THE INCREMENTAL BENEFITS OF THE NEAREST NEIGHBOR FORECAST OF U.S. ENERGY COMMODITY PRICES

A Thesis

by

OLGA KUDOYAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2010

Major Subject: Agricultural Economics

The Incremental Benefits of the Nearest Neighbor Forecast of U.S. Energy Commodity Prices

THE INCREMENTAL BENEFITS OF THE NEAREST NEIGHBOR FORECAST OF U.S. ENERGY COMMODITY PRICES

A Thesis

by

OLGA KUDOYAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	James W. Richardson
	Henry L. Bryant
Committee Member,	Michael F. Speed
Head of Department,	John P. Nichols

December 2010

Major Subject: Agricultural Economics

ABSTRACT

The Incremental Benefits of the Nearest Neighbor Forecast of U.S. Energy Commodity Prices. (December 2010)

Olga Kudoyan, B.S., Armenian State Agrarian University

Co-Chairs of Advisory Committee: Dr. James W. Richardson,
Dr. Henry L. Bryant

This thesis compares the simple Autoregressive (AR) model against the k-Nearest Neighbor (k-NN) model to make a point forecast of five energy commodity prices. Those commodities are natural gas, heating oil, gasoline, ethanol, and crude oil. The data for the commodities are monthly and, for each commodity, two-thirds of the data are used for an in-sample forecast, and the remaining one-third of the data are used to perform an out-of-sample forecast. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to compare the two forecasts. The results showed that one method is superior by one measure but inferior by another. Although the differences of the two models are minimal, it is up to a decision maker as to which model to choose.

The Diebold-Mariano (DM) test was performed to test the relative accuracy of the models. For all five commodities, the results failed to reject the null hypothesis indicating that both models are equally accurate.

NOMENCLATURE

AIC	Akaike Information Criterion
AR	Autoregressive
ARMA	Autoregressive Moving Average
DKNAW	Dynamic K-Nearest-Neighbor Naïve Bayes With Attribute Weighting
DM	Diebold-Mariano
k-NN	k-Nearest-Neighbor
MAE	Mean Absolute Error
MBR	Memory-Based Reasoning
MSE	Mean Squared Error
MSPE	Mean Squared Prediction Error
NYSE	New York Stock Exchange Energy Index
OLS	Ordinary Least Squares
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
STAR	Smooth Transition Autoregressive

TABLE OF CONTENTS

Page

ABSTRACT iii

NOMENCLATURE iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES ix

1. INTRODUCTION: THE IMPORTANCE OF RESEARCH 1

2. LITERATURE REVIEW 6

3. DATA 18

4. MODELS 20

    4.1 AR Model 20

        4.1.1 AIC Calculations and Model Selection 24

        4.1.2 One-Step-Ahead Out-of-Sample Forecast 25

        4.1.3 Two-Step-Ahead Out-of-Sample Forecast 27

    4.2 k-NN Model 27

        4.2.1 In-Sample Forecast 28

        4.2.2 Out-of-Sample Forecast 32

5. METHODOLOGY 34

    5.1 MAE Calculations 34

    5.2 RMSE Calculations 34

Page

5.3 Analysis of Forecasting Accuracy of Models 35

6. ESTIMATION AND RESULTS 38

7. CONCLUSIONS 56

REFERENCES 61

APPENDICES 64

VITA 70

LIST OF FIGURES

FIGURE Page

1 Graphical Illustration of AIC Scores for Natural Gas In-Sample Forecast 41

2 Graphical Illustration of AIC Scores for Heating Oil In-Sample Forecast 41

3 Graphical Illustration of AIC Scores for Gasoline In-Sample Forecast 42

4 Graphical Illustration of AIC Scores for Ethanol In-Sample Forecast 42

5 Graphical Illustration of AIC Scores for Crude Oil In-Sample Forecast 43

6 Graphical Illustration of MAEs for Natural Gas 44

7 Graphical Illustration of MAEs for Heating Oil 44

8 Graphical Illustration of MAEs for Gasoline 45

9 Graphical Illustration of MAEs for Ethanol 45

10 Graphical Illustration of MAEs for Crude Oil 45

11 Graphical Illustration of Predicted vs. Actual Values for Natural Gas One-Step-Ahead Out-of-Sample Forecast 46

12 Graphical Illustration of Predicted vs. Actual Values for Heating Oil One-Step-Ahead Out-of-Sample Forecast 46

13 Graphical Illustration of Predicted vs. Actual Values for Gasoline One-Step-Ahead Out-of-Sample Forecast 47

14 Graphical Illustration of Predicted vs. Actual Values for Ethanol One-Step-Ahead Out-of-Sample Forecast 47

LIST OF FIGURES (CONTINUED)

FIGURE Page

15 Graphical Illustration of Predicted vs. Actual Values for Crude Oil One-Step-Ahead Out-of-Sample Forecast 48

16 Graphical Illustration of Predicted vs. Actual Values for Natural Gas Two-Step-Ahead Out-of-Sample Forecast 49

17 Graphical Illustration of Predicted vs. Actual Values for Heating Oil Two-Step-Ahead Out-of-Sample Forecast 49

18 Graphical Illustration of Predicted vs. Actual Values for Gasoline Two-Step-Ahead Out-of-Sample Forecast 50

19 Graphical Illustration of Predicted vs. Actual Values for Ethanol Two-Step-Ahead Out-of-Sample Forecast 50

20 Graphical Illustration of Predicted vs. Actual Values for Crude Oil Two- Step-Ahead Out-of-Sample Forecast 51

LIST OF TABLES

TABLE Page

1 Summary of Models 40

2 MAE and RMSE Calculations for One-Step-Ahead and Two-Step-Ahead Out-of-Sample Forecasts 52

3 Diebold-Mariano Test Results 54

1. INTRODUCTION: THE IMPORTANCE OF RESEARCH

For most people, the world is full of many opportunities to succeed. However, those opportunities would be more achievable if one has the ability to predict the future. Forecasting is the tool that makes the future more or less predictable and prepares people for upcoming changes.

Forecasting has always been an important component of running businesses. Recently, business forecasting has been accomplished by using more scientific endeavors, with different theories and methods designed to forecast different types of data. Businesses have tried to focus on key factors in business production and extrapolate from available information to accurately project future costs, revenues, and opportunities. According to a survey by the Hudson Institute (an Ohio-based Answer Think Consulting Company that specializes in studies of business planning), for every billion dollars of revenue, the average business in the United States spends more than 25,000 person-days on forecasting activities.

Forecasts are very common in every industry. Even government agencies make predictions to support their operations. For instance, the Energy Information Administration (EIA) predicts that between 2008 and 2035 energy consumption will increase by 14%, with a 0.5% annual growth rate. U.S. fuel consumption has significantly increased within the past 25 years, although the imports have decreased

---

This thesis follows the style of the *American Journal of Agricultural Economics.*

(EIA, 2010). In 2006, about 35% of U.S. oil and 30% of natural gas production was coming from Federal Lands (EIA, 2010). In contrast, over the next 10 years, it is projected that domestic production will increase by 47% for oil and 37% for natural gas (EIA, 2010).

Certain industries are sometimes hard to predict because of their sudden changes. However, scientists have still found ways to deal with predicting behavior when there are difficult factors involved. For example, the rapid fluctuations in the price of oil in 2008 surprised many people. However, Edward Morse (Managing Director of Louis Capital Markets) thinks this behavior is not unusual and that commodity markets are cyclical by nature, with a history filled by sudden turning points (Morse, 2009). Edward further comments that this behavior generally makes it difficult to forecast prices; however, he thinks that commodity markets will remain lower over the next few years than they have been over the past five years (Morse, 2009).

According to Safavi (2000), there are three business forecasting models: cause-and-effect, judgmental, and time series. The cause-and-effect model assumes a cause that determines an outcome. It shows that a similar cause in the future is likely to yield a similar effect. This model depends on historical data and also assumes that the cause-and-effect relationship is more or less stable and can be quantified. On the other hand, the judgmental model attempts to forecast when there is no useful historical data. A business can use this model to project sales for a brand new product where the available historical data has become obsolete.

This thesis, however, concentrates on the third forecasting model: the time-series. The model makes forward projections based on the data's historical performance. It utilizes the behavior of data patterns from the past and assumes it will continue similarly into the future. For this model, you only input the available historical data into the forecasting formulas, and it generates the desired predictions. This model has proven to be useful when forecasting based on historical data that has smooth and stable patterns. Even when jumps and anomalies occur, the time series model may still prove to be useful, as long as it is possible to account for the anomalies.

As there is no reliable and widespread technology to utilize renewable energy commodities, the world market will continue to heavily depend on non-renewable energy commodities. Among those non-renewable commodities are natural gas, heating oil, gasoline, and crude oil, which have always been an important factor for the economy. Major newspapers cover stories about those commodity prices on a daily basis. Moreover, the linkage of the commodity prices to the other macroeconomic variables, such as stock market index and exchange rates, also draws attention by mass media. For example, Standard & Poor (S&P) reported in 2008 that it expected the average U.S. household to spend 6.7% of its income on energy that year – the same portion spent on average in 1971. In the early 1980s, in contrast, energy costs accounted for 7.9% of U.S. household income. Also, historically, increases in economic activity have correlated strongly with increased energy use. In recent years, there has only been a weak connection between economic growth and consumption of energy in the United States. Gross Domestic Product (GDP) has grown rapidly while energy use has grown at

relatively modest levels, particularly since the mid-1980s (Clausssen et al., 2001). Major news agencies cover stories about energy commodity prices and their corresponding stock markets on a daily basis. For instance, Bloomberg covers the New York Stock Exchange Energy Index (NYSE) on a daily basis. The NYSE Energy Index was introduced to give investors and issuers a more detailed summary of the energy segment inside the NYSE marketplace.

In this thesis, two well-known forecasting techniques are compared: simple Autoregressive (AR) and k-Nearest-Neighbor (k-NN) approaches. The thesis will show the incremental benefits of k-NN over the AR model. More specifically, we will see how the k-NN forecasting approach can be used in forecasting multivariate distributions and how accurate the results are for a nonparametric model. The research concentrates on the price forecasts of five energy commodities: natural gas, crude oil, heating oil, gasoline, and ethanol.

The results of this thesis can be used in real world situations in several ways. First, those who have investments in any of the energy commodities will be interested in the outcome, since the thesis will be forecasting future prices for actual commodities. Second, the industry representatives will be interested in the results because they can make adjustments to their business strategies by forecasting with better methods presented here. Finally, the thesis will affirm the usefulness of the k-Nearest-Neighbor forecasting method and generate grounds for further studies.

In the Literature Review section, I will discuss some of the most relevant research related to different forecasting methods, including nearest neighbor forecasts. In

the Data section, I will describe the data on which my research is based and what software was used to make various calculations and analyses. In the Models section, I will describe the two types of forecasting models utilized in my thesis: AR and k-NN. In this section, I will further describe the specific methodologies used in each model. In the Analysis of Forecasting Accuracy of Models, I will describe the technique used to test the relative forecasting accuracy of the models. Finally, in the Results section, I will compare the calculation results in both AR and k-NN models and make conclusions and necessary recommendations.

2. LITERATURE REVIEW

The Autoregressive (AR) model was first introduced by G. Undy Yule in 1926, which led to an approach of time-series investigation. According to Yule, the past observation of a variable can explain its movements in the present. Since the introduction of this AR(p) model, the study of time-series analysis started to develop thoroughly. The use of AR models became more practical in the mid-1960s when computer technologies advanced. Nowadays, the AR(p) model is used in the analyses of fundamental higher econometrics. The most widely used tests, such as stationary, Granger causality, and cointegration, are built on the basis of the AR(p) model (Liew et al., 2003).

Continuous change in energy commodity markets has raised an increasing interest for investigation of different econometric models that will provide trustworthy price forecasts. Cuaresma et al. (2004) has explored electricity spot-price predictions by exclusively concentrating on linear univariate models, such as autoregressive moving average models, models with unobserved components and with jumps. As a base model, the paper used the simple first order autoregressive process [AR(1)]. The second model concentrated on the systematic seasonal variation that was found in electricity prices. The intercept in the base models is changing based on the hour of the day, day of the week and month of the year. The third model considered time-varying intercept using an autoregressive moving average model. The fourth model, which was crossed ARMA with time-varying intercept, was implemented to achieve more flexibility by allowing electricity spot-price in hour z to depend upon price realizations in hour $s \neq z$. The fifth

model used in the paper is ARMA process with jumps. This model was used because one of the notable characteristics of electricity spot prices is the percent of price spikes. The last model used is an unobserved components model. Using the above mentioned models, Cuaresma et al. (2004) performed 168-hour-ahead (one week) forecast. Two measures of forecast error Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used. The Diebold-Mariano test was applied to measure the significance of observed differences in forecasting power across models. The results showed that an hour-by-hour modeling strategy significantly improved the forecasting performance of linear univariate time-series models and assessing the process of arrival of price spikes can also result in better forecasting.

Staudenmayer and Buonaccorsi (2005) have addressed the estimation of the parameters in linear AR models in the presence of additive and uncorrelated measurement errors. They have established the asymptotic properties of naive estimators which ignore measurement error. The methods reviewed by the group included ones that required no information about the measurement error variances and compared the various estimators both theoretically and via simulations. While reviewing existing estimators that account for measurement error, the team has developed new estimators. Staudenmayer and Buonaccorsi (2005) have set the stage for two major research areas. First, they call for further development of procedures for valid inferences regarding the AR parameters in various sample sizes where measurement error exists. The current estimators all have substantial small-sample bias, which is an important consideration. The team also thought that another area for future study would be to take advantage of

an estimated equations-based approach in situations where the measurement errors are correlated.

Pekarova and Pekar (2006) studied the analysis of natural fluctuations and long-term trends in annual discharge time-series of the Danube River and a stochastic prediction of Danube River discharge at the Turnu Severin station for the next 20 years using linear autoregressive (AR) models. The models used included one based on harmonic functions (hidden periods) and linear autoregressive models (AR), autoregressive moving average (ARMA) and seasonal autoregressive integrated moving average model (SARIMA). To obtain the long-term annual discharge prediction, the team used the following steps: (1) take the time-series of logarithms of annual discharge and center it; (2) remove the harmonic component from the time-series; (3) remove the autoregressive component from the residuals; (4) test the correctness of the model specification; and (5) predict the annual discharge, to specify the appropriate confidence intervals. To choose the appropriate model, the team used sum of squared residuals and Akaike Information Criterion (AIC).

Several authors have compared the forecasting performance of a simple AR model to more complex forecasting methods. Liew et al. (2003) have tested the adequacy of the linear autoregressive (AR) time-series model. Their study was based on the real exchange rates of Asian economies. The team identified two important findings. First, they found the research showed empirical evidence that the AR model is inadequate in characterizing the behavior of Asian real exchange rates. Second, they found that the behavior of the real exchange rates' linearity has been formally rejected in

favor of the nonlinear Smooth Transition Autoregressive (STAR) model. They further found that, taken together, the supremacy of nonlinearity over linearity in the data generating process of those real exchange rates warrants the use of a linear framework in empirical modeling. The group thought that the statistical testing procedures in the context of Asian exchange rates could amount to inappropriate policy conclusions. The researchers found that estimating Exchange Rates Model in the form of linear autoregressive (AR) and disregarding the presence of nonlinearity will yield a wrong model, and thereby provide incorrect policy recommendations. Therefore, they concluded that the linear model is valid only when the formal linearity test result fails to provide evidence on the existence of nonlinearity.

Christini et al. (1995) employed parametric models to determine if a more appropriate heart rate (HR) dynamics modeling structure existed. The linear AR, autoregressive moving average (ARMA), the nonlinear polynomial autoregressive (PAR), and bilinear (BL) models were used for an HR time-series obtained from nine subjects. Model orders were determined by the Akaike Information Criteria (AIC). The researchers found that the BL model best represented the heart rate (HR) dynamics, as its residual variance was significantly ($p < 0.05$) smaller than that of the corresponding AR model for nine out of nine data sets. They observed that in all cases, smaller residual variance than either the ARMA or PAR models were present in the BL model. Additionally, the team found that the BL model's residual variance was found by the Priestley statistic to be significantly smaller than that of the AR model. The team observed that the residual variances of the ARMA and PAR models were significantly

smaller than those of the AR model fits for fewer subjects, and for those subjects the reduction in residual variance was less than that seen for the BL model. Researchers concluded that the apparent superiority of the nonlinear BL model suggested that future HR studies should put greater emphasis on nonlinear analysis.

The k-Nearest-Neighbor method was invented six decades ago and was analyzed by statisticians in the early 1950. k-NN was implemented in the early 1960s and was widely used in the pattern recognition field for more than three decades. The whole idea behind k-NN is the memory-based reasoning (MBR), which is the result based on similar situations that occurred in the past. The model selection using k-NN method is based on past experience. The center of the k-NN technique is the similarity: how is the new situation similar to the one that occurred in the past? Another key concept related to the k-NN method is that it combines the information from its neighbors (Berry and Linoff, 2004).

Several researchers tested the performance of a comparatively new method, k-Nearest Neighbor forecast, with respect to widely used methods such as simple Autoregressive, GARCH and ARIMA. For example, Bordignon and Lisi (2001) presented a technique to measure prediction intervals for chaotic data that were both clean and noisy. Their work is based on point forecasts and employs the simple idea of using nearest neighbors to give an estimate of local variability. Combining the empirical distribution of the prediction error and the nearest neighbor forecast allowed them to build prediction intervals. The procedure in the paper does not take into consideration any distributional assumption because it is basically computational. The authors

estimated local variance and percentile of the prediction error distribution using the nearest neighbor approach. Due to the approach used in the paper, it has become possible to forecast values with a given probability. The paper demonstrated the use of the nearest neighbor approach and the empirical distribution of the prediction error to estimate prediction intervals. The results were encouraging because, although the method was really simple, it worked very well with noise-free data and showed promising results for noisy data.

Several problem areas were discovered by researchers when using k-NN regression and different approaches were discussed to overcome those problems. Jaditz and Riddick (2000) analyzed the general algorithm for a k-Nearest-Neighbor forecast. Their paper emphasized how useful the k-NN regression is, but at the same time pinpoints the major decisions that a researcher should make using the approach, as well as discusses several problem areas. Jaditz and Riddick (2000) estimated a multivariate distribution. They also provided a more flexible code for k-NN than the one which came with standard packages and showed numeric examples of how to use the code.

Similarly, Jiang et al. (2009) have presented three main weaknesses of using the k-Nearest-Neighbor approach and have come up with three approaches for solving the issues. After many data comparisons, which were done in four groups, in three of the groups certain k-NN algorithms were compared. Lastly, in the fourth group, the hybrid approach was compared to each of the approaches. The three weaknesses of using the k-NN approach that the group identified were: (1) the Euclidean distance function which was used to measure the difference or similarity between two instances, (2) the

neighborhood size being assigned as an input parameter, and (3) the simple voting being the base of the class probability estimation. The classification accuracy was obtained using cross-validation. Various k-NN algorithms were tested on the same training sets. They also were evaluated on the same test sets. The group also made comparisons of related k-NN algorithms through a two-tailed t-test with a 95% confidence level. After their review of some of the k-NN algorithms, the research group formulated a hybrid algorithm called "dynamic k-Nearest-Neighbor naïve Bayes with attribute weighting" (DKNAW).

Azmi et al. (2010) in his research compared the performance of five forecasting models to forecast flood and drought warning with the real time operation. Based on several research studies that were suggesting the data fusion approach showed a better forecast result than using a single forecast approach, Azmi et al. studied a comparative measurement of data fusion including simple and weighted averaging. He did two case studies and in each case study used multiple linear regression, non-parametric k-Nearest-Neighbor regression, conventional multilayer perception, and an artificial neural network improved for extreme value forecasting. Azmi et al. found that using a mixture of data could considerably improve the forecast rather than using a single model. Besides the fact that it is better to use fusion data, he also came to the conclusion that data fusion by the k-NN method outperforms common methods by improving forecasts through decreasing the bandwidth of combined forecast and error of point forecast in both case studies.

The k-NN method has also enjoyed a few applications in economics. Barkoulas, Baum, and Chakraborty (2003) employed a nonlinear, nonparametric method to model stochastic interest rates. They applied a nonlinear autoregression to the data series by using a locally weighted regression, or *loess,* estimation model. Locally weighted regression is a way to estimate regression surface using multivariate smoothing procedures; it is an alternative way of calculating moving averages by locally fitting the function of independent variables. This is a nearest neighbor method and estimates out-of-sample forecasting performance with a measure of root mean squared error (RMSE). They compared the forecasting performance of the nonparametric fit to the performance of two linear models: the autoregressive (AR) model and the random-walk-with-drift model. The paper used nonstructural and univariate approaches, letting the data determine the regression function. The research approach was based on the historical behavior of individual securities' yield series to model nonlinearities. The paper provided evidence that the nonparametric fit was generated using the locally weighted regression was significantly improved compared to the simple linear model, which was chosen as a benchmark.

Gençay (1999) applied two methods which captured the nonlinearities in the conditional mean in studying how to forecast the spot foreign exchange rate returns. Those methods were the nearest neighbor and feedforward network regressions. The paper study concentrates on the linear and nonlinear predictability of the daily spot exchange rates from the simple forms of technical trading rules, also called moving average rule. To analyze the predictability of spot foreign exchange rate returns, Gençay

(1999) looked at the simple technical analysis methods of past sell-buy signals. Here, the random walk and the GARCH(1,1) models were used as benchmarks. The number of neighbors was chosen based on the cross-validation method which minimizes the mean squared error (MSE). Gençay (1999) compared two nonparametric and two parametric conditional mean estimators. The results showed that in contrast to feedforward, the k-Nearest-Neighbor regression (nonparametric model) provided significantly correct signs and turned out to be a more accurate forecasting method. The Diebold-Mariano test was applied in the end to prove the statistical significance of the predictions. In contrast to k-NN regression, the benchmark models did not generate significant sign predictions. The Diebold-Mariano test showed statistically insignificant results for parametric models.

Mizrach (1992) implemented a multivariate setting for nonlinear modeling of exchange rates in the European Monetary System (EMS). Three European currencies (franc, lira, and mark) were used at a daily frequency throughout the whole floating exchange rate period to forecast their multivariate nonlinear model. For that purpose, nearest neighbor generalization was employed. He found that exchange rates were greatly affected by multivariate information. He also found that the multivariate model performed much better than a univariate model in- and out-of-sample. Mizrach saw a 4-5% reduction in mean squared errors (MSE) from a range of about 750 returns daily. Using his own test statistic (from 1991), Mizrach concluded that the MSE was not statistically significant. Finally, to test his conclusion, the forecast was cross-validated by reversing the estimation and forecast samples. He saw that the model forecasting better into the late 1980s was predicting poorly backwards into the 1970s. Mizrach

concluded that the time-series representation for the 1980s was very different from the 1970s.

Nowadays, numerous papers investigate how accurate are the results of the forecasting methods that are being tested using different test statistics. Diebold and Mariano (1995) proposed tests of the null hypothesis of no difference in the accuracy of two competing forecasts. Their approach was based on the predictive performance and accuracy measures that can be tailored to a particular decision-making situation. Diebold and Mariano (1995) emphasized that the economic loss associated with a forecast may be poorly assessed by the usual statistical metrics. For example, forecasts are used to guide decisions, and the loss associated with a forecasting error is induced directly by the nature of the decision problem at hand. The team has stressed that their tests are valid for various types of loss functions. For instance, the loss function does not need to be quadratic, nor symmetric or continuous. The various tests utilized by the team have been applied to exchange rate forecasting. The series that Diebold and Mariano (1995) forecasted, and measured monthly, was the three-month change in the nominal Dollar/Dutch guilder end-of-month spot exchange rate. They reviewed two forecasts: the "no change" (0) forecast associated with a random-walk model, and the forecast implicit in the three-month forward rate. As for the point estimates, the random-walk forecast was found to be more accurate. The mean absolute error of the random-walk forecast was observed to be lower than the forward market forecast. They concluded that forecast errors can be non-Gaussian, non-zero mean, and at the same time correlated. Several

studies consider the question of which test statistics is better and what improvements should be made to overcome shortcomings of each test.

Harvey et al. (1997) studied the comparison of prediction records given two sources of forecasts of the same quantity. They compared the behavior of two tests and of modifications of those two tests to overcome the shortcomings in the original formulations. The two tests were Morgan-Grange-Newbold and Diebold-Mariano. They found that the Morgan-Granger-Newbold test was considerably less flexible than the Diebold-Mariano test, but it still had useful properties in one particular case: the null distribution of the statistics is known exactly in finite samples in the presence. However, as the Diebold-Mariano test showed, the Morgan-Granger-Newbold test can be quite over-sized in the case of two-step-ahead prediction even when the samples are really small and quite weak, and especially when heavy-tailed distribution of the forecast errors is present. Like the Morgan-Granger-Newbold test, the Diebold-Mariano test also showed discouraging results, but in simulation. They analyzed the performance of the two tests (Diebold-Mariano vs. Morgan-Granger-Newbold), made modifications to the tests to overcome the shortcomings in the original formulations and found that the modified test had better performance than the originals.

In the same way, Robledo and Zapata (2003) in their article provided an experimental evaluation of two tests: Diebold-Mariano (DM) and Stock and Watson[1] (SW), for choosing a forecasting model of quarterly data from the wheat market and

[1] Common trends test for the possibility of cointegration among nonstationary vector processes of integrated order one (Robledo et al., 2003)

evaluated how those models perform when small samples are used in an out-of-sample forecast by implementing the Monte-Carlo experiment. They mention that the model choice is usually based on the lowest MSE. Models are updated using fixed, rolling, and recursive schemes. In their paper, they used the Dickey-Fuller test to evaluate unit-roots. They concluded that the tests of differences in MSEs showed that one model was not better than the other. It was especially true for wheat exports, where the assumption of even 15% would not warrant that a new model should be adopted. The Monte Carlo experiment showed that the Diebold-Mariano test has better size, but the Stock and Watson test has better power. The Stock and Watson test dominated over the DM test even at the 5% level. They advised that the DM test would be a better criterion choice for forecasters who want to be conservative in evaluating two alternative forecasting models.

3. DATA

The research concentrates on five energy commodity prices. These commodities are natural gas, ethanol, crude oil, heating oil, and gasoline. The data for each commodity are monthly. Natural gas data covers January 1976 to April 2010 of U.S. average wellhead price, which was estimated from the New York Mercantile Exchange (NYMEX) futures closing price of near-month delivery of Henry Hub, and prevailing cash market prices (spot prices) at 5 major trading hubs: Henry Hub, LA; Carthage, TX; Katy, TX; Waha, TX; and Blanco, NM. The prices are expressed in dollars per thousand of cubic feet. Ethanol data covers from June 1989 to March 2010. The price data used here is expressed in dollars per gallon and are based on the average of the prices from 33 U.S. cities.[2] Crude oil data covers January 1986 to June 2010. The data are monthly. It was taken from Cushing, OK WTI[3] Spot Price FOB and are expressed in dollars per barrel. Heating oil data covers the period June 1986 to June 2010. The data are based on New York Harbor No. 2 Heating Oil[4] Spot Price FOB and are expressed in cents per gallon. Gasoline data covers December 2005 to June 2010. The data prices are based on

[2] Phoenix, Los Angeles, San Francisco, Denver, Bettendorf, Cedar Rapids, Boise, Chicago, Decatur, Pekin, Indianapolis, Kansas City, Wichita, Louisville, Lexington, New Orleans, Detroit, Niles, Minneapolis, Fargo, Lincoln, Omaha, Albuquerque, Las Vegas, Upstate NY, Cincinnati, Portland, Memphis, Nashville, Houston, Richmond, Seattle, Milwaukie. (Hart's Oxy Fuel News, 2010)

[3] West Texas Intermediate, also known as Texas Light Sweet. WTI is produced in Texas and South Oklahoma. Price from WTI serves as a reference for many other crude streams and is traded in Cushing, Oklahoma. WTI crude oil is used as a benchmark in oil pricing and is the primary commodity of NYMEX's future contracts. (EIN, 2010)

[4] A distillate fuel oil for use in atomizing type burners for domestic heating or for medium capacity commercial-industrial burner units, with distillation temperatures between 540-640 degrees Fahrenheit at the 90-percent recovery point; and the kinematic viscosities between 1.9-3.4 centistokes at 100 degrees Fahrenheit as defined in ASTM Specification D396-92. (EIN, 2010)

New York Harbor Conventional Gasoline[5] Regular Spot Price FOB and are expressed in cents per gallon.

All the data were taken from the Energy Administration's website except for ethanol, which was taken from Hart's Oxy-Fuel News. The time periods were chosen based on data availability. Further, within each commodity's timeframe, the sample period is split into two sections. The first section accounts for two-thirds of the data and is used for the in-sample forecast. The second section accounts for the remaining one-third of the data and is used for the out-of-sample forecast. Two types of models will be generated to perform one-step-ahead and two-step-ahead forecasts. Those models are the simple Autoregressive (AR) model and the k-Nearest-Neighbor (k-NN) model. The results from both models will be compared to determine which model forecasts better.

[5] Finished motor gasoline not included in the oxygenated or reformulated gasoline categories. Excludes reformulated gasoline blendstock for oxygenate blending (RBOB) as well as other blendstock.

4. MODELS

The thesis analyzes the simple Autoregressive (AR) model in comparison to k-Nearest-Neighbor (k-NN) forecast; thus two types of models are performing point forecast evaluation. The first model evaluated in the thesis is the simple AR model. As mentioned in the Data section of this thesis, for each method, twelve models will be evaluated. Each of these twelve models will have four sub-models. For instance, for the first sub-model of Model 1 (Model 1-1), we have one lag with an intercept. For Model 1-2, we have one lag without an intercept. For Model 1-3, we have one lag with intercept and seasonal harmonic variables. For Model 1-4, we have one lag without an intercept and with a seasonal component. The mathematical formulas of these four sub-models are illustrated below:

$$Model\ 1\_1 = \alpha_0 + \beta_1 dlnP_{t-1} + \varepsilon_t$$

$$Model\ 1\_2 = \beta_1 dlnP_{t-1} + \varepsilon_t$$

$$Model\ 1\_3 = \alpha_0 + \beta_1 dlnP_{t-1} + \sin_T + \cos_T + \varepsilon_t$$

$$Model\ 1\_4 = \beta_1 dlnP_{t-1} + \sin_T + \cos_T + \varepsilon_t$$

The same convention of intercepts and seasonality applies for Models 2 through 12. The only difference is the number of price change lags (e.g., two lags for Models 2-1, 2-2, 2-3, 2-4; three lags for Models 3-1, 3-2, 3-3, 3-4; and so on).

4.1 AR Model

After the data preparation is over, we are moving to the basic part of the research. The first thing we need to do is a simple OLS regression.

Consider Model 1-1. The regression function for the model has an intercept and one lag. Mathematically, the regression equation can be written as follows:

$$\widehat{dlnP_t} = \alpha_0 + \beta_1 dlnP_{t-1} + \varepsilon_t$$

where $\widehat{dlnP_t}$ is the predicted price in period *t*, α_0 is the intercept, β_1 is the regression coefficient, $dlnP_{t-1}$ is the change in the price, and ε_t is the error term. The primary purpose of regression analysis is to predict the value of dlnP, given the knowledge of its lags. The regression analysis returns the values for the intercept α_0 and the coefficients $\beta_1, \ldots, \beta_n$ (which are unknown parameters). For a natural gas in-sample forecast, we consider a sample of 275 observations, i.e., two-thirds of the whole data. Certainly, the equation will not exactly fit our sample data; deviations (errors) will appear. The variable ε_t in the models is assigned as the error term where the mean value of ε_t is zero:

$$\widehat{\Delta P_t} = \alpha_0 + \beta_1 \Delta P_{t-1},\ \varepsilon_t = \widehat{\Delta P_t} - \Delta P_t$$

As a result, the average difference between the observed values (ΔP_t) and the predicted values ($\Delta \widehat{P_t}$) will be zero.

Certain assumptions are necessary to calculate α_0 and β_1. We can use the least squares method to calculate the unknown parameters if the error on the models (ε_1, ε_2, $\varepsilon_3, \ldots$, ε_t) are random and distributed independently. The Markov theorem states that α_0 and β_1 are the BEST unbiased linear estimates. BEST means that the variance of the error terms is at minimum.

For Model 1-1 and Model 1-2, we use simple linear regression. The formulas used for these models to calculate the coefficient and the parameter are similar to those of simple linear regression.

According to Brennan (1960), to find the values for α_0 and β_1, we should first solve the following pair of normal equations:[6]

$$(1)\sum dlnP_t = t * \alpha_0 + \beta_1 \sum dlnP_{t-1}$$

$$(2)\sum (dlnP_t * dlnP_{t-1}) = \alpha_0 \sum dlnP_{t-1} + \beta_1 \sum (dlnP_{t-1})^2$$

Then, we multiply (1) by ∑dlnP and (2) by t and subtract (1) from (2) to eliminate α_0 and find the value for β_1, thus

$$\beta_1 = \frac{\mathrm{t} * \sum \mathrm{dlnP_t} * \mathrm{dlnP_{t-1}} - \sum \mathrm{dlnP_t} * \sum \mathrm{dlnP_{t-1}}}{\mathrm{t} * \sum (\mathrm{dlnP_{t-1}})^2 - (\sum \mathrm{dlnP_{t-1}})^2}$$

After substituting this value for dlnP_lag1 in (1), we get the value for α_0:

$$\alpha_0 = \frac{\sum \mathrm{dlnP} - \mathrm{a1} * \sum \mathrm{dlnP_{lag1}}}{t}$$

---

[6] Let $D = \varepsilon_1{}^2 + \varepsilon_2{}^2 + \varepsilon_3{}^2 + \cdots + \varepsilon_t{}^2 = (\mathrm{dlnP_1} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_1}})^2 + (\mathrm{dlnP_2} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_2}})^2 + \cdots + (\mathrm{dlnP_t} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_t}})^2$

D is to be minimized with respect to α_0 and β_1:

(1) $\frac{\partial D}{\partial \alpha_0} = 0$

(2) (2) $\frac{\partial D}{\partial \beta_1} = 0$

(3) $\frac{\partial D}{\partial \alpha_0} = -2(\mathrm{dlnP_1} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_1}}) - 2(\mathrm{dlnP_2} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_2}}) - \ \ldots - 2(\mathrm{dlnP_t} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_t}}) = 0$

(4) $\frac{\partial D}{\partial \beta_1} = -2\mathrm{dlnP_{(t-1)_1}}(\mathrm{dlnP_1} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_1}}) - 2\mathrm{dlnP_{(t-1)_2}}(\mathrm{dlnP_2} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_2}}) - \ldots - 2\mathrm{dlnP_{(t-1)_t}}(\mathrm{dlnP_t} - \alpha_0 - \beta_1 \mathrm{dlnP_{(t-1)_t}}) = 0$

Dividing both sides of each equation by 2 and combining like terms gives

(5) $-\sum_{t=14}^{n} dlnP_t + \alpha_0 n + \beta_1 \sum_{t=14}^{n} \mathrm{dlnP_{(t-1)_t}} = 0$

(6) $-\sum_{t=14}^{n} dlnP_t * \mathrm{dlnP_{(t-1)_t}} + \alpha_0 \sum_{t=14}^{n} \mathrm{dlnP_{(t-1)_t}} + a1 \sum_{t=14}^{n} (\mathrm{dlnP_{(t-1)_t}})^2 = 0$

(Brennan, 1960)

After the value of the constant and the slope coefficient were calculated, we substituted them into the Model 1-1 regression formula. Given the values for $\mathrm{dlnP}_{(t-1)_t}$, this equation provides the estimated values of $\widehat{dlnP_t}$.

Above was the calculation of the constant and the slope coefficient given simple linear regression. For the remaining of the models Model 1-3 through Model 12-4, we cannot use simple linear regression equations to calculate the unknown parameters; thus, we use the equations for multiple linear regression.

The multiple linear regression method is different from the simple regression in that it can have two or more explanatory variables for calculating the predicted values.

Consider the case of Model 2-1:

$$\widehat{dlnP_t} = \alpha_0 + \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \varepsilon_t$$

As in the case of simple linear regression, the calculated values of the constants α_0, β_1 and β_2 give the BEST unbiased linear estimates of $\mathrm{dlnP_t}$ (Brennan, 1960). In the formula shown above, α_0 is the intercept, and β_1 and β_2 are the regression coefficients. The error term ($\varepsilon_t = \widehat{dlnP_t} - dlnP_t$) is considered to be random with a mean of zero. As in the case of simple linear regression, here we also use the method of least squares. Because we are using linear functions of two variables ($\mathrm{dlnP_{lag1}}$ and $\mathrm{dlnP_{lag2}}$), we are dealing with three constants and three normal equations:

(3) $\sum dlnP_t = \alpha_0 n + \beta_1 \sum dlnP_{t-1} + \beta_2 \sum dlnP_{t-2}$

(4) $\sum dlnP_t * dlnP_{t-1} = \alpha_0 \sum dlnP_{t-1} + \beta_1 \sum dlnP_{t-1}{}^2 + \beta_2 \sum dlnP_{t-1} * dlnP_{t-2}$

(5) $\sum dlnP_t * dlnP_{t-2} = \alpha_0 \sum dlnP_{t-1} + \beta_1 \sum dlnP_{t-1} * dlnP_{t-2} + \beta_2 \sum dlnP_{t-2}{}^2$

The procedure utilized to arrive at the above normal equations is the same as in Footnote 6.

To solve this system of equations for α_0, β_1 and β_2, we first treat two of the equations, for instance, (3) and (4), to eliminate one variable (α_0). Next, we treat equations (4) and (5) together and multiply the members of equation (5), so that we again eliminate α_0 (Brennan, 1960).Next, we will get two equations with two unknown variables. We have already shown the procedure of solving two equations with two unknown variables. We then calculate the value for β_1, substitute the calculated value into either of the equations (4) or (5), and get the value for β_2. We can obtain values for α_0 after substituting the values for β_1 and β_2 in any of the three equations above. As a result, we end up with all the values for the unknown parameters.

Once the values for the unknown parameters are calculated, we are able to calculate fitted values for the in-sample portion of our forecast ($\widehat{dlnP}$). After getting the summaries for all 48 models, we can move forward to AIC calculations to select the best model. Refer to Appendix A for the R code.

4.1.1 AIC Calculations and Model Selection

After the regression results are obtained, we need to measure the goodness-of-fit of our models. Akaike Information Criterion (AIC) will be used to make the selection among the models. When using AIC, the model with the lowest AIC score is considered to be the best. So, we estimate each model and calculate the AIC score for each. These calculations will make use of all of the observations in the initial sample. AIC calculations are based on the projected values (i.e., "$\widehat{dlnP}$") and actual observed values

("dlnP") of the dependent variable for all observations in the initial sample (except for the omitted variables due to the lags in the models). Since the most lags we considered is 12, we calculate the AIC for all models being evaluated using observations $dlnP_{14}$ through the end of the in-sample-forecast. We lose one observation due to differencing and 12 observations due to AR lags. Thus, a total of 13 observations are lost. An important point here is that we need to use the same number of residuals across different models. If we use different numbers of residuals across different models, the models with more observations/residuals in the sum of squared residuals will look worse simply because they have more residuals being summed. When calculating AICs, we use actual observations of dlnP and the projections ($\widehat{dlnP}$).

The formula for AIC calculation is:

$$AIC = 2p + n[\ln(\frac{RSS}{n})]$$

where p is the number of parameters and changes for each model; n is the number of observations (same for each model); and RSS is the sum of squared residuals

$$RSS = \sum_{t=i}^{n} {\varepsilon_1}^2 + {\varepsilon_2}^2 + {\varepsilon_3}^2 + \cdots + {\varepsilon_t}^2,$$

where $\varepsilon_t = \widehat{dlnP_t} - dlnP_t$ (Akaike, 1981).

Once the AIC scores for all the 48 models are calculated, we chose the model with the lowest AIC score for an out-of-sample forecast.

4.1.2 One-Step-Ahead Out-of-Sample Forecast

Model selection does not only depend on the goodness-of-fit but also on the degree of objectiveness of the analyses. If a model is the best in the in-sample forecast,

that does not necessarily mean that the out-of-sample forecast will be accurate. Hence, many researchers are evaluating the performance of the best model from an in-sample forecast and an out-of sample forecast to help choose the best model. When we say out-of-sample forecast, we mean that the predicted observations are not used for model fitting. The whole idea is that we divide the data into two sub-periods. As mentioned above, the first period includes only two-thirds of the data and is called estimation sub-sample. The second period includes the remaining one-third of the data and is called forecasting sub-sample. Suppose we have N data points x_1, x_2, …, x_N. The whole data are divided into two parts. The first part is $\{x_1,\ldots,x_n\}$ and the second part is $\{x_{n+1},\ldots.,$ $x_N\}$. Here, n is the initial forecast origin and n=2N/3.

According to AIC scores, Model 3-4 was selected for natural gas, i.e., no constant, three lags with the seasonal component:

$$dlnP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \beta_3 dlnP_{t-3} + sin_t + cos_t + \varepsilon_t$$

To perform one-step-ahead out-of-sample forecast, we need a rolling estimation window. Consider the natural gas example where the in-sample forecast ends at period 275. To forecast observation 276, we will be using only information available through observation 275. Then, by incrementing the estimation period from 14 to 276, we will re-estimate the model and use those estimates and data trough observation 276 to forecast observation 277 and so forth. For the out-of-sample forecast, the same regression analyses are performed as for the in-sample-forecast. Refer to Appendix A for the R code.

4.1.3 Two-Step-Ahead Out-of-Sample Forecast

Again, consider the case of natural gas. Since we are now doing a two-step-ahead forecast, we have prices not only for periods 14 to 275 but also for period 276. So, to perform a forecast, we need to take the forecasted price for period 276 and replace it with the actual price in period 276:

$$dlnP_{276} = \widehat{dlnP_{276}}$$

The reason for the replacement is so that the next prediction, i.e., period 277, is based on the predicted value of the previous period instead of the actual value. The regression model for the natural gas two-step-ahead forecast will look like:

$$\widehat{dlnP_{277}} = \beta_1 \widehat{dlnP_{276}} + \beta_2 dlnP_{275} + \beta_3 dlnP_{274} + sin_{277} + cos_{277} + \varepsilon_{277}$$

where $\widehat{dlnP_{277}}$ is the two-step-ahead forecast and $\widehat{dlnP_{276}}$ is the predicted value from the one-step-ahead forecast. Refer to Appendix A for the R code.

4.2 k-NN Model

As was discussed in the Literature Review section, one of the central concepts of k-NN is Memory Based Reasoning (MBR). The biggest advantage of MBR is that it uses the data as is: it does not account for the format of the record, which is the case in many other data mining techniques. The key ingredients of MBR are the distance function, the combination function, and the number of neighbors. The distance function is calculating the distance between any two records. The combination function combines results from several neighbors to give a prediction.

There can be two different approaches to k-NN. In one case, the combination function is used, which classifies or categorizes the data into several categories by

assigning classification codes to the data. For example, "gender" is a categorical data. In another case, the data is continuous, and no classification is necessary.

There are three ways to compute the distance function: Manhattan distance summation, normalized summation, and Euclidean distance. Most researchers use Euclidean distance, and it is used here as well.

4.2.1 In-Sample Forecast

As in case of the AR model and with the k-NN as well, we are using the same number of observations. There are two basic steps that we need to follow to calculate k-NN: calculate the distance function and choose the number of neighbors. The minimum number of neighbors is 2, but more neighbors may give a more accurate result. In our research, we have considered the use of 5 to 40 neighbors and then decided on the best number of neighbors by calculating the MAE for each of the models. Once the best number of neighbors is selected, we need to perform one-step-ahead and two-step-ahead forecasts. But before that, we need to do an in-sample k-NN forecast.

For an in-sample forecast, the choice of RHS variables is based on the AIC scores for the AR model. The best model that was chosen based on the AIC scores for AR model was used in k-NN calculations. The AR and k-NN approaches that are being tested use the same information to generate forecasts. Thus, the regression function for the in-sample forecast is similar to the one for the AR in-sample forecast:

$$dlnP_t = \alpha_0 + \sum_{i=1}^{J} \beta_i dlnP_{t-i} + \varepsilon_t \quad \text{for J=1,…,12}$$

Here, t=14,…, y, where y is the last in-sample forecast.

The fundamental part in k-NN is the selection of neighbors based on which the predicted values are generated. One can always perform the analysis manually by gradually increasing the number of neighbors but it is a long process and using software would be very helpful. Before selecting the appropriate number of neighbors, we need to calculate the distance function.

There are several methods that can be used to measure the distance between vectors but the most common distance function, which is also used in our code, is the Euclidian distance, which is defined as:

(6) $d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{t}(x_i - y_i)^2}$

where $\vec{x}$ and $\vec{y}$ are vectors of length m and x_i is the i^{th} component of vector $\vec{x}$ (also called the base period). Despite its wide use, Euclidian Distance has two major weaknesses:
(1) If one of the inputs has a wider range than the other, it can dominate the others. For example, if there are two vectors K and L, and K can take values from 1 to 100 and L can take values from 1 to 10, then K's influence on the distance is bigger since it takes higher values. However, the value of 10 in L can be more influential than the values of 100 in K (Kidron and Klein, 2007)
(2) If the amplitude of changes in one of the attributes of each vector is significantly larger or smaller than that of the others, the distance function can reflect a wrong interpretation of changes in each attribute. For example, consider there are two different series: one measuring the speed of the wind, and the second the direction of the wind. Only when the speed of the wind is more than 40, is it considered a hurricane, while the wind direction can be from 0 to 360. The Euclidian Distance does not show that,

although 40 is a small number, it has a more important meaning than 360 (Kidron and Klein, 2007).

To overcome these problems, researchers who are using Euclidian Distance need to use normalization. By normalization we mean that each attribute in the vector should be normalized based on its average and standard deviation. First, we should define:

$$\underline{\vec{x}} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_l) \; where \; \underline{x}_i = \frac{x_i - \overline{x_l}}{\sigma_i}$$

where $\overline{x_l}$ is the average of the attribute i, i=1,…,t, over all time-series elements, and σ_i is the standard deviation. So, if we replace the above equation in (1), we will have the following:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{t}(x_i - y_i)^2} = \sqrt{\sum_{i=1}^{t}(\frac{x_i - \overline{x_l}}{\sigma_i} - \frac{y_i - \overline{y_l}}{\sigma_i})^2} = \sqrt{\sum_{i=1}^{t}\frac{(x_i - y_i)^2}{\sigma_i}}$$

Now that we know how to calculate the distance, it is time to explain the logic behind the k-NN regression. Consider natural gas data. Our in-sample forecast covers periods 14-275 and out-of-sample forecast covers periods 275-410. After the model selection, we need to decide on the number of neighbors. Assume that the base period is 275. For the base period of 275, we are attempting to forecast the value of $dlnP_{275}$ while the values for all the lags (3 lags for natural gas) are already available. For each period before 275, we calculate the Euclidean distance from base value of each lag to each period's value. Euclidean distance calculation for $dlnP_{lag1}$ is:

$$d\left(\overrightarrow{dlnP_{lag1_{275}}}, \overrightarrow{dlnP_{lag1_f}}\right) = \sqrt{(dlnP_{lag1_{275}} - dlnP_{lag1_f})^2}$$

where f stands for periods 14 through 274. After Euclidean Distance calculations are completed for all the lags in all periods, for each period, we average all the distances. The average Euclidean distance is:

$$\frac{\sqrt{\left(dlnP_{lag1_{275}} - dlnP_{lag1_f}\right)^2} + \sqrt{\left(dlnP_{lag2_{275}} - dlnP_{lag2_f}\right)^2} + }{3}$$

$$+ \frac{\sqrt{(dlnP_{lag3_{275}} - dlnP_{lag3_f})^2}}{3}$$

where f stands for periods 14 through 275.

Now we are ready to test and choose the appropriate number of neighbors for our model. To evaluate the appropriate number of neighbors, we consider possible number of neighbors from 5 through 40. Consider the case of 5 neighbors. Given the distances calculated above, we identify the periods with the 5 lowest distances. Then, for those 5 periods, we average their corresponding dlnP values. The resulting average will be the predicted value of $dlnP_{275}$ (the base period) or:

$$\widehat{dlnP} = \frac{\sum_{i=1}^{k} dlnP_i}{k}$$

where $dlnP_t$ corresponds to the dlnPs of the 5 observations with the lowest distances and k is the number of neighbors considered. Refer to Appendix A for the R code.

By repeating the same process for neighbors 6 through 40, we end up with 36 different predictions for $dlnP_{t+1}$. Then, we calculate MAEs for each of the 36 cases.

The MAE is calculated using the formula:

$$MAE_{Knn} = \frac{1}{n} * \sum|r_1 + r_2 + r_3 + \cdots + r_n|\,,$$

where n is the number of observations we are using for an in-sample forecast of k-NN, and $r_1 + r_2 + r_3 + \cdots + r_n$ are the residuals from the model.

The appropriate number of neighbors is chosen based on the lowest MAE. For natural gas, for example, it turned out that k-NN, which has 20 neighbors, has the lowest MAE. Thus, for the out-of-sample forecast, 20 neighbors were used.

Unlike the AR model, in k-NN, the sample is rolled one observation forward and the second observation in the forecast sample is studied in the same manner as described for the first forecast observation above. Thus, new k-NN predictions are never based on the old ones. This procedure is repeated until all observations are covered in the forecast sample.

4.2.2 Out-of-Sample Forecast

As in the case of the AR model, and for the k-NN model, we will do the out-of-sample forecast to compare the results later. When performing the k-NN out-of-sample forecast, changes need to be done to the right-hand side and left-hand side variables. In contrast to k-NN in-sample forecast, where the left-hand side and right-hand side variables were including only in-sample portions of the data, out-of-sample forecasts include the whole data from period 14 to t. One-step-ahead and two-step-ahead forecasts are performed in the same way as in the in-sample portion of the research. The only difference is in the periods we are using for the forecast. Assume B_t is the vector of attribute values for observation t (here attributes refer to the RHS variables), then, the predicted values (responses) from one-step-ahead and two-step-ahead out-of-sample forecasts would be conditioned on the values of attributes in time t:

$$dlnP_{t+1} = f(B_t), \quad dlnP_{t+2} = f(B_t)$$

where $dlnP_{t+1}$ stands for the responses or LHS variables in one-step-ahead forecasts and $dlnP_{t+2}$ stands for the responses in two-step-ahead forecasts. The process of calculating the predictions is the same as in the in-sample forecast section of k-NN. After out-of-sample one-step-ahead and two-step-ahead forecasts are calculated, we need to calculate MAEs for both forecasts and compare the results to the ones from the AR model. Refer to Appendix A for the R code.

5. METHODOLOGY

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) methods are used in the thesis to compare the results from the AR and k-NN models.

5.1 MAE Calculations

In statistics, Mean Absolute Error (MAE) is calculated to determine how close the predicted values are to the actual values. The lower the MAE, the closer the forecasted values are to the actual observations. MAE is calculated as:

$$MAE = \frac{1}{n}\sum_{t=i}^{n} |w_i - \widehat{w_i}|$$

where n is the number of observations, w_i represents the actual values, $\widehat{w_i}$ is the forecasted values from the models, and t is the periods.

5.2 RMSE Calculations

Usually, it is better to explain the results of the research using Root Mean Squared Error (RMSE) rather than MAE. The reason for this is that the RMSE results are expressed in the same unit as the data and can be considered as a representative of the error size. MAE is also measured in the same unit as the actual data, but it is smaller to some extent than the RMSE. For some researchers, it is easier to understand the MAE. RMSE is calculated using the following formula:

$$RMSE_t(l) = \sqrt{MSE_t(l)}$$

where l is the number of forecast periods and MSE is the Mean Squared Error. In statistics, MSE measures the average of error square. MSE is calculated by the following formula:

$$MSE = \frac{SSE}{n}$$

where n is the number of observations and SSE is the Sum of Squared Errors, which is calculated using the following formula:

$$SSE = \sum_{t=1}^{n} (y - \bar{y})^2$$

where y represents the actual observations and $\bar{y}$ is the forecasted observations. Once the RMSE is calculated, we are ready for the k-Nearest-Neighbor forecast. After the k-NN forecasting results are ready, we will compare them with the results from the simple AR model and, based on the comparisons, we will make our conclusions.

5.3 Analysis of Forecasting Accuracy of Models

After making a prediction with a specific model, one needs to determine how accurate the forecasting results are because people use forecasts to make important decisions and, if the model is not accurate, those decisions will also be inaccurate. We use the Diebold-Mariano test to determine the relative forecasting accuracy of our models. The idea behind the test is that, given forecasts from two models and the associated forecast errors, we can estimate the expected loss or, in other words, its accuracy associated with each of the forecasts. Here, the null hypothesis is that both models have the same forecast accuracy:

$H_0: E[g(e_{it})] = E\left[g(e_{jt})\right]$ or $E[d_t] = 0$

$H_1: E[g(e_{it})] \neq E\left[g(e_{jt})\right]$ or $E[d_t] \neq 0$

where $g(e_{it})$ and $g(e_{jt})$ are the functions of forecast errors from the AR and k-NN models, and d_t is the loss differential; $d_t = [g(e_{it}) - g(e_{jt})]$.

The quality of a forecast is judged on a particular loss function. Two popular forms of loss function are: squared error loss: $g(e_{it}) = (e_{it})^2$, and absolute error loss: $g(e_{it}) = |e_{it}|$. In the research, we will be using the squared error loss function.

The Diebold-Mariano test statistic is based on the loss differential explained above and is as follows:

$$S = \frac{\bar{d}}{(\widehat{avar}(\bar{d}))^{1/2}} = \frac{\bar{d}}{(\widehat{LRV}_{\bar{d}}/T)^{1/2}}$$

Where $\widehat{avar}$ is the asymptotic variance of $\bar{d}$.

$$\widehat{avar}(\bar{d}) = \frac{1}{T - T_0} \sum_{j=-q}^{q} k(j,q)\hat{\Gamma}_j$$

where k(.) is a kernel function and k(.) = 0; j > q, j =1,2, depending on the loss function, and $\hat{\Gamma}_j$ is the j^{th} autocovariance function estimate.

$$\bar{d} = \frac{1}{T_0} \sum_{t=t_0}^{T} d_t$$

$$LRV_{\bar{d}} = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j, \qquad \gamma_j = cov(d_t, d_{t-j})$$

where LRV is the estimate of the long-run variance for $\sqrt{T}\bar{d}$, which is used in statistics because of serial correlation between the sample of loss differentials when h>1. Under the null hypothesis of no significant differences, the DM test statistic is distributed as

standard normal distribution and the two-sided test is used as a default. The null hypothesis of equal predictive accuracy of DM shows

$$S \overset{A}{\sim} N(0,1)$$

So, at the 5% significance level, we reject the null hypothesis if

$$|S| > 1.96$$

It is also possible to compute the one-sided test. Refer to Appendix A for the R code.

6. ESTIMATION AND RESULTS

The calculations performed for this research were completed in R computer software. However, before inputting the data into R for forecasting, Microsoft Excel was used to better organize the data.

First, numbers were assigned to the periods in our monthly data starting from one. Second, the data were logarithmically transformed. For commodity prices, it is important to run them through logarithmic transformation because the transformed prices tend to more closely conform to the types of econometric assumptions we like to make (normally distributed innovations, constant variance of innovations, etc.). To use logarithmic transformation, we use the following formula:

$$lnP_t = ln(P_t),$$

where P is the original level's price observation. We use $\ln(P_t)$ in the modeling. Third, the logarithmically transformed data were differenced. Because of differencing, with the addition of each lag, we lose one observation. For instance, when we are doing the log differencing, which is $dnlP_t = lnP_t - lnP_{t-1}$, we are losing one observation. With the first differencing, which is $dlnP_{lag1} = lnP_{t-1} - lnP_{t-2}$, we lose one more observation, thus eliminating a total of two observations. Losing one observation at each lag, we end up with 13 missing observations at the 12^{th} lag. This is illustrated in Appendix B. For each of the components of our models to have equal period lengths, we need to eliminate the first 13 periods (months). Therefore, for natural gas, for example, we consider that our data starts in February 1977 instead of January 1976. Fourth, 12 lags were generated. After differencing the data, we need to decide how many lags our model will need to

have. To do that, we will test 12 different AR models estimated by Ordinary Least Squares (OLS). Model 1 will have one lag. Model 2 will have two lags. Model 3 will have three lags and each successive model will have one additional lag than the previous model. As a result, Model 12 will have 12 lags.

$$dlnP_t = \alpha_0 + \sum_{i=1}^{J} \beta_i dlnP_{t-i} + \varepsilon_t \quad \text{for J=1,…,12}$$

where $dlnP_{t-i}$ is the dependent variable in period *t-i*, α_0 is the intercept; β_i represent price change sensitivity in periods *t-i* to price changes in period *t* or, in other words, regression coefficients, and ε_t is the vector of disturbances or the error term. For each of the 12 lags, $dlnP_{t-i} = lnP_{t-i} - lnP_{t-i-1}$. Lastly, seasonal harmonic variables were added to capture the cyclical patterns. If seasonality is not considered in the regression, the estimation error may increase and the coefficients of variables that are correlated with the seasonality may be biased. By inserting seasonal harmonic variables in the form of sine and cosine, we insure smoother transitions than if we were to use dummy variables. To add seasonal components, we need to calculate sine and cosine for each t period and add to the model. The formula to calculate sine and cosine in period t is as follows:

$$Sin_t = Sin(2* \Pi * t/12),\ Cos_t = Cos(2*\Pi*t/12)$$

where Π=3.14159 and t is the corresponding period for each commodity.

Now that we know how the calculations were done, it is time to look at the results. Table 1 summarizes model selections for each of the five commodities. You can see from the table that, according to the AR model for natural gas, based on the lowest AIC score, Model 3-4 was chosen to be the best. The remaining calculations were based

on that model. k-NN for natural gas was also calculated based on Model 3-4. For heating oil, Model 2-4 had the lowest AIC score; for gasoline, again Model 2-4 turned out to be the best; for ethanol – Model 5-4; and for crude oil Model 5-2. Figures 1-5 show the AIC scores for all five commodities. For natural gas more lags the model has, the better is the AIC score. This does not mean that the very last model is always the best. For each commodity, one can notice a pattern that is the same throughout all the models. Consider Figure 2. Within each model, sub-models 2 and 4 are preferred over sub-models 1 and 3. Sub-models 2 and 4 do not have intercepts; therefore, models without intercept are preferred for heating oil analysis. The same logic is used to explain the pattern for the AIC scores of the rest of the commodities.

Table 1. Summary of Models

Natural Gas	$dnlP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \beta_3 dlnP_{t-3} + \sin T_t + \cos T_t + \varepsilon_t$
Heating Oil	$dnlP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \sin T_t + \cos T_t + \varepsilon_t$
Gasoline	$dnlP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \sin T_t + \cos T_t + \varepsilon_t$
Ethanol	$dnlP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \beta_3 dlnP_{t-3} + \beta_4 dlnP_{t-4} + \beta_5 dlnP_{t-5} + \sin T_t + \cos T_t + \varepsilon_t$
Crude Oil	$dnlP_t = \beta_1 dlnP_{t-1} + \beta_2 dlnP_{t-2} + \beta_3 dlnP_{t-3} + \beta_4 dlnP_{t-4} + \beta_5 dlnP_{t-5} + \varepsilon_t$

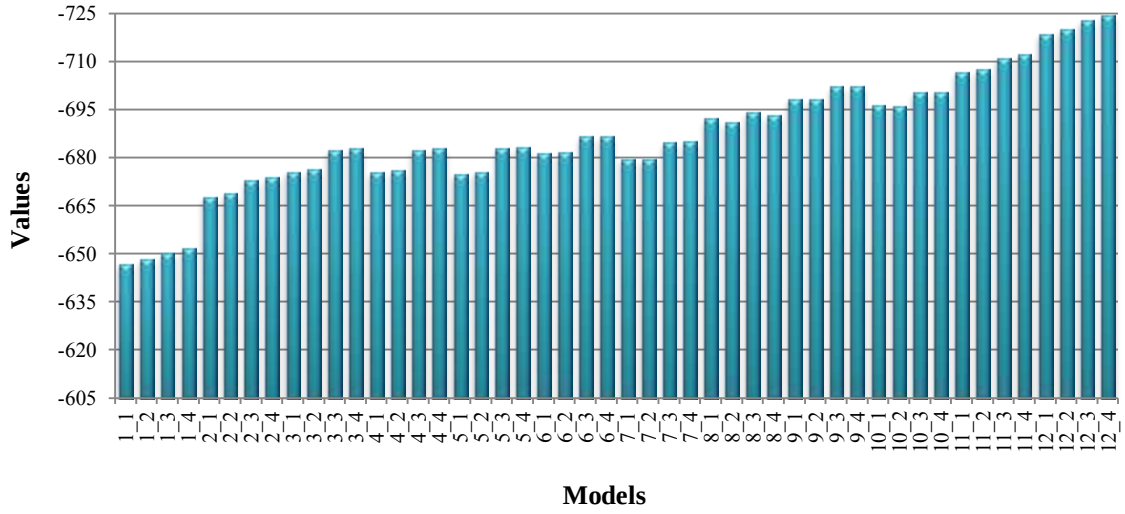

Figure 1. Graphical illustration of AIC scores for natural gas in-sample forecast

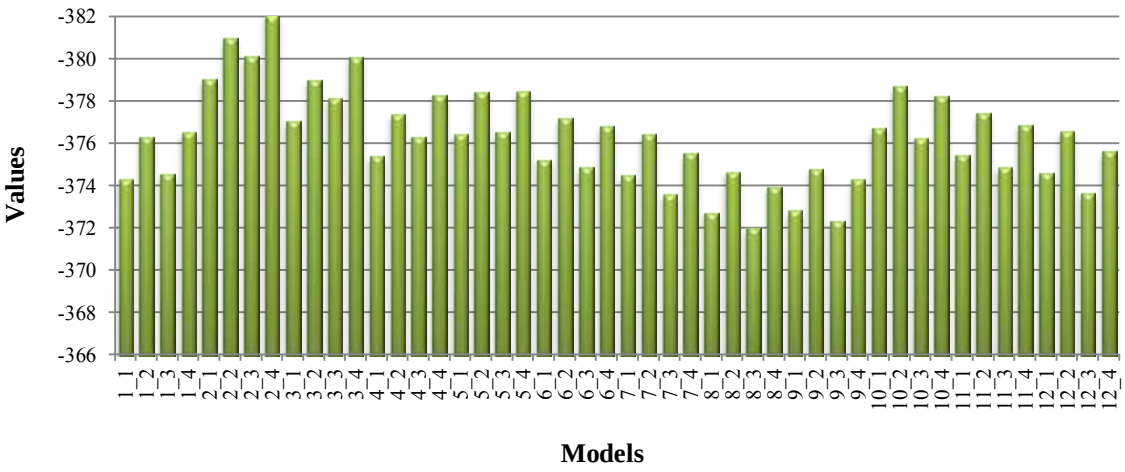

Figure 2. Graphical illustration of AIC scores for heating oil in-sample forecast

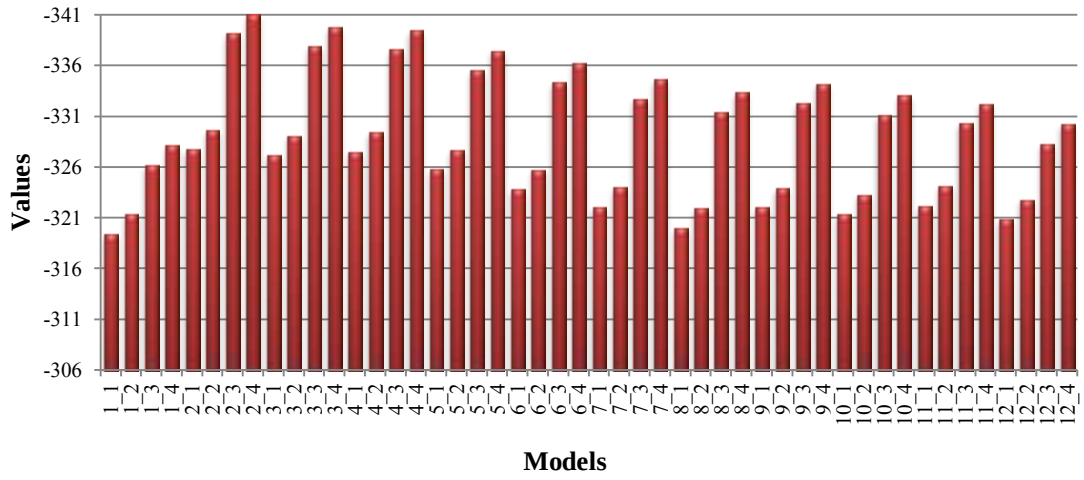

Figure 3. Graphical illustration of AIC scores for gasoline in-sample forecast

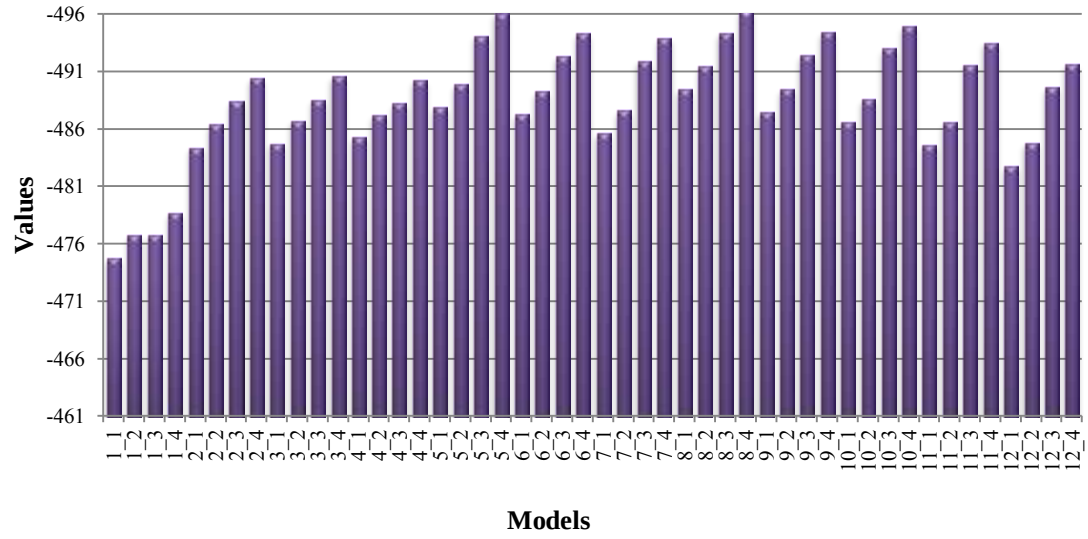

Figure 4. Graphical illustration of AIC scores for ethanol in-sample forecast

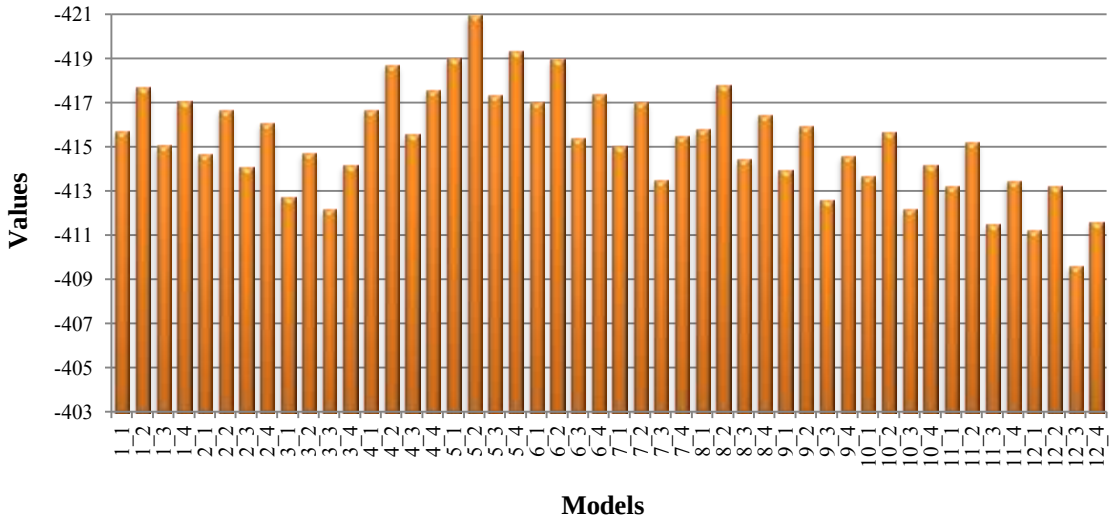

Figure 5. Graphical illustration of AIC scores for crude oil in-sample forecast

Figures 6-10 graphically illustrate MAEs for each commodity, based on which the best number of neighbors was chosen for each model. Based on the results for natural gas, 24 neighbors were chosen. For heating oil – 37 neighbors, for gasoline – 22, for ethanol – 38 and for crude oil – 28 neighbors were chosen. The in-sample and out-of-sample portions of k-NN forecast were based on the number of neighbors with the lowest MAE. The graphs in Figures 11-15 demonstrate the actual vs. predicted values for both AR and k-NN one-step-ahead out-of-sample forecasts for each of the five commodities.

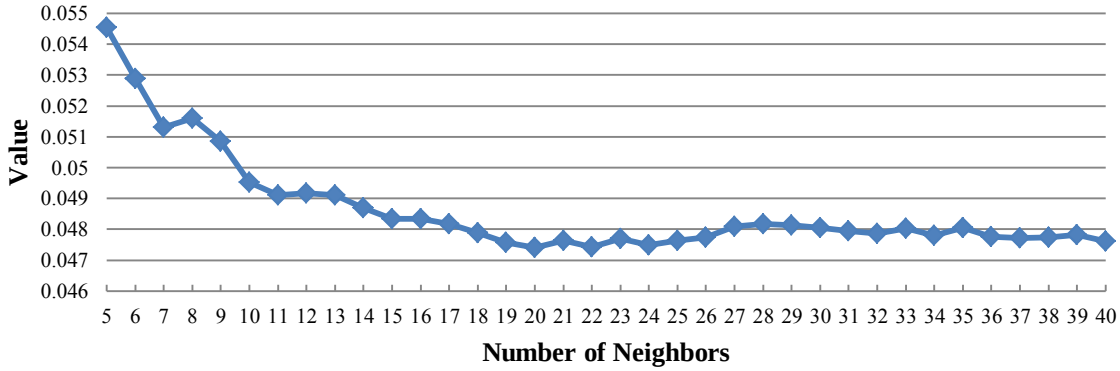

Figure 6. Graphical illustration of MAEs for natural gas

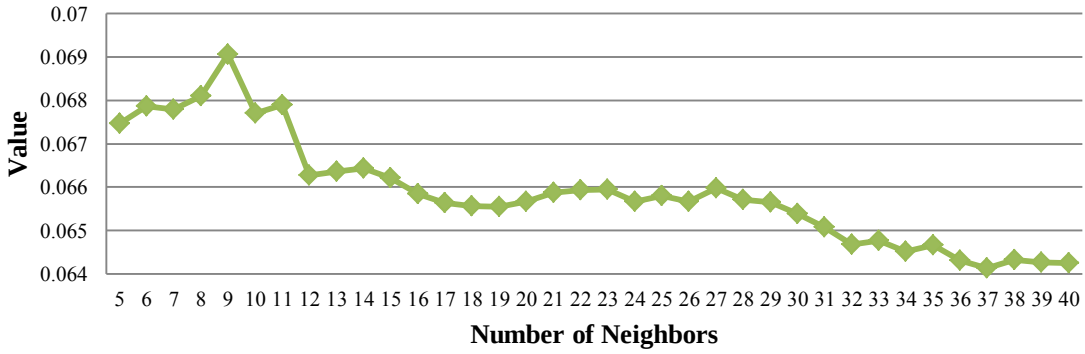

Figure 7. Graphical illustration of MAEs for heating oil

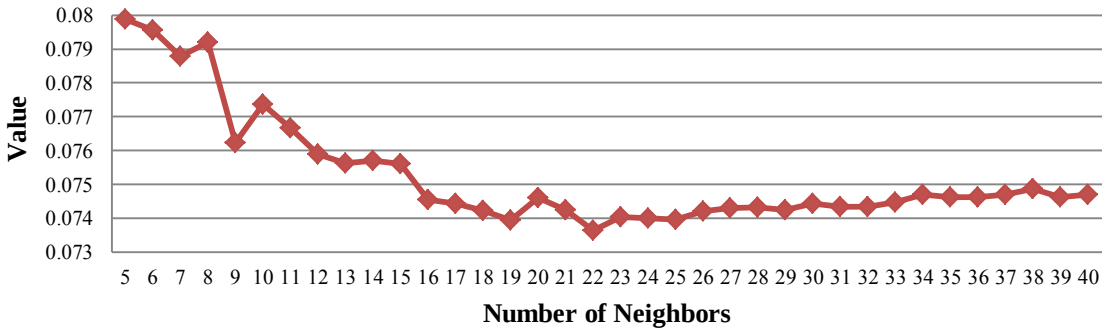

Figure 8. Graphical illustration of MAEs for gasoline

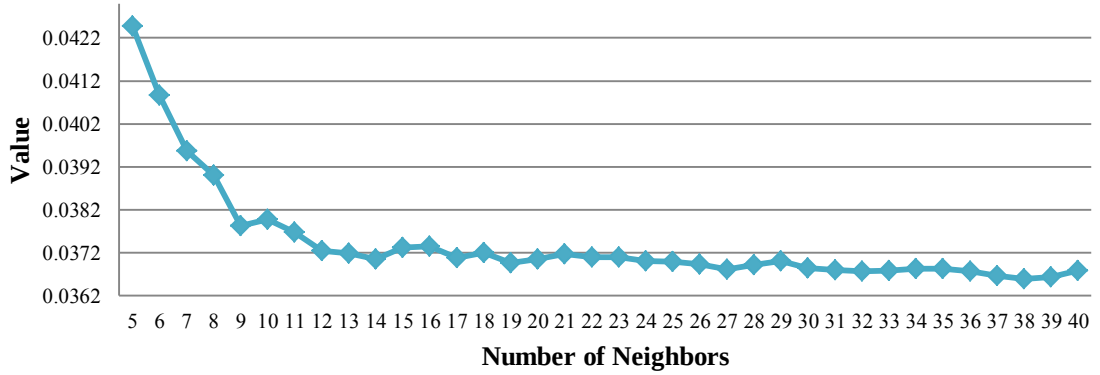

Figure 9. Graphical illustration of MAEs for ethanol

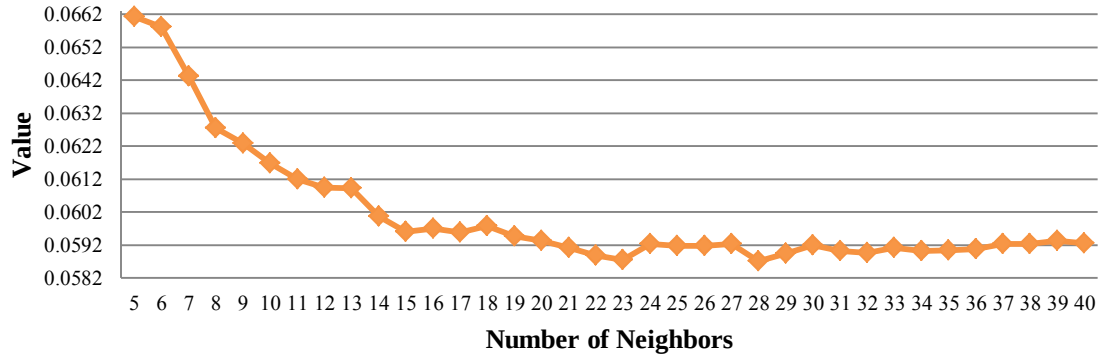

Figure 10. Graphical illustration of MAEs for crude oil

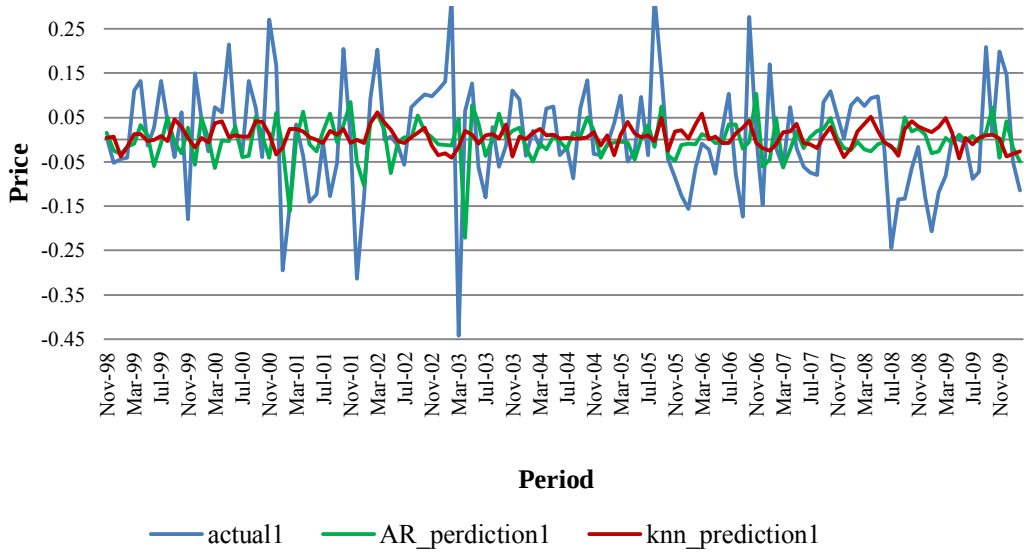

Figure 11. Graphical illustration of predicted vs. actual values for natural gas one-step-ahead out-of-sample forecast

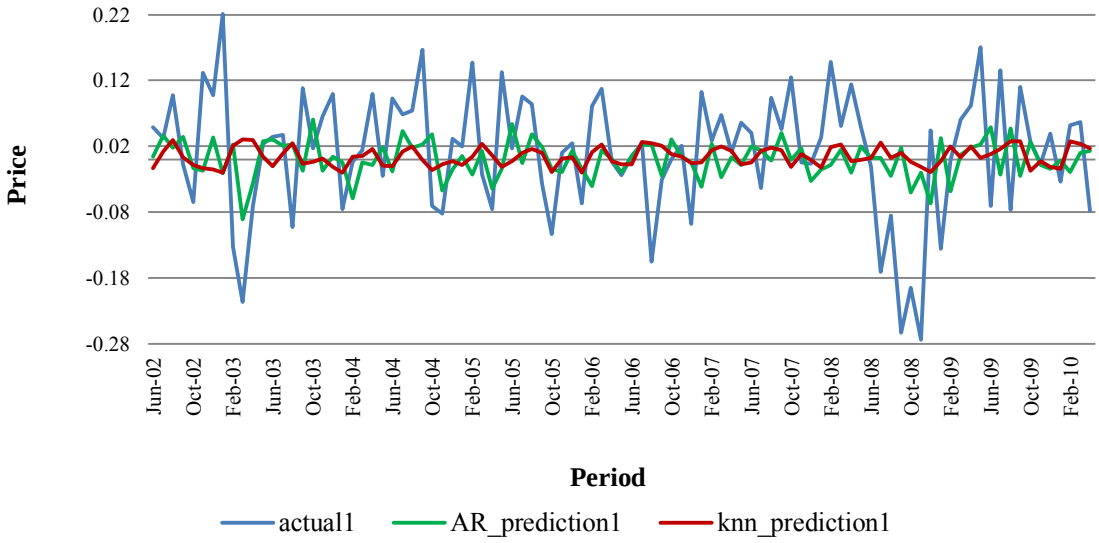

Figure 12. Graphical illustration of predicted vs. actual values for heating oil one-step-ahead out-of-sample forecast

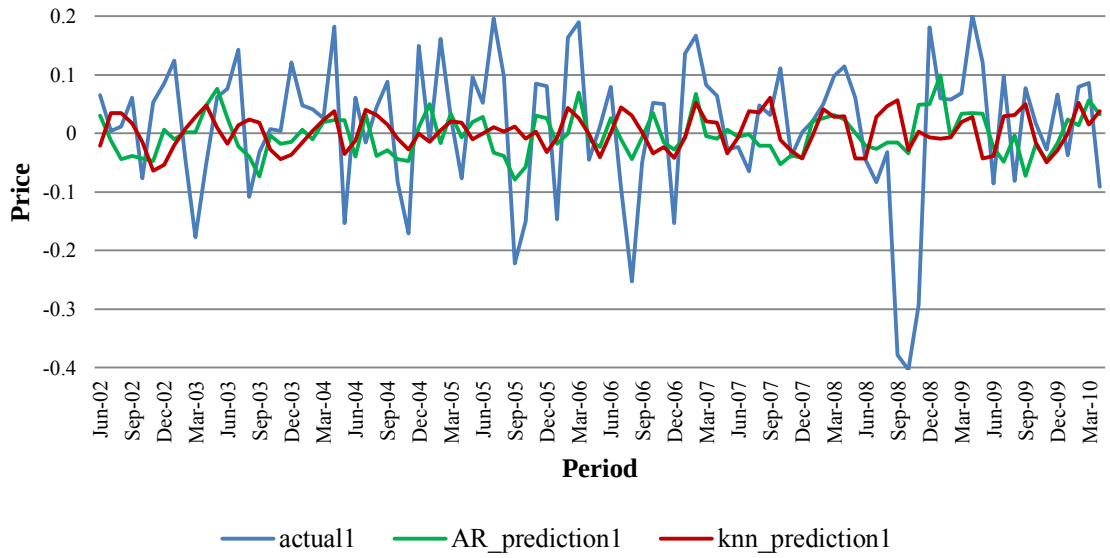

Figure 13. Graphical illustration of predicted vs. actual values for gasoline one-step-ahead out-of-sample forecast

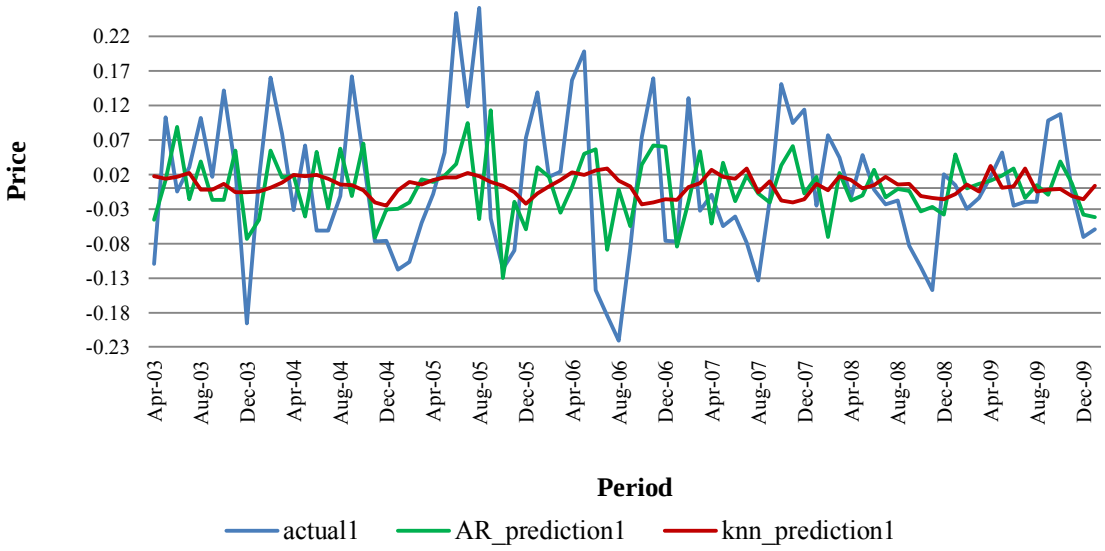

Figure 14. Graphical illustration of predicted vs. actual values for ethanol one-step-ahead out-of-sample forecast

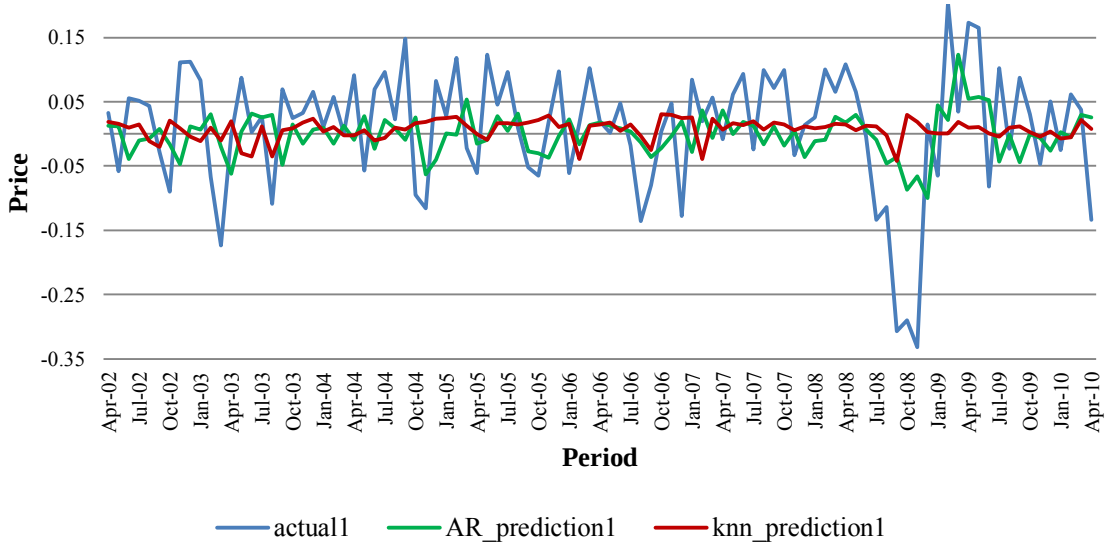

Figure 15. Graphical illustration of predicted vs. actual values for crude oil one-step-ahead out-of-sample forecast

Figures 16-20 graphically illustrate two-step-ahead out-of-sample forecast results for both AR and k-NN models vs. actual values. After the calculations of predicted values from the AR model and predicted values from the k-NN model are completed, it is time to do summary statistics for two types of models. For that purpose, we have calculated the MAEs and the RMSEs of out-of-sample forecast for each type of model.

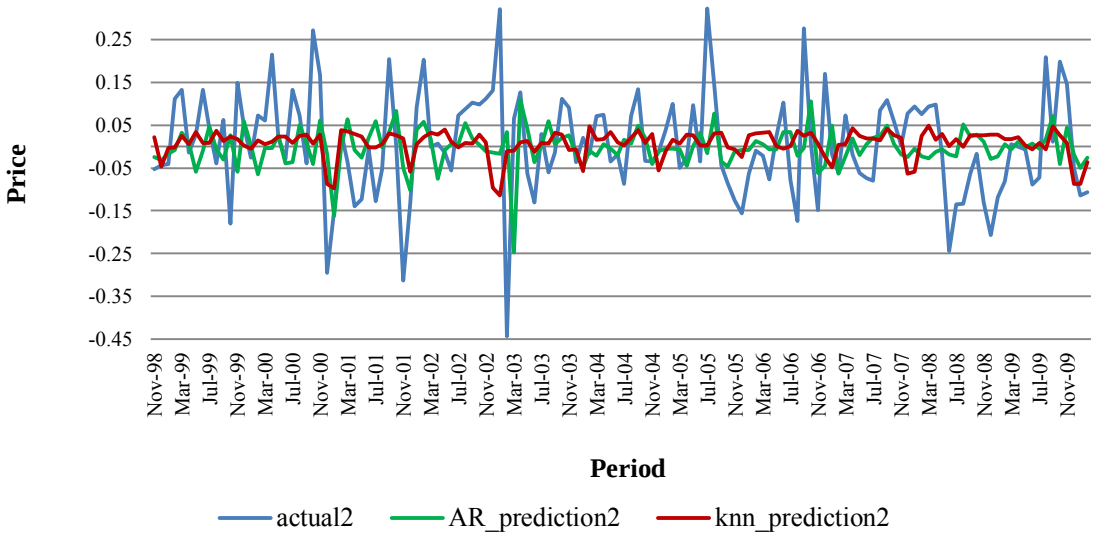

Figure 16. Graphical illustration of predicted vs. actual values for natural gas two-step-ahead out-of-sample forecast

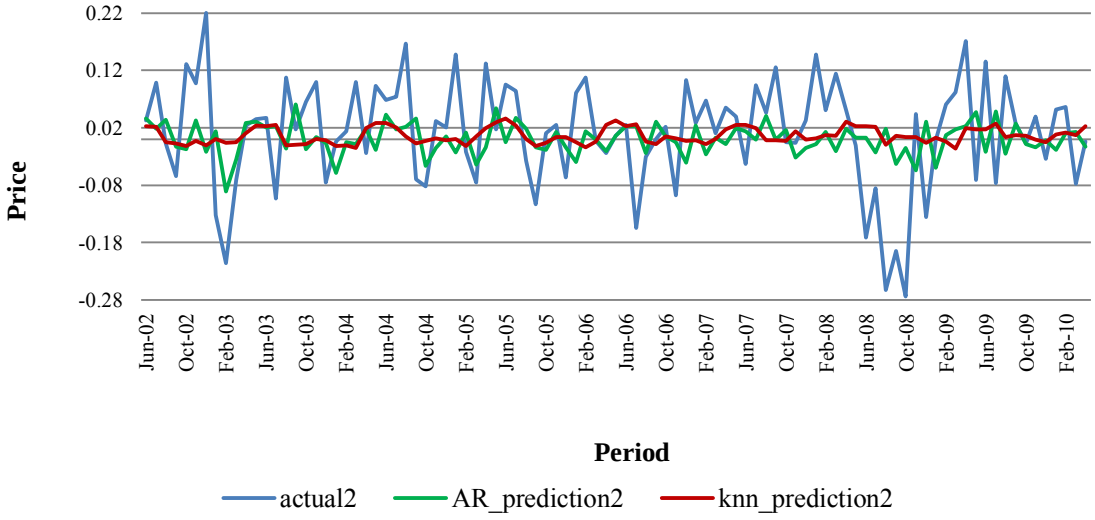

Figure 17. Graphical illustration of predicted vs. actual values for heating oil two-step-ahead out-of-sample forecast

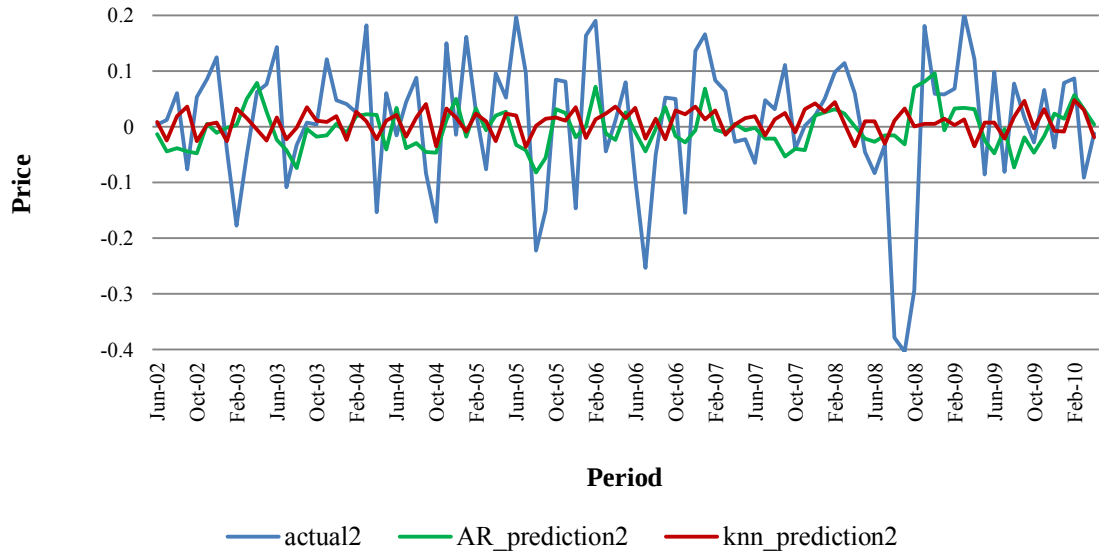

Figure 18. Graphical illustration of predicted vs. actual values for gasoline two-step-ahead out-of-sample forecast

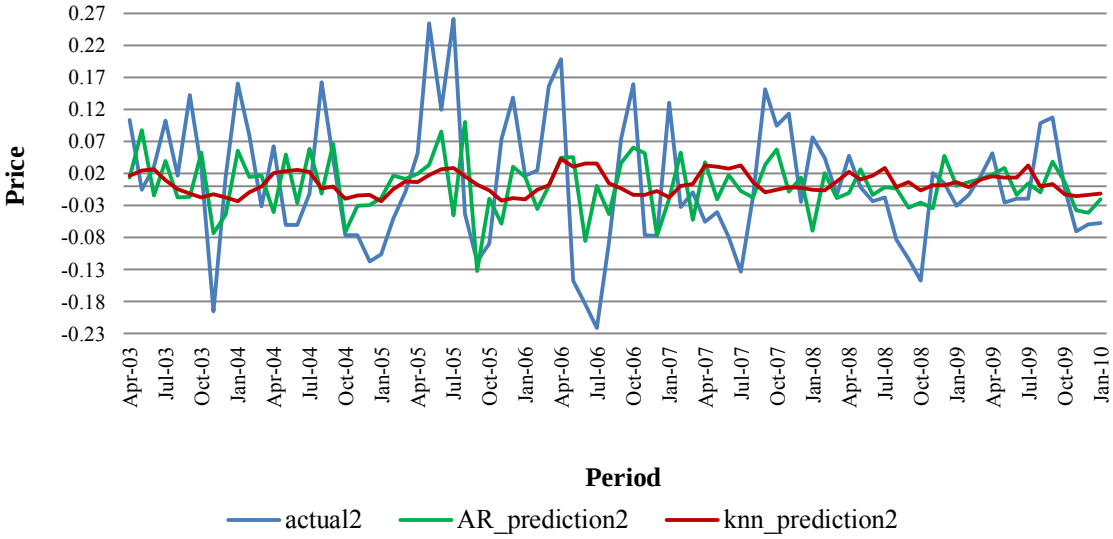

Figure 19. Graphical illustration of predicted vs. actual values for ethanol two-step-ahead out-of-sample forecast

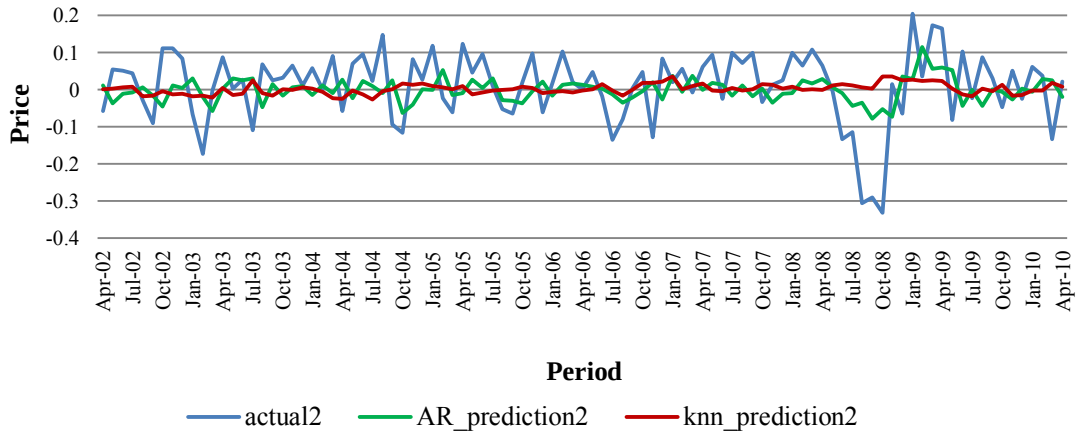

Figure 20. Graphical illustration of predicted vs. actual values for crude oil two-step-ahead out-of-sample forecast

Table 2 summarizes the MAEs and RMSEs for the out-of-sample forecast of AR and k-NN models of five energy commodities. According to the MAE results of natural gas, it is obvious that the AR model results are better, though the difference is not large. For one-step-ahead forecast, MAE for the AR model is 0.0907626 vs. 0.09201676 for k-NN, and for the two-step-ahead forecast the MAE is0.09141592 vs. 0.0946592, respectively. This means that if a decision maker were to choose the best model based on MAEs, he or she would likely choose AR, because the MAEs are lower for both one-step-ahead and two-step-ahead forecasts. Unlike MAEs, RMSEs show slightly different results. Here, k-NN performs better than the AR model: for AR, one-step-ahead RMSE is 0.1216967 vs. 0.1201075 for k-NN and for AR two-step-ahead 0.1221922 vs. 0.1216654 for k-NN. If a decision maker were to choose the best model based on RMSE results, he or she would most likely choose the k-NN model. Although the difference

between the two models is very small, it is up to the decision maker as to which model to choose.

Table 2. MAE and RMSE Calculations for One-Step-Ahead and Two-Step-Ahead Out-of-Sample Forecasts

Crude Oil		
	AR	**k-NN**
MAE 1-Step Ahead	0.07705724	0.07164995
MAE 2-Step Ahead	0.0768801	0.07401227
RMSE 1-Step Ahead	0.0937631	0.09467704
RMSE 2-Step Ahead	0.09376737	0.09714399
Ethanol		
	AR	**k-NN**
MAE 1-Step Ahead	0.0739671	0.0807191
MAE 2-Step Ahead	0.07359046	0.0812664
RMSE 1-Step Ahead	0.09456147	0.1001802
RMSE 2-Step Ahead	0.09411202	0.1011792
Gasoline		
	AR	**k-NN**
MAE 1-Step Ahead	0.0876068	0.09409706
MAE 2-Step Ahead	0.08730985	0.09157353
RMSE 1-Step Ahead	0.1127708	0.1202504
RMSE 2-Step Ahead	0.1132086	0.1190735
Heating Oil		
	AR	**k-NN**
MAE 1-Step Ahead	0.07474962	0.07345258
MAE 2-Step Ahead	0.07431657	0.07236612
RMSE 1-Step Ahead	0.096269	0.09548149
RMSE 2-Step Ahead	0.09616	0.09389789
Natural Gas		
	AR	**k-NN**
MAE 1-Step Ahead	0.0907626	0.09201676
MAE 2-Step Ahead	0.09141592	0.0946592
RMSE 1-Step Ahead	0.1216967	0.1201075
RMSE 2-Step Ahead	0.1221922	0.1216654

In case of heating oil results, both MAEs and RMSEs of out-of-sample forecasts show a better performance for the k-NN model than the AR model: MAE for AR one-step-ahead returned 0.07474962 vs. k-NN's 0.07345258 and 0.07431657 vs. k-NN's 0.07236612 for two-step-ahead forecasts. RMSEs of AR one-step-ahead returned 0.096269 vs. k-NN's 0.09548149 and 0.09616 vs. k-NN's 0.09389789 for two-step-ahead forecasts. Again, although k-NN performs better for both MAE and RMSE, the results are so close that it is up to the decision maker which method to choose.

The picture is a little different for gasoline prices. Here, the AR model turned out to be the best for all the measurements. MAE of AR one-step-ahead returned 0.0876068 vs. k-NN's 0.09409706 and 0.08730985 vs. k-NN's 0.09157353 for the two-step-ahead forecast, while RMSEs for AR one-step-ahead returned 0.1127708 vs. k-NN's 0.1202504 and 0.1132086 vs. k-NN's 0.1190735 for two-step-ahead.

Ethanol appeared to show the same performance as gasoline in terms of k-NN results being a little worse than AR results. MAE one-step-ahead for AR returned 0.0739671 vs. 0.0807191 for k-NN, and AR two-step-ahead returned 0.07359046 vs. 0.0812664 for k-NN. RMSEs for AR one-step-ahead forecasts returned 0.09456147 vs. k-NN's 0.1011802 and 0.09411202 vs. k-NN's 0.1011792 for the two-step-ahead forecast.

Finally, results for crude oil out-of-sample forecast appeared to be inferior by the RMSE measure, but superior by the MAE method. MAE for AR one-step-ahead forecast returned 0.07705724 vs. 0.07164995 for k-NN, and for two-step-ahead AR returned 0.0768801 vs. k-NN's 0.07401227. RMSE for AR one-step-ahead returned 0.0937631

vs. k-NN's 0.09467704 and 0.09376737 vs. 0.09714399, respectively, for two-step-ahead.

Now that we know the results from MAE and RMSE calculations of out-of-sample forecasts, we are now ready to formally test predictive accuracy of the models. As was mentioned in the previous sections, we have used the DM test to test forecasting accuracy of the model. Table 3 summarizes the results for DM test statistics and also includes p-values. According to DM test results, the p-values for the natural gas one-step-ahead forecast is 0.734. The interpretation of the results depends on the significance level that we will choose. If we assume the 5% significance level, then the p-value =0.734 > 0.05. Therefore, we do not reject the null hypothesis; thus, the forecasting models are equally accurate. Given the 10% significance level, again we can say that the p-value =0.734 > 0.1. The MSE from the AR model is not significantly different than MSE from the k-NN model. The same would be true with the 1% significance level: p-value= 0.0734 > 0.01. Therefore, the forecasting models are equally accurate. The p-value for the two-step-ahead forecast is 0.918; thus, given significance levels of 1%, 5%, and 10%, we fail to reject the null hypothesis.

Table 3. Diebold-Mariano Test Results

	One-Step-Ahead Forecast		Two-Step-Ahead Forecast	
	DM	**p-value**	**DM**	**p-value**
Natural Gas	0.3397	0.734	0.103	0.918
Heating Oil	0.2067	0.8362	0.8558	0.3921
Gasoline	-1.5282	0.1265	-1.4652	0.1429
Ethanol	-1.1348	0.2565	-1.2883	0.1976
Crude Oil	-0.1449	0.8848	-0.541	0.5885

In the case of heating oil, the p-value for the DM test of one-step-ahead forecast is 0.8362, which is again not falling into the rejection region given 1%, 5%, or 10% significance levels. For two-step-ahead forecasts, the p-value is 0.3921, which is also out of the rejections region given the significance levels chosen. Therefore, we conclude that the two forecasting models are equally accurate.

For gasoline, the results are the following: the p-value for the one-step-ahead DM test is 0.1265, which is again, given significance levels, 1%, 5%, and 10%, is not falling into the rejection region. The DM test statistic is not significantly different than zero. For two-step-ahead forecasts, the p-value from the DM test is 0.1429, which is within the acceptance range; thus, we fail to reject H_0.

In terms of ethanol, the results are similar to the above models. The p-value for the one-step-ahead DM test is 0.2565. So we do not reject H_0. For two-step-ahead forecasts, the p-value is 0.1976. Given the significance levels above, the two forecasting models are equally accurate.

In case of crude oil, the results are similar to the ones described above. The p-value for the one-step-ahead DM test is 0.8848. Again, we fail to reject H_0 and the forecasting models are equally accurate. The p-value for the two-step-ahead forecast is 0.5885. The DM test statistic is not significantly different than zero.

7. CONCLUSIONS

The thesis concentrated on the use of two forecasting models: simple Autoregressive (AR) and k-Nearest-Neighbor (k-NN) models. Monthly data for five energy commodities were divided into two parts. The first two-thirds of the data were used to perform in-sample one-step-ahead and two-step-ahead forecasts and the remaining one-third of the data were used for out-of sample one-step-ahead and two-step-ahead predictions. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used in calculations to decide which model is better. Outcomes of those tests showed that one model can be superior by one measure, but inferior by another and vice versa. The Diebold-Mariano (DM) test was used to test the accuracy of the AR and k-NN models. The test indicated that for all five commodity models, the MSE from the AR model is not significantly different than the MSE from the k-NN model; thus, forecasts are equally accurate and we do not reject the null hypothesis.

Barkoulas et al. (2003) compared the forecasting performances of AR and random-walk-with drift models to the k-NN method to test the relationship of U.S. T-bill and bond yields. The data used in the research consisted of short- and long-term U.S. Treasury interest rates: monthly observations on the Federal Funds rate, 3-month, 6-month, and 12-month U.S. Treasury bill yields, and yields on 5-year and 10-year Treasury bonds. Their interest was in the out-of-sample one-step-ahead forecasting of the non-parametric method, which was measured by RMSE. Similar to this thesis, where we used two-thirds of the observations in the in-sample forecast and one-third of the observations in the out-of-sample forecast, Barkoulas et al. (2003) used 48 observations

(two-thirds of the data) from each yield-change series for forecasting purposes and the first 24 observations (one-third of the data) for the in-sample forecast. Analogous to Barkoulas et al. (2003), in this thesis, I also chose the AR order for the linear model based on the AIC scores from 48 estimated models. In contrast to Barkoulas et al. (2003), where Euclidean distance was implemented to find the nearest neighbors and then each of the neighbors was inversely weighted by their Euclidean distance, in this thesis, all the observations were assigned a zero weight and the nearest neighbors were determined only based on the Euclidean distance function.

Barkoulas et al. (2003) used the Granger and Newbold test to measure the forecasting accuracy of the models where the null hypothesis was testing no difference in the forecasting accuracy of the linear and nonlinear models. To formally evaluate model performance, we applied the Diebold–Mariano (DM) test to evaluate the hypothesis that the forecasting models are equally accurate. Also, Barkoulas et al. (2003) used a one-sided test as an alternative, while a two-sided test was used as an alternative hypothesis in this thesis. On the basis of the RMSE forecasting criterion, Barkoulas et al. (2003) found that only in few cases AR fitted better than the LWR and for most of the sample series LWR model's forecast performance was statistically superior to that of the AR model. LWR failed to successfully predict stock returns, but it was useful in prediction of conditional mean changes in interest rate series. The results of the thesis found that k-NN method was superior by one measure but inferior by the other measure. K-NN method showed little benefit over AR. In all the cases, DM test statistics failed to

reject the null hypothesis, stating that the MSE from the AR model was not significantly different than the MSE from the k-NN model.

Gençay (1999) studied the predictability of spot foreign exchange rate returns from the past buy-sell signals of the simple technical trading rules by comparing the performance of two parametric and two non-parametric forecasting models: random-walk and GARCH(1,1) and feedforward networks and the nearest neighbors regressions, respectively. Gençay (1999) used data consisting of logarithmically transformed daily spot rates for the British pound, Deutsche mark, French franc, Japanese yen, and the Swiss franc. Similar to this thesis, Gençay (1999) used one-third of the data to perform out-of-sample one-, five- and ten-step ahead forecasts, and the same training set was used to calculate forecasts for all the models. Unlike this thesis, where numbers of neighbors were calculated based on the Euclidean distance function and the optimal number of neighbors was chosen based on the smallest MAE, Gençay (1999) used a cross-validation method, which minimizes the MSE. In contrast to Barkoulas et al. (2003) and this thesis, where RMSE was used as a measure of performance, Gençay (1999) used Mean Squared Prediction Error (MSPE) and sign predictions. To evaluate the statistical significance of the out-of-sample predictions, the DM test was applied to all of the currencies with a null hypothesis of equal accuracy for the two competing forecasts. The research by Gençay (1999) showed that the Nearest Neighbor model gave significant gains over not only parametric models, but also over the feedforward network model. For the Nearest Neighbor regression, MSPE turned out to be 12.8% smaller than the random-walk model across all the currencies. The average sign prediction of the

nearest neighbor forecast was 62% in comparison to feedforward's 58% and 50% of GARCH(1,1). Dissimilar to Gençay's (1999) findings, in this thesis, nearest neighbor was not performing better across all the five commodities. Instead, results showed that the k-NN method was superior by one measure but inferior by another, and these distinctions vary across all five commodities. The reason for this could be the qualitative differences between the series that Gençay (1999) was using and the energy price series. Also, it could be that Gençay (1999) was comparing the performance of k-NN models to non-linear models rather than to AR model. Based on the DM test results, Gençay (1999) findings indicated statistical significance of non-parametric models. In this thesis, Diebold-Mariano (DM) test statistics showed equal accuracy for both AR and k-NN models.

The findings of this thesis can be extended in several ways. First, further investigation is needed for multiple-step-ahead forecasting horizon using the k-NN estimation method. Second, promising results of k-NN as a forecast generating mechanism give ground for the use of alternative non-parametric methods to be employed as a forecasting tool for U.S. energy commodity prices. Third, a cross-validation method should be implemented to decide on the optimal number of neighbors since this method prevents overfitting in noisy environments and uses a certain number of in-sample observations rather than the entire in-sample. Also, the improved Distance function (as discussed in the thesis) or any other distance function can be tested in selection of optimal neighbors. Fourth, other test statistics should be used to compare the forecasting accuracies of the models (Stock and Watson, Granger-Newbold) and the

results could be compared to the finding of this thesis. Finally, future research of LWR can be applied in studying the energy commodity prices of other countries.

REFERENCES

Akaike, H. 1981. "Likelihood of a Model and Information Criteria." *Journal of Econometric* 16: 3-14.

Azmi, M., S. Araghinejad, and M. Kholghi, 2010. "Multi Model Data Fusion for Hydrological Forecasting Using K-Nearest Neighbor Method." *Iranian Journal of Science and Technology Transaction B-Engineering* 34(B1): 81-92.

Barkoulas, J., F. Baum, and A. Chakraborty, 2003. "Nearest Neighbor Forecasts of U.S. Interest Rates." *International Journal of Banking and Finance* 1(1):119-135.

Berry, M. J.A. and G. S. Linoff, 2004. *Data Mining Techniques: For Marketing, Sales and Customer.* Indianapolis: Wiley Publishing, Inc.

Bordignon, S. and F. Lisi, 2001. "Predictive Accuracy for Chaotic Economic Models." *Economic Letters* 70: 51-58.

Brennan, M. J. 1960, 1973. *Preface to Econometrics*. Oklahoma City, OK: South-Western Publishing Company.

Christini, D. J., F. D. Bennett, K.R. Lutchen, H. M. Ahmed, J. M. Hausdorff, and N. Oriol, 1995. "Application of Linear and Nonlinear Time-Series Modeling to Heart Rate Dynamics Analysis." *IEEE Transactions on Biomedical Engineering* 42(4): 411-415.

Claussen E., V.A. Cochran and D.P. Davis, 2001. *Climate Change: Science, Strategies and Solutions.* Arlington, VA: The Pew Center on Global Climate Change.

Cuaresma, J. C., J. Hlouskova, S. Kossmeier, and M. Obersteiner, 2004. "Forecasting Electricity Spot Prices Using Linear Univariate Time Series Models." *Applied Energy* 77: 87-106.

Diebold, F. X., and R. S. Mariano, 1995. "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics* 13(3): 253-263.

Gencay, R. 1999. "Linear, Non Linear and Essential Foreign Exchange Rate Prediction with Simple Technical Trading Rules." *Journal of International Economics* 47: 91-107.

Harvey, D., S. Leybourne, and P. Newbold, 1997. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13: 281-291.

Jaditz, T. and L. A. Riddick, 2000. "Time-Series Near-Neighbor Regression." *Studies in Nonlinear Dynamics & Econometrics* 4(1): 35-44.

Jiang, L., D. Wang, Z. Cai, S. Jiang, and X. Yan, 2009. "Scaling up the Accuracy of K-Nearest-Neighbour Classifiers: A Naïve-Bayes Hybrid." *International Journal of Computers and Applications* 31(1): 36-43.

Kidron, A., and S. T. Klein, 2007. "An Information Retrieval Approach to Predicting Meteorological Data." *International Journal of Modeling and Simulation* 27(1): 1-18.

Liew, V. K., T. T. Chong and K. Lim, 2003. "The Inadequacy of Linear Autoregressive Model for Real Exchange Rates: Empirical Evidence from Asian Economies." *Applied Economics* 35: 1387-1392.

Mizrach, B. 1992. "Multivariate Nearest-Neighbour Forecasts of EMS Exchange Rates." *Journal of Applied Econometrics* 7: S151-163.

Morse, E.L. 2009. "Making the Most of Cheap Oil." *Foreign Affairs*. www.foreignaffairs.com

Pekarova, P., and J. Pekar, 2006. "Long-Term Discharge Prediction for the Turnu Severin Station (the Danube) Using a Linear Autoregressive Model." *Hydrological Processes* 20: 1217-1228.

Robledo, C.W., and H. O. Zapata, 2003. "Measuring Predictive Accuracy in Agribusiness Forecasting." *Journal of American Academy of Business* 2(2): 486-491.

Safavi A. 2000. "Choosing the Right Forecasting Software and System." *Journal of Business Forecasting Methods & Systems* 13(2): 6-14

Staudenmayer, J., and J. P. Buonaccorsi, 2005. "Measurement Error in Linear Autoregressive Models." *Journal of the American Statistical Association* 100(471): 841-852.

U.S. Energy Information Administration (EIA). 2010. Monthly Wholesale and Retail Prices of Energy Commodities. www.eia.doe.gov

APPENDICES

APPENDIX A

R CODES FOR NATURAL GAS

```
                                        ***
#regress dlnP[T] on the 1st, 2nd, and 3rd differences
(dlnP_lag1,dlnP_lag2,dlnP_lag3,dlnP_lag4) with intercept, add the
seasonal harmonic, display results
simple_model3_3=lm(dlnP~dlnP_lag1+dlnP_lag2+dlnP_lag3+sinT+cosT,subset=
isf_start:isf_end)
print(summary(simple_model3_3))
#project in-sample prices
dlnP_hat3_3= fitted.values(simple_model3_3)

#create the same model above without intercept, display results
simple_model3_4=lm(dlnP~dlnP_lag1+dlnP_lag2+dlnP_lag3-
1+sinT+cosT,subset=isf_start:isf_end)
print(summary(simple_model3_4))
#project in-sample prices
dlnP_hat3_4= fitted.values(simple_model3_4)

                                        ***

#project one-step-ahead out-of-sample forecast
prediction1=predict.lm(mymodel,newdata=data.frame(dlnP_lag1+dlnP_lag2+d
lnP_lag3-1+sinT+cosT))[(oosf_start+1):(oosf_start+1)]
cat("out-of-sample predicted dlnP:",prediction1,"\n")
cat("actual observed dlnP:",dlnP[(oosf_start+1):(oosf_start+1)],"\n")
actual1 <- dlnP[(oosf_start+1):(oosf_start+1)]

#replace the actual value with the predicted value
dlnP_lag1_alt=dlnP_lag1
dlnP_lag1_alt[(oosf_start+2):(oosf_start+2)]=prediction1

#project two step ahead out-of-sample forecast
prediction2=predict.lm(mymodel,newdata=data.frame(dlnP_lag1_alt+dlnP_la
g2+dlnP_lag3-1+sinT+cosT))[(oosf_start+2):(oosf_start+2)]
cat("out-of-sample predicted dlnP:",prediction2,"\n")
cat("actual observed dlnP:",dlnP[(oosf_start+2):(oosf_start+2)],"\n")
actual2 <- dlnP[(oosf_start+2):(oosf_start+2)]

#create an empty vector to store the results
results_prediction1=0*oosf_start:oosf_end
results_actual1=0*oosf_start:oosf_end
results_prediction2=0*oosf_start:oosf_end
results_actual2=0*oosf_start:oosf_end

#create a loop for periods 275:410
for (t in oosf_start:oosf_end)
{
mymodel= lm(dlnP~dlnP_lag1+dlnP_lag2+dlnP_lag3-1+sinT+cosT,subset=14:t)
#print(summary(mymodel))
```

```
#project in-sample prices
mymodel_hat= fitted.values(mymodel)
#project one-step-ahead out-of-sample forecast
prediction1=predict.lm(mymodel,newdata=data.frame(dlnP_lag1
+dlnP_lag2+dlnP_lag3-1+sinT+cosT))[(t+1):(t+1)]
#cat("out-of-sample predicted dlnP:",prediction1,"\n")
#cat("actual observed dlnP:",dlnP[(t+1):(t+1)],"\n")
actual1 <- dlnP[(t+1):(t+1)]

#replace the actual value with the predicted value
dlnP_lag1_alt=dlnP_lag1
dlnP_lag1_alt[(t+2):(t+2)]=prediction1

#project two step ahead out-of-sample forecast
prediction2=predict.lm(mymodel,newdata=data.frame(dlnP_lag1
_alt+dlnP_lag2+dlnP_lag3-1+sinT+cosT))[(t+2):(t+2)]
#cat("out-of-sample predicted dlnP:",prediction2,"\n")
#cat("actual observed dlnP:",dlnP[(t+2):(t+2)],"\n")
actual2 <- dlnP[(t+2):(t+2)]

cat(t,",",prediction1,",",actual1,",",prediction2,",",actual2,"\n")

#store the results in the empty vector created above
results_prediction1[(t-(oosf_start-1)):(t-(oosf_start-1))]=prediction1
results_actual1[(t-(oosf_start-1)):(t-(oosf_start-1))]=actual1
results_prediction2[(t-(oosf_start-1)):(t-(oosf_start-1))]=prediction2
results_actual2[(t-(oosf_start-1)):(t-(oosf_start-1))]=actual2}

cbind(results_prediction1,results_actual1,results_prediction2,results_a
ctual2)
```

***

```
#define in sample training (RHS) and response (LHS) data for the in-
sample kNN analysis
RHS=cbind(dlnP_lag1[isf_start:isf_end]+dlnP_lag2
[isf_start:isf_end]+dlnP_lag3[isf_start:isf_end]-1+sinT
[isf_start:isf_end]+cosT[isf_start:isf_end])
LHS=dlnP[isf_start:isf_end]

#create a loop for the knn out-of-sample forecast using different
number of neighbors
for (i in 5:40){
knn=knn.reg(train=RHS,test=NULL,y=LHS,k=i, algorithm="VR")
myresiduals=knn$residuals
MAE_knn=1/length(myresiduals)*sum(abs(myresiduals))
print(cbind(i,MAE_knn))
}

#create an empty vector to store the results
knn_results_prediction1=0*oosf_start:oosf_end
results_actual1=0*oosf_start:oosf_end
knn_results_prediction2=0*oosf_start:oosf_end
results_actual2=0*oosf_start:oosf_end
```

```
#create a loop for periods 275:410
for (t in oosf_start:oosf_end)
{
      #redefine LHS and RHS parts of KNN.reg function to include all
      periods
LHS_new1=dlnP[14:t]
      RHS_new1=cbind(dlnP_lag1[14:t]+dlnP_lag2[14:t]+dlnP_lag3[14:t]-
      1+sinT[14:t]+cosT[14:t])

      #for out_of_sample forecast of knn add the test component to the
      function
      test_oosf1=data.frame(dlnP_lag1[(t+1):(t+1)]+dlnP_lag2[(t+1):(t+1
      )]+dlnP_lag3[(t+1):(t+1)]-1+sinT[(t+1):(t+1)]+cosT[(t+1):(t+1)])

      #calculate the 1 step ahead knn using the number of neighbors
      with the lowest MAE
      knn_prediction1=knn.reg(train=RHS_new1,test=test_oosf1,y=LHS_new1
      ,k=20,algorithm="VR")

actual1=dlnP[(t+1):(t+1)]
      #redefine LHS and RHS parts of KNN.reg function for t+2 forecasts
LHS_new2=dlnP[15:t]
      RHS_new2=cbind(dlnP_lag1[14:(t-1)]+dlnP_lag2[14:(t-
      1)]+dlnP_lag3[14:(t-1)]-1+sinT[14:(t-1)]+cosT[14:(t-1)])

      test_oosf2=data.frame(dlnP_lag1[(t+1):(t+1)]+dlnP_lag2[(t+1):(t+1
      )]+dlnP_lag3[(t+1):(t+1)]-1+sinT[(t+1):(t+1)]+cosT[(t+1):(t+1)])

      #calculate the 2 step ahead knn using the number of neighbors
      with the lowest MAE
      knn_prediction2=knn.reg(train=RHS_new2,test=test_oosf2,y=LHS_new2
      ,k=20,algorithm="VR")

actual2=dlnP[(t+2):(t+2)]

      #print the prediction and actuals, check screen output against
      results that are stored above (see if the storage is working
      correctly)
      cat("pred1:",knn_prediction1[4][[1]],"actual1:",dlnP[(t+1):(t+1)]
      ,"pred2:",knn_prediction2[4][[1]],"actual2:",dlnP[(t+2):(t+2)],"\
      n")

#store the results in the empty vector created above
      knn_results_prediction1[(t-(oosf_start-1)):(t-(oosf_start-
      1))]=knn_prediction1[4][[1]]
results_actual1[(t-(oosf_start-1)):(t-(oosf_start-1))]=actual1
      knn_results_prediction2[(t-(oosf_start-1)):(t-(oosf_start-
      1))]=knn_prediction2[4][[1]]
results_actual2[(t-(oosf_start-1)):(t-(oosf_start-1))]=actual2
}
print(cbind(knn_results_prediction1,results_actual1,knn_results_predict
ion2,results_actual2))
```

***

```
# calculate the Diebold-Mariano test to compare the forecast accuracy
of two models
DM_one_step_ahead=dm.test(results_prediction1-
results_actual1,knn_results_prediction1-results_actual1, h=2, power=2)
print(DM_one_step_ahead)

DM_two_step_ahead=dm.test(results_prediction2-
results_actual2,knn_results_prediction2-results_actual2, h=2, power=2)
print(DM_two_step_ahead)
```

***

APPENDIX B

ILLUSTRATION OF ELIMINATED OBSERVATIONS IN NATURAL GAS DATA

Date	Periods	Price	lnP	lnP(t-1)	dlnP	dlnP_lag1	dlnP_lag2	dlnP_lag3	dlnP_lag4	dlnP_lag5	dlnP_lag6	dlnP_lag7	dlnP_lag8	dlnP_lag9	dlnP_lag1	dlnP_lag1	dlnP_lag1	sinT	cosT
Jan-76	1	0.54	-0.616															0.5	0.8660254
Feb-76	2	0.54	-0.616	-0.6162	0													0.866025	0.5
Mar-76	3	0.54	-0.616	-0.6162	0	0												1	6.13E-17
Apr-76	4	0.55	-0.598	-0.6162	0.018	0	0											0.866025	-0.5
May-76	5	0.55	-0.598	-0.5978	0	0.018349	0	0										0.5	-0.8660254
Jun-76	6	0.58	-0.545	-0.5978	0.053	0	0.018349	0	0									1.23E-16	-1
Jul-76	7	0.58	-0.545	-0.5447	0	0.05311	0	0.018349	0	0								-0.5	-0.8660254
Aug-76	8	0.6	-0.511	-0.5447	0.034	0	0.05311	0	0.018349	0	0							-0.866025	-0.5
Sep-76	9	0.6	-0.511	-0.5108	0	0.033902	0	0.05311	0	0.018349	0	0						-1	-1.84E-16
Oct-76	10	0.62	-0.478	-0.5108	0.033	0	0.033902	0	0.05311	0	0.018349	0	0					-0.866025	0.5
Nov-76	11	0.63	-0.462	-0.478	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0	0				-0.5	0.8660254
Dec-76	12	0.64	-0.446	-0.462	0.016	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0	0			-2.45E-16	1
Jan-77	13	0.67	-0.4	-0.4463	0.046	0.015748	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0	0		0.5	0.8660254
Feb-77	14	0.71	-0.342	-0.4005	0.058	0.04581	0.015748	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0	0	0.866025	0.5
Mar-77	15	0.75	-0.288	-0.3425	0.055	0.057987	0.04581	0.015748	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0	1	1.19E-15
Apr-77	16	0.77	-0.261	-0.2877	0.026	0.054808	0.057987	0.04581	0.015748	0.016	0.03279	0	0.033902	0	0.05311	0	0.018349	0.866025	-0.5
May-77	17	0.77	-0.261	-0.2614	0	0.026317	0.054808	0.057987	0.04581	0.015748	0.016	0.03279	0	0.033902	0	0.05311	0	0.5	-0.8660254
Jun-77	18	0.82	-0.198	-0.2614	0.063	0	0.026317	0.054808	0.057987	0.04581	0.015748	0.016	0.03279	0	0.033902	0	0.05311	3.68E-16	-1
Jul-77	19	0.83	-0.186	-0.1985	0.012	0.062914	0	0.026317	0.054808	0.057987	0.04581	0.015748	0.016	0.03279	0	0.033902	0	-0.5	-0.8660254

VITA

Name: Olga Kudoyan

Address: 332 Blocker Bldg.,2124 TAMU
College Station, TX 77843-2124

Email Address: o_kudoyan@yahoo.com

Education: M.S., Agricultural Economics. Texas A&M University, 2010

B.S., Agribusiness and Marketing. Armenian State Agrarian University, 2007

Certificate in Agribusiness and Marketing. Agribusiness Teaching Center, 2007