisor	PhDr. Ladislav Krištoufek, Ph.D.
Proposed topic	Efficiency, predictability and liquidity in the commodity futures markets

Topic characteristics Efficient market hypothesis asserts, that we should not be able to profit in financial markets using price forecasting or predict prices at all. However in some cases, we are able to create good forecasting model, that predicts prices accurately. Such behavior would inevitably violate efficient market hypothesis, unless we were unable to profit on it because of other factors, such as liquidity.

In my thesis, i would like to further look into problems of forecasting of the futures market. In the first place, my goal will be to create model that is able to accurately predict market prices and then compare my findings with traded volume and open interest. Furthermore, i would like to examine whether the same behavior persists in different commodity markets. To do so, i will use standard methods used for short term forecasts, such as autoregressive moving average model.

Hypotheses

1. Price of futures can be predicted in specific cases
2. Liquidity of futures can be approximated by traded volume and open interest.
3. There is a trade-off between liquidity and the ability to predict the price of futures, therefore the efficient market theory holds.
4. The relation between price forecasting and liquidity is same for all commodities, which means that this attribute is specifically for futures in general.

Methodology

Outline

1. Introduction
2. Literature review, description of future contracts and its prediction
3. Description of possible methods
4. Application methods on various commodity type datasets
5. Results
6. Conclusion

Core bibliography

1. Wooldridge , J. M. (2008): "Introductory Econometrics: A Modern Approach." *South Western College*
2. Wei, W. W. S.(2005): "Univariate and Multivariate Methods (2nd Edition)." *Pearson*
3. Murphy, J. J. (1986): "Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications." *Prentice Hall Press*
4. Tomek, W. G. (1997): "Commodity futures prices as forecast." *Review of Agricultural Economics* **19(1)**: pp. 23–44.,Agricultural & Applied Economics Association
5. Fama, E., K. French(1987): "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums and the Theory of Storage" *The Journal of Business* **60(1)**: pp. 55–73.,The University of Chicago Press
6. Kellard, N., P. Newbold, T. Rayner, & C. Ennew (1999): "The relative efficiency of commodity futures markets" *Journal of Futures Markets* **19(4)**: pp. 413–432.

Author Supervisor

Chapter 1

Introduction

Physical way of trading commodities carries a lot of inconvenience and risk, related to the storing of given commodity. Because of that, most of the investors, who are willing to invest in commodities, prefer to trade commodities using some of the instruments. Among the most popular instruments, used for trading commodities, are without doubt futures contracts.

Popularity of commodity futures among investors has been growing since the year 1990. This trend has intensified even more in recent 15 years and we can expect that the growing popularity of commodities will continue in near future as well. Few reasons why investors switch to commodities or at least include them into their portfolios, are suggested in Jensen & Mercer (2011) . First one reflects poor performance of traditional assets like equities and real - estate, which failed the expectations of the investors during the Dot-com bubble and events related to real - estate market, causing the "2008 financial crisis". These incidents caused distrust to established financial instruments among investor and forced them to look around for new stable assets worth investing in, such as commodities. As described in Gorton & Rouwenhorst (2006), commodities are indeed effective in providing diversification of both stock and bond portfolios. However, perhaps the most important reason for the popularity of commodities as investment asset are advances in research of commodity investment instruments, which resulted in growing trust of investors to invest in these.

For most risk averse investors, especially when investing in futures, market efficiency is very important feature. Efficient markets are considered as transparent and trustworthy, therefore it is an important area to study. Since publication of Samuelson (1965), when Samuelson first suggested the relation of commodity futures prices as predictor of spot prices in the future as result of market efficiency, importance of this relation has been discussed extensively. Being able to predict future prices of spots is not only very important for speculative purposes, but is crucial especially for managing risk, which is highly appreciated by hedgers as well as market participants who in fact require given commodity for carrying out their business.

Goal of this thesis is to show on returns of several commodity futures, divided into four main categories, whether there is market failure or not. To do so, we will use group of ARMA models and perform out of sample forecast of returns and then compare them to white noise. Under the efficient market hypothesis, there should not statistically be a significant difference between them. However, for some commodities, forecasting performance of ARMA model can be significantly better than white noise. Such behavior does not necessarily mean market failure, but is certainly extraordinary and require further analysis. One way how to explain this phenomenon can be examining traded volume of given commodity and look for its relation to forecast error. Apart of that, this thesis will also examine several possible factors, which might also have effect on difference of ARMA forecast errors and white noise, such as affiliation to commodity group, size of the sample and size of the forecast error series.

Structure of this work is as follows, second chapter sums important findings about efficient market hypothesis and papers which already studied the issue of commodity market efficiency. Third chapter is thoroughly describing suitable methods for out - of - sample forecasting. Then, in fourth chapter, main concern will be application of suitable method on selected commodities, with short commentary. Last chapter is dedicated to discussion of these results and further analysis of problematic results. Then follows brief conclusion.

All calculations and analysis are performed using R language scripts and RStudio as developing environment. To facilitate the computations and to provide more efficient performance, programming the analysis is divided into three scripts. First one is used for general analysis of commodity group, second is specially modified to do detailed analysis of one ARMA model for one commodity, while last one is adjusted for generating graphic representations. All scripts used for the analysis are available upon request.

Chapter 2

Literature review

2.1 Efficient market hypothesis

In the most basic way, as described in Lo (2007), the Efficient market hypothesis suggests, that if market is efficient, then is impossible to gain abnormal profit from trading on that market. The market prices imidiately reflect all available relevant informations, thus making it impossible for investors to buy undervalued and sell overvaulted items.

Among the first researchers who examined this property of financial markets were Samuelson in his paper Samuelson (1965) . Samuelson stated so called martingale definition of market efficiency. As he examined the commodity futures, he observed that, under some assumptions, prices of commodity futures follows martingale. This suggests relation between commodity futures and spot prices as effect of efficient markets. Results of his research are that prices of futures at time t are predicting prices of spot prices for time t. Further he claimed, that on efficient markets futures prices are best unbiased estimator of forthcoming futures prices and spot prices.

Apart of Samuelson, one of the most influential researcher in this area is Fama, who discovered efficient market hypothesis independently on Samuelson on stock prices using random walk definition Fama (1965). He deeper examined issue of market efficiency in Fama (1970) and further extented this work in Fama (1991). As result of his extended research, he suggested to classify market efficiency types to weak form, semi-strong form, and strong form.

Weak form of market efficiency suggests, that prices fully reflect information about the market that is contained in historical prices. Therefore no investor relying purely on historical prices cannot gain extra profit and market prices cannot be predicted using technical analysis. This form of market efficiency implies, that price does not follow any pattern and is strictly random.

Semi strong efficiency suggests that prices immediately reflect both historical prices and all publicly available information. This imply that even predicting prices using fundamental analysis will not abnormally benefit the investors. However, investors can make profit using insider information, which are not available to public.

And finally, market under strong form of efficiency does reflect immediately historical prices and both public and private information related to the market. This would imply, that even with insider information, the investor is unable to make extraordinary profit by any means.

2.2 Evidence of efficiency on commodity futures markets

Algieri & Kalkuhl (2014) thoroughly examined the role of futures market in stabilizing spot prices. Authors analyzed on set of commodity futures what factors may have effect on forecast errors and differences between realization and prediction of future spot prices. Findings of their analysis are as follows, relevant factors that drive forecast errors up are a high level of realized price volatility, the lack of liquidity in the market, and a longer contract maturity horizon. Further, they found out that maize, soybean and wheat markets are not fully informational efficient and investors can profit on them.

Kaminsky & Kumar (1990) examined excess returns in seven different commodity market over the 1976 - 1988 period. When taking in account this whole period, the markets showed no signs of inefficiency for short - term forecast horizont and the excess returns were statistically insignificant. However, further detailed analysis of sub periods showed that for some commodities, like cocoa and wheat, there are excess returns significantly positive. Further they stated that empirical rejection of efficient market hypothesis does not necessarily imply market failure and can happen from different reasons. One explanations suggested by the authors is that returns reflects non zero risk premium instead of market failure. Second possible explanation is, that the processes generating spot prices are changing.

In Wang & Ke (2005), authors studied efficiency of soybean and wheat markets in China. As result, they found out long term equilibrium between soybean futures and soybean spot prices. The long - run efficiency is implied by soybean futures prices being reliable predictor of spot prices. Furthermore, they studied short term weak efficiency in soybean futures market. In this part of the study, they discovered that for some soybean spot contracts there can still be inefficiency in short term and for such spot prices, futures are not good price predictor. On the other hand, wheat markets in China are generally inefficient, because wheat futures prices are not co-integrated with any wheat spot prices. Suggested explanation is in market failure, which caused by manipulation from large traders and by government regulation.

Authors of Kristoufek & Vosvrda (2014) examined market efficiency of 25 commodity futures across various groups such as metals, energies, softs, grains and other agricultural commodities. Their findings are, that affiliation to some group of commodity have great impact on efficience, because commodity futures belonging to same group share several common properties. Further, the authors found out that energy futures seems to be the most efficient, while agricultural commodities, especially livestock, are the least efficient group.

The paper Tomek (1997) discuss differences between of econometrics models and futures markets as forecasts for spot prices. Because new information can arise between the time the forecast is made and the price is realized, Tomek suggested that futures price can be an unbiased forecast of spot price, but it is likely to have large variance of forecast error. Further, he suggests that forecasting spot prices using econometrics models can have smaller variances of forecast error than usage of futures prices. In the paper, it is also mentioned that different commodity markets are likely to differ in their efficiency.

In Fama & French (1987) examined two different ways how to view differences between present commodity futures and spot prices. While theory of storage views this difference as result of expenses related to storing given commodity, alternative view splits a future price into expected premium and a forecast of the spot price. This idea was further extended in Kellard *et al.* (1999). In addition, authors of Kellard *et al.* (1999) examined influence of forecasting period on efficiency and found out that at 28 days forecast horizont, the commodities were relatively less efficient then for 56 days forecast horizont. In short run, markets are not efficient, because there is limited information about the period. On the other hand, in long run, efficiency is present as the markets have enough time to include all relevant information into consideration.

Paper Chinn & Coibion (2014) examines the predictive content of commodity futures prices for group of energy, agricultural, precious and base metals commodities. They found out that there are large differences both across and within these groups. In general, both precious and base metals are poor predictors of subsequent price changes. On the other hand, energy and agricultural futures are quite good predictors, with exception of oil futures. Further, they study influence of liquidity on the predictive content of commodity futures prices. They used traded volume as a measure of liquidity instead of open interest, since it proved to be reliable measurement of how many contracts have been traded relative to contracts outstanding. As a result, they found out that market liquidity has little influence on price changes.

Relations between market predictability and liquidity is further discussed in Chordia *et al.* (2008). Authors state hypothesis about existence of predictability in short run. There are other financial market attributes such as liquidity which are related to market returns. They examined NYSE (New York Stock Exchange) during the 1993 - 2002 period and found out that liquidity plays important role in creation of efficiency. Conclusion is, predictability is severely diminished during high - liquidy period. Therefore, liquidity improves efficiency of given market.

Chapter 3

Methodology

This chapter discuss ways of forecasting commodity futures returns and methods how to evaluate its forecasting performance. First part of this chapter is dedicated to introducing ARMA models and explaining how to use them, second part is describing issue of forecasting methods and its differences, while last part focuses on methods how to select best performing model. Although all methods described in this chapter can be easily further extended for multiple steps forecasts, for needs of this work, one - step ahead forecasts will be sufficient, since we are interested in maximising forecasting accuracy. It is reasonable to expect that one-step ahead forecasts will be more accurate than two and more step forecasts. Forecasting accuracy is crucial for making any conclusions about market efficiency and therefore for purposes of this thesis.

3.1 ARMA models

Autoregressive moving average (ARMA) family of models are group of models commonly used to do financial time series analysis. The model ARMA(q, p) consists of two parts. As suggested in Wooldridge (2012), ARMA models can be easily described by equations, as it consists of two main components. First is autoregressive (AR) part of order q, described by equation 3.1

$$y_t = \alpha + \sum_{i=1}^{q} \beta_i y_{t-i} + \epsilon_t \tag{3.1}$$

Second part is moving average (MA) of order p, described by equation 3.2

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i \epsilon_{t-i} + \epsilon_t \tag{3.2}$$

When both parts are put together, we obtain ARMA(q,p) model, as decribed by equation 3.3

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i \epsilon_{t-i} + \sum_{i=1}^{q} \beta_i y_{t-i} + \epsilon_t \tag{3.3}$$

Main reason for using ARMA models is their simplicity. ARMA models are easy to deploy and interpret because of the their linearity, yet they provide reasonably accurate results. Linearity of ARMA models also ensure that methods used for regression are not very demanding for computational power and are quick to carry out. This can be deciding factor especially when using large data samples. In case of more complicated linear or nonlinear methods, it would most likely take immense time to compute and obtain desired results.

3.1.1 Assumptions for using ARMA models

In order to have unbiased and consistent estimate when using ARMA models, we need to assume stationary of given time serie. As we can seen in figure 3.1, prices of CBOT Wheat Futures over ten years period (10.1.2004 - 10.1.2014) does not seem to be stationary.

Figure 3.1: CBOT Wheat Futures

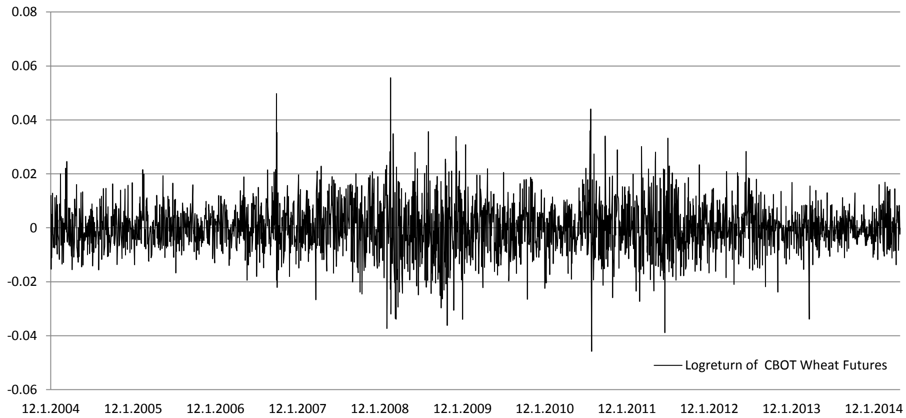
Source: CME Group Inc.

To resolve this issue, we can transform using first differencing of prices to obtain returns, as suggested inWooldridge (2012). It is also suitable, in order to further stabilize variance of return series, to use logarithmic transformation.

$$logreturn_t = ln(price_t) - ln(price_{t-1}) \tag{3.4}$$

Figure 3.2 shows CBOT Wheat Futures logreturns series over the same period after the transformation.

Figure 3.2: Logreturn of CBOT Wheat Futures



Source: CME Group Inc.

Resulting log return series appears to be stationary around zero. To further confirm this, we can use Augumented Dickey Fuhler test for detecting unit root, as described in Said & Dickey (1984). ADF test statistics follows t distribution. The structure of the test is as follows:

$H_0 : ADF = 0$; data needs to be further differenced and are not stationary

$H_1 : ADF < 0$; data appears to be stationary and does not need further differencing.

We are interested in rejecting the null hypothesis, therefore negative values are appreciated. Testing logreturns of wheat futures yields $ADF = -13.7$ which is enough for rejection of the null hypothesis for all considered levels of significance.

On the other hand, testing prices of CBOT Wheat Futures yields $ADF = -2.64$ which is not enough to reject null hypothesis at 95% significance level. This suggests that differencing was necessary.

It is reasonable to assume that this transformation will transform series of prices to stationary logreturn series for various types of commodity futures, not only for wheat.

3.2 Model selection methods

While there exists several methods how to select optimal configuration of q and p parameters in ARMA(p,q) such as cross validation or Akaike information criterion mentioned in Hastie *et al.* (2009), for this work simple one step ahead out - of - sample analysis will be sufficient method of choosing model with best forecasting performance. This method will ensure that any possible pattern within the data is detected. In addition, apparent advantage of using this method are lower demands for computational power. It is likely, that using more complicated methods would be enormously time consuming if able at all.

3.2.1 In-sample and Out - of - sample analysis for one - step ahead forecasts

As described in Hastie *et al.* (2009), we considering general model $y_t = f(X_t)$. Be X_t vector of explanatory variables (in case of ARMA model it will be either or both MA and AR parts of some order) and be y_t dependent variable for time $t = 1 \dots h$, where time h is the most recent observation.

By doing simple in - sample analysis we use all available data to construct the model. Then, we compare the models fitted values to real values previously used to estimate the model. Output will be sequence of differences between fitted values and real values, the forecast error series. While this procedure is simple and easy to deploy, doing so can give us very unreliable results because this method is very susceptible to over fitting and therefore the model will be describing irrelevant white noise instead of true long term pattern

More suitable way to select best performing forecasting model is out - of sample analysis. At first, it is similar to in - sample method. We use all available data from time 1 to time h to construct the model. Then, unlike the in - sample method, we try to use this model to forecast values of y for time $h + 1$. In case of general model, the forecast error is shown in equation 3.5.

$$e_{h+1} = y_{h+1} - \hat{f}(X_{h+1}) \tag{3.5}$$

Then, we compare our forecasted value to the real value when it is available. This procedure should be repeated in order to obtain forecast error series. While this procedure can give us realistic view on the models performance, it is very time consuming because we need to wait some time till the real $h + 1$ data are available.

One solution to this problem is simulating this procedure on smaller sample. This method is also sometimes called pseudo out - of - sample forecast. We select given some historical time $k, k < h$ and regress the model on sample $1 \dots k$. Then, we can immediately compare our forecasted value to real value. In order to obtain forecast error series, this procedure should be repeated for each time $k + 1, k + 2 \ldots$ and finally for h. Output of this method are errors for each time in $k \ldots h$. However, there are several different approaches how to carry this out, namely expanding window and rolling regression with or without re-estimation.

3.2.2 One step ahead rolling window regression with re-estimation vs expanding window

When performing out - of - sample analysis using expanding window, we regress again whole dataset for each time shift. Therefore, we obtain first forecast error from models using dates $1 \ldots k$, second from $1 \ldots k+1$, then repeat and finally using whole sample $1 \ldots h$. Each time we need to regress the model again, once our simulated forecast is compared to real value and then the new information is included into the regression.

On the other hand, when carrying out rolling window method, we shift whole window of the sample we are using, within the dataset. Hence, we obtain first forecast error from models using time $1 \ldots k$, then from $2 \ldots k + 1$, then repeat and finally $h - k \ldots h$. This procedure can be repeated with only one regression in the beginning, or we can re-estimate the model each time we add new observation. Apparent drawback of using rolling window without re-estimation is loss of relevance for large error series which may result in loss of overall accuracy.

To conclude, rolling window is better at capturing relevant time frame than expanding window, because the internal structure of time series can vary in time. For our needs it is more suitable to use rolling window with re-estimation.

3.3 Accuracy measurements - comparing forecast error series

From previous section, we are able to generate forecast error series using simulated out - of - sample method. Now we need to compare the each considered specification of ARMA (p,q) in terms of forecasting performance. This can be done by using forecast error series generated for each ARMA(p,q) specification. Methods how to do so are discussed further in this section.

3.3.1 Average of forecasting error methods

One of the simplest methods of measuring accuracy of forecast is suggested in Wooldridge (2012). Basis of this method is using simple average of errors to compare two error series. In order to get reliable information about the accuracy of forecast, simple average of forecast errors by itself is not enough, because the error series is symmetric and usually have both positive and negative values. Therefore, we need to either square each error or get it's absolute value. This procedure will result into obtaining series of absolute errors

$$|\hat{e}|_{n+1} = |y_{n+1} - \hat{y}_{n+1}| \tag{3.6}$$

or squared errors:

$$\hat{e}^2_{n+1} = (y_{n+1} - \hat{y}_{n+1})^2 \tag{3.7}$$

To obtain information about the performance of our method, we simply calculate average absolute error or average squared error by summing the series of these series and dividing by number of observations. Resulting values are mean are absolute error (MAE):

$$MAE = \frac{\sum\limits_{t=1}^{n} |y_{n+1} - \hat{y}_{n+1}|}{n} \tag{3.8}$$

or rooted mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum\limits_{t=1}^{n} (y_{n+1} - \hat{y}_{n+1})^2}{n}} \tag{3.9}$$

We can say, that lower MAE or RMSE generally means that the model has better forecasting accuracy. While having this property in common, there are some differences in interpreting these two measurements. In some cases when comparing two models to each other, one model can have lower MAE while having higher RMSE than the second model. This issue is caused by the very nature of these measurements. By using MAE, we adjust to each difference between forecast and real value same weight, while in case of using RMSE, we take squared difference. Therefore, large differences between forecasts and real values have higher weight in case of RMSE. Generally, we can say that MAE is better in measuring overall accuracy, while RMSE is better at capturing substantially large forecast errors.

3.3.2 Diebold - Mariano test

Although MAE or RMSE measurements can roughly suggest which model might be better, they cannot tell us whether this difference is statistically significant. In order to obtain reliable results, we need to perform formal statistical test. Suitable method is, as suggested in Diebold & Mariano (2002), Diebold Mariano test which uses proper statistical inference to compare two error series. Structure of the test is as follows:

$H_0 : DM = 0$; two models have same accuracy

$H_1 : DM \neq 0$; two models have different accuracy

The DM test statistics is asymptotically $N(0, 1)$ distributed. Therefore, if DM statistics falls outside interval $(-z_{\alpha/2}, z_{\alpha/2})$, we can reject H_0 and we know that one model is better than the other in terms of forecasting accuracy.

Input to this test can be both absolute and squared error series. Similarly to the forecasting error average absolute version of this test is better for comparing overall accuracy of given models, while squared version is more sensitive for rejection when large errors occurs in the analysis. For purposes of this work, absolute error series version of the DM test is used.

Chapter 4

Application methods on various commodities

This chapter is dedicated to application of methods from previous chapter on various commodity futures datasets. Generally, we can divide commodity futures into several sub categories. First group of commodity futures contains agricultural products. This group is further divided to three subgroups: grains (wheat, corn, soybean), soybean related products (soybean, soybean oil, soybean meat) and livestock (live cattle, and lean hogs). Next examined group contains commodities related to energy industry. Here can be found crude oil futures, heating oil futures and natural gas futures. Third important group of commodity futures consists of precious metals, namely silver, gold, platinum and palladium.

All data used are from CME group Inc. (Chicago Mercantile Exchange & Chicago Board of Trade) exchanges, as they own one of the largest market places for precious metals (NYMEX), energy futures (NYMEX) as well as agricultural products futures (CBOT). Other commodity market places approximately follows same trends and the prices are in general very similar.

4.1 Specifications for the analysis

First, we need to select proper method of model selection. To analyze market efficiency of given futures market, optimal way seems to be testing selected specification of ARMA model against simple white noise on selected data frame. There are several parameters which must be specified in order to do proper analysis. As described in previous chapter, suitable option seems to be pseudo out - of - sample analysis with rolling window for one step ahead forecasts. Using this method with combination of Diebold Mariano test, we can see whether there is or is not any pattern in the data over given time period and whether this pattern is significantly different from the noise.

4.1.1 Time frame

Then, we need to select time frame: starting date and period over which are we trying to forecast values. In this work the examined period is starting from 10.1. 2004 and forecasts are to date 10.1. 2014. We are taking one trading year as approximately 250 days, hence the length of forecast error series is about 2500 observations. Such length will provide us with enough information, which are required to make reasonable conclusions about the time period.

For each commodity there will be ARMA models of maximum combined order 3. This number should be enough, because ARMA models of larger orders tend to suffer from over fitting and are usually worse than white noise in terms of out - of - sample forecasting performance. Although both MAE and RMSE are included into the table, these measures are only indicative and main measure will be DM statistic. Of course, if DM statistic is positive, MAE of such model should be lower than MAE of white noise.

We are interested in rejecting null hypothesis, therefore we are looking for positive values around 1.96 which allows us reject null hypothesis at 95% level of significance. On the other hand, negative such as -1.92 values indicate that white noise is significantly performing better than given model. Generally, All values above 1.00 are interesting and should be further examined.

One way how to further examine such data is shortening the error time series and then examine forecast error sub sequence derived from the original error series, like Kaminsky & Kumar (1990) suggested. We assume, there can be patterns that are related to same smaller period than the ten years period we originally selected. If this period is large and significant enough, it may prevail in the results anyway. Considering smaller time frame, in which is given ARMA model significantly better than white noise, will inevitably increase DM statistics of analysis performed over whole ten years period. To decide whether increased values of DM statistic are indicators of such behavior will require adjusting the time frame and further analysis in general.

4.1.2 Window size

Another important parameter when using rolling window regression is size of the window. For analysis carried out in this work are used 250, 500 and 750 window sizes. These values should approximately have number of observation similar to 1, 2 and 3 trading years. We can expect that with larger window size values of DM statistics will fluctuate around zero and the test will tend to be inconclusive as the model tries to fit larger sample. If there is any pattern in the data, it is reasonable to expect, that it will prevail at all sizes of window, even the smaller ones.

4.2 Precious metals

This group of commodity futures consists of gold futures, silver futures, platinum and palladium futures. Used data are from New York Mercantile Exchange (NYMEX) , exchange owned by CME group. This marketplace is suitable representative of precious metal market because is among the largest, if not the largest, precious metals market places in the world. Prices on other large exchanges, such as Singapore Mercantile or London Metal Exchange, usually follow similar development of prices. All precious metals contract units are in Troy Ounce which is approximately 31.1 grams. Stevenson (2010)

4.2.1 Gold

According to CME webpage, size of this type contract is 100 troy ounces. Symbol of this contract is GC1. Results of the analysis are below in table 4.1.

Table 4.1: Analysis of gold futures (GC1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.252	0.00913	0.01278	-0.274	0.00913	0.01276	-0.23	0.00912	0.01275
ARMA(1,0)	-0.257	0.00913	0.01279	-0.265	0.00913	0.01276	-0.217	0.00912	0.01275
ARMA(0,2)	-0.654	0.00917	0.01284	-0.538	0.00915	0.01279	-0.439	0.00914	0.01278
ARMA(1,1)	-0.518	0.00916	0.01282	-0.442	0.00914	0.01278	-0.365	0.00914	0.01275
ARMA(2,0)	-0.585	0.00916	0.01284	-0.502	0.00915	0.01279	-0.427	0.00914	0.01278
ARMA(0,3)	-1.043	0.00921	0.01289	-0.769	0.00918	0.01282	-0.537	0.00915	0.01279
ARMA(1,2)	-1.03	0.00921	0.01289	-0.656	0.00916	0.01281	-0.49	0.00915	0.01278
ARMA(2,1)	-0.957	0.0092	0.01288	-0.716	0.00917	0.01281	-0.677	0.00917	0.0128
ARMA(3,0)	-0.903	0.00919	0.01289	-0.716	0.00917	0.01282	-0.511	0.00915	0.01279
White noise	-	0.0091	0.01274	-	0.0091	0.01272	-	0.0091	0.01272

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

Analysis of gold futures shows that we do not have enough evidence to reject hull hypothesis. Further we can see, that with increasing order of both MA and AR, the performance of models is getting worse. Size of rolling window does not have noticeable influence on the results. Using white noise to predict this time series seems to be the best option because values of DM statistics are always negative and both RMSE and MAE are lower for the noise than for the ARMA models. In other words, for forecasting of gold futures returns we can not find better model than noise and therefore this market is most likely efficient.

4.2.2 Silver

Standard size of one silver futures contract (SI1) on NYMEX is 5000 troy ounces. Results of analysis are in table 4.2.

Table 4.2: Analysis of silver futures (SI1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.681	0.01645	0.02341	-0.237	0.01639	0.02332	-0.403	0.01641	0.02334
ARMA(1,0)	-0.524	0.01643	0.0234	-0.185	0.01639	0.02332	-0.361	0.01641	0.02333
ARMA(0,2)	-1.445	0.01653	0.02349	-0.841	0.01645	0.02338	-0.662	0.01644	0.02336
ARMA(1,1)	-1.164	0.0165	0.02346	-0.443	0.01641	0.02333	-0.62	0.01643	0.02335
ARMA(2,0)	-1.243	0.01651	0.0235	-0.678	0.01644	0.02337	-0.606	0.01643	0.02336
ARMA(0,3)	-2.017	0.01658	0.02356	-1.097	0.01648	0.02343	-0.793	0.01645	0.02336
ARMA(1,2)	-1.711	0.01655	0.02352	-0.878	0.01646	0.0234	-0.826	0.01645	0.02336
ARMA(2,1)	-1.586	0.01654	0.02353	-0.84	0.01645	0.02338	-0.791	0.01645	0.02336
ARMA(3,0)	-1.782	0.01656	0.02355	-1.013	0.01647	0.02342	-0.784	0.01645	0.02337
White noise	-	0.01638	0.02331	-	0.01637	0.02327	-	0.01637	0.02328

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

Returns of silver futures show similar behavior like in case of gold futures. Comparing ARMA models to white noise will again always yield negative DM statistics. With increasing order of both AR and MA part, the DM statistics tend to be more negative and size of the window does not seem to matter, similarly to gold futures. Therefore, we can assume that there will be no model of higher order that would be better performing than white noise and therefore we can say silver futures market is also likely efficient.

4.2.3 Platinum

NYMEX standard for this commodity contract is 50 troy ounces. Symbol of this contract type is PL1. Table 4.3 shows results of analysis of platinum.

Table 4.3: Analysis of platinum futures (PL1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.543	0.01075	0.01503	-0.458	0.01071	0.01497	-0.389	0.0107	0.01495
ARMA(1,0)	-0.511	0.01075	0.01503	-0.406	0.01071	0.01496	-0.344	0.01069	0.01495
ARMA(0,2)	-0.893	0.01079	0.01508	-0.557	0.01072	0.01499	-0.48	0.01071	0.01496
ARMA(1,1)	-1.051	0.0108	0.01509	-0.625	0.01073	0.01497	-0.576	0.01072	0.01497
ARMA(2,0)	-0.891	0.01079	0.01508	-0.551	0.01072	0.01499	-0.476	0.01071	0.01496
ARMA(0,3)	-1.204	0.01082	0.01515	-0.731	0.01074	0.01504	-0.525	0.01071	0.01499
ARMA(1,2)	-1.568	0.01085	0.01517	-0.832	0.01075	0.01502	-0.664	0.01072	0.01498
ARMA(2,1)	-1.171	0.01081	0.01515	-0.77	0.01074	0.01503	-0.709	0.01073	0.015
ARMA(3,0)	-1.131	0.01081	0.01514	-0.702	0.01074	0.01504	-0.542	0.01071	0.01499
White noise	-	0.0107	0.01501	-	0.01067	0.01497	0	0.01066	0.01496

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

Similarly to gold and silver futures, for all measured models the DM statistics are negative. Although we can observe tendency of DM statistic to grow with increasing size of window, it does not necessarily mean that higher widow size will have better forecasting performance. More straightforward explanation is, that with larger widow size the model will take more observations into consideration and will be more similar to white noise, and therefore the DM statistic will most likely converge to zero as size of window goes to larger values. Taking this in account, we can conclude that platinum futures market is efficient as well.

4.2.4 Palladium

Last analysed precious metal futures traded on NYMEX are palladium futures. Standard size of this palladium contract (PA1) is 100 troy ounces. Results of analysis for paladium futures are in table 4.4.

Table 4.4: Analysis of palladium futures (PA1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.477	0.01588	0.02219	-0.316	0.01588	0.02211	-0.191	0.01587	0.0221
ARMA(1,0)	-0.43	0.01588	0.02219	-0.269	0.01588	0.02211	-0.175	0.01586	0.02209
ARMA(0,2)	-0.788	0.01591	0.02219	-0.183	0.01587	0.02211	-0.188	0.01587	0.02209
ARMA(1,1)	-1.141	0.01595	0.02229	-0.568	0.01591	0.02217	-0.411	0.01589	0.02212
ARMA(2,0)	-0.786	0.01591	0.02219	-0.149	0.01586	0.02211	-0.178	0.01587	0.0221
ARMA(0,3)	-0.62	0.01589	0.02226	0.051	0.01584	0.02212	0.016	0.01585	0.0221
ARMA(1,2)	-0.719	0.0159	0.02224	-0.658	0.01591	0.02219	-0.505	0.0159	0.02214
ARMA(2,1)	-0.774	0.01591	0.02226	-0.436	0.01589	0.02216	-0.179	0.01587	0.02211
ARMA(3,0)	-0.803	0.01591	0.02225	-0.009	0.01585	0.02211	0.02	0.01585	0.02209
White noise	-	0.01583	0.02212	-	0.01585	0.02208	-	0.01585	0.02208

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

We can observe similar behavior to platinum futures, with increasing size of window the ARMA models tend to improve. However, unlike platinum, DM statistic of some ARMA models for palladium are slightly positive. Although we cannot reject null hypothesis right away, this case still requires more attention. In similar way as Kaminsky & Kumar (1990) suggested, we can expect existence of reasonably large sub period within the original ten years' time period, where the AR(3) and MA(3) is able to outperform the white noise. Using simple methods like bisection we should be able to identify such period within the original time frame if it exists. Table 4.5 shows results of analysis for sub period of 2010 - 2012.

Table 4.5: Analysis of palladium futures on sub period

	500 window size		
Model	DM	MAE	RMSE
ARMA(3,0)	1.825	0.0177	0.0228
ARMA(0,3)	1.851	0.01769	0.02281
White noise	-	0.01791	0.0229

Source: computation for skript parameters: w = 500, s=5, z = 2

For this script configuration, we can reject null hypothesis of DM test for reasonably large significance level. In other words, both AR (3) and MA(3) models outperformed white noise in terms of forecasting accuracy for the period of 2010 - 2012, using data from 2009 - 2010 (rolling window uses dataset of 500 observations behind the forecasted values). This case is quite extraordinary, and such behavior would violate EMH. Therefore, this case needs our further attention. Reasonable step would be analyzing liquidity of the marked during this period, which will be concern of the next chapter.

4.3 Agricultural - Grains

In this group of commodities, the contract size are usually in bushel unit of volume in US Customary System.According to Stevenson (2010), one US bushel unit is approximately 0.03524 cubic meters. Main concern of our analsis are corn, wheat and soybean futures. Further, we analysed soybean products such as soybean meat and soybean oil.

4.3.1 Corn

Contract size is usually 5000 bushels, symbol of this futures on Chicago board of trade market place is C1. Results of analysis for corn futures can be found in table 4.6.

Table 4.6: Analysis of corn futures (C1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.216	0.01512	0.02035	-0.186	0.01508	0.02029	-0.248	0.01509	0.02029
ARMA(1,0)	-0.17	0.01511	0.02035	-0.167	0.01508	0.02029	-0.257	0.01509	0.02029
ARMA(0,2)	-0.546	0.01515	0.02039	-0.39	0.0151	0.02033	-0.263	0.01509	0.02031
ARMA(1,1)	-0.758	0.01517	0.02041	-0.391	0.0151	0.02033	-0.401	0.01511	0.02031
ARMA(2,0)	-0.509	0.01515	0.02039	-0.421	0.0151	0.02033	-0.297	0.0151	0.02031
ARMA(0,3)	-0.695	0.01516	0.02044	-0.587	0.01512	0.02037	-0.536	0.01512	0.02034
ARMA(1,2)	-0.984	0.01519	0.02046	-0.795	0.01514	0.02036	-0.577	0.01512	0.02033
ARMA(2,1)	-1.212	0.01522	0.0205	-0.728	0.01513	0.02037	-0.699	0.01514	0.02035
ARMA(3,0)	-0.754	0.01517	0.02044	-0.694	0.01513	0.02037	-0.567	0.01512	0.02035
White noise	-	0.01509	0.0203	-	0.01506	0.02027	-	0.01507	0.02027

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

We can observe that DM statistic are further being pushed to negative values with both growing order of ARMA model and size of window. Neither MAE or RMSE of any ARMA model are lower than white noise for all selected configurations. Therefore it is reasonable to say that corn futures returns cannot be described by any ARMA model and appear to not follow any pattern. This implies market efficiency for this type of futures contract.

4.3.2 Wheat

Again, contract size is 5000 bushels. Contract symbol is W1. Results are in table 4.7.

Similarly to case of corn futures, we are not able to observe any positive value of DM statistic, and average measures of ARMA error series are in each case worse than white noise . Apparent difference to corn case is that with increasing size of window DM statistics tent to be growing. However, even with 750 size of window they does not get positive, therefore it is reasonable to assume that the values of DM statistics will converge to zero, instead of continuous growth and possible rejection of null hypothesis. We can conclude that wheat futures were overalls efficient during this time period.

Table 4.7: Analysis of wheat futures (W1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.563	0.01644	0.02183	-0.239	0.01642	0.02177	-0.182	0.01639	0.02175
ARMA(1,0)	-0.455	0.01642	0.02181	-0.239	0.01642	0.02177	-0.234	0.0164	0.02175
ARMA(0,2)	-0.869	0.01647	0.02191	-0.242	0.01642	0.02182	-0.341	0.01641	0.02178
ARMA(1,1)	-0.707	0.01645	0.02189	-0.497	0.01645	0.0218	-0.493	0.01642	0.02178
ARMA(2,0)	-0.762	0.01646	0.0219	-0.233	0.01642	0.02182	-0.371	0.01641	0.02178
ARMA(0,3)	-1.299	0.01651	0.02195	-0.444	0.01644	0.02186	-0.689	0.01644	0.02182
ARMA(1,2)	-1.525	0.01653	0.02198	-0.199	0.01642	0.02183	-0.434	0.01642	0.02179
ARMA(2,1)	-1.305	0.01651	0.02195	-0.482	0.01644	0.02186	-0.415	0.01641	0.02181
ARMA(3,0)	-1.287	0.01651	0.02196	-0.395	0.01644	0.02186	-0.689	0.01644	0.02182
White noise	-	0.01638	0.02175	-	0.0164	0.02173	-	0.01637	0.02172

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

4.3.3 Soybean

Contract size is again 5000 bushel, while contract symbol is S1. Results can be found in table 4.8.

Table 4.8: Analysis of soybean futures (S1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.497	0.01305	0.01776	-0.182	0.013	0.01769	-0.115	0.013	0.01768
ARMA(1,0)	-0.493	0.01305	0.01776	-0.175	0.013	0.01769	-0.101	0.013	0.01768
ARMA(0,2)	-0.807	0.01308	0.01782	-0.362	0.01302	0.01772	-0.373	0.01302	0.01771
ARMA(1,1)	-0.816	0.01308	0.01782	-0.305	0.01301	0.01772	-0.299	0.01302	0.01772
ARMA(2,0)	-0.883	0.01309	0.01783	-0.363	0.01302	0.01772	-0.363	0.01302	0.01771
ARMA(0,3)	-1.231	0.01312	0.01787	-0.504	0.01303	0.01776	-0.382	0.01303	0.01773
ARMA(1,2)	-0.924	0.01309	0.01786	-0.629	0.01305	0.01775	-0.54	0.01304	0.01774
ARMA(2,1)	-1.134	0.01311	0.01786	-0.539	0.01304	0.01775	-0.821	0.01307	0.01776
ARMA(3,0)	-1.048	0.0131	0.01787	-0.423	0.01303	0.01775	-0.377	0.01302	0.01773
White noise	0	0.013	0.01769	0	0.01298	0.01766	0	0.01299	0.01765

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

Soybean futures shows similar behavior to corn and wheat. No model shows better forecasting performance than white noise. With increasing size of window the DM statistics are closer to zere. Again, this is probably just result of convergence to zero, not evidence of improvement.

4.3.4 Soybean oil

Units used for measuring volume of this commodity are pounds, where one pound is approximately 0.453 Kg. Standardized contract size of soybean oil (BO1) is 60,000 pounds. Results for this contract type are in table 4.9.

Table 4.9: Analysis of soybean oil futures (BO1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.331	0.01217	0.01639	-0.077	0.01214	0.01634	-0.059	0.01215	0.01633
ARMA(1,0)	-0.376	0.01217	0.0164	-0.085	0.01214	0.01634	-0.066	0.01215	0.01633
ARMA(0,2)	-0.573	0.01219	0.01643	-0.248	0.01216	0.01636	-0.297	0.01217	0.01635
ARMA(1,1)	-0.713	0.01221	0.01643	-0.268	0.01216	0.01636	-0.199	0.01216	0.01634
ARMA(2,0)	-0.666	0.0122	0.01643	-0.264	0.01216	0.01637	-0.301	0.01217	0.01635
ARMA(0,3)	-0.902	0.01223	0.01647	-0.369	0.01217	0.01639	-0.365	0.01218	0.01637
ARMA(1,2)	-1.193	0.01226	0.0165	-0.486	0.01218	0.0164	-0.429	0.01219	0.01637
ARMA(2,1)	-0.76	0.01221	0.01645	-0.469	0.01218	0.01638	-0.255	0.01217	0.01636
ARMA(3,0)	-0.905	0.01223	0.01647	-0.368	0.01217	0.01639	-0.362	0.01218	0.01637
White noise	-	0.01214	0.01633	-	0.01213	0.01632	-	0.01214	0.01631

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

4.3.5 Soybean meat

For measuring soybean meat are used short tons, one short ton is approximately 907,2 Kg. Size of standard SM 1 Contract is 100 short tons. Results for soybean meat futures can be found in table 4.10.

Table 4.10: Analysis of soybean meat futures (SM1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.39	0.01427	0.01932	-0.133	0.01421	0.01923	-0.137	0.01421	0.01922
ARMA(1,0)	-0.419	0.01427	0.01933	-0.14	0.01421	0.01923	-0.152	0.01421	0.01922
ARMA(0,2)	-0.707	0.0143	0.01938	-0.278	0.01423	0.01926	-0.345	0.01423	0.01926
ARMA(1,1)	-1.138	0.01434	0.01941	-0.315	0.01423	0.01926	-0.346	0.01423	0.01925
ARMA(2,0)	-0.7	0.0143	0.01938	-0.313	0.01423	0.01926	-0.372	0.01423	0.01926
ARMA(0,3)	-0.978	0.01433	0.01943	-0.431	0.01424	0.01929	-0.543	0.01425	0.01928
ARMA(1,2)	-1.467	0.01438	0.01944	-0.578	0.01426	0.01928	-0.503	0.01425	0.01926
ARMA(2,1)	-1.383	0.01437	0.01946	-0.65	0.01426	0.01928	-0.583	0.01425	0.01927
ARMA(3,0)	-1.001	0.01433	0.01943	-0.512	0.01425	0.01929	-0.579	0.01425	0.01928
White noise	-	0.01423	0.01926	-	0.0142	0.01922	-	0.0142	0.01922

Source: computation for skript parameters: (v = 3, w = 250/500/750, s=11, z = 10)

Both soybean oil and soybean meat display similar behavior. DM tests are inconclusive for every ARMA model and they are converging to zero with increasing window size. There is no reason to doubt the efficiency of both markets over given time period.

4.4 Agricultural - Livestock

This group consists of live cattle futures and lean hog futures. Standardized contract size is 40,000 pounds, which is approximately 18 tons. Symbol of the live cattle contract is LC1 and symbol of the lean hog contract is LN1.

4.4.1 Live cattle

Table 4.11: Analysis of live cattle futures (LC1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	0.045	0.00696	0.0092	0.095	0.00696	0.0092	0.1	0.00696	0.0092
ARMA(1,0)	-0.009	0.00697	0.0092	0.094	0.00696	0.0092	0.088	0.00696	0.00919
ARMA(0,2)	-0.17	0.00699	0.00921	0.011	0.00697	0.0092	0.006	0.00697	0.00919
ARMA(1,1)	-0.123	0.00698	0.00921	0.036	0.00697	0.0092	0.045	0.00696	0.00918
ARMA(2,0)	-0.176	0.00699	0.00921	-0.03	0.00698	0.0092	0.006	0.00697	0.00918
ARMA(0,3)	-0.399	0.00701	0.00923	-0.207	0.00699	0.00922	-0.133	0.00698	0.0092
ARMA(1,2)	-0.368	0.00701	0.00924	-0.14	0.00699	0.00922	-0.086	0.00697	0.0092
ARMA(2,1)	-0.351	0.007	0.00924	-0.095	0.00698	0.00921	-0.059	0.00697	0.00919
ARMA(3,0)	-0.369	0.00701	0.00924	-0.175	0.00699	0.00922	-0.099	0.00698	0.0092
White noise	-	0.00697	0.00923	-	0.00697	0.00923	-	0.00697	0.00923

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

As we can see in table 4.11, DM statistic is positive for ARMA (0,1) even in case of 250 window size and is increasing with growing size of the window. Both MAE and RMSE are lower for ARMA models than they are in case of white noise. Although this is good step, DM statistic is nowhere near levels required for rejecting null hypothesis. This could either mean that this behavior is over the whole period but is non-significant, or that this type of commodity can be predicted on smaller than ten years' time period. To decide which case are live cattle futures we specify the script for suitable sub period, using bisection. Results of this extended analysis are in table 4.12.

Table 4.12: Analysis of live cattle futures on sub period

	500 window size		
Model	DM	MAE	RMSE
ARMA(0,1)	0.323	0.00807	0.01025
ARMA(1,0)	0.38	0.00806	0.01025
ARMA(0,2)	-0.062	0.00811	0.01026
ARMA(1,1)	0.443	0.00806	0.01024
White noise	0	0.0081	0.01028

Source: computation for skript parameters: w = 500, s = 4, z = 1

By performing further analysis we can observe that values of DM statistics are usually around zero. Best performance of ARMA (1,1) model was around 2010 - 2011 period, when DM statistic was around 0,44 . However, even this value is not enough to reject null hypothesis. We can conclude that this market is unpredictable by any ARMA model on time frame longer than one year.

4.4.2 Lean hog

Table 4.13: Analysis of lean hog futures (LN1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	1.01	0.01092	0.01473	1.018	0.01091	0.0147	1.032	0.01091	0.01468
ARMA(1,0)	0.917	0.01093	0.01471	1.031	0.01091	0.01467	1.145	0.0109	0.01465
ARMA(0,2)	0.588	0.01097	0.01473	1.051	0.0109	0.01463	1.046	0.01091	0.0146
ARMA(1,1)	0.526	0.01097	0.01475	1.017	0.01091	0.01464	1.243	0.01089	0.01458
ARMA(2,0)	0.483	0.01098	0.01474	1.063	0.0109	0.01462	1.117	0.0109	0.01458
ARMA(0,3)	0.305	0.01099	0.01476	0.909	0.01092	0.01464	1.058	0.0109	0.0146
ARMA(1,2)	0.108	0.01101	0.0148	0.908	0.01092	0.01464	1.082	0.0109	0.01459
ARMA(2,1)	-0.002	0.01103	0.01481	0.87	0.01092	0.01465	0.96	0.01091	0.0146
ARMA(3,0)	0.134	0.01101	0.01479	0.861	0.01092	0.01465	0.981	0.01091	0.01459
White noise	-	0.01103	0.01483	-	0.01101	0.01483	-	0.01101	0.01483

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

Results of analysis displayed in table 4.13 suggests, that this product is exceptional, because we can observe that DM statistics are for ARMA (1,0) above 1 even for 250 size of window. With increasing widow size values of DM statistics are generally growing across all orders of ARMA. Particularly interesting is that for each examined window size the best ARMA model has different configuration of orders. While for size of 250 best model is ARMA (1,0), in case of 500 window size the best model is ARMA (2.0) and for 750 window size it is ARMA(1,1). By doing further analysis, we can observe similar behavior like in live cattle care but DM statistic are generally substantially more over zero in this case. Some smaller periods within analyzed ten year time frame have lower DM statistics than aggregate results. To compensate this, in average there should exist also smaller time period within the ten year time frame, where the results are above average. Such period in this case is between 2005 to 2008 (Table 4.14), where we can reject null hypothesis for all analyzed ARMA models.

And we can even further extend for ARMA models of higher orders to see that this effect eccurs even at such orders of ARMA (Table 4.15).

In any case, we can strongly reject null hypothesis for all examined situation on the sub period. Therefore, we can conclude that any out of ARMA models, even of high orders, are better at forecasting performance than random white

Table 4.14: Analysis of lean hog futures on sub period

	750 window size		
Model	DM	MAE	RMSE
ARMA(0,1)	2.092	0.01023	0.01336
ARMA(1,0)	2.173	0.01022	0.01334
ARMA(0,2)	2.166	0.01022	0.01333
ARMA(1,1)	1.998	0.01024	0.01335
ARMA(2,0)	2.112	0.01023	0.01334
ARMA(0,3)	1.967	0.01024	0.01334
ARMA(1,2)	2.059	0.01023	0.01334
ARMA(2,1)	2.052	0.01023	0.01334
ARMA(3,0)	1.85	0.01025	0.01336
White noise	-	0.01044	0.01362

Source: computation for skript parameters: w = 750, s = 10, z = 3

Table 4.15: Extended analysis of lean hog futures on sub period

	750 window size		
Model	DM	MAE	RMSE
ARMA(5,0)	2.007	0.01024	0.01335
ARMA(0,5)	2.148	0.01022	0.01332

Source: computation for skript parameters: w = 750, s = 10, z = 3

noise. Such behavior would necessarily violate Efficient market hypothesis and require further attention to provide satisfactory explanation.

4.5 Energy

This group of futures consists of crude oil, heating oil and natural gas futures. Examined time frame is ten years period from 10.1. 2004 to 10.1. 2014, therefore size of forecast error series is 2500 observation.

4.5.1 Crude oil

Crude oil contract (CL1) are in US barrel units, where one barrel is approximately 35,2 liters Stevenson (2010) and standardized size of one NYMEX contract is 1000 barrels. Results can be found in table 4.16.

As we can see, the DM statistic is increasing with growing size of window. However, even at 750 window size, it did not cross zero. Although it is possible that at even higher window size the DM statistic values will be positive, it is highly unlikely that this such values will be enough to reject null hypothesis even at smaller time period. Therefore, it is safe to conclude that this market is efficient.

Table 4.16: Analysis of crude oil futures (CL1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.534	0.01575	0.02191	-0.23	0.01569	0.02185	-0.087	0.01565	0.02182
ARMA(1,0)	-0.556	0.01575	0.02191	-0.244	0.01569	0.02185	-0.106	0.01565	0.02182
ARMA(0,2)	-1.078	0.0158	0.02197	-0.718	0.01574	0.02191	-0.495	0.01569	0.02187
ARMA(1,1)	-1.341	0.01583	0.022	-0.355	0.0157	0.02188	-0.43	0.01568	0.02187
ARMA(2,0)	-1.047	0.0158	0.02197	-0.672	0.01573	0.02191	-0.47	0.01569	0.02187
ARMA(0,3)	-1.934	0.01589	0.0221	-1.013	0.01577	0.02198	-0.639	0.01571	0.02191
ARMA(1,2)	-1.501	0.01585	0.02204	-1.017	0.01577	0.02198	-0.472	0.01569	0.02188
ARMA(2,1)	-1.701	0.01587	0.02204	-0.972	0.01576	0.02194	-0.55	0.0157	0.0219
ARMA(3,0)	-1.917	0.01589	0.02208	-0.995	0.01577	0.02197	-0.596	0.0157	0.02191
White noise	-	0.0157	0.02186	-	0.01567	0.02181	-	0.01564	0.0218

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

4.5.2 Heating oil

This commodity futures are traded in US gallon unit, which is approximately 3,785 liters. Standardized size of one contract is 42 000 gallons of heating oil. Symbol used for this commodity is HO1.

Table 4.17: Analysis of heating oil futures (HO1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
White noise	0	0.01537	0.02053	0	0.01536	0.02051	0	0.01535	0.02049
ARMA(0,1)	-0.433	0.01541	0.02058	-0.1	0.01537	0.02052	0.029	0.01535	0.0205
ARMA(1,0)	-0.487	0.01542	0.02058	-0.125	0.01537	0.02052	0.024	0.01535	0.0205
ARMA(0,2)	-0.748	0.01545	0.02063	-0.344	0.01539	0.02054	0.035	0.01535	0.02051
ARMA(1,1)	-1.473	0.01552	0.02072	-0.226	0.01538	0.02055	-0.125	0.01536	0.02051
ARMA(2,0)	-0.741	0.01545	0.02064	-0.309	0.01539	0.02055	0.052	0.01534	0.02051
ARMA(0,3)	-0.695	0.01544	0.02065	-0.176	0.01538	0.02056	-0.007	0.01535	0.02055
ARMA(1,2)	-1.855	0.01556	0.02079	-0.401	0.0154	0.0206	-0.112	0.01536	0.02054
ARMA(2,1)	-1.481	0.01552	0.02071	-0.592	0.01542	0.02064	-0.12	0.01536	0.02055
ARMA(3,0)	-0.885	0.01546	0.02067	-0.203	0.01538	0.02057	0.023	0.01535	0.02054
White noise	-	0.01537	0.02053	-	0.01536	0.02051	-	0.01535	0.02049

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

In table 4.17, we can observe rapid increase of DM statistic with increasing window size. At size of 750, DM statistic is positive for some ARMA models. By further investigating this issue, we can select more specific parameters and analyze shorter time period within the original one where we would be able to reject null hypothesis. By using simple bisection for searching best values of DM statistic, we can obtain values displayed in table 4.18.

We can observe at 750 rolling window (w = 750, s=10, z = 1) that during year 2005, there was one year frame where we can reject null hypothesis at

Table 4.18: Analysis of heating oil futures on sub period

	750 window size		
Model	DM	MAE	RMSE
ARMA(0,2)	1.269	0.01904	0.02462
ARMA(2,0)	1.336	0.01903	0.02461
ARMA(2,1)	1.375	0.019	0.02456
White noise	0	0.0192	0.02484

Source: computation for skript parameters: (w = 750, s=10, z = 1)

smaller levels of significance and this case certainly requires further examination.

4.5.3 Natural gas

Because of gaseous state of this commodity, unit of one contract is British thermal units (BTU), which measures resulting calorific value of used natural gas. One unit of BTU is equivalent of 1,0546 Giga Joules or, under standard temperature and pressure, 28,26 cubic meters of natural gas as described in Stevenson (2010). Standard size of one natural gas futures contract(NG1) is 10,000 million British thermal units.

Table 4.19: Analysis of natural gas futures (NG1)

	250 window size			500 window size			750 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.598	0.02361	0.03115	-0.408	0.02355	0.03107	-0.45	0.02355	0.03104
ARMA(1,0)	-0.619	0.02361	0.03114	-0.477	0.02356	0.03107	-0.498	0.02355	0.03104
ARMA(0,2)	-1.722	0.02372	0.03124	-0.795	0.02359	0.0311	-0.637	0.02356	0.03108
ARMA(1,1)	-0.773	0.02363	0.03117	-0.604	0.02357	0.0311	-0.575	0.02356	0.03103
ARMA(2,0)	-1.471	0.0237	0.03123	-0.827	0.0236	0.03111	-0.635	0.02356	0.03108
ARMA(0,3)	-2.515	0.0238	0.03134	-1.219	0.02364	0.03113	-0.948	0.0236	0.03112
ARMA(1,2)	-2.209	0.02377	0.03135	-1.117	0.02363	0.03116	-0.79	0.02358	0.03111
ARMA(2,1)	-1.355	0.02369	0.03125	-0.828	0.0236	0.03112	-0.73	0.02357	0.03109
ARMA(3,0)	-2.219	0.02377	0.03132	-1.133	0.02363	0.03115	-0.891	0.02359	0.03112
White noise	-	0.02355	0.03111	-	0.02351	0.03106	-	0.0235	0.03102

Source: computation for skript parameters: v = 3, w = 250/500/750, s=11, z = 10

As can be seen in table 4.19, we can observe that none of ARMA models outperformed random white noise in terms of forecasting accuracy. Window of 750 observation shows even worse results than window of 500 observations. There is no reason to expect that this behavior would change with further increase in window size. Assuming that, we can say that natural gas futures market is efficient.

Chapter 5

Discussing results

In this chapter, we will discuss the results of the analysis from previous chapter. First, we will examine factors which may have effect on differences between ARMA and white noise forecast errors. Such factors are order of ARMA model, size of rolling window, size of size of forecast error series, and affiliation to some commodity group. Then, we will focus on explaining cases, where the results were significantly better than white noise, which would violate the weak form of market efficiency.

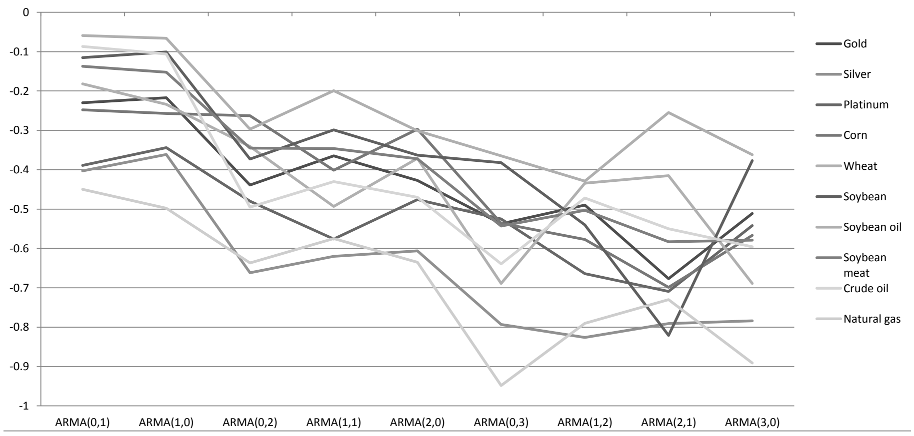
5.1 Influence of parameters on results

5.1.1 Order of ARMA models

As assumed, we can generally observe that with increasing order of both MA and AR or combined order including both parts, the DM statistic tend to further go to negative values. Reasonable explanation is, that out - of - sample analysis is particularly sensitive to over fitting. As evidence, we can observe that over fitting will usually prevail even at quite small orders. Often the best models among the ARMA group are either AR(1) or MA(1). Furthermore, we can observe that models with combination of both MA and AR part and same combined order tend to be worse than models with only either MA or AR part of the same combined order, which is likely again caused by over fitting of the model.

Figure 5.1 shows, that with increasing order of ARMA, the DM statistics is being pushed further into negative values. Result of diplaying several different ARMA models with same combined order is the staircase - like structure of lines.

Figure 5.1: Negative effect of ARMA order on DM statistics for selected commodities



Source: computations

5.1.2 Window size

Almost in all cases we can observe that with increasing window, the forecasting performance of given ARMA models improved with growing size of rolling window. Correctness of the expectation about fluctuation of DM statistics around zero with window size growth is somewhat questionable. Although it does explain DM improvement as result of increasing window size in many cases, it does not explain behavior of lean hog futures and palladium futures, where already positive DM statistics further improved with growing window size. In case of lean hog, the only commodity futures that had DM statistic over 1.00 of some model even for 250 window size, we can see that the value of DM statistic is growing to even more positive values in case of 750 window size. Similar behavior can be observed with palladium futures. We can substantially increase the window size and show whether this behavior persists.

Taking lean hog futures as example with positive DM statistics for most of the models, we are interested whether the DM statistics further improve even with higher sizes of windows such as 1500 observations (6 trade years). Results can be seen in table 5.1.

Natural gas is example of futures contract with lower DM statistic on 750 window size than on 500 size. By further extending the window size we want to find out, whether this decreasing trend continue even for 1500 window size, as displayed in table 5.2.

To make reasonable conclusion about influence of window size to our analysis, we need to take Crude oil as example of futures contract with increasingly

Table 5.1: Analysis of lean hog futures - 750 and 1500 window size

	750 window size			1500 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	1.032	0.01091	0.01468	1.103	0.0109	0.01467
ARMA(1,0)	1.145	0.0109	0.01465	1.215	0.01089	0.01464
ARMA(0,2)	1.046	0.01091	0.0146	1.166	0.0109	0.01461
ARMA(1,1)	1.243	0.01089	0.01458	1.311	0.01088	0.01458
ARMA(2,0)	1.117	0.0109	0.01458	1.239	0.01089	0.01458
ARMA(0,3)	1.058	0.0109	0.0146	1.197	0.01089	0.0146
ARMA(1,2)	1.082	0.0109	0.01459	1.296	0.01088	0.01458
ARMA(2,1)	0.96	0.01091	0.0146	1.192	0.01089	0.01459
ARMA(3,0)	0.981	0.01091	0.01459	1.109	0.0109	0.01459
White noise	-	0.01101	0.01483	-	0.01101	0.01483

Source: computation are for skript parameters: (v = 3, w = 750/1500, s=11, z = 10)

Table 5.2: Analysis of natural gas futures - 750 and 1500 window size

	750 window size			1500 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.45	0.02355	0.03104	-0.239	0.02352	0.03104
ARMA(1,0)	-0.498	0.02355	0.03104	-0.258	0.02353	0.03103
ARMA(0,2)	-0.637	0.02356	0.03108	-0.359	0.02354	0.03105
ARMA(1,1)	-0.575	0.02356	0.03103	-0.181	0.02352	0.03102
ARMA(2,0)	-0.635	0.02356	0.03108	-0.425	0.02354	0.03106
ARMA(0,3)	-0.948	0.0236	0.03112	-0.74	0.02357	0.03109
ARMA(1,2)	-0.79	0.02358	0.03111	-0.447	0.02354	0.03106
ARMA(2,1)	-0.73	0.02357	0.03109	-0.421	0.02354	0.03106
ARMA(3,0)	-0.891	0.02359	0.03112	-0.759	0.02358	0.03109
White noise	-	0.0235	0.03102	-	0.0235	0.03103

Source: computation are for skript parameters: (v = 3, w = 750/1500, s=11, z = 10) negative DM statistics we want to show, whether such behavior will prevail even at window size of 1500. Results for crude oil are in table 5.3.

Table 5.3: Analysis of natural gas futures - 750 and 1500 window size

	750 window size			1500 window size		
Model	DM	MAE	RMSE	DM	MAE	RMSE
ARMA(0,1)	-0.087	0.01565	0.02182	-0.057	0.01565	0.0218
ARMA(1,0)	-0.106	0.01565	0.02182	-0.061	0.01565	0.0218
ARMA(0,2)	-0.495	0.01569	0.02187	-0.159	0.01566	0.02181
ARMA(1,1)	-0.43	0.01568	0.02187	-0.058	0.01565	0.02181
ARMA(2,0)	-0.47	0.01569	0.02187	-0.137	0.01565	0.02181
ARMA(0,3)	-0.639	0.01571	0.02191	-0.289	0.01567	0.02183
ARMA(1,2)	-0.472	0.01569	0.02188	-0.176	0.01566	0.02181
ARMA(2,1)	-0.55	0.0157	0.0219	-0.23	0.01566	0.02182
ARMA(3,0)	-0.596	0.0157	0.02191	-0.289	0.01567	0.02183
White noise	-	0.01564	0.0218	-	0.01564	0.02179

Source: computation are for skript parameters: (v = 3, w = 750/1500, s=11, z = 10)

By further extending the window size, we can observe that the value is always slowly growing. Possible explanation lies in the very nature of ARMA model estimation. With increasing window size, each regression have larger sample and the ARMA model takes more observation into consideration and tries to describe pattern of the data on larger time period. When there is some significant long term pattern among the data, with increasing sample size the model will be more precise at describing the model and detecting the pattern.

If the best fitted model of each ARMA configuration is significantly different from white noise, the DM statistics will tent to grow even at high window sizes and slowly converge to some positive value. This is case of lean hog futures and palladium futures measured on appropriate sub period.

On the other hand, if the best fitted model of each ARMA configuration is similar to white noise, with increasing window size the DM statistic will converge to zero, which would explain growth of DM statistic in case of Crude oil, Natural gas and likely as well in majority of remaining commodities.

To conclude influence of window size, we can say that with increasing window the DM statistics will converge to the value of DM statistic of best fitted ARMA models.

5.1.3 Error series size

Because of the assumption about existence of shorter time frame within the ten years period where rejection of null hypothesis is possible, we carried out separate analysis for each commodity futures when DM statistic of some model exceeded zero. Results of additional analysis corresponds with findings mentioned in Kaminsky & Kumar (1990). Using bisection algorithm and one trade year as minimal time period, we found in cases of Palladium, Lean hog, live cattle and heating oil on some sub period with noticeably higher DM statistics than for the original ten years period.

In case of Palladium we found out, that the largest sub period where we are able to reject null hypothesis is two trade years (forecast error series have 500 elements, DM of AR3 = 1,825), in case of lean hogs, such sub period are three trade years (forecast error series have 750 elements, DM of AR1 = 2,173). On the other hand, even at smallest considered sub period size (one trade year, 250 elements of forecast error series) we are not able to reject null hypothesis in case of live cattle (smallest DM statistic have ARMA (1,1), 0,443) and heating oil (closest to rejection of null hypothesis was DM of ARMA (2,1) = 1,375).

It is reasonable to assume, that in case of heating oil futures, we should be able to find some sub period smaller than 250 elements of error series. Furthermore, we can generalize this assumption for every of commodity futures where DM test is inconclusive. In the other words, we can expect existence of small enough sub period, where the rejection is possible all commodity futures.

However, these sub periods can be quite small and not very frequent, therefore the results of DM test will be always insignificant. Knowing this, we can assume that this phenomena does not break the market efficiency, because such sub periods are hard to detect and we can be never sure how long will they last. Under these conditions investors are not able to systematically make exceptional profits which means that markets are efficient.

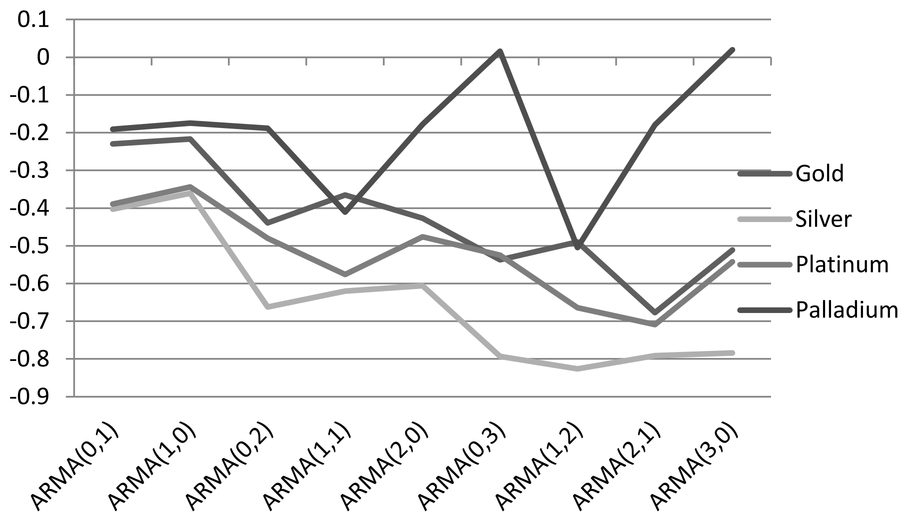
Exception are commodities, where white noise is better for every forecasted value in the sub period. However, nonexistence of single observation with ARMA model being better than noise will most likely led to rejection of DM test as result of DM statistic being strongly negative. White noise will be in that significantly better than any ARMA model, which imply market efficiency. Only exception would be case when any ARMA model would be almost same as white noise, but slightly worse for every forecast. This case is extremely unlikely. This behavior of test statistic implies that DM test statistics of ARMA model and white noise is quite reliable measurement of market efficiency.

Values around zero can imply that given model is not significantly better than noise or that ARMA model is significantly underperforming for some sub period while being significantly better than noise in another sub period. Either way, if the period of good performance is small enough, market efficiency is not violated, because we may be never sure about time stability of this occurrence.

5.1.4 Commodity group affiliation

From the results, it is clear that some commodity group share some similar properties in terms of market efficiency. For example group of precious metal futures is generally efficient, with exception of palladium. Gold, silver and platinum have negative values of DM statistic for all ARMA models and window sizes. We can notice that these three futures are not very sensitive to growth of window size. While palladium shares the behavior of this group for 250 and 500 window size, for 750 size we can observe abnormally high DM statistic for ARMA (3,0) and ARMA (0,3), which even grows with restricting the analysis for sub period. In figure 5.2, we can observe that for precious metals, the pattern of how DM statistic changes with different ARMA models is similar for most commodities withing the group. The only exeption are again palladium futures.

Figure 5.2: Precious metals - structure of DM statistic with change of ARMA configuration

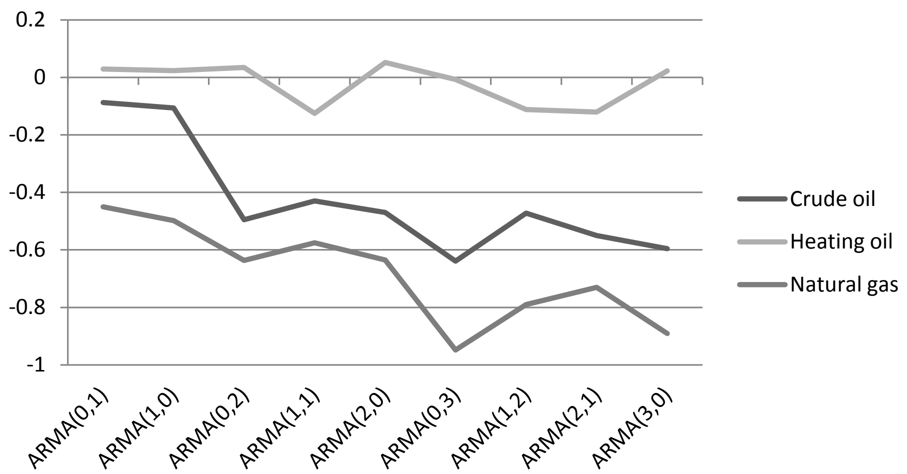


Source: computations

For most energy futures is typical high sensitivity for window size growth. Values of DM statistics grow from values around -0,5 at 250 window size to values around zero for higher window sizes. Only exception is natural gas, which is have strongly negative values for all examined window sizes.

In figure 5.3, we can observe that energy futures follows similar pattern in changes of ARMA order with exeption of heating oil futures.

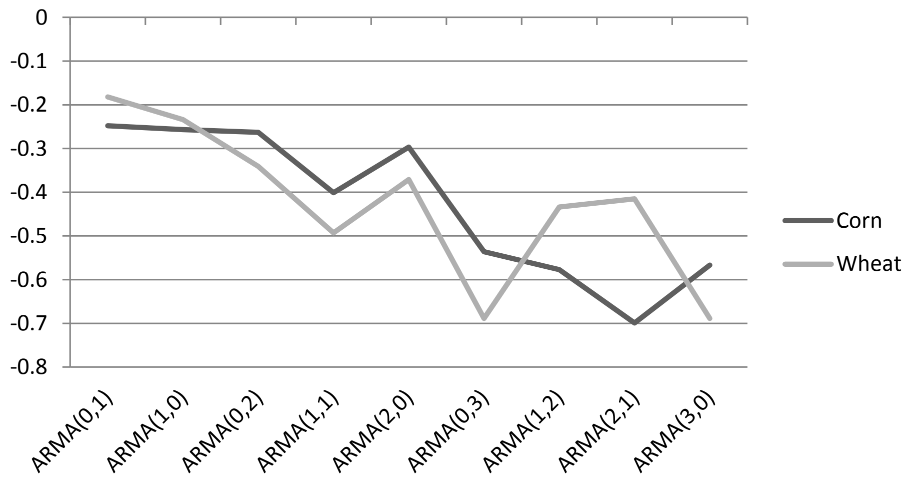
Figure 5.3: Energy futures - structure of DM statistic with change of ARMA configuration



Source: computations

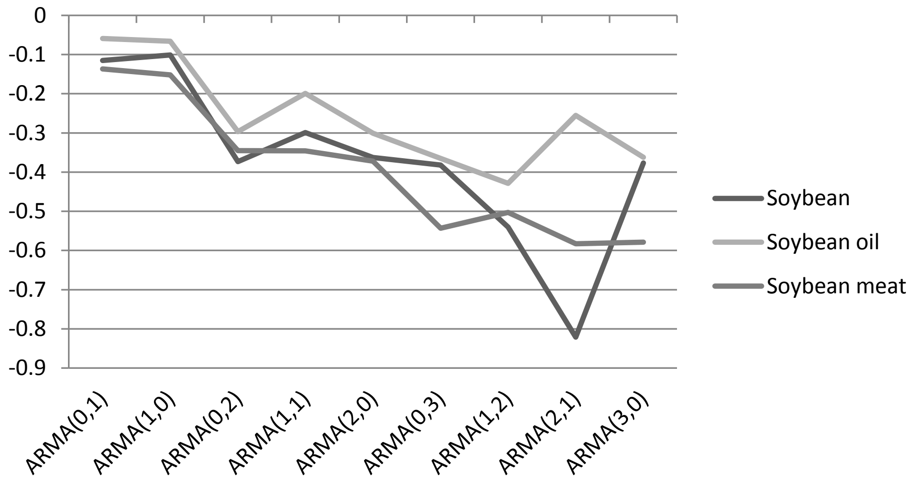
Figure 5.4 shows that both corn and wheat have similar reaction on ARMA order change. Surprisingly, neither soybean and other soybean related futures follows this pattern. On the other hand, as can be seen in figure 5.5, soybean related futures as well as soybean futures appears to be following mutual pattern.

Figure 5.4: Wheat, corn - structure of DM statistic with change of ARMA configuration



Source: computations

Figure 5.5: Soybean related futures - structure of DM statistic with change of ARMA configuration



Source: computations

We can say, that most efficient commodity group are precious metal futures with exceptions of palladium followed by energy futures and grains. Least efficient are livestock futures, as the only group of examined futures with mostly possitive DM statistics. Worth mentioning is, that commodities with exeptional pattern, such as heating oil and palladium, aswell shows signs of inefficiency.

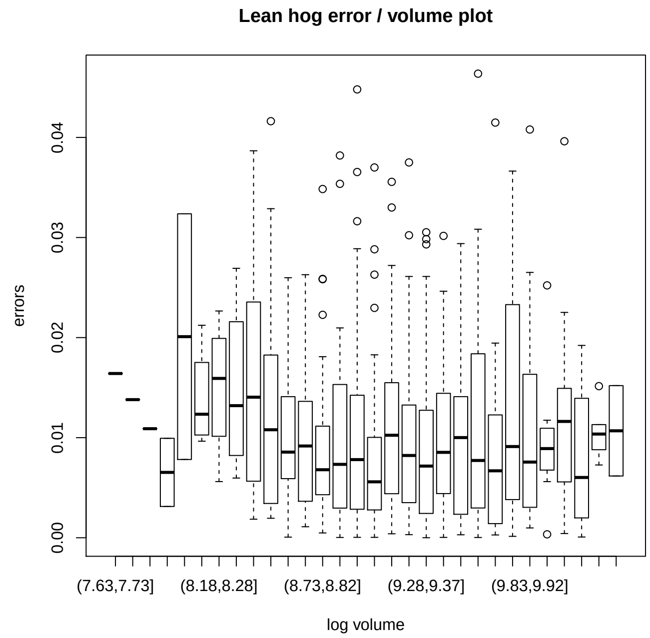
In general, these results approximately confirms findings mentioned in Kristoufek & Vosvrda (2014) about market efficiency of commodity groups.

5.2 Market liquidity

Before we can make any conclusion about efficiency of the three problematic futures(lean hog in figure 5.6, palladium in figure 5.7, heating oil in figure 5.8), it is appropriate to examine market liquidity during the sub period where rejection of null was possible. To do that, we will graph absolute forecast error over the traded volume for given sub period. As control check, we take natural gas futures as example of commodity with efficient market(figure 5.9).

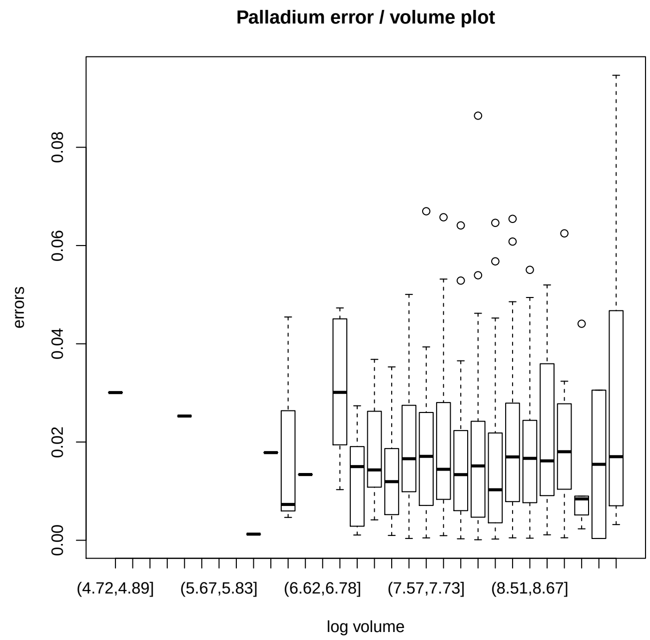
From the figures can be seen that there is no evident relation between error size and traded volume of given commodity. Mean of absolute error is approximately the same regardless the volume for all selected commodity types. Box plots for palladium, lean hogs and heating oil have similar structure as surely efficient natural gas. This is another evidence that volume has no relation to forecast error and market efficiency. We can conclude that there are no

Figure 5.6: Analysis of lean hog market liquidity over predictable sub period - plot of ARMA (3,0) absolute errors over traded volume



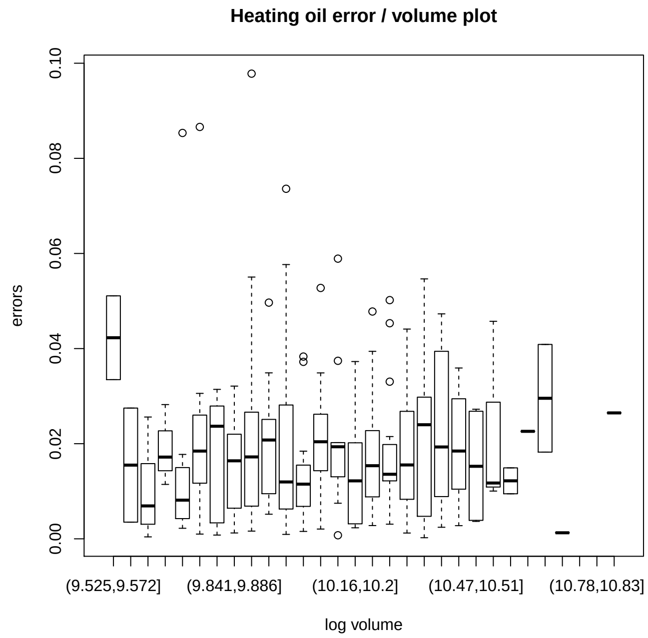
Source: computation for scrip parameters a = 3 b = 0 w = 750, s = 10, z = 3

Figure 5.7: Analysis of palladium market liquidity over predictable sub period - plot of ARMA (3,0) absolute errors over traded volume



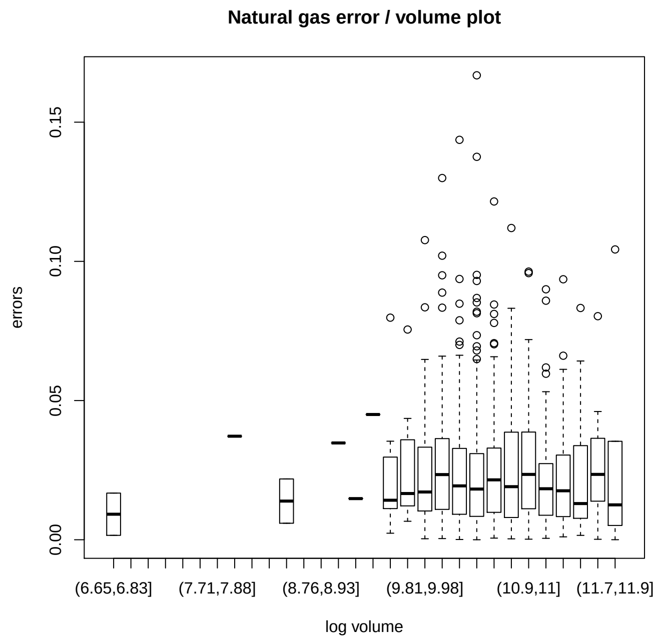
Source: computation for scrip parameters a = 3, b = 0, w = 500 , s=5, z = 2

Figure 5.8: Analysis of heating oil market liquidity over predictable sub period - plot of ARMA (2,1) absolute errors over traded volume



Source: computation for script parameters a = 2, b= 1, w = 750, s = 10, z = 1

Figure 5.9: Analysis of natural gas market liquidity over predictable sub period - plot of ARMA (1,0) absolute errors over traded volume



Source: computations for script parameters a = 1, b= 0, w = 750, s = 11, z = 5 liquidity restriction that would prevent the investors from making extraordinary profit.

Although there does not seem to be any relation between error size and traded volume and we can predict the return using better than white noise, the efficient market hypothesis is not necessarily violated if the sub period when we are able to predict returns of this commodities appears at random and is small enough, which is likely the case of heating oil. Under that, we would not be able to forecast when such window of predictability happen and we have no certainty when it will end. On the other hand, this sub period in cases of lean hog and palladium is substantial, three trading years for lean hog and two for palladium. In case of palladium, from historical development can be seen that sub period of size is anomaly. In that case we no reason to doubt market efficiency, unless similar behavior will became regular for future sub periods. Lean hog is quite different in this concern. Because the DM statistic is bordering rejection for whole ten years period, we can expect regular appearance of predictable sub periods. With no liquidity restrictions, we have no reason to insist on market efficiency for this type of futures contract.

Chapter 6

Conclusion

Main fidings of the work are that ARMA models provide better out - of - sample forecasts than whine noise for some commodities on suitable time period and test statistic of Diebold - Mariano test between ARMA models and white noise is reliable measurement of market efficiency.

Size of combined order of ARMA model have mild impact on forecasting performance, lower models are usually better. This is caused by over fitting of the model. Apart of that, models of same combined order have similar forecasting performance.

On the other hand, size of window in case of rolling window method of measuring forecasting performance have considerable effect on forecasting performance. With increasing size of the window, DM statistics will converge to the value of DM statistic of best fitted ARMA models, which can be some positive value, like in case of lean hog, or zero, like most of other commodities. This explains why DM statistic is usually growing with increasing size of window.

Size of period, over which we try to forecast is the determining factor for measuring forecast performance. We can almost always find small enough period, where the null hypothesis of DM test is rejected. This behavior however does not break the market efficiency if the sub period is small enough and appearing at random. Affiliation to commodity group have questionable effect on market efficiency, mainly because most commodity futures shows evidences of market efficiency anyways. However, there are some apparent similarities among the commodity groups.

Examination of market liquidity shows little to none effect of traded volume on forecast error over predictable sub period. This implies that there is no restriction in liquidity of given market in problematic periods. While for heating oil and palladium there exists reasonable arguments for not rejecting EMH, lean hog futures shows strong evidences of market inefficiency, which cannot be explained by any explanation included in extend of this work.

Natural way how to extend this work is using more complicated models to carry out the analysis. Another possible direction how to extend this thesis is to analyse how much randomly does the predictable sub periods appear, which might be crucial in analyzing general market efficiency of palladium futures and heating oil futures.

Bibliography

ALGIERI, B. & M. KALKUHL (2014): "Back to the futures: An assessment of commodity market efficiency and forecast error drivers." *Technical report*, University of Bonn - Center for Development Research.

CHINN, M. D. & O. COIBION (2014): "The predictive content of commodity futures." *Journal of Futures Markets* **34(7)**: pp. 607–636.

CHORDIA, T., R. ROLL, & A. SUBRAHMANYAM (2008): "Liquidity and market efficiency." *Journal of Financial Economics* **87(2)**: pp. 249–268.

DIEBOLD, F. X. & R. S. MARIANO (2002): "Comparing predictive accuracy." *Journal of Business & economic statistics* **20(1)**.

FAMA, E. F. (1965): "The behavior of stock-market prices." *Journal of business* pp. 34–105.

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

FAMA, E. F. (1991): "Efficient capital markets: Ii." *The journal of finance* **46(5)**: pp. 1575–1617.

FAMA, E. F. & K. R. FRENCH (1987): "Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage." *Journal of Business* pp. 55–73.

GORTON, G. & K. G. ROUWENHORST (2006): "Facts and fantasies about commodity futures." *Financial Analysts Journal* **62(2)**: pp. 47–68.

HASTIE, T., R. TIBSHIRANI, J. FRIEDMAN, T. HASTIE, J. FRIEDMAN, & R. TIBSHIRANI (2009): *The elements of statistical learning*, volume 2. Springer.

JENSEN, G. R. & J. M. MERCER (2011): "Commodities as an investment." .

KAMINSKY, G. & M. S. KUMAR (1990): "Efficiency in commodity futures markets." *Staff Papers-International Monetary Fund* pp. 670–699.

KELLARD, N., P. NEWBOLD, T. RAYNER, & C. ENNEW (1999): "The relative efficiency of commodity futures markets." *Journal of Futures Markets* **19(4)**: pp. 413–432.

KRISTOUFEK, L. & M. VOSVRDA (2014): "Commodity futures and market efficiency." *Energy Economics* **42**: pp. 50–57.

LO, A. W. (2007): "Efficient markets hypothesis." .

SAID, S. E. & D. A. DICKEY (1984): "Testing for unit roots in autoregressive-moving average models of unknown order." *Biometrika* **71(3)**: pp. 599–607.

SAMUELSON, P. A. (1965): "Proof that properly anticipated prices fluctuate randomly." *Industrial management review* **6(2)**: pp. 41–49.

STEVENSON, A. (2010): *Oxford Dictionary of English.* Oxford Dictionary of English. OUP Oxford.

TOMEK, W. G. (1997): "Commodity futures prices as forecasts." *Review of Agricultural Economics* **19(1)**: pp. 23–44.

WANG, H. H. & B. KE (2005): "Efficiency tests of agricultural commodity futures markets in china." *Australian Journal of Agricultural and Resource Economics* **49(2)**: pp. 125–141.

WOOLDRIDGE, J. (2012): *Introductory econometrics: A modern approach.* Cengage Learning.