Forecasting Energy Prices Using Data Mining Methods

Pejman Bahramian Far

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Eastern Mediterranean University
February 2017
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Mustafa Tümer
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Economics.

Prof. Dr. Mehmet Balcılar
Chair, Department of Economics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Economics.

Prof. Dr. Mehmet Balcılar
Supervisor

Examining Committee

1. Prof. Dr. M. Akif Bakır
2. Prof. Dr. Mehmet Balcılar
3. Prof. Dr. Zeynel A. Özdemir
4. Prof. Dr. Sevin Uğural
5. Asst. Prof. Dr. Çağay Çoşkuner

ABSTRACT

Energy prices have been playing an increasingly significant role in the world economy since all elements involved in this area are considered as a major input for the production. The energy prices as it affect economic variables in the world, is influenced by economic activities of great countries. Indicatively, oil prices which are a major energy index globally are affected by economic activities of great countries, and when such activities are on the decrease, the economy of the industrial countries slips into recession.

The energy market is a complex market which does not follow the random walk process. There are many reasons behind the complexity of the energy market such as political situation, etc. Therefore prediction of this type of market is a difficult task. This study aims to investigate, model and forecast the whole US energy market as an important energy market in the world using different machine learning methods. Besides that, the effect of the US inflation on the volatility of the energy market has as well examined.

Keywords: Forecasting, Neural Networks, US Energy Market, LPPL Models, Data mining methods

ÖZ

Enerji piyasaları karmaşık bir yapıya sahiptir ve bu piyasalarda oluşan fiyatlar rastsal yürüyüş sürecini takip ederler. Bu karmaşıklığın ardındaki sebepler arasında siyasi gündemin bile dahil olduğu bir çok faktör yer almaktadır. Dolayısıyla enerji piyasalarının öngörülmesi oldukça zordur. Bu çalışmanın amacı Amerikan enerji piyasasını öngörü amacıyla modellemektir. Amerikan enerji piyasasının incelenmesindeki en önemli neden Amerikan ekonomisinin global bir öneme sahip olmasıdır. Bu çalışmada, yukarıda bahsedilenlerin yanı sıra Amerikan enflasyonu ve buna bağlı olarak enerji piyasasının oynaklığı ile olan ilişkisi de incelenmiştir.

Enerji ürünleri üretimde ana girdiler olduğundan dünya ekonomisinde gittikçe büyüyen bir role sahiptir. Enerji fiyatları dünya ekonomisindeki bir çok makroekonomik değişkeni etkilemekte büyük ekonomilerin aktiviteleri bu fiyatlar üzerinde etkili olmaktadır. Büyük ekonomilerin aktivitelerindeki yavaşlama ve yükselmeler petrol fiyatalar üzerinde etkili olmakta ve tüm diğer ekonomileri de etkilemektedir. Petrol gibi enerji fiyatlarının öngörülmesi bu nedenle tüm ekonomiler için büyük öneme sahiptir.

Anahtar Kelimeler: Öngörü, Yapay Sinir Ağı, Amerika Birleşik Devletleri Enerji Piyasası, LPPL Modeli, Veri Madenciliği

To My Family

ACKNOWLEDGMENT

I would like to express my deepest appreciation to my supervisor, Prof. Dr. Mehmet Balcilar, who patiently guided me from the initial phases of this study. His continues supervision motivated me and put me back on track when I occasionally went astray, making this task an interesting learning experience. Without his encouragement and supports, it would not have been possible to complete this thesis in such a short time.

Special thanks go to my life, Andisheh, whose unconditional support and encouragement were amazing her faith in me made me feel confident and proud of my work.

I also thank all the academic and administrative staff of the Economics Department of the Eastern Mediterranean University for their friendship and for providing good environment and facilities to study.

Finally, my heartfelt gratitude and absolute respect go to my parents, Khosro Bahramianfar and Azar Bahramian, for their support and understanding in all my educational endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Energy market as a top level factor in economic and social activities lies at the heart of the global economy. Infractions in the supply of energy and critical energy price increases can cause economic and political displacement and recessions. Indeed, There is a significant link between most recessions in the past forty years and supply disruptions in the Middle East.

People listen when energy forecasters talk about future energy production and prices. Individuals, companies, and nations are learned to make plans based on their assessments and mainly tend to trust on the forecasts. But believe in wrong predictions, especially for long-term horizons, is mostly an inefficient point of view. Since policymakers believe misguided projections and they make a decision based on incorrect policy alternatives.

In the other hand, good forecasting can help policymakers to adopt appropriate policies to form the stable economy. This essay aims to consider all aspects of US energy markets using a suite of robust advanced methods.

To this end, Chapter 2 investigates the ability of Log Periodic Power Law models to identify bubble (s) and its corresponding termination point (s). As it is quite clear, in the recent centuries, economists worldwide have been confounded due to the

Increased occurrence and the unpredictable speculative bubbles on financial markets. Therefore the investigation of such a phenomenon in US energy market is crucial.

Chapter 3 targets at the automatic specification of the optimal number of clusters for an unlabeled energy data set through five classes of various evolutionary techniques. In the recent decades, cluster analysis has been applied in different fields such as machine learning, artificial intelligence, pattern recognition, spatial database analysis, textual document collection, image segmentation, sociology, psychology, and archeology. Hence appropriate clustering methodology in US energy market is necessary to define some potential regimes in the market.

Chapter 4 covers the application of some nonlinear methods in electricity forecasting. Electricity cannot be saved while power system stability depends upon a constant balance between production and consumption. These items give rise to severe price volatility and sudden changes known as spikes. Therefore good forecast needs good models to capture complex behavior in US electricity market.

Chapter 5 attempts to contribute to the literature on forecasting inflation by evaluating the performance of a suite of non-linear models. As opined by economists, over a short period and medium period, oil prices are capable of driving some variation in inflation. Since international inflation rates move together according to Neely & Rapach (2011), then international factors which include commodity prices like oil, could have a significant drive on inflation.

Hence, the central bank should be informed about the expected future inflation in association with publicized inflation target to set its policy actions related to the inflation targeting framework. Therefore proper forecasting technique needs to be applied to get reliable forecasts.

All in all, this essay attempts to inspect all corners in US energy market as an influential economy in the world. A precise understanding of the whole nature of energy market in US which incorporates with complexity, leads to have better point of view in order to apply proper policies for US policy makers and rest of the world.

Chapter 2

BUBBLE DETECTION IN US ENERGY MARKET: APPLICATION OF LOG PERIODIC POWER LAW MODELS

2.1 Introduction

In the recent centuries, economists worldwide have been confounded due to the increased occurrence and the unpredictable speculative bubbles on financial markets. These crashes have convulsed the belief in the capitalist financial system and disintegrated the lives of millions of people. A notable example is a birth of the financial crisis of 2008 as one of the worst crashes in memories where losses in potential GDP in consequence of the crash predicted to the amount of 7.6 trillion U.S. dollars only in the United States (www.bettermarkets.com, 2012).

The speculative bubble commonly has been defined as an asset experiences high volume trades at considerably deviated prices from intrinsic values (Smith et al., 1993). The several crashes that happened in 20th and early 21st centuries have refreshed the discussion on the reasons and implications of speculative bubbles. Nobel laureate Robert Shiller (1987) following the 1987 stock market crash, suggested that many investors are induced by emotion rather than rationality and established himself as a promoter of the significance of behavioral factors in financial markets.

Marx, (1867) claimed that bubbles, crashes, and crises are the inevitable consequences of the capitalist system. He believed when the ultimate failure point of the contradictions between the mode of production and the development of productive forces is reached, these crises will be increasingly severe. Therefore, he linked the inevitability of crises to the final failure of the capitalist system.

Schumpeter (1942/2014), expanding on Marx's theories, with a different view on the consequences of market crashes, believed that crashes are an essential component of the evolving economy. Through introducing the concept of creative destruction, he proposed that new technologies must inevitably be rendered for some outdated technologies which have a decrease or a crash in their value. However, even the inevitability of the crashes and crises does not necessarily mean the unpredictability of them and thereby impossible to moderate.

Minsky (1974), another influential economist, following the Keynesian economics reasons that although crashes are inevitable in financial markets, the effects could be dampened through government and central bank actions. Many economists argue that regulatory mismatches affect the excessive speculation.

In addition to Shiller (1987), Stiglitz (2010) and Krugman (1999), institutional economists like Ostrom (1990), Acemoglu (2012) and North (1997) also discussed the significance of institutional factors for governing the behavior of market actors. They noticed that the strategic regulation could help to avoid the excessive speculation in financial markets.

Kindleberger (1978/2011) introduces a model to describe the development of financial crises. In his model, which is heavily influenced by Minsky's (1974) financial instability hypothesis, he emphasizes the relation between financial crises and the business cycle and especially highlights the supply of credit which increases during an economic progression and decreases during an economic recession.

Kindleberger and Minsky assume that the events through which financial crisis happens are followed after some exogenous shocks to the macroeconomic system, called displacement, which in turn, leads to an altered economic outlook. During this phase of economic prosperity, investors are more optimistic about the future and borrowing is more desired while lenders' risk rate is decreased at the same time. This situation leads to overvalued asset prices with even sensitivity to small exogenous shocks which can cause a quick reverse in the economic outlook.

Reinhart & Rogoff (2009), with their crisis sequence, gather empirical findings of previous works to reveal the financial crises. The authors offer that crises often are occurring after financial liberalization which acts as a triggering factor. The following happening is a boom in lending and asset prices, after which weaknesses appear on bank balance sheets. Then the central bank through credit extension begins to support the institutions. As a result, the central bank cannot control the currency, and the currency crash occurs which often leads to increased inflation. At this stage following the currency crash, the banking crisis either reaches its peak or gets worse as the economy approaches sovereign default.

The efficient-market hypothesis (EMH), which was developed by Fama (1965), states that financial markets are efficiently informative since the assets' price reflect all information regarding that particular asset. As a matter of fact, in an efficient market, prices at any point in time represent the best estimates of intrinsic values, and all crashes occur as a result of exogenous variables. This relationship between price and intrinsic value suggests that the inherent value of an asset and its price increase together even during periods of price rising. However, as mentioned, the common definition of a speculative bubble implies the substantial deviation of the asset prices from their intrinsic value in a given period.

This definition clearly contradicts the hypothesis of the efficient markets; meaning that the occurrence of the bubbles may be considered as one of many signs that can violate the claim of efficient financial markets. Thus this inefficiency of financial markets requires a model with the ability to identify speculative bubbles and to predict their end.

Sornette et al. (1996) with inspiring of earlier work on the prediction of earthquakes could satisfy this need to some extent through quantification of the asset price dynamics leading up to crash. The authors suggest that during speculative bubbles, financial time series exhibit the same patterns and properties as a seismic activity which leads up to a critical point; signifying the end of a bubble or in the case of seismic activity implies the beginning of an earthquake. They claim that since all speculative bubbles occur as a result of endogenous market dynamics, therefore, they are in contradiction with the efficient-market hypothesis.

They offer that during a speculative bubble, asset prices increase as a power law decorated with log-periodic fluctuations; meaning that in this period increasing of asset prices is faster than exponential and systematically fluctuate around this faster-than-exponential increase. Henceforth the log-periodic power law model is called the LPPL-model which suggests that as the bubble approaches its peak then the magnitude of the oscillations decrease in size and when the magnitude turns to zero it signifies the end of the bubble. This critical point should not necessarily be a crash, as the model simply predicts the most probable point in time for a change in regime which can be a change in growth rate of asset prices.

Rodrigue et al. (2009) made a significant contribution to the study of crash prediction through indicating that asset prices follow a specific pattern during a speculative bubble and also in its consequences. These patterns are similar to those identified by Sornette et al. (1996) but with a major difference about the empirical short-run fluctuations around the price growth during a bubble, which Rodrigue has not considered.

Rodrigue recognizes four stages of the bubble, in the stealth phase as a first step, a few investors who have access to better information than others can realize the essential appreciation potential of the asset prices. In phase two (awareness stage) many institutional investors come to this realization, hence, a large inflow of money occurs which drives the prices further up.

Eventually, the public becomes aware of the potentials in the assets through media reports; this takes us to the third phase, where the psychological factors affect the price through herding and positive feedback. This unsustainable growth leads to a bull trap where before the market crashes a smaller fall precedes an ephemeral return to normal.

The bull trap and the go back to normal can be described by investor mentality where the little crash followed by denial before the asset prices decay.

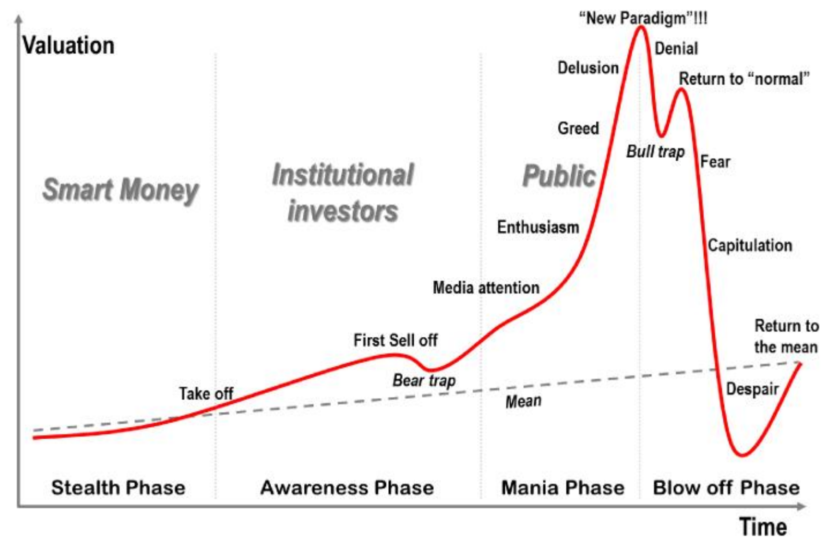


Figure 1. The Different Phase of Bubble, Source: Rodrigue et al. (2009).

2.2 Methodology and Data

2.2.1 Data Description

This study aims to apply LPPL models to identify bubble (s) and its corresponding termination point (s) $[t_c]$ in the US energy market using three US major energy prices namely, The daily US dollar crude oil price of West Texas Intermediate (WTI), the US dollar natural gas price and the US dollar coal price covering the 1987:05:15_2015:01:30 periods on daily basis. Descriptive statistics for logarithmic form of all three indexes are given in Table 1 while, Figure 2 plots them for the sample period[1].

Table 1. Descriptive Statistics of Log Form of Data

Series	Min	Max	Mean	Median	S.D	Skewness	Kurtosis	JB
Crude oil	2.24	4.94	3.51	3.26	0.73	0.45	1.73	730.99*
Natural Gas	0.95	2.51	1.48	1.42	0.38	0.46	2.03	544.64*
Coal	3.19	5.20	3.82	03.67	0.47	0.79	2.55	824.63*

*p < .05.

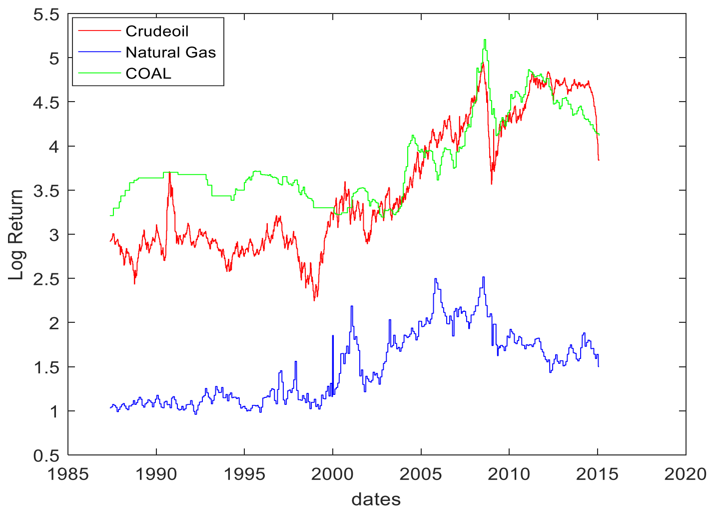


Figure 2. US Energy Market Prices in Sample Period.

[1] All price indexes have been tested by z-scores (more than 3 or less than -3) to identify the possible outliers. Based on our findings none of them are involved with observations which might be outliers.

As shown in Table 1, based on JB test result, the normal distribution has rejected for all series and the estimated Skewness and Kurtosis indicate that all series are right skewed and also are more peaked in comparison with the normal distribution. Therefore all series are with a high probability of extreme values against the normal distribution which is confirmed by Figure 3.

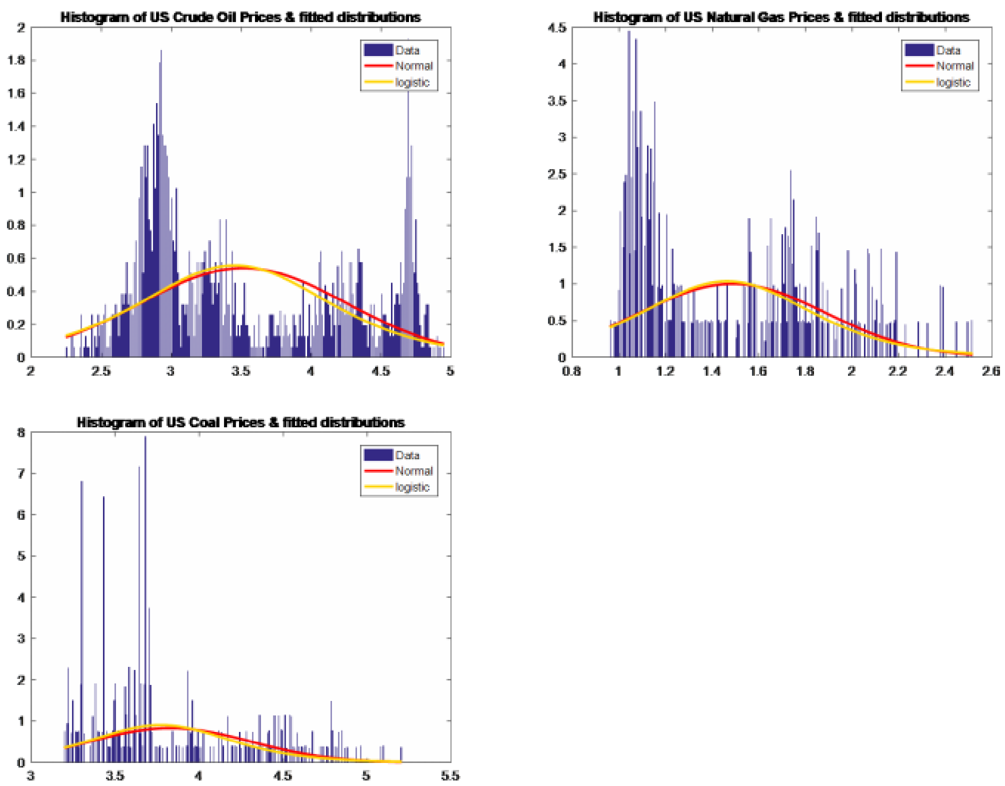


Figure 3. Histogram of US Energy Market Prices and Fitted Distributions.

2.2.2 Methodology

Before discussing in details about the nature and components of the LPPL, It is crucial to differentiate between endogenous and exogenous market sections, since the oscillation and the faster-than-exponential growth part of the model, only feature endogenous market sections, i.e. speculative bubbles.

Johansen & Sornette (2010) described exogenous market part as an outcome of external shocks, frequently triggered by political influence. For instance, the stock market of Russia declined by 20 percent in December 2014, due to decreasing crude oil prices and sanctions placed on the Russian government by the international communities.

Thus, it becomes difficult to predict exogenous market divisions through the use of economic modeling. These factions in the market can also be grouped as endogenous, provided it become difficult to analyze them by exogenous factors, and alternatively speculations become the order of the day. In a situation whereby speculations become a driving force, that would gear up asset prices to rise, as we perceive speculative bubbles. It cannot be overemphasized that, LPPL model only works well with speculative bubbles and it can be used to forecast exogenous-induced bubbles.

There are two main divisions of the LPPL model, which are features by oscillating movements and secondly by faster-than-exponential growth. These features are crucial for the conception of the model, and they will be discussed in details in the following sections. The oscillating design of the LPPL model is more complicated to describe in theory Thus; several researchers have investigated the model and solely describe the circumstances of the design as being statistically proven.

Sornette et al. (1996) pointed out that, the oscillation share similarity with Elliot's wave principle as postulated by Elliot (1938/2012) and this is often employed in the technical analysis. According to him, crowd or collective investor psychology, rotate between pessimism and optimism, both of these produce an observable design in price changes. One crucial distinguishing fact about Elliot Wave principle and the oscillation of the LPPL model is that the wave principles propounded by Elliot presume that, all cycles comprise five waves with two preparatory waves and three spontaneous waves.

The fundamental structure of the LPPL model failed to identify any particular amount of waves in a bubble sequence. Rather Sornette et al. (1996) suggest that the oscillation in a speculative bubble should decline in amplitude as the price change move towards the regime switch. The hypothesis suggests that the regime switch should only materialize when the amplitude of the oscillation revolve around or to zero.

One of the fundamental structures of LPPL model is the idea of positive feedback. According to Sornette et al. (2013) positive feedback simply means, when the price increases/rises, the investor would act to buy more, because of the speculation of a future rise in price. Shiller (2000) argued that an observer of a market might notice this during its speculative bubble when positive feedback becomes superior, the outcome would tend towards the self-reinforcing circle, which pushes the market out of equilibrium. This circle simply means that, as price increases, the demand for asset would also increase until it rises to its decisive point, a point where the regime switching exists. The structure of the positive feedback put a light on the faster-than-exponential-growth of a speculative bubble in asset prices.

Shiller (1984) and Nofsinger (1999) pointed out the fundamental reason why positive feedback circle manifests. According to them, it is due to the psychological occurrence, which can also be called herding behavior. They define herding as a class of investors trading in a particular line of commerce over time, as the outcome of the merchant reacting to the common attitude among their colonies. Herding behavior here simply built on learning by example; that is when investors ignore their beliefs and rather act like other investors in the market.

Johansen (2000), Geraskin and Fantazzini (2013) explain the possibility of two different types of merchants in a financial market. The first, being characterized by its rational expectations, while the other by its unreasonable expectation, which can also refer to as noise –merchant. However, the rational investor group brings about negative feedback, while the noise merchant group are responsible for the herding behavior, which is being influenced by external factors and the attitude of another merchant in the market and their social webs.

This equation mostly measures the designs of the LPPL model. The LPPL model was formulated by Sornette (1996), and it was defined as:

$$p(t) = A + B\ (t_c - t)^z + C(t_c - t)^z \cos(\omega \log(t_c - t) +\ \varphi) \quad (2.1)$$

Where p depicts asset price, which is a function of t and the critical point was represented by t_c, which is most predictable time that describe a switch in regime. The power law constituent of the function is defined by:

$$B\ (t_c - t)^z \qquad (2.2)$$

The above equation represents the faster-than-exponential growth of time series and hence the positive feedback structure. The constituent measures the increasing log-periodic oscillation of the model is characterized by:

$$C(t_c - t)^z \cos(\omega \log(t_c - t) +\ \varphi) \qquad (2.3)$$

The function of the parameter z ∈ [0, 1], in the model is to restrain the strength of the feedback structure and the amplitude of the oscillation and ω ∈ [4.8,7.9] (Johanson *et. al.* 2010) shows the frequency of the oscillations as shown the effect of both in Figure 4.

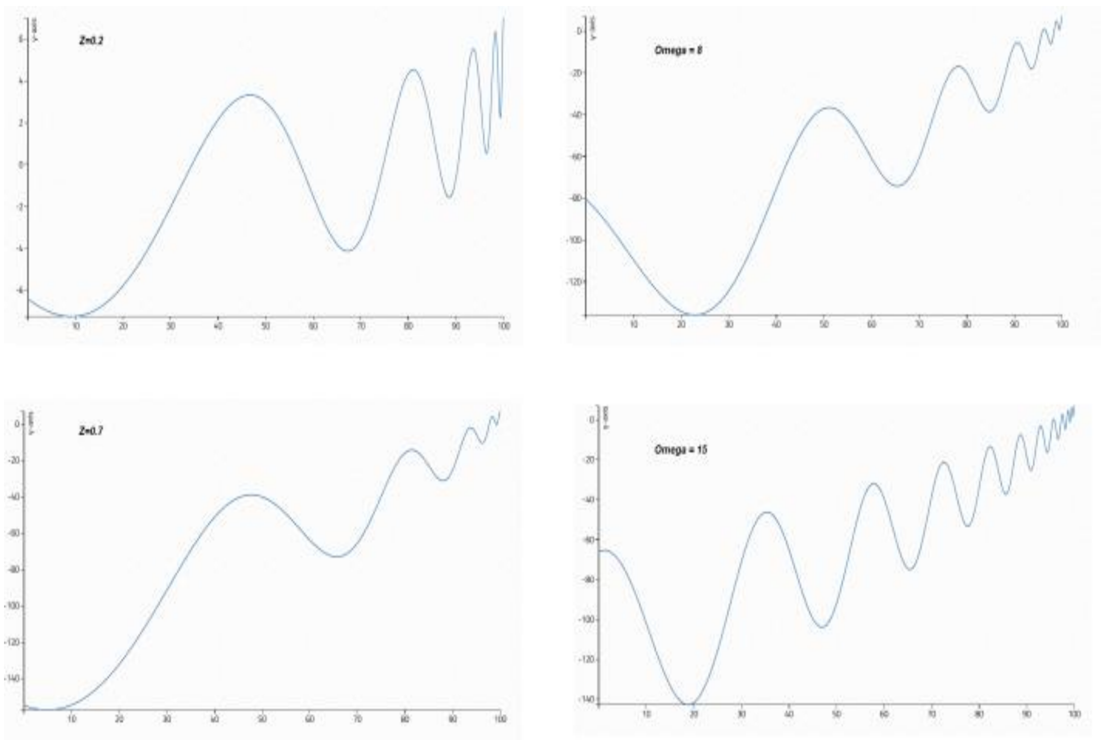


Figure 4. The Effect of **z** & **ω**. Source: Own Figure.

Equation (2.1) comprise three linear parameters, A, B, C and four non-linear parameters t_c z,ω, φ, which are to be evaluated. Since the aim of the model is to forecast the limit of the speculative bubble, we seek to discover the best evaluation techniques of the t_c (critical values). The previous researcher on the subject matter commenced their analysis by subordinating the 3 linear coefficients into the 4 nonlinear coefficients. This decreases the amount of parameter that should be evaluated concurrently.

The protagonist then employed the techniques of nonlinear least square (NLS) to obtain the optimal estimation of the critical value (t_c), (Johansen et al. 2000).The purpose of the NLS like the OLS is to help estimate a cost function, that is, the sum of square error (SSE) as shown in (1.4).

$$SSE(t_c, \omega, \varphi, A, B, C) = \sum_{i=1}^{N}[p(t) - A + B\ (t_c - t)^z + C(t_c - t)^z \cos(\omega \log(t_c - t) + \ \varphi)]^2$$

(2.4)

Cost minimization of the multivariate nonlinear cost function is a difficult venture due to the existence of the multiple local minima.

Therefore, optimization based on the actual equation necessitates derivation of the overall minimum employing some metaheuristic techniques, such as genetic algorithms or taboo search.

These evaluation techniques are quite demanding because they necessitate several iterations for spotting the overall minimum. Also, the optimization algorithm might be confined at local minima, which give no assurance that the local minimum estimated was not the actual overall minimum. This fact implies that, the forecasting of the critical point (t_c) might be acutely biased.

Filimonov & Sornette (2013) for this reason came up with a different version of the equation. The equation was expanded in term of its cosine term of the actual equation and rewords the equation as presented below.

$$p(t) = A + B\ (t_c - t)^z + C(t_c - t)^z cos(\omega\, log(t_c - t))\, cos\varphi + \ C(t_c - t)^z sin(\omega\, log(t_c - t))\, sin\ \varphi)\ ) \qquad (2.5)$$

The author, therefore, rearranges the equation (1.1) as:

$$\boldsymbol{p}(t) = A + B\,(t_c - t)^z + C1(t_c - t)^z \cos(\omega \log(t_c - t)) \cos\varphi + \; C2(t_c - t)^z \sin(\omega \log(t_c - t)) \sin\;\varphi\,)\;) \qquad (2.6)$$

Taken cognizance of the (2.6), the equation comprises three distinct nonlinear parameters t_c ,ω, z and 4 linear parameters (A, B, C_1, C_2). By modifying the original equation, the researcher introduces no additional constraint in the equation. The modified cost function to be estimated is displayed in (2.7).

$$\boldsymbol{S}SE(t_c, \omega, \varphi, A, B, C_1, C_2) = \; \sum_{i=1}^{N}(p(t) - (A + B\,(t_c - t)^z + C1(t_c - t)^z \cos(\omega \log(t_c - t)) \cos\varphi + \; C2(t_c - t)^z \sin(\omega \log(t_c - t)) \sin\;\varphi\,)\,)\,)^2 \qquad (2.7)$$

This modification indicates two severe implications. First, the changes decline the difficulty of the attachment process, due to the optimization problem that is transformed into 3-dimensional spaces from 4-dimensional spaces. Second, this have better value because the newly transformed cost function now have single minima, and this well fit the model, since it is for the empirical analysis. The model stability was significantly enhanced. Due to this modification, the complexity of the taboo search was eliminated and gives way for a simple algorithm. The Gauss – Newton algorithms was employed and can be used for the estimation without eroding the robustness of the model.

From the information given above, it becomes easy to conclude that, the newly proposed equation by Filimonov and Sornette (2013) generates more convenient, efficient and stable estimation process, compared to that of the original formula. On this premise, this study base its estimation procedure and techniques on the credibility of (1.6).

2.2.2.1 Bubble Selection

Before fitting the LPPL-equation, it is certainly important firstly to choose which time series or bubbles are going to be analyzed. The common practice in prior studies is to either select bubbles based on historical context or through identifying bubbles with the drawdown methodology described by Johansen & Sornette (2010). They defined drawdown as a steady decrease in the price of an asset across several consecutive days that is, the cumulative loss from the last local maximum to the next minimum. The authors choose which bubbles fit with the model by first identifying the largest drawdowns in the available market and after that fitting the LPPL-equation to the period prior this drop. The main problem of identifying through drawdowns is related to not taking into account the fact that not all speculative bubbles end in a crash. As a matter of fact, the model forecasts the regime shifts not necessarily crashes. Accordingly, the drawdowns methodology excludes the speculative bubbles which do not end in a crash, and equally at the same time includes bubbles that lack any real interest in the model. Besides that fitting, the LPPL-equation would generate deceptive fits whereas the model is only applicable to speculative bubbles. Hence, this study tries to focus on the selection of bubbles where their historical background makes them particularly attractive.

Besides that to rectify selection criteria, the authors benefited from the bubble index which is an open source tool developed by Taylor Trott[2], to discover the likelihood of market bubble at any given time.

The bubble index algorithm with the aid of Sornette's researches on market crashes determines the solidity of the LPPL oscillations in time series data. The algorithm

[2] More information about bubble index is available in: https://thebubbleindex.com .

engulfs the idea of non-parametric, (H, q) analyses which proposed by Wei-Xing Zhou and Didier Sornette (2002). To this end, first the time series are fitted by the LPPL model [Equation (2.1)] then by using the estimated parameters from the model and making the assumption that the critical time (t_c) is to be one month in the future, the (H, q) derivative is obtained by :

$$D_q^H f(x) \triangleq \frac{f(x)-f(qx)}{[(1-q)x]^H} \tag{2.8}$$

Where:

$$D_q^H y(x) = t^{z-H}[B' + C'g(t)] \tag{2.9}$$

$$B' = -B\frac{(1-q^m)}{(1-q)^H} \quad , C' = \frac{C}{(1-q)^H} \ , \ g(x) = C_1 cos(\omega\, log(x)) + C_2 sin(\omega\, log(x))$$

With $C_1 = 1 - q^m\, cos(\omega\, log(x))$, $C_2 = sin(\omega\, log(x))$.

The bubble index finds the (H, q) derivative and then executes a search for the best parameters of H and q. In this study, H varies between [-1,1] and q changes between 0 and 1.Then with the total defined (H, q) derivatives a Lomb Periodogram is applied to detect the strongest periodogram signal in a specific frequency range. If Lomb frequency $\omega Lomb$ is concentrated near to specific frequency, hence the existence of Log- Periodicity is approved. The Bubble index attempts to form a daily index by contrasting the current price by the LPPL function.

2.2.2.2 Estimation

Since determining the initiation of the speculative behavior is difficult, therefore, in fitting the LPPL-equation to a historical time series, it is important to consider at what time interval the estimation process is going to conduct. In many prior studies, the starting date of the analyzed time interval is quite arbitrarily selected to be where the first LPPL signatures appear. The estimation is conducted then by fitting the

LPPL-equation from the starting date up to some times after the peak which covers the peak date plus a prediction interval. However, the used data includes price data from the starting date until the peak or sometimes before the peak.

To eliminate the arbitrariness of selecting one specific starting date, this study uses rolling window technique through which the estimation is allowed to be iterated with both different start dates and different end dates. The application of rolling window is also consistent with the suggestions of Sornette et al. (2013) to make the anticipations statistically more robust.

In various sub-samples, the model is estimated iteratively from the start date to the end date by varying the equation parameters where in each iteration the data range covers the rolling start date up to close one month before the actual peak of the bubble. Hence, this study sets the last observed date one month before the real peak of the bubble, and the start date is rolling with increments of twenty days, and end date moved forward two days.

Applying the method to a rolling window of estimation fits more, with acceptable characteristics in comparison with many prior studies which make the fitting procedure less sensitive to input values.

Practically the estimation process is a complex procedure which by using a Gauss-Newton algorithm, as a method for solving nonlinear least squares problems, solves the minimization problem of Equation (2.6).

Minimization of the equation (2.7) reveals two fits with the lowest RMSE in each iteration; therefore, using rolling window ends up with thousands of fits, both good and bad.

2.2.2.3 Filtration

In this study to eliminate the severe fits of the estimation results, some constraints in order to do filtering are applied.

Hence, in this study, the guideline proposed by Filimonov & Sornette (2013) is adopted. Based on their guideline the parameter for the frequency of oscillations, ω, can lie between 3 and 15. Through imposing this constraint the fits where the oscillations are either too short or too long are not allowed, since they are obvious mismatches for capturing the LPPL-signatures.

For the parameter of feedback and amplitude of oscillations, z, no other further constraint imposed, except the accepted values between 0 and 1.

During a speculative bubble, asset price increases up to the critical point which is denoted by A as the highest point of the fit.

Obviously from the equation can also be found that; B must take negative values which should be restricted less than 0. As the asset prices are not mean reverting during a speculative bubble, therefore, some constraints are introduced on the augmented Dickey-Fuller and Phillips-Perron values which filter out the stationary fits of the model and only permits for the non-stationary ones.

2.2.3 Empirical Findings

Figures 5, 6, 7 demonstrate the drawdown analysis for crude oil prices, natural gas prices, and coal prices respectively. All graphs denote the existence of a chaotic situation which needs to capture by LPPL models.

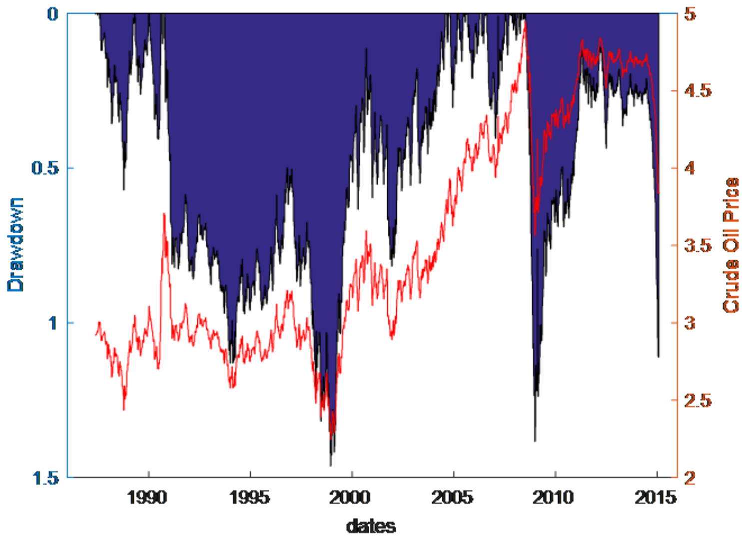


Figure 5. Drawdown US Crude Oil Prices.

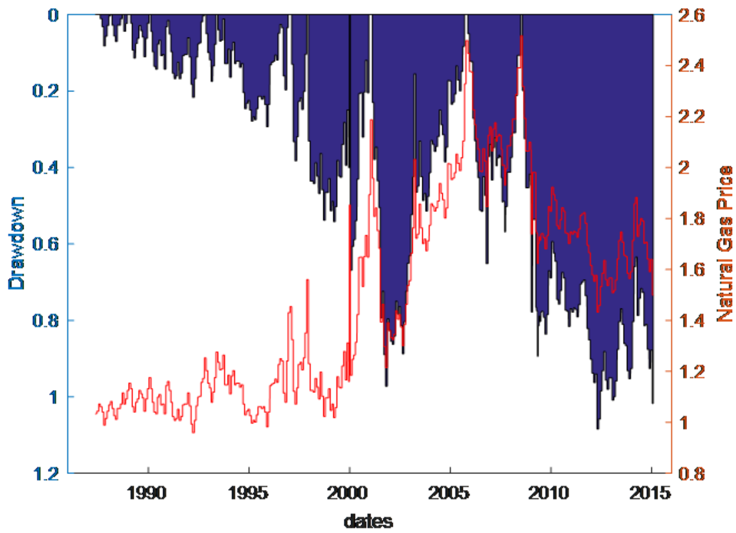


Figure 6. Drawdown US Natural Gas Prices.

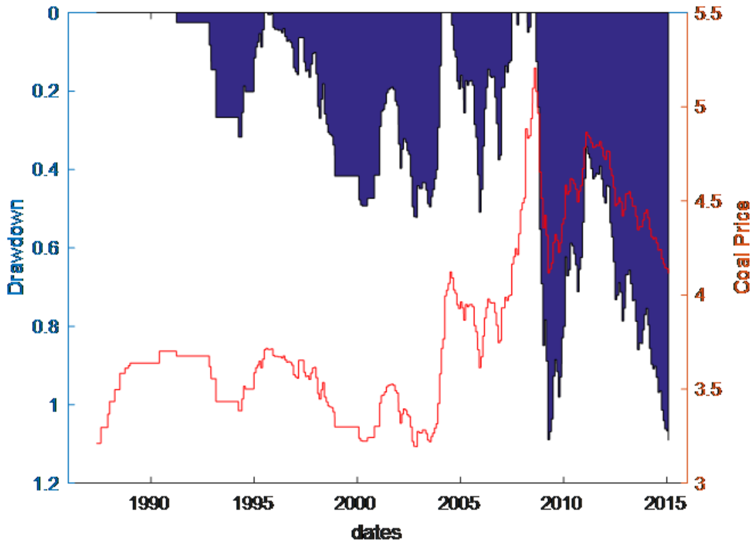


Figure 7. Drawdown US Coal Prices.

For crude oil prices and natural gas, four major drawdowns detected, while there is just two maximum loss for coal prices from the last local maximum to the next minimum. Given the backdrop of drawdown analysis, motivate us to modify the open source bubble index tool based on our sample data set.

The results of bubble index can enforced selection criteria in the fitting of LPPL models. Figures 8, 9, 10 indicate the bubble index in US energy market. Note that to increase the reliability of the bubble index algorithm; the process has been repeated (78 times) for a different number of sample windows.

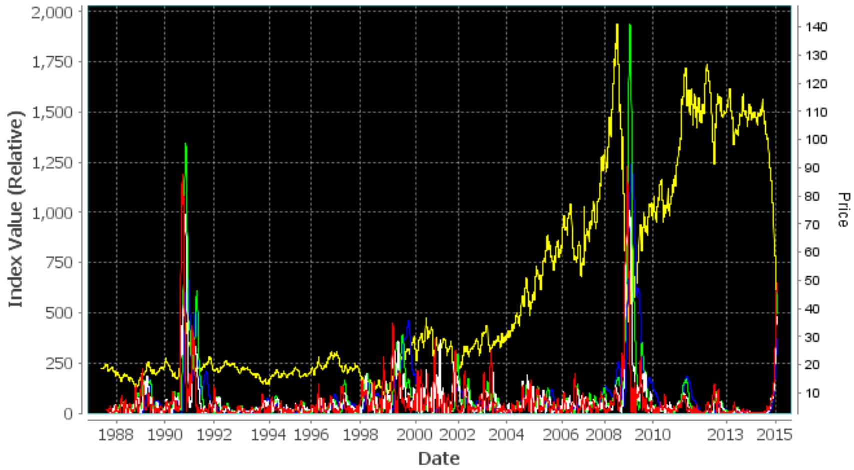


Figure 8. The Bubble Index: Crude oil.

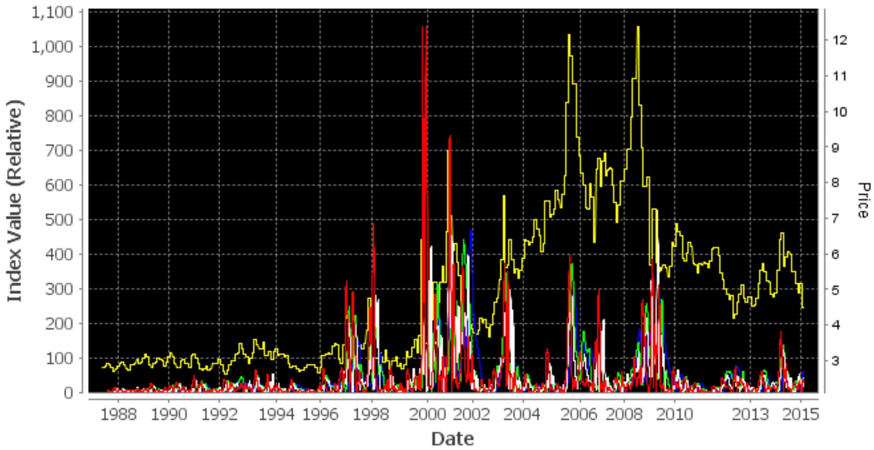


Figure 9. The Bubble Index: Natural Gas.

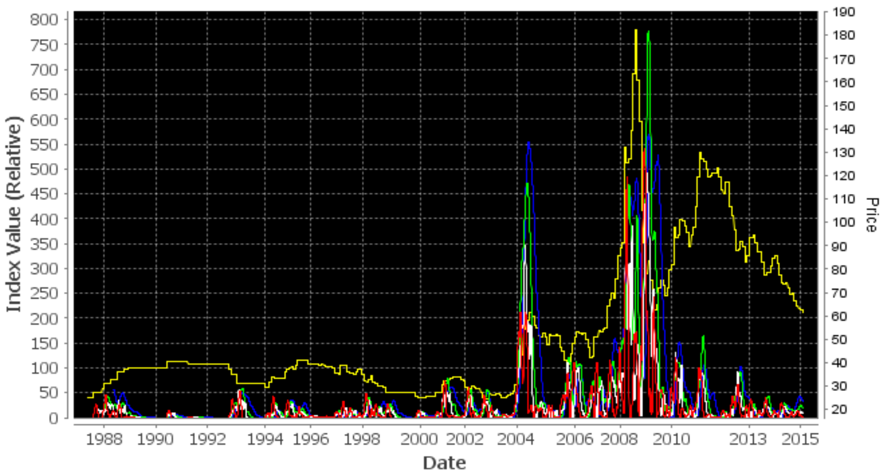


Figure 10. The Bubble Index: Coal.

Based on the results of two different methods, some the potential bubble which are selected to adopt by LPPL models are given in Table 2.

Table 2. Selected Bubbles and Corresponding Real Regime Shifts.

Series	**Crude Oil**			**Natural Gas**			**Coal**		
#	Year	From	To	Year	From	To	Year	From	To
1	**1990**	14-Jun-1990	4-Oct-1990	**1999**	28-Sep-1999	4-Jan-2000	**2004**	27-Jun-2004	30-Jul-2004
2	**2000**	19-Apr-1999	4-Sep-2000	**2001**	29-Mar-2000	22-Jan-2001	**2008**	2-Nov-2007	26-Jul-2008
3	**2008**	27-Dec-2007	22-Jul-2008	**2005**	4-Oct-2004	24-Dec-2005	**2011**	1-Nov-2010	1-Mar-2011
4	**2011**	30-Mar-2010	30-Apr-2011	**2008**	3-Oct-2007	14-Aug-2008	-----		

Figure 11 shows a typical LPPL estimation for US crude oil market in this study at 2008. The median forecasted wreck date depicted in red line and the 80 percent confidence interval[3] in gray. The bold black lines depicted the last observed date. The other estimation's plots are available in Appendix A.

[3] After adjusting last observed date, we ignored those whose fit parameter t_c (the critical time) was greater than last observed date, since the predictive strength of the LPPL model falls off considerably at longer time scales. We used the remaining fits to calculate the 80% confidence interval of t_c.

Table 3. The Events Classified as Bubble in US Crude Oil Market.

Year	Description	Real regime shifts	Estimated t_c	80% Confidence Interval
1990	The war between US and Iraq due to the occupation of Kuwait by Iraq navigated to raise the indefiniteness in the supply of oil which causes the bubble of 1990.	28-Sep-1990	21-Sep-1990	20-Aug-1990–24-Nov-1990
2000	. --------------------	2-Aug-2000	17-Aug-2000	26-Jul-2000 - 13-Oct-2000
2008	According to Conway (2009) the energy crisis experienced was due to the hike of about 400 percent increase in oil prices between 2003 and 2008. Some of the many factors that contributed to this hike in oil prices discussed as follows; for instance, the persistent tension experience in the Middle East, diminishing the value of the dollar, unjustifiable speculation in oil price, a frequent disturbance over the peak oil, etc.	3-Jul-2008	20-Jun-2008	25-May-2008–2-Aug-2008[4]
2011	In the US housing market, after the peak in the mid of 2006 the house prices begins to fall. This happening made a vicious circle in the way that lender increased the mortgage rates and difficult to renovate the loans which speeded up the fall in house prices. As a result, the value of the mortgage, backed derivatives lost their value which navigated to decline in the interest of the global investors and also the confidence to the US financial markets. Due to the risk of recession for the US and the world economy; the Fed reacted by the expansionary fiscal and monetary policies which led to higher liquidity and lower interest rates this policy made the dollar cheaper that caused to massive financial inflows to the crude oil market and many commodities market which induced the bubble.	29-Apr-2011	14-Apr-2011	28-Mar-2011–4-May-2011

[4] Sornette (2009) in his analysis, carried out both ex-post and ex-ante evaluation of the bubble. In his ex-post evaluation, 80 percent confidence interval of wreck dates spreading from 17th May to 14th July was discovered. Besides that due to Sornette suggestion the last observed date in the rolling window is adjusted at (2008-05-23), 41 days before the real change shift.

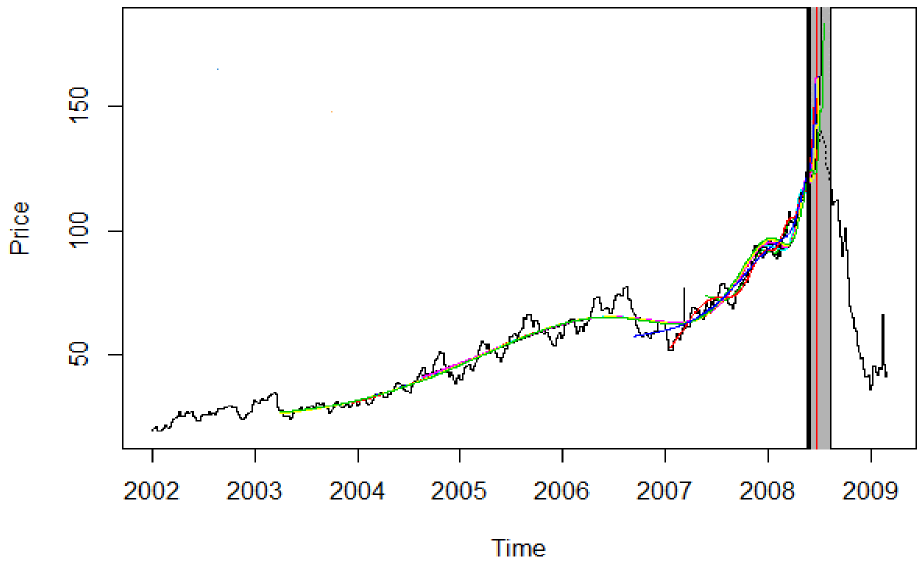


Figure 11. LPPL Estimation Crude Oil at 2008.

Table 3, 4 and 5 summarized the events which are classified as a bubble in US crude oil, US natural gas and US coal markets respectively. Typical LPPL estimation for US natural gas market is shown in Figure 12, while, the other estimation plots are available in Appendix B.

2.3 Conclusion

This study aims to investigate the ability of LPPL models to identify bubble (s) and its corresponding termination point (s) $[t_c]$ in US energy market using three US major energy prices namely, The daily US dollar crude oil price of West Texas Intermediate (WTI) and the US dollar natural gas price and also the US dollar coal price covering the 1987:05:15_2015:01:30 periods on daily basis.

To this end, first the drawdown analyses and the bubble index tool are adopted in order to identify the potential events that can classified as bubble. Then the LPPL model by means of rolling windows is applied, to forecast termination points (t_c).

To raise the reliability of the estimation process, the new constraint on the Augmented Dickey-Fuller and Philips–Perron values are introduced to the fits to accept fits that are nonstationary. Our findings reinforce the fact that energy market prices during bubble periods oscillate with decreasing amplitude around a faster-than-exponential growth. Beside that our results retell that the application of the LPPL model in US energy market are error-free where the actual regime shift date encloses by the confidence interval of termination point.

Table 4. The Events classified as Bubble in US Natural Gas Market.

Year	Description	Real regime shifts	Estimated t_c	80% Confidence Interval
1999	The price of natural gas and crude oil are related. They are substitutes in consumption and complements in production. During the 1980s and 1990s based on new series of problems in the natural gas industry; the capacity of suppliers expanded due to after the growth in demand for gas, which led to a gas bubble, but there was more producible gas than the market's demand. Although regularly market analysts predict the end of the bubble as only a few years further but the bubble refused to burst; so since it extended over time some called it the "gas sausage." Until the late 1990s, the problem of large gas inventories threatening the market and keeping down prices did not disappear.	28-Dec-1999	15-Feb-2000	15-Dec-1999 –15-Mar-2000
2001	Due to the global recession in 2001 and also the freezing weather in January 2001, the natural gas price started to increase therefore the bubble in 2001 happened.	15-Jan-2001	16-Feb-2001	17-Dec-2000 – 28-Feb-2001
2005	In late 2005, the severe supply disruptions of U.S. gas due to hurricanes Katrina and Rita led to sharp price increases. During the several months of disrupted production, prices stayed high until the improvement the effects of the hurricanes in early 2006.	14-Oct-2005	4-Oct-2005	26-Sep-2005 – 30-Oct-2005
2008	In the late 2007 and early 2008, increasing demand caused prices rise rapidly, and then in mid-2008 an economic crisis drove prices down quickly. Besides in 2008 and 2009 many newly-discovered natural gas fields caused a glut of gas that put additional pressure on prices to shift downward.	1-Jul-2008	4-Jun-2008	26-May-2008 – 11-Jul-2008

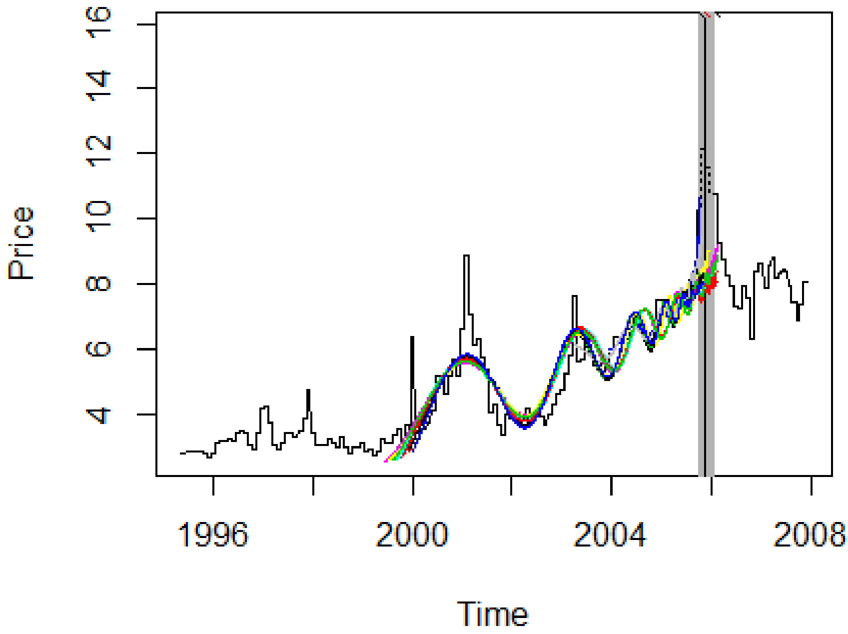


Figure 12. LPPL Estimation Natural Gas at 2005.

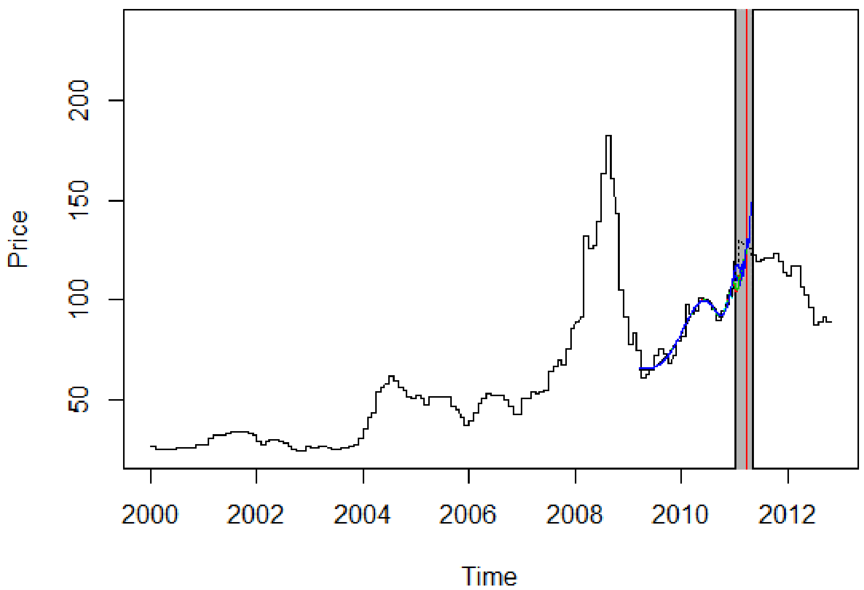


Figure 13. LPPL Estimation Coal at 2011.

Table 5. The Events classified as Bubble in US Coal Market.

Year	Description	Real regime shifts	Estimated t_c	80% Confidence Interval
2004	--	30-Jun-2004	4-Jul-2004	25-Jun-2004 – 16-Jul-2004
2008	Despite that 2008 was not a great year for coal consumption in the US, but domestic coal prices were increased due to increasing in fuel surcharges by the transportation sector regarding the massive rise in oil prices and increase in the eastern coal spot market prices in regards to the growing demand internationally for U.S. coal.	31-Jul-2008	11-Jul-2008	26-Jun-2008 – 30-Jul-2008
2011	Two reasons exist for believing the likely rise of the coal prices: First, many of recent studies offer that attainable and useful coal may be less than what has been assumed. In fact, the peak of world coal production may be only years after. Moreover, second, the rapid growth in global demand largely driven by China. China as the world's biggest producer of coal and also it is the greatest consumer so its effects on future coal prices should not be disesteemed.	31-Jan-2011	25-Feb-2011	28-Dec-2011 **-** 3-Mar-2011

Chapter 3

AUTOMATIC CLUSTERING IN US ENERGY MARKET

3.1 Introduction

Clustering means, partitioning an unlabeled data set into groups, which each group called a "cluster" containing of similar objects that are different to themes of other groups. In the recent decades, cluster analysis has been applied in various fields such as machine learning, artificial intelligence, pattern recognition, spatial database analysis, textual document collection, image segmentation, sociology, psychology, archaeology, and education, and economics, marketing and business (Evangelou *et. al*. 2001).

Data clustering algorithms are divided into two types; hierarchical and partitional, where each type has plenty of subtypes and various methods for finding the clusters (Leung *et. al*.2000). In hierarchical type, a sequence of clustering shown in the output; which each group considered as a partition of the data set (Leung *et. al*.2000). Hierarchical algorithms are in the form of bottom-up which called agglomerative or top-down which called as a divisive. In the agglomerative algorithm, each element begins as a separate cluster and combines them successively in larger clusters, while the whole set is considered in divisive algorithms and then divided into the smaller successive cluster.

Since there is no need for specifying the number of classes as *a priori* and freeness of the initial conditions counted as the main advantages of hierarchical algorithms; however the primary drawback is their stability meaning data points located in a cluster cannot move to another cluster. Moreover, they may fail to separate overlapping clusters due to inadequate information about the general shape or size of the clusters (Jain *et. al.*1999).

On the other hand, partitional clustering algorithms try to break down the data set directly into a set of disjoint clusters and optimize certain criteria. The criterion function may confirm the local structure or the global structure of the data through the assignment of clusters to peaks in the probability density function. The general criteria minimize dissimilarity of the samples across each group while maximizing the disparities of various clusters.

The advantages and disadvantages of the partitioned algorithms are opposite of hierarchical algorithms. Crisp and fuzzy are two different models under which clustering can also be performed; the drawback of crisp clustering is the disorganization among the clusters with no overlapping. Therefore, any pattern may belong to only one class while in fuzzy clustering, a schema may belong to all the categories with a certain fuzzy membership rate. It is notable to mention that a comprehensive survey of different clustering methods is available in (Jain *et. al.*1999).

Researchers around the world are coming up with new regular basis algorithms related to the problem of partitional clustering, to face the cumulative complication of large data set. Therefore, it beseems almost impractical to contain the extensive and multi-facets of clustering in the literature scope of this study. Instead, this study focuses on the field of evolutionary partitional clustering. The evolutionary approach considers the data clustering as an optimization problem and solves it through an evolutionary search heuristic such as genetic algorithm (GA) inspired by Darwinian evolution and genetics (Holland.1975).

The main idea is to generate a throng of candidate solutions for the optimization problem. Candidate solutions are chosen based on a fitness function, which measures their quality with respect to the optimization problem. In GAs, the alteration contains mutation and crossover; to discover solutions in the vicinity of existing solutions and reunite information between different candidate solutions respectively. Indeed the approach iteratively refines the solutions by conversion and selection of good ones for the next repeat.

These algorithms are capable of coping with local optimum by simultaneously preserving, reuniting, and comparing different candidate solutions; while in contrast, local search heuristics such as the simulated annealing algorithm (SA) (Selim *et. al*.991) are clearly weak to cope with local optimum and only refine a single candidate solution. Algorithms like the *K*-means use deterministic local search and from the beginning position of the search always converge to the nearest local optima (MacQueen.1967).

In the past few years extensive research works have adopted the application of the evolutionary computing methods in complex data clustering; however, in many of them, the determination of the optimal number of clusters has not been reported. Most of the existent clustering techniques which use evolutionary algorithms, accept *K* as an input for the number of classes instead of assigning the same on the run. Nonetheless, practically in many situations, the proper number of groups in an unhandled data set may be unknown or even not possible to specify approximately; for example, in the high-dimensional feature vectors dataset visualization of the dataset to trace its number of clusters may be impossible in practice.

It is worth mentioning that to specify the optimal number of clusters in a dataset; the traditional approach uses some special contrived statistical–mathematical function to check the clustering validity index for assessing the partitioning quality for a range of cluster numbers. A good clustering validity index determines global optimum at the exact number of classes in a data set; nevertheless, it is costly due to several clustering experiments for various possible cluster numbers.

In the applied evolutionary learning frameworks, a variety of test solutions appear with different cluster numbers along with the cluster center fits for the same dataset. Among all of them, a global validity index measures the quantitative correctness of each possible grouping (e.g., the CS [Chou *et al*...2004] or (DB) [Davies *et. al*.1979]). Then through mutation and selection mechanisms, the fittest solutions overcome the population, and the bad ones deleted. Finally, when the validity index indicates a near-optimal partitioning of the data set as the best option, then the evolution of the solutions stops. Therefore one runt of the evolutionary optimization algorithm contains the optimal number of classes as well as the accurate cluster center coordinates.

This method is a suitable clustering validity index dependent. An inefficient index may yield many false clusters, even when the actual number of clusters may be so manageable, however, a perfect choice can automate the entire process of clustering in the proposed algorithm.

In the traditional methods to apply in the clustering like iterative K-means algorithm, the user has to specify the number of the cluster in advance. Also, this algorithm is strongly data dependent, and since it follows the greedy approach, therefore the initial condition has a significant role in eliminating the suboptimal solutions.

Hence, this study adopted a novel approach which was developed by Das *et. al.* (2008) to track a twofold objective. First, it targets the automatic specification of the optimal number of clusters for an unlabeled energy data set through five classes of various evolutionary techniques, with utilizing a new representation scheme for the search variables to assign the optimal number of clusters. Moreover, it intends to evaluate the application of considered methodology in bubble detection strategy.

We have compared the performance of five automatic clustering methods, based on Genetic Algorithm (GA), particle swarm optimization (PSO) and harmony search (HS) and differential evolution (DE) and artificial bee colony (ABC) in the case of the unlabeled energy data set.

3.1.1 Problem Definition

The combination of different features as a common set of attributes represent a pattern or data points. Let X_{n*d}, *a* profile data matrix with *n* *d*-dimensional row vectors. Each element $X_{i,j}$ in $\overline{X_\iota}$ correspond to the j_{th} (*j=1,2,…,d*) real value feature of the i_{th} (*i=1,2,…,n*) pattern. A partition *C* = *{C*1, *C*2,.., *CK}* of *K* classes is determined using partitional clustering , with maximum likeness of the schemas in the same cluster. Three properties are necessary for each partition.

I. At least one pattern should assign to each cluster.
II. No common pattern should exist between two different groups.
III. Each data points should be attached to one cluster.

Due to different ways of partitioning for a given data set, and also to maintain all the characteristics above, a fitness function must specify, and then the problem tries to find the optimal cluster (C^*).

$$\text{Optimize } f\,(X_{n*d}, C^*) \tag{3.1}$$

Where $\boldsymbol{C}^*$ the optimal cluster from the set of **C** and $\boldsymbol{f}$, is a mathematical function that guarantees the goodness of a cluster based on the distance measure.

3.1.2 Similarity Measures

Since clustering is adopted based on some homology measures, therefore, identifying the appropriate similarity measures has the important role in clustering. The Euclidean distance as one of the popular similarity measures has been widely used to evaluate similarity such that between any two $\boldsymbol{d}$ dimensional patterns, $\overline{\boldsymbol{X_\iota}}$ and $\overline{\boldsymbol{X_\jmath}}$ is given by:

$$d\,(\bar{X}_i, \bar{X}_j) = \sqrt{\sum_{p=1}^{d} (X_{i,p} - X_{j,p})^2} = \|\bar{X}_i - \bar{X}_j\| \quad (3.2)$$

It is notable to mention; the Euclidean distance derived from the general form of distance measure known as Minowsky metric (Jain *et al*...1999), which defined as

$$d^{\,\alpha}(\bar{X}_i, \bar{X}_j) = \|\bar{X}_i - \bar{X}_j\|^{\alpha} \qquad (3.3)$$

The Minowsky metric[5] is inefficient to make a cluster for the high dimensional data set. Since in the high-dimensional data, the distance between the clusters is increasing. Therefore the concept of near and far in clustering became weaker. Besides that, in this method, tendency of the large-scale features to dominate over the other traits can be solved by normalizing the features over a common range. (Jain *et. al*.1999)

3.1.3 Clustering Validity Indexes

As we mentioned before, in this study, we tried to define the automatic cluster(s) in the unlabeled energy data set, to do so, a statistical, mathematical function needed to appraisal for the results of the clustering algorithm. A cluster validity indexes are utilized to provide the mentioned concern. Using the cluster validity index guarantees first the optimal number of cluster(s) and second it finds out the corresponding best partition(s). Validity indexes consider two aspects of partitioning. First known as cohesion where denotes that, the data points in each cluster should be similar as much as possible which fitness variance of pattern in each cluster indicates the cluster‘s cohesion. Moreover, the second one is separation such that clusters should well disassemble.

[5] When α =1, the measure is known as the Manhattan distance.

The distance between the centers of the cluster indicates the cluster separation. In crisp clustering, many of validity indexes are available (e.g. Dunn's index (DI) [Dunn.1974], the Calinski–Harabasz index [Calinski *et. al.*1974], the DB index, and the CS measure. All indexes are optimized in nature, therefore; it is a good idea to associate any optimization algorithm like GA, etc. in the adoption process. Next section reports only one validity index (DB) which used in this study.

3.1.4 DB Index

This index is determined as a function of the ratio of the sum of within-cluster scatter to between-cluster separation which uses both the clusters and their sample means. The within i_{th} cluster scatter $\boldsymbol{S_{i,q}}$ and the between $\boldsymbol{i_{th}}$ and $\boldsymbol{j_{th}}$ cluster distance $\boldsymbol{d_{ij,t}}$ respectively, are defined as the following.

$$S_{i,q} = [\frac{1}{N_i} \sum_{\vec{X} \in C_i} \|\vec{X} - \vec{m_i}\|_2^q]^{1/q} \quad (3.4)$$

$$d_{ij,t} = \left\{\sum_{p=1}^{d} |m_{i,p} - m_{j,p}|^t\right\}^{1/t} = \|\vec{m}_i - \vec{m}_j\|_t \quad (3.5)$$

Where $\vec{\boldsymbol{m}}_{\boldsymbol{i}}$ is the $\boldsymbol{i_{th}}$ cluster center q, $t \geq 1$, q is an integer, and q and t are selected independently. The number of elements in i_{th} cluster $\boldsymbol{C_i}$ denotes by $\boldsymbol{N_i}$ **.**The ratio ($\boldsymbol{R_{i,qt}}$) can be defined as:

$$R_{i,qt} = \max_{j \in K, j \neq i} \left\{\frac{S_{i,q} + S_{j,q}}{d_{ij,t}}\right\} \quad (3.6)$$

Finally, the DB is defined as:

$$DB(K) = \frac{1}{K} \sum_{i=1}^{K} R_{i,qt} \quad (3.7)$$

Note that a valid optimal partition shows by smallest *DB(K).*

3.1.5 Metaheuristic Algorithms

In this section, we summarized the description of five evolutionary techniques which have been adopted in this study.

3.1.5.1 Artificial Bee Colony (ABC)

Artificial bee colony (ABC) is an optimization technique recently developed in 2005 by Karaboga who inspired by the behavior of honey bees. ABC algorithm as a swarm intelligence algorithm mimics the sagacious foraging act of honey bees. Swarm which is a set of honey bees is responsible for doing the tasks which can be accomplished via social collaboration.

Algorithm includes three kinds of bees. The first type is accountable for searching for food around the food source and assigning the information of the memory about food sources at the same time to the second type of the bees which is called onlooker bees. The onlooker bees are inclined to choose healthy food source that has higher quality. The scout bees as the third type they are a few ones coming out of the working bees, which drop their food sources and look for the new ones. ABC algorithm has been employed to solve variety kinds of problems. In this algorithm, employed bees constitute the first half of the swarm and the second half is comprised of the onlooker bees.

3.1.5.2 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively attempting to improve the solution of a candidate on a measured quality. The problem is solved by having a pool (population) of candidate solutions called particles. These particles are moved around in the search space according to simple mathematical formulae calculated over the particle's position and velocity. The particles movement which defined by the local best-known position of each particle is instructed to the best-known position in the search space as well as the whole swarm's best-known position.This process immediately is updated to find the best solution.

Kennedy, Eberhart and Shi (1995,1998) are originally linked with PSO; it was first intently used to simulate social behavior which is represented in a stylized movement of organisms in a bird flock or fish school. The durability of the algorithm is simplified and was observed for optimal functioning. Different aspects of PSO and swarm intelligence were discussed in Kennedy and Eberhart's book, while a wide survey of PSO applications was made by Poli (2007).

The work of Bonyadi and Michalewicz (2016) recently revealed a comprehensive review both theoretical and experimental works on PSO. The main advantage of PSO is that this method is assumptions free or at least a few assumptions needed for problem optimization, while the algorithm can search huge spaces of candidate solutions. Categorically, PSO does not use the gradient of the optimized problem, indicating that PSO does not require a differentiable optimization problem which is in contrast to the requirement of the classic optimization methods such as quasi-newton and gradient descent method.

3.1.5.3 Harmony Search (HS)

The Harmony Search as a relatively new music-based metaheuristic optimization algorithm was initially developed by Zong Woo Geem et al. in 2001. The scientists have found the sizable connection between music and finding an optimal solution to a tough problem.

The inspiration of the HS is not from the natural phenomena, but it is inspired by the aim of the musical process to search for a perfect pleasing harmony determined by aesthetic standards. In the composition of a harmony, the musicians usually try different feasible mixtures of the music pitches stored in their memory. This kind of efficient search for a perfect harmony is similar to the process of finding the optimality in an optimization problem where the optimal solution should be the best available solution to the problem under the desired objectives and given constraints.

As a matter of fact, both processes search to detect the best. Such similarities between two processes resulted in proposing such a successful algorithm. HS algorithm transforms the qualitative improvisation process into some quantitative rules through idealization. Then by searching for a perfect harmony, turns the beauty and harmony of music into an optimization procedure. Since its first appearance in 2001, due to its advantages and effectiveness demonstrated it had gained exceptional research success in different applications. During the recent years, it has been applied to solve many optimization problems such as water distribution networks (2010), groundwater modeling, control, function optimization, energy-saving dispatch, vehicle routing, etc.

3.1.5.4 Differential Evolution (DE)

Differential evolution (DE) which was proposed by Storn and Price (1997), commonly known as a metaheuristic method optimizes a problem by trying to improve a candidate solution iteratively according to a given measure of quality. Such methods make few or no assumptions and are capable of searching vast spaces of candidate solutions to optimize the problem. However, in metaheuristics such as DE, there is no guarantee for finding the optimal solution.

For the standard optimization methods such as gradient descent and quasi-newton, the optimization problem should be differentiable while for DE used for multidimensional real-valued functions, this is not needed, and it does not use the gradient of the problem. Therefore, DE can also be applied to optimization problems which are noisy, or change over time, or even are not continuous, etc. (2011).

In problem optimization, DE maintains a throng of candidate solutions and combines the existing ones based on its simple formulae to generate new candidate solutions and then preserves of the candidate solution with the best score or fitness. So by this way, the optimization problem looks like a black box that only provides a given quality measure and therefore the gradient is not required. Perfect surveys exist on the multi-faceted research aspects of DE[6].

3.2 Methodology

Das et al. (2008) modified the DE algorithm using new chromosome representation such that for $\boldsymbol{n}$ data points, with $\boldsymbol{d}$ dimensions by a user-specified maximum number of clustersK_{Max} .

A vector of real numbers of dimension $K_{Max} + K_{Max} * d$ is called as a chromosome. K_{Max} is the positive numbers between 0 and 1 to controls the activation of each cluster. And the second K_{Max} represents cluster centers.

To check for activation of j_{th} cluster center in i_{th} chromosome an activation threshold $T_{i,j}$ is defined such that:

6 GA has been discussed by detail in chapter 5.

"IF $T_{i,j}$ > 0.5 **THEN** the j_{th} cluster center

$\vec{\boldsymbol{m}}_{\boldsymbol{i,j}}$ is **ACTIVE**

ELSE $\vec{\boldsymbol{m}}_{\boldsymbol{i,j}}$ is **INACTIVE "(**Das *et al* (2008)**)**

After creation of new offspring chromosome, the T values are controlled using the above rule to select the active cluster centroids. The advantage of using this method is that validity index can be used as a fitness function. Since clustering as the most important unsupervised learning problem deals with forming logical structure to extract the hidden properties in the unlabeled data set. Hence, this Study contributes to the literature in four ways. First of all, we generalized the same methodology which was proposed by Das *et al.*(2008) for another four classes of evolutionary algorithms (Namely PSO, HS, GA, and ABC) to apply in US energy market prices. Secondly, we examined the long-horizon data set which contains three major price index namely crude oil prices, coal prices and natural gas prices over the period 1987:05:15_2015:01:30 on a daily basis[7]. Thirdly, to the best of our knowledge, this essay is the first remarkable study on the energy price classification using the automatic clustering method. Moreover, the last contribution is proposing a novel approach by using the modern data mining method to detect the termination point of a bubble in US energy market.

[7] Data descriptions are available in chapter 2.

3.3 Empirical Findings

Using five automatic clustering methods have revealed that in the US energy data set only two clusters out of 10 user specified[8] clusters can be assigned as shown in the figures 14-19.

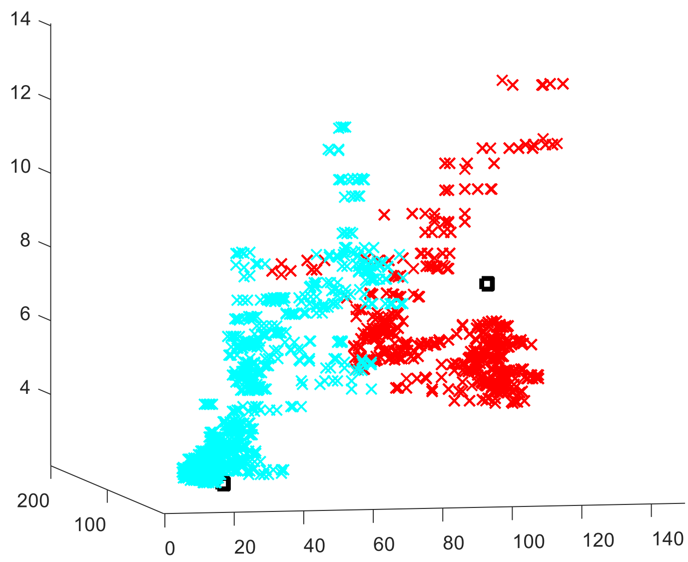


Figure 14. DE Result.

Table 6. DE Properties.

Centers	X(Crude Oil)	Y (Coal)	Z (Natural Gas)
M1	110.48	109.15	7.36
M2	21.77	31.94	2.61

As shown in Tables 11 the performance of PSO and GA respectively are outperformed by every other algorithm. They have a lower cost for clustering of employed data set. On the cluster centers in each algorithm, two different classes can determined.

[8] In our study, each algorithm has been iterated 200 times with same population size which was set to 100.

The first class shows high price level (which the probability of the existence of the bubble is high) while the second class represents low price level in energy data set.

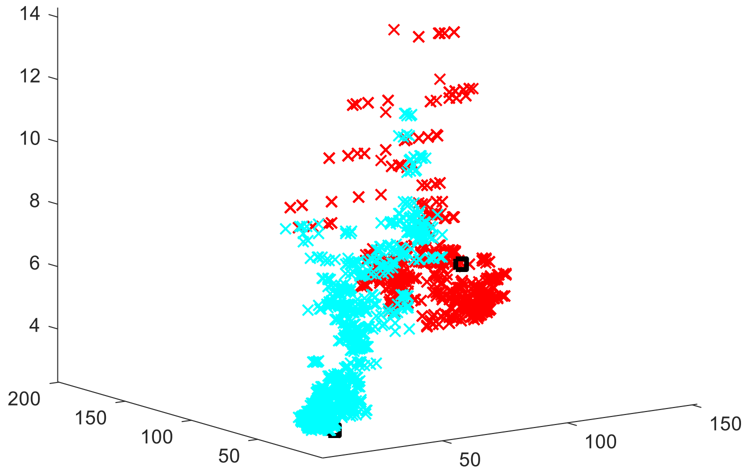


Figure 15. ABC Result.

Table 7. ABC Properties.

Centers	X(Crude Oil)	Y (Coal)	Z (Natural Gas)
M1	114.56	108.94	5.86
M2	21.55	29.55	2.61

The chance of existence bubble in high-level class is greater than low-level class. Therefore the procedure has been repeated for this class to classify the new dataset. Using the proposed method iteratively in newly extracted class is useful to identify the switching interval between two different regimes (classes) which contain the critical point for the bubble (high price level change).

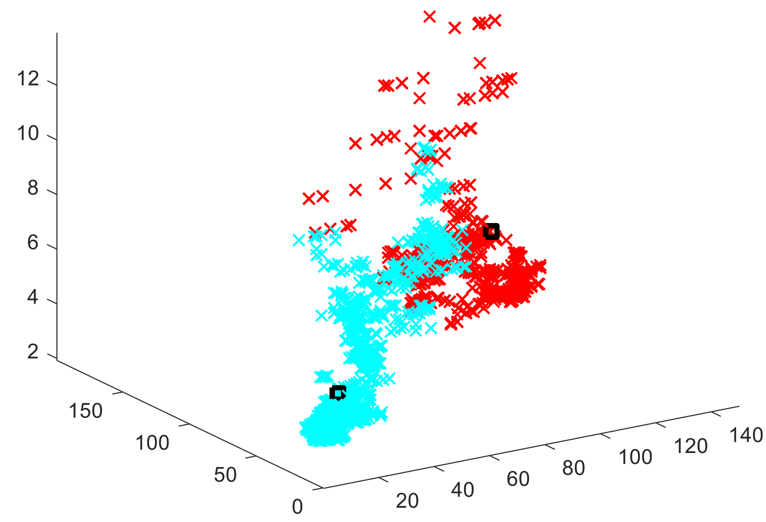


Figure 16. GA Result.

Table 8. GA Properties.

Centers	X(Crude Oil)	Y (Coal)	Z (Natural Gas)
M1	113.26	109.03	6.50
M2	20.91	32.08	4.23

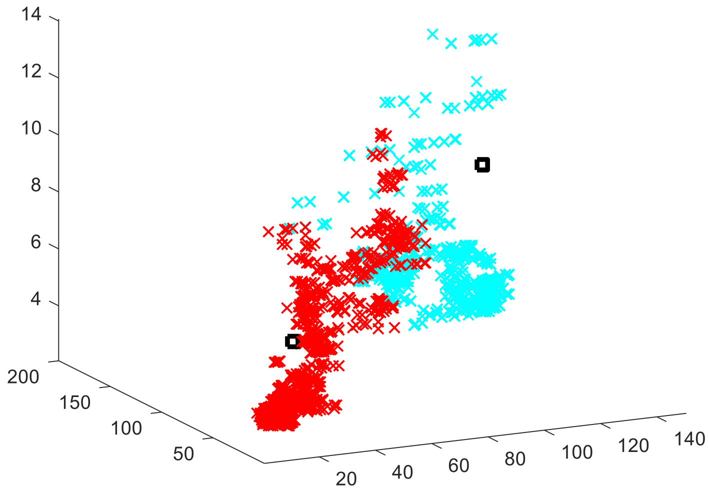


Figure 17. HS Result.

Table 9. HS Properties.

Centers	X(Crude Oil)	Y (Coal)	Z (Natural Gas)
M1	118.37	110.57	9.19
M2	21.33	30.08	5.58

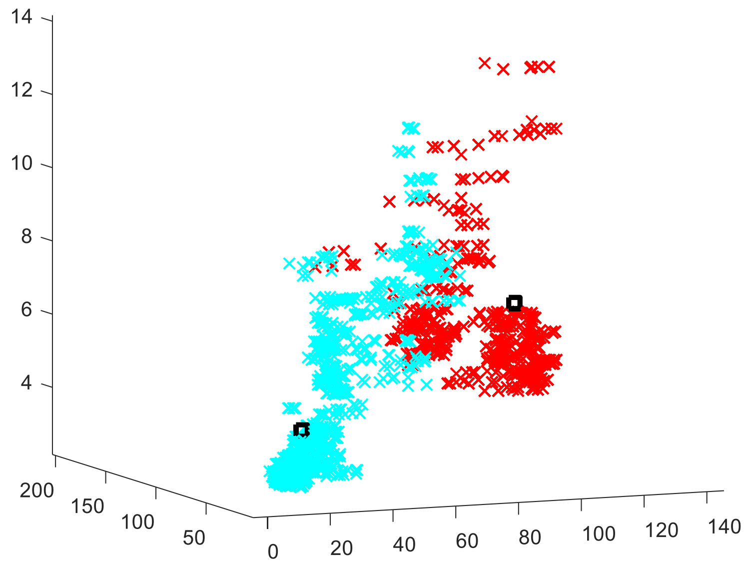


Figure 18. PSO Result.

Table 10. PSO Properties.

Centers	X(Crude Oil)	Y (Coal)	Z (Natural Gas)
M1	112.52	108.44	6.53
M2	20.27	31.99	4.19

Note that to get robust results rescaling the variables which involved in the data set is recommended. Adoption the proposed method for high-level class reveals the following results as shown in Table 12.

Table 11. Assigned Clusters, Cost and Rank of Each Method in Clustering.

Model	Number of Assigned Clusters	Best Cost	Rank
PSO	2 out of 10	0.462	1
GA	2 out of 10	0.462	2
HS	2 out of 10	0.467	5
ABC	2 out of 10	0.464	4
DE	2 out of 10	0.463	3

Table 12. US Energy Market Clustering Based on Bubble Phenomena.

Algorithm	ABC	GA	PSO	HS	DE
Termination point Horizon	24:06:2008 To 02:07:2008	14:06:2008 To 28:08:2008	30:06:2008 To 30:08:2008	30:05:2008 To 02:06:2008	25:06:2008 To 15:08:2008

The results of Table 12 are comparable with results of LPPL models which are discussed in chapter one. As we highlighted there, the conduction of LPPL patterns in energy market (in the case of crude oil prices) released the confidence interval for termination point spanning from 25:05:2008 till 02:08:2008.

While Sornette (2009) discovered, 80 percent confidence interval of wreck dates spreading from 17^{th} May to 14^{th} July, despite the fact that, the real change happened in 03:07:2008. Based on the findings, similarity detection in crude oil particularly and in the energy market is approvable.

Note that since we applied high-frequency data set with different scales, the concept of a bubble in this methodology (Since this method has a universal point of view to classify the database) is determined by the whole observation. For example, US energy prices have experienced a high level of changes in prices during 2008, about the other frequencies (i.e. 1990)[9].

Hence this method is not able to detect the other extreme shifts in the price level. To overcome this drawback, first rescaling all variables (transforming to the percent change) to have similar range is vital and second using different time frames in the data set as a pre-process is suggested. In this way, for each time frame (i.e. Three years) the proposed model should apply to define the class of data set then after definition for each class the appropriate policy can be adopted.

All in all, the proposed method, has a potential to detect real extreme shifts in the energy market, albeit, pre-analyses for the data set is a very vital task to achieve robust and reliable results.

3.4 Conclusion

In the recent decades, cluster analysis has been applied in different fields such as machine learning, artificial intelligence, pattern recognition, spatial database analysis, textual document collection, image segmentation, sociology, psychology, and archeology.

Extensive research works have adopted the application of the evolutionary computing methods in complex data clustering; however, in many of them, the determination of the optimal number of clusters has not been reported. This study targets at the automatic specification of the optimal number of clusters for an unlabeled energy data set through five classes of various evolutionary techniques,

[9] The range of crude oil prices in 1990 is not as well as range of price at 2008

with utilizing a new representation scheme for the search variables to assign the optimal number of clusters.

Moreover, it intends to evaluate the application of considered methodology in bubble detection strategy. We have compared the performance of five automatic clustering methods, based on GA, particle swarm optimization (PSO) and harmony search (HS) and differential evolution (DE) and artificial bee colony (ABC) in the case of the unlabeled energy data set.

We found that the performance of PSO and GA respectively are outperformed by any other algorithms. Besides that, adopting the proposed method to label energy market, reveals two different classes, while generalization of this approach using the iterative algorithm, detect the correct interval for the crude oil bubble at 2008.

Chapter 4

OUT OF SAMPLE FORECASTING OF US ELECTRICITY PRICES, USING GMDH

4.1 Introduction

Many countries have taken the path of power market liberalization over the past two decades. Due to this process which is based on the concept of differentiation in services and infrastructures, the power industry has changed from the systematized structure to an open competitive market setting (Kirschen and Strbac, 2004, Weron, 2006).

Electricity as a salable commodity at any market rates is a very particular case since the electricity demand is weather and business cycle dependent. The price of electricity is inelastic, at least over the short time horizons, due to Incuriosity of the consumer about the current price of electricity. Moreover, electricity cannot be saved while power system stability depends upon a constant balance between production and consumption. These items give rise to severe price volatility and sudden changes known as spikes.

There are many techniques to forecast electricity prices, but only some of them are well suited to deal with price fluctuations. A whole range of models which have been proposed in the literature are listed as follow: Zhou et al., 2006 were applied ARIMA and seasonal ARIMA models, while autoregressions with heteroscedastic methods were adopted by Garcia et al., 2005. Conejo et al., 2005 used dynamic regression' (or ARMAX). Vector Autoregression model (VAR) with exogenous effects are applied by Panagiotis and Smith, 2008, and Misiorek et al., 2006 adopted threshold AR and ARX models to forecast price electricity.

In this study the performance of a suite of non-linear models (namely, Locally Linear Model Tree (LoLiMot) Neuro-fuzzy Model, Multi-Layered Perceptron (MLP) Artificial Neural Network (ANN), Nonlinear Autoregressive (NAR) ANN, and Group Method of Data Handling (GMDH)) in forecasting US daily electricity prices are evaluated.

The instruction of this study is as follow: in section 2 a description of the data set is introduced. Section 3 deals with the methodology. Section 4 discussed the empirical findings, and finally, section 5 concludes.

4.2 Data Description

We used the logarithmic form of daily US dollar-weighted average price (power index)[10] of wholesale electricity price of Mid-Columbia (Northwest hub)[11] which covering 2001:01:05_2016:08:16 period and calculated by EIA. The data for this study was sourced and obtained from EIA (*Energy Information Administration*). The power index which used in this study is developed using the following formula

$$I = \sum \frac{P*V}{T} \qquad (4.1)$$

Where P stands for a price of the individual transaction, V denotes the volume of individual transaction and T represents the total volume of all qualifying transactions. Figure 19 shows the hub position in US and Figure 20 demonstrate the plot of the dataset which used.

As shown in Figure 20 the extreme volatility in the price leads to having high complexity. This idea is enforced by Figure 21 and Table 13 which both report non-normality in the data, hence, this fact motivates us to use a class of nonlinear models to map complex phenomena and catch a better performance in forecasting of US electricity prices.

[10] Applied power indices in this study are obtained per se from transactions administrated on the ICE (Intercontinental Exchange for prompt or "day ahead" markets in North American power) platform representing approximately 70 percent of next day trading activity.
[11] “Mid-C is a power trading hub for the Northwest U.S. containing the control areas of three public utility districts in Washington that run hydroelectric projects on the Columbia River. The three PUDs are Grant, Douglas and Chelan. Hydro projects include Wells, Rocky Reach, Rock Island, Wampum and Priest Rapids dams”.

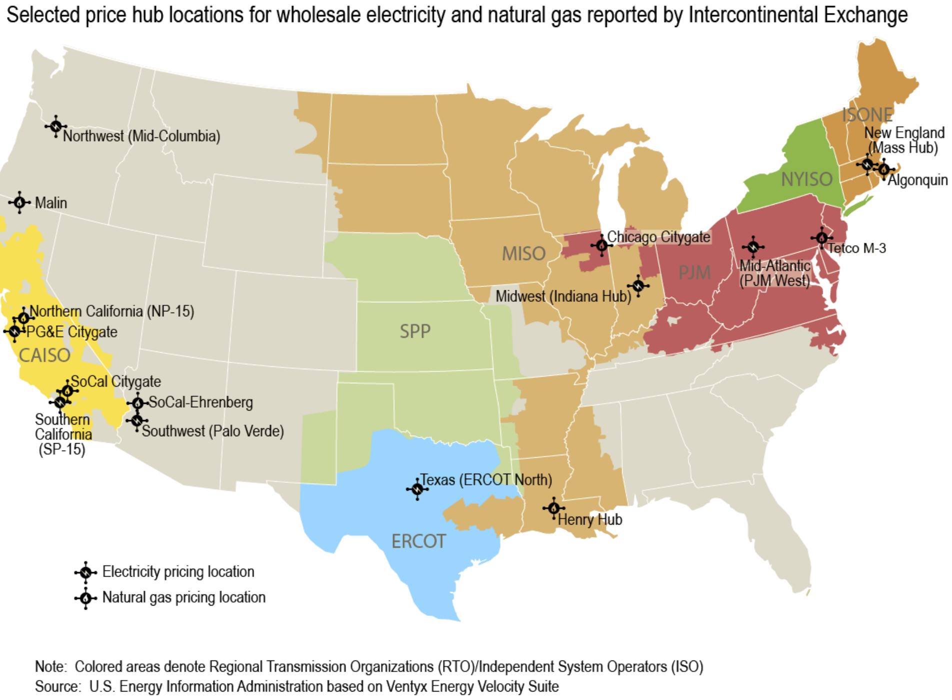


Figure 19. Hub Positions in the US.

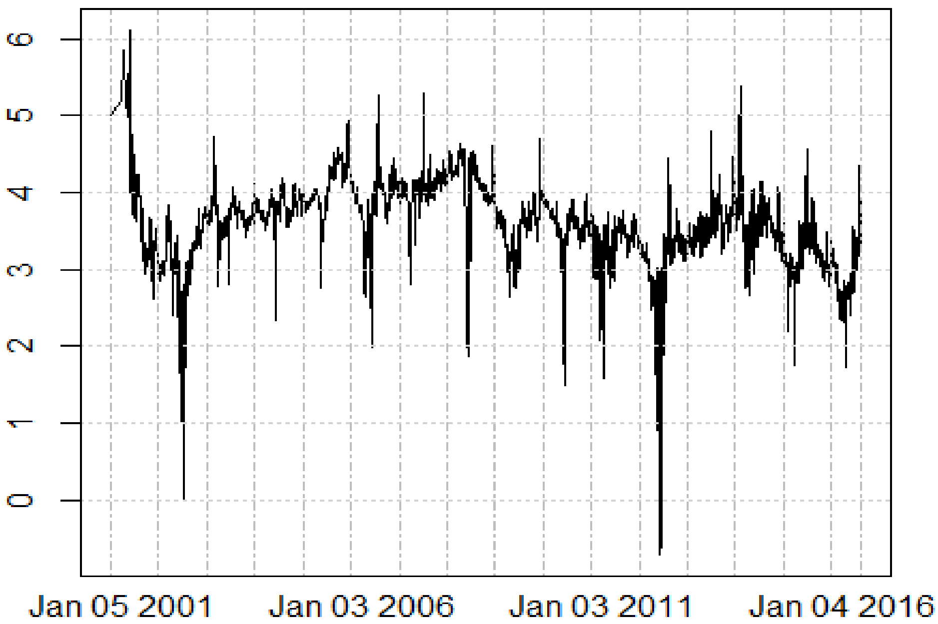


Figure 20. Logarithmic Form of Daily US Dollar-Weighted Average Price of Mid-Columbia.

Table 13 and Figure 21 represent the statistical properties of logarithmic form of weighted average price of electricity.

Table 13. Descriptive Statistic of Log Price.

Mean	3.569
Median	3.609
Maximum	6.109
Minimum	-0.713
Std. Dev.	0.519
Skewness	-0.921
Kurtosis	7.999
Jarque-Bera	4457.192
Probability	0.000000

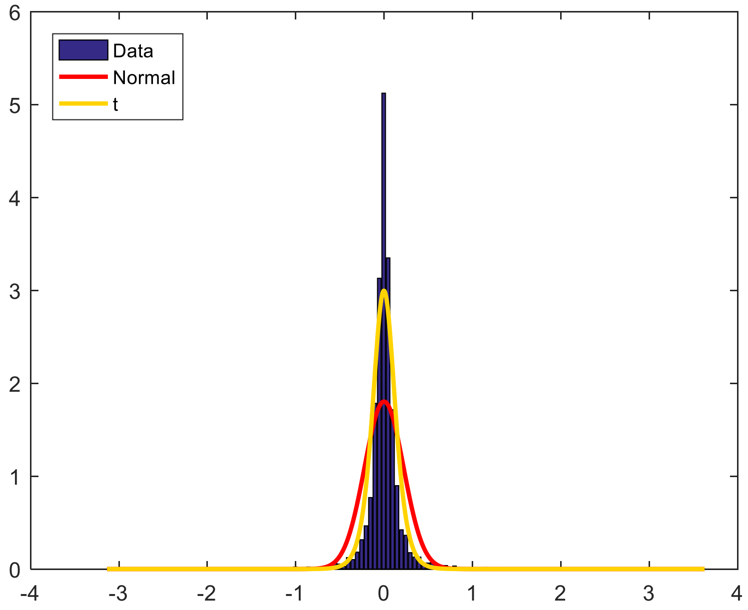


Figure 21. Histogram of Log-Return Data & Fitted Distributions.

Figure 21 shows that the data are not distributed normally. Moreover, the dataset are not fitted by t-student as well. The series are more peaked in comparison with both distribution. Therefore the series is fat tiled with high probability of extreme values against the normal and t distributions.

4.3 Methodology

In this study the performance of a suite of non-linear models (namely, Locally Linear Model Tree (LoLiMot) Neuro-fuzzy Model, Multi-Layered Perceptron (MLP) Artificial Neural Network (ANN), Nonlinear Autoregressive (NAR) ANN, and Group Method of Data Handling (GMDH)) in forecasting US daily electricity prices evaluated. The models are used to evaluate 1, 6, 12, 24 and 30 step ahead forecasts over an out-of-sample period of 2016:05:11–2016:08:16.

We contribute to the literature on forecasting electricity prices in three ways. First, we applied a different class of machine learning techniques to cover the nonlinearity and complexity in US electricity prices. Second, we used a high-frequency data set (on a daily basis) to forecast 30 days ahead. Third, we developed the Group Method of Data Handling (GMDH) method as a recursive forecasting tool to apply in the US out of sample forecasting of electricity prices. The methodology of all methods has explained in chapter four except the GMDH which discussed in the following.

4.3.1 Group Method of Data Handling (GMDH)

GMDH technique is more heuristic and self-organizing in nature; it’s modeling approach are created from neurons in the form of active networks in a recalling generation of populations of contending models of growing difficulty, corresponding validation as well as the selection model until an ideal complex model that is not very straightforward and not very perplexing have been figured it out.

GMDH runs systematically like pattern as a feed forward network in a quadratic node transfer framework manner, where the coefficients are gotten via a regression technique. (Farlow, S.J., 1984)

Ivakhnenko (1971) is the brain behind the GMDH model. The core of the method is to offer systematically an approach that operates in a feed forward pattern which is hinged on a quadratic node function and the coefficients of this model are arrived at by the famous regression technique (Karabulut *et. al.* 2006) given that it's a multivariate analysis approach for system difficult to recognize and not easy to explain. GMDH is employed to prevent the difficulty in getting a priori setting with the algebraic method of the progression.

Thus, this goes to show that GMDH can handle a system that is complicated regardless of full information about the system. In a scenario were factors are approximated with the aid of LS (least square) technique could be grouped on the general outset and partial outset to represent the combinatorial (COMBI) and multi-layered iterative algorithms (MIA), reciprocally (Muller and Lemke, 2000).

The application of GMDH method has gained wide coverage; this is basically due to the increased use of the self-organizing heuristic network (Nariman-Zadeh *et. al.* 2003).

Series of attempt exist in the past on how to put together a population-based stochastic search approach like the artificial neural network (ANNs), with the emergence of numerous methods, that assist in difficulty in having huge search space .(Porto, 1997). In (Yao, 1999) the presentation of the review of the evolutionary method of ANNs is given.

Genetic algorithms employ the feed-forward GMDH approach NNs, the excellent assortment of connection with successive layers are being utilized by the neuron (Vasechkina *et. al*. 2001).

Using a quadratic polynomial through the GMDH algorithm, standard forms of neurons within associated layers and pairs are known to generate new layers further. Linking of outputs to inputs could be modeled with such information. Acceptably, recognition issue could be perceived as a determinant of a function $\hat{f}$ that is used instead of real one, f with the purpose of computing an output $\hat{y}$ of a unique input vector X which is next to its real target y.

Given a different M observations of multi-input–single-output data with the actual object, the following expression is assumed:

$$y_i = f(X_{i1}, X_{i2}, \dots, X_{in}) \quad (i = 1,2, \dots., M) \tag{4.2}$$

For any specific input vector, a GMDH type NN is trained to predict the target values as expressed below:

$$\hat{y}_i = \hat{f}(X_{i1}, X_{i2}, \dots, X_{in}) \quad (i = 1,2, \dots., M) \tag{4.3}$$

GMDH style NN is expressed below to minimize the square of the error value:

$$\sum_{i=1}^{M} [\hat{f}(X_{i1}, X_{i2}, \dots, X_{in}) - \hat{y}_i]^2 \longrightarrow min \tag{4.4}$$

An assembled discrete form of the Volterra functional series which describes the following equation gives the overall relation between output and input parameters:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \text{ [12]} \quad (4.5)$$

Moreover, the next expression represents a full algebraic arrangement using a system of partial quadratic polynomials with only two parameters (neurons):

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 {x_i}^2 + a_4 {x_j}^2 + a_5 x_i x_j \quad (4.6)$$

Specifically, the general arithmetic correlation of input and output parameters of the equation (4.5) above is formed from a reversely linked network of neurons of a utilized partial quadratic sketch. To minimize the diversity among the determined output $\hat{y}$ and the real output y given any mate of input parameters x_i , x_j the coefficients a_i of (4.6) are clearly indicated through regression approaches.

In real sense, the notably hierarchy of polynomials is based on using the expressed quadratic form of (4.6) that have obtainable constants using the least-squares logic. Through the entire process of set of output–input data pairs, the constants the quadratic functions G_i are optimally obtained in output form as follows:

$$E = \frac{\sum_i^M (y_i - G_i)^2}{M} \longrightarrow min \quad (4.7)$$

In the GMDH approach, the selected n input parameters with associated probabilities are assigned two non-aligned parameters that yield the regression polynomial in (4.6) which properly fits the dependent observations by least-squares logic. This least-squares methodology through the estimate of the multiple-regression sense yield the below expression:

$$a = (A^T A)^{-1} A^T \quad (4.8)$$

[12] The expression called as a Kolmogorov–Gabor polynomial (Fonsecaet *et. al.* 1996).

Where A stands for the subsequent matrix for each row of M data trebles. For all the array of M data trebles, the above equation indicates the vector of the best coefficients of (4.6).

It is notable to denote that this method can be replicated for every neuron of the last hidden layer together with the relatedness format of the NN. The output of such standard equations is likely used to enhance deflections and more powerfully in increasing the propensity of the formulas mentioned above (Lin *et. al* 2006).

The training process of NNs with linked coefficients or weights adopted by the stochastic techniques is presently proved to be premier to established gradient-based method. Neurons in all layers of most GMDH mod of NNs are connected to those in the closest layer.

4.4 Empirical Studies

Table 14 shows the RMSE of each model over the forecast horizons h ($h \in \{1,6,12,24,30\}$). Close competition exists between the performance of GMDH model and the performance of LoLiMoT model in all forecast horizons.

However, on the whole, LoLiMoT outperforms GMDH except at medium horizon ($h=6$). In all horizons, the MLP comes the third model, while the NAR has the worst forecasting performance. The performance of the NAR is improved gradually due to the rise of complexity and dynamics in nature of data on the latter horizons.

In the last selected horizon (h=30) the performance of LoLiMoT and GMDH are better than MLP and NAR which are ranked by lowest RMSEs. Figure 22 also shows the performance of all models in one step ahead forecast.

Table 14. Root Mean Squared Errors & Ranks.

Models	**GMDH**		**LoLiMoT**		**NAR**		**MLP**	
Step *(h)*	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
1	0.34	2	0.32	1	0.90	4	0.34	3
6	0.32	1	0.32	2	0.87	4	0.47	3
12	0.33	1	0.33	2	0.81	4	0.48	3
24	0.34	2	0.30	1	0.78	4	0.54	3
30	0.35	2	0.31	1	0.81	4	0.56	3

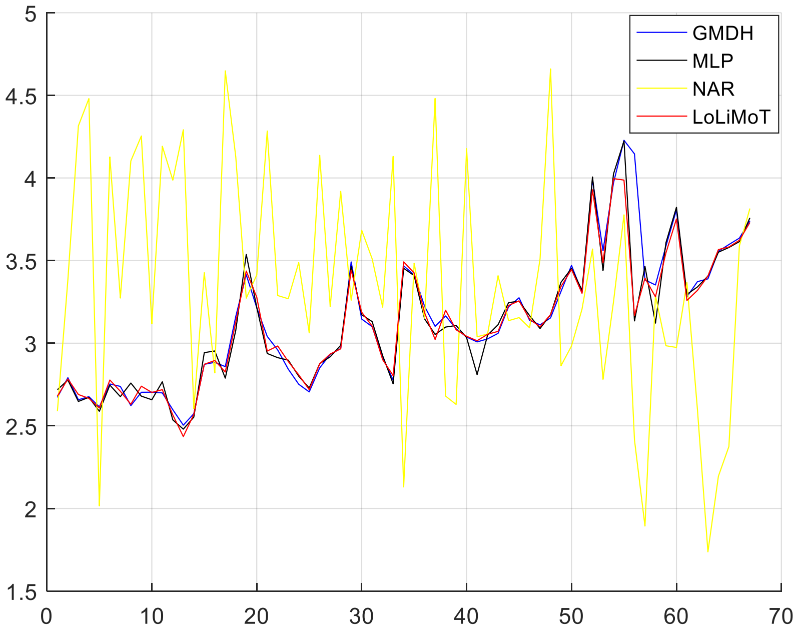


Figure 22. One-Step-Ahead Electricity Price Forecasts.

4.5 Conclusion

Electricity as a commodity which is saleable at market rates is a very particular case since the demand for electricity depends on weather and business cycle conditions. There are many techniques to forecast electricity prices, but only some of them are well suited to deal with price volatility. Non-linear approaches recently have been the center of attention in the literature, in forecasting economic variables including inflation (Balcilar, Gupta, & Miller 2012).

This study aims to evaluate the performance of a suite of non-linear models (LoLiMot-MLP-NAR-GMDH) in forecasting US daily electricity prices. The models are used to evaluate 1, 6, 12, 24 and 30 step ahead forecasts over an out-of-sample period of 2016:05:11– 2016:08:16.

This study contributes to the existing literature on forecasting electricity prices in three ways. First, we applied different classes of machine learning techniques to cover the nonlinearity and complexity in US electricity prices. Secondly, we used a high-frequency data set (on a daily basis) to forecast 30 days ahead.

Thirdly, we developed the Group Method of Data Handling (GMDH) method as a recursive forecasting tool to apply in the US out of sample forecasting of electricity prices. The outcomes of this research reveal that the LoLiMoT model and the GMDH model outperform by two other models in all forecast horizons which retells the potential of these two methods in forecasting of the complex high-frequency data set.

Chapter 5

OUT OF SAMPLE FORECASTING US INFLATION USING NON-LINEAR MODELS

5.1 Introduction

The energy prices as they affect economic variables in the world are influenced by economic activities of great countries. Indicatively, oil prices which are a major energy index globally are affected by economic activities of great countries, and when such activities are on the decline, the economy of the industrial countries slips into recession. When such countries plunge into economic recession, production decreases and this results in low demand for oil input and subsequent drop in the global oil demand. Conversely, oil price increase in these countries can lead to inflation which results in low oil productivity.

Changes in the oil price can cause the main fluctuations in the inflation rate, and these fluctuations are liable to lead to economic changes which directly affect overall economic performance. Researchers have revealed the significant impact of the oil price to determine the price inflation for the consumers due to the direct impact of oil on many production activities and in consuming products.

The significant implications of the oil price shocks on inflation during the 1970s and 1990s are indicative of this fact. Historical references recorded a rise in the oil price levels from $3 per barrel before 1973 to almost $40 per one in 1979 and also a high rise from $15 in 1998 to almost $140 in 2008. The 1982 base year consumer price index developed by Bureau of Labor Statistics of the United States of America equivalently revealed the rise from 41.10 at the beginning of 1972 to 86.30 at the end of 1980 and later a rise from 164.30 in January of 1999 to 214.82 in April 2008. (S.K.Sek, et. al. 2015)

Notwithstanding, evidence of a strong correlation between inflation and oil price could vary over time periods. In the mid-1980s, such evidence was thought to have disappeared. In the work of Evans and Fisher (2011), lack of proof of oil price in direct relation to inflation since the mid-1980s was detailed and a sample period from 1982 to 2008 is inferred.

Also, Chen and Wen (2011) by using data from 1985 to 2011 detected the same results reported by Evans and Fisher (2011) and consequently confirmed that the shocks in the oil price lack any significant effect on the inflation trend, but this is not without a transitory effect through headline or core inflation.

As opined by economists, over a short period and medium period, oil prices are capable of driving some variation in inflation. Since international inflation rates move together according to Neely & Rapach (2011), then international factors which include commodity prices like oil, could have a significant drive on inflation. In the long run, monetary policy according to economist could offset shocks.

To this end, the central bank should be informed about the expected future inflation in association with publicized inflation target to set its policy actions related to the inflation targeting framework (Green 1996). Monetary policy instruments are applied according to this context, to make the inflation forecasts closer to the targets (Khan and Croce 2000).

Inflation targeting regime necessitates the existence of capable forecasting models as well as the transmission mechanisms of monetary policy (Leiderman and Svensson 1995). Non-linear approaches recently have been the center of attention in the literature, in forecasting economic variables including inflation (see for example Binner et al. 2005; Teräsvirta 2006; Chauvet, Lima & Vasquez 2002; Balcilar, Gupta, & Miller 2012).

The majority of the studies have compared the performance of non-linear techniques to that of their linear counterparts in forecasting. Indeed the failure of the linear time series models such as AR and VAR in the detection of non-linear relationships in macroeconomic variables resulted to the increased reliance on non-linear models in macroeconomic forecasting. Moreover, sophisticated time series models are increasingly used as more robust benchmarks in economic forecasting (Marcellino 2007).

The inflation forecasts can be improved with taking the non-linearity into account for the macroeconomic data (Hillman 1998). Some scholars like Ascari and Marrocu (2003) used both linear (NAIRU Phillips curve) and nonlinear (threshold autoregressive) models for forecasting US inflation. They found the outperforming of the non-linear models in capturing the distribution of inflation, while the Phillips curve's forecast performed superior regarding mean-squared forecast errors over the medium to long term horizon.

In another study, Stock and Watson (1998) compared the autoregressions performance with that of exponential smoothing; artificial neural networks; and smooth transition autoregression models in forecasting US macro variables. Their findings proved the better performance of autoregression models with unit root pretests than other models. Also, the authors observed the improvement of the forecast performance when combined with other models' forecasts.

Chauvet, Lima, and Vasquez (2002) compared the linear and non-linear models in the forecasting ability of Brazilian GDP growth. Besides, the authors observed that the structural breaks in the data properly represent the Brazil's business cycle.

In the study of forecasting Euro inflation, Binner et al. (2005) investigated the performance of linear and non-linear models which showed that linear models are restricted in the presence of non-linear events in the data. The authors found that the non-linear (neural networks) model generates better in- and out-of-sample forecasts comparing to the linear (i.e. VAR and ARIMA) ones for Euro inflation.

Balcilar, Gupta, and Miller (2012) compared the out-of-sample performance of linear with non-linear models in the study of forecasting housing prices of US and Census regions. The authors proved that whereas the linear autoregressive model does better in point forecasts at shorter horizons; the non-linear smooth transition autoregressive model beats the linear autoregressive model at longer horizons.

This study attempts to contribute to the literature on forecasting inflation by evaluating the performance of a suite of non-linear models (including, Multi-Layered Perceptron (MLP) Artificial Neural Network (ANN), Locally Linear Model Tree

(LoLiMoT) Neuro-fuzzy Model, Nonlinear Autoregressive (NAR) ANN, and Genetic Algorithm (GA)-based forecasting), as well as a simple forecast combination model (MEAN) in forecasting US inflation. We use the seasonally adjusted month-on-month percentage change in consumer prices over the period 1950:01–2016:05 (the out-of-sample forecasting period being 1990:08–2016:05[13]). In addition to that, root means squared error statistic and the Diebold and Mariano (1995) test have been applied to compare the performance of the different models in forecasting. The starting point in out of sample horizon is defined by considering the bubble in the crude oil market. The logic behind this selection is the high correlation between the US inflation and crude oil prices (0.76) rather than the other US energy market prices (Natural Gas and Coal prices, 0.71 and 0.62 respectively).

In what follows, Section II provides the methodology. Then we reveal the empirical results in the next section (Section III). Finally, Section IV deals with the conclusion.

5.2 Methodology and Data

5.2.1 Forecasting models

The evidence showed that the non-linear models outperform regression models in modeling both linear and non-linear relationships via logistic functions (Bishop 1995). Many good reasons are available to the point that not-linearity can better characterize the data-generating process for inflation.

Factors like policy changes, structural breaks, frequent general price shock, asymmetric business cycles and many others, lead to non-linear dynamics. It would

[13] Since the bubble in energy prices is considered as a main source of inflation‘s variation, the adopted starting point for an out of sample horizon is the bubble on crude oil market due to invasion of Kuwait by Iraq.

be surprising to think that the relationship between the current and future values of economic variables such as inflation would be linear.

Recent studies in which dynamic stochastic general equilibrium (DSGE) models were used in the macro economy, showed the firm predict of DSGE models in non-linear dynamics among the macroeconomic aggregate variables (see, e.g. Milani 2012).

In the presence of nonlinear dynamics in macroeconomic, nonlinear models do not improve forecasting ability. Faust & Wright (2013) indicated that failures of forecasting inflation using linear models made researchers react by arguing that non-linear models be only useful for forecasting in some periods. In the literature, about forecasting inflation rate, only some studies exist that have found a role for parametric non-linear models.

However, there are existing researches which discuss the non-parametric model's usefulness in forecasting just for a small niche of macroeconomic variables (Calhoun and Elliott 2012).

This study considers a suite of machine learning algorithms without any particular parametric form to investigate the usefulness of non-linear models in forecasting inflation. The non-linear behavior of dynamical systems makes it difficult for many of the well-established classical techniques in time series forecasting, thus, compelling them to make assumptions concerning the structure of the underlying stochastic process. These assumptions eventuate in certain parameterizations which may not be supported by the underlying time series.

Accordingly, the need for sharp tools is apparent in forecasting non-linear time series and model the nonlinear relationship between output and input variables without any prior knowledge regarding the data. Recently machine learning models, also called data-driven models, as a class of non-linear models are considered as serious contenders to classical statistical models in forecasting time series (Lachtermacher and Fuller 1995; Teräsvirta, van Dijk, & Medeiros 2005).

These models have proved the ability and capacity of modeling nonlinearity and chaos and only use historical data to learn the stochastic dependency between the past and the future. In forecasting time series data with non-linear dynamics, the machine learning algorithms could achieve prosperity which reflects their capacity of universal function approximation (White 1988, 1989). Their ability in approximating non-linear functions could well demonstrate advantages compared to the parametric models (Hansen, McDonald, and Nelson 1999).

5.2.1.1 Multi-Layered Perceptron (MLP)

The multi-layered perceptron (MLP) network as a good example of the supervised learning network possesses a significant advantage in the management of non-linear functions. The main particles of the MLP are nodes which depend on the properties being modeled. To define the weights of each processing node the error which got through the comparison of actual and desired output is the main standard. An ANN is made up of input and an output layer, and also one or more hidden nodes layers. The predictive ability of ANN is mainly affected by the number of input and hidden nodes.

In the simplest form of computational algorithms weight coefficients minimize the objective function of the network, therefore in each training process weight coefficients are defined as below:

$$w_{k+1} = w_k - a_k g_k \quad (5.1)$$

Where w_k stands for weight coefficient of network, g_k is gradient of network's output error and a_k is learning coefficient of network. Recently in computational stage a different class of algorithms such as Conjugate Gradient, Quasi Newton or modern method of Levenberg-Marquardt are proposed that increase the speed of computation to reach gradient error and at the same time reduce the mass of computation. In this study The MLP design is completed with one hidden layer each with 10 sigmoid neurons and one output layer and the Levenberg–Marquardt method is used for computational stage.

5.2.1.2 Non-Linear Autoregressive Model (NAR)

The following input-output relation denotes a non-linear autoregressive model of order P

$$y(t) = f((y(t-1), \ldots . y(t-b), d(t-1), \ldots . . d(t-b)) \quad (5.2)$$

Where d are the targets for the time series that we want to predict, y shows the past predicted values by the model; b is the output order which is shown the number of previous estimations fed into the model, and f represents a nonlinear function.

Despite the MLP, NAR has a different topology. In this type of network, no adjustment needed for building a loop because it is a recursive network which uses the output as an input in the next step. Similar to MLP, NAR also updates the network training function according to Levenberg Marquardt optimization technique.

To simplify the problem of an outlier for the network the input and target value means are normalized, and they mapped into the interval (-1, 1). In this study, NAR is formed with five lag as feedback delays and with one hidden layer with ten neurons and after that the result of each part of the data put into the loop for finding the 24 steps ahead forecasting, in the form of univariate and recursively.

5.2.1.3 Genetic Algorithm (GA)

The genetic algorithm as a tool for numerical optimization in the forecasting process has been inspired by the process of natural selection. The general Algorithm (G.A) was pioneered by Holand (1975) and explained by Goldberg (1989a). Genetic algorithm (GA) employs H chromosomes $g_{h,t} \in H$, that are binary strings divided into N genes $g_{h,t}^{n}$ each of them encodes a candidate parameter $\theta_{h,t}^{n}$ for the argument θ^{n} when a chromosome h ∈ {1,…H} is at time t ∈ {1,..., T} it can be determined as:

$$g_{h,t} = \{g_{h,t}^{1}, \dots\dots, g_{h,t}^{N}\} \qquad (5.3)$$

As a result of each gene n ∈ {1,…, N}has its length equal to integer Ln and is a string of binary entries (bites).

$$g_{h,t}^{n} = \{g_{h,t}^{n,l}, \dots\dots, g_{h,t}^{n,Ln}\}, g_{h,t}^{n,l} \in \{0,1\} for\ each\ j \in \{1, \dots\dots Ln\} \quad (5.4)$$

Equation 1.4, encodes θ^{n}an integer, since $\sum_{l=0}^{Ln-1} 2^{j} = 2^{Ln} - 1$, therefore easily specific gene of. $g_{h,t}^{n}$ Decodes by:

$$\theta_{h,t}^{n} = \sum_{l=1}^{Ln} \frac{g_{h,t}^{n,l}\, 2^{l-1}}{2^{Ln}-1} \qquad (5.5)$$

Implementation of GA contained some evolutionary operators such as procreation, mutation, and crossover. New generations of chromosomes are generated in crossover step, and the process will be continued until some an old and new generation become equal. In the next step which is called mutation, each binary entry in each chromosome should have a chosen δ_m probability to be swapped and at this stage binary representation reaches its maximum efficiency.

Mutation of bites will lead to new different arguments which are substantially different from original ones. And one chromosome from the old generation can be transformed into the new generation in the procreation stage. Hence, GA can easily evaluate arguments to find the best solution based on the fitness function. In this research, the GA based forecasting designed on five lags and intercept as cost function and with the maximum number of iteration equal to 100. And for procreation of parents the **Roulette-wheel** is used**.**

5.2.1.4 LoLiMoT (Locally Linear Model Tree)

The strategy used in LoLiMoT is a divide and conquer method which is the most important factor for its success. LoLiMoT divides a complex modeling problem into smaller and simpler problems which can be solved by linear models independently, so its approach is dividing the input space into small linear subspaces with fuzzy validity functions.

This particular network is composed of two sets of parameters: nonlinear parameters and narrow parameters. The nonlinear part contains means and deviations of Gaussians invalidity functions. Since the overall model is a fuzzy-neuro network, there are linear and nonlinear parameters belonging to each neuron and these two parts should be multiplied; the whole model is a summation of the multiplications.

The overall prediction equation for the LoLiMoT, which includes a linear and non-linear part, is given by:

$$\hat{y}_t = \sum_{i=1}^{M} \hat{y}_{i,t}\, \emptyset_i(x_t) \qquad (5.6)$$

The linear part of this network is represented by:

$$\hat{y}_{i,t} = \omega_{i0} + \omega_{i1}\, y_{t-1} + \cdots + \omega_{ip}\, y_{t-p} \qquad (5.7)$$

In equation1.6, $\emptyset_i(x_t)$ denotes the nonlinear part $\mu_i(x_t)$ which is calculated from

$$\mu_i(x_t) = \exp\{-\frac{1}{2}\, (\frac{(x_{t-1} - C_{1i})^2}{\sigma_{1i}{}^2} + \frac{(x_{t-2} - C_{2i})^2}{\sigma_{2i}{}^2} + \cdots . + \frac{(x_{t-p} - C_{pi})^2}{\sigma_{pi}{}^2})\} \qquad (5.8)$$

Which can be normalized to obtain $\emptyset_i(x_t)$ as follow:

$$\emptyset_i(x_t) = \frac{\mu_i(x_t)}{\sum_{i=1}^{M} \mu_i(x_t)} \qquad (5.9)$$

The LoLiMoT's training with constructions is its excellent strongest property its algorithms are as follows:

I. Firstly, the model starts with a single locally linear neuron, and the least square technique estimates local optimization of narrow parameters.

II. In this step, it finds the worst neuron according to local loss functions, calculates MSE for each of locally linear neurons, and afterward finds the worst performing neuron.

III. In this step, all divisions checked. The worst neuron refined further. (By checking the error or total cost function for the overall model).

IV. In this step, the best division validated. If it results in a reduction of loss functions on training or validation, then the neuron is updated, and we move back to step 2 and continue again.

The smoothing factor that indicates which area should be used for fitting the Gaussian membership function is equal to 1/3.

Since the LoLiMoT network is more efficient by adding a global feedback state to inputs, feedback is done by adding network's output column to input in regression matrix. In this recurrent model due to adding one column, the network gets more complex, and the least square should be modified.

4.2.1.5 Forecast Combination Method

Special characteristic that makes forecast combination method different from others is the way it uses historical data and information to get a mix of prediction to the point where the weight given to each could change during time. We have some forecast sampling methods like (the simple forecasts MEAN, the discounted DISC, principle component PC method) some of them need a holdout period to compute the weights used to combine the individual model forecasts.

Therefore we have to use the first P_0 observations from out of sample time period as the primary holdout period (Rapach *et.al*. 2010). Mixture of forecasts such as y_{t+h}^{h} created at time t and $y^{\wedge h}_{CB,t+h}|$ t are a linear combination of each forecast models.

$$y^{\wedge h}_{CB,t+h}|\,t = \sum_{i=1}^{n} w_{i.t}\; y^{\wedge h}_{i,t+h}|\,t \qquad (5.10)$$

Where $\sum_{i=1}^{n} w_{i.t} = 1$. And the weights $\{wi.t\}_{i=1}^{n}$ are estimated and individual out of sample forecast and there, are y_{t+h}^{h} observations available from the beginning of holdout out of sample time t. we compute combination forecasts over the post-holdout sample period per combining method, which at the end gives us a total of

$P_h = P - (h-1) - P_0$ combination forecasts $\{ y^{\wedge h}_{CB,t+h} | \, \mathrm{t} \}_{t=R+P_0}^{T-h}$ available for evaluation.

We used the simple combination forecasts calculate combination forecast regardless of historical performance for the individual forecasts. According to Stock and Watson (1999,2003,2004) simple combining methods perform good in forecasting inflation and output growth through applying number of potential predictors.

Later on (2004) they noticed that there's slight difference in performance of the mean and trimmed mean forecast while especially median has higher relative MSFE than both of them. The mean combination forecast includes $W_{i,t} = \frac{1}{n}$ (i=1… n) in the main equation so the simple combining methods do not require a holdout out-of- sample period.

4.2.1.6 The Method of Evaluation of Accuracy of Forecasting Models

Different methods on how to evaluate the accuracy of forecasting models have attracted the attention of many economists in recent decades. In this study, RMSE and DM-test were used for the comparison of the forecasting error. When the error is less, the forecast is better. DM-test can be calculated as follows; suppose that the forecasting error of model j^{th} is $e_{t,j}$ and the function of this case is $g(e_{t,j})$ then the null hypothesis would be:

$$E[g(e_{t,j})] = E[g(e_{t,i})] \quad (5.11)$$

Alternatively, E $[d_t] = 0$ since that $d_t = E[g(e_{t,j})] - E[g(e_{t,i})]$

That means the power of accuracy in both forecast model should be equal (null hypothesis) otherwise one model outperforms ($\mathbb{H}_1$ alternative hypothesis). Suppose that $\{d_t\}_{t=1}^{T}$ is the mean of difference series of the error function. Therefore we have

$$\text{DM}= \frac{\bar{d}}{\sqrt{\frac{2\pi f^{\wedge}{}_{d\,(0)}}{T}}} \quad (5.12)$$

Where $\frac{2\pi f^{\wedge}{}_{d\,(0)}}{T}$ is the variance of $\{d_t\}_{t=1}^{T}$ and $f^{\wedge}{}_{d\,(0)}$ is spectral density of the loss differential at frequency 0. The DM test was proposed by Diebold and Mariano (1995) as a way to contrast the forecasting abilities of two competing models. This test analyzes whether the average loss differences between the two models are significantly different from zero.

5.2.2 Data Description

We measure inflation rates as the seasonally adjusted (using the Census X-12 method) month-on-month growth rates of Consumer Price Index (CPI) data for the US (see Figure 23), obtainable from the DataStream (EMU account). Essentially, our sample covers the period 1950:01–2016:05, i.e. a total of 796 observations. We estimate the models using an in-sample period of 1950:01–1990:08 (i.e. 487 observations), while, the out-of-sample forecasting period is 1990:09–2016:05 (i.e. 309 observations).

Frequently researchers divide the sample into three sets in building ANNs for forecasting time series (Namely; training, validation, and test sets). In this Study data from 1950:01 to 1986:07 used as the training set (440 observations), data from 1986:08 to 1990:01 adopted as the validation set (40 observations), and data from 1990:02 to 2016:05 introduced as the test set (316observations)

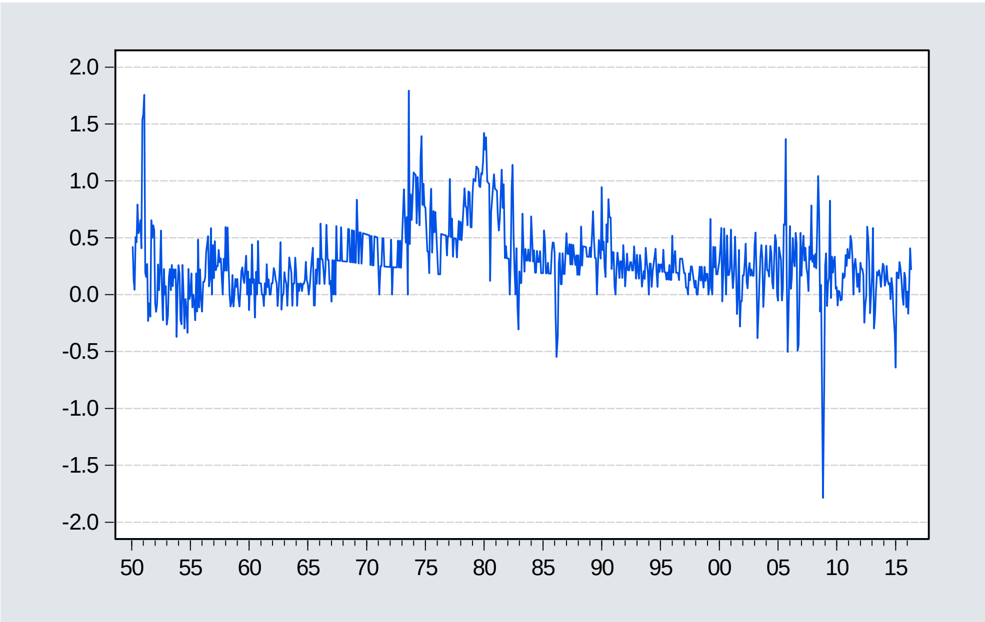


Figure 23. Monthly Inflation Rate (Seasonally Adjusted), US, 1950:01-2016:05

5.3 Empirical Studies

Table 15 shows the RMSE of each model over the forecast horizons h ($h \in \{1,4,12,24\}$) as well as associated ranking. The LoLiMoT model outdoes that of the other models across all forecast horizons as it generates the smallest RMSE in all cases. In very short horizon (h=1), the MLP places on the rank two, while the GA and NAR have the worst forecasting performances.

In middle horizon (h=12), combination (Mean) model has better performance than the NAR and the GA, while the LoLiMoT and MLP still are outperformed by other models. In long horizon ($h=24$), the GA has not captured well the pattern of the dataset which denotes the existence of the complex situation and dynamic phenomenon in the US inflation data. In this horizon, the performance of NAR improved, but the Mean‘s performance became worse due to the imperfect behavior of GA, while again the LoLiMoT and MLP are placed in rank 1 and two respectively. The full findings reveal that the LoLiMoT model‘s performance is outstanding to that of another model.

Table 15. Root Mean Squared Errors and Ranks.

Models	**GA**		**LoLiMoT**		**NAR**		**MLP**		**Mean**	
Step *(h)*	**RMSE**	**Rank**	**RMSE**	**Rank**	**RMSE**	**Rank**	**RMSE**	**Rank**	**RMSE**	**Rank**
1	0.37	4	0.28	1	0.38	5	0.28	2	0.28	3
4	0.48	5	0.28	1	0.40	4	0.29	2	0.29	3
12	1.61	5	0.29	1	0.41	3	0.31	2	0.81	4
24	6.82	5	0.30	1	0.40	3	0.33	2	5.31	4

Table 16 states DM test results of each model over the forecast horizons h ($h \in \{1,4,12,24\}$).The null hypothesis in DM test which is the equality in the performance of two forecast models has been examined, using 5 % level of significance, The negative sign of DM test reports that the first model is better than the second model while the positive sign retells that the second model outperforms by the first example.

At h=1 results state that no model outruns the LoLiMoT and MLP models in forecasting US inflation. At h= 4, 12, 24 LoLiMoT model has better performance in comparison to MLP while the MLP outperformed by NAR. At medium and long horizon except for the LoLiMoT and MLP, the other models ‘performances are relatively equal since the null hypothesis of DM test is not rejected at 5 percent level.

Table 16. DM Test Results, Critical Values =± 1.96 at 5% Level.

Models / **Step *(h)***	**LoLiMoT MLP**	**LoLiMoT GA**	**LoLiMoT NAR**	**MLP GA**	**MLP NAR**	**NAR GA**
1	-3.00*	-1.39	-3.71*	-1.39	-2.25*	-1.39
4	-2.95*	-1.40	-3.79*	-1.40	-2.23*	-1.42
12	-2.85*	-1.43	-3.66*	-1.43	-2.26*	-1.48
24	-3.01*	-1.45	-3.49*	-1.45	-2.17*	-1.52

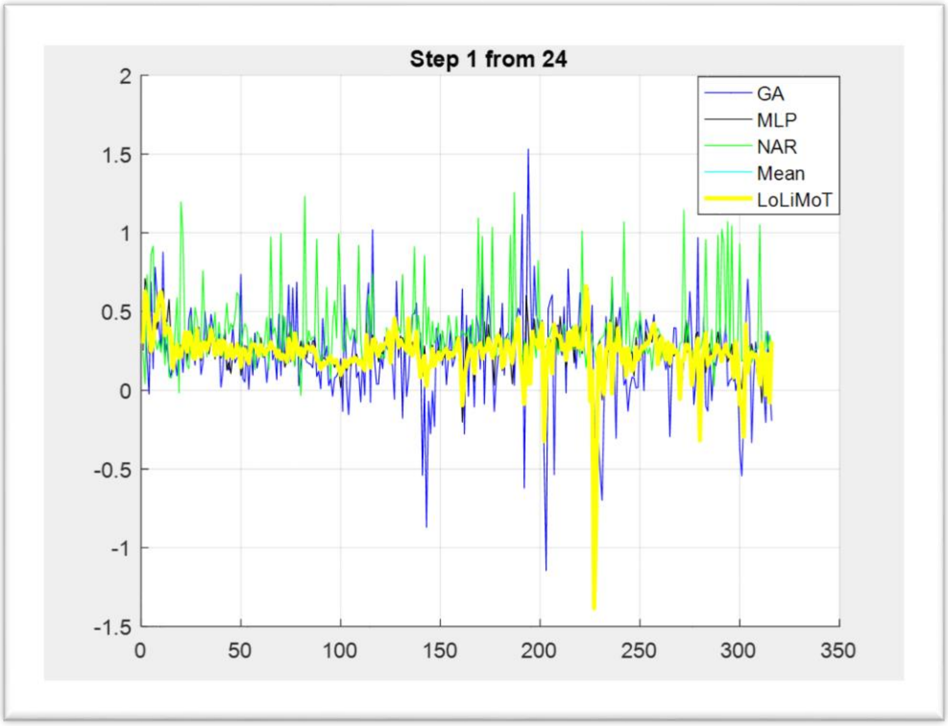


Figure 24. One-Step-Ahead Inflation Rates Forecasts.

5.4 Conclusion

As opined by many studies, energy prices and inflation have a significant connection. Therefore adequate reaction in inflation targeting approach as part of an objective monetary policy to energy market signals should be taken into account. More recently Ben Bernanke [14] (2008) argued that rise in prices of globally traded commodities such as energy prices have been a main determining factor to the inflationary experience of the 2000's former to the financial crises at 2008.

Inflation targeting regime necessitates the existence of capable forecasting models as well as the transmission mechanisms of monetary policy (Leiderman and Svensson 1995). In the literature, non-linear approaches have been recently the center of

[14] In a Speech entitled:" Outstanding Issues in the Analysis of Inflation"; presented at the Federal Reserve Bank of Boston's 53^{th} Annual Economic Conference, Chatham, MA, June 9, 2008; this opinion has been expressed.

attention in forecasting economic variables including inflation (Balcilar, Gupta, & Miller 2012).

This study contributes to the literature on forecasting inflation by evaluating the performance of a suite of non-linear models (including, Multi-Layered Perceptron (MLP) Artificial Neural Network (ANN), Locally Linear Model Tree (LoLiMoT) Neuro-fuzzy Model, Nonlinear Autoregressive (NAR) ANN, and Genetic Algorithm (GA)-based forecasting), as well as a simple forecast combination model (MEAN) in forecasting US inflation.

We use the seasonally adjusted month-on-month percentage change in consumer prices over the period 1950:01–2016:05 (the out-of-sample forecasting period being 1990:08–2016:05). In addition to that, root means squared error statistic and the Diebold and Mariano (1995) test have been applied to compare the performance of the different models in forecasting.

The results reveal that the LoLiMoT model outdoes that of the other models across all forecast horizons. The DM test proves the outcomes at 5% level of significance. All in all, the LoLiMoT and MLP have better performances in forecasting of US inflation. Besides that, selection of 1990:08, which covers the bubble in the crude oil market as the starting point of out of sample horizon, and allows all models to capture and lay out the complex phenomena in the data set better than other counterparts.[15]

[15] We examined two different starting point for out of sample horizon (1990:02, 1989:10), but the results of 1990:08 was outperformed specially in case of GA.

Chapter 5

CONCLUSION

As mentioned in Chapter 1, Energy market lies at the heart of the global economy. Energy as a core of economic and social activity, particularly in developed countries affects all corners of the economic worldwide. Energy costs involve not only companies but also the cost of living of citizens. A practical understanding of behavior in energy market leads to make appropriate policies by policy makers. The energy market in the US as a third large industry plays the significant role in world energy structure. Therefore investigation of all corners of this market is an essential factor to an effective style to stabilize decisions. Therefore, this essay with the aid of different advanced methodologies aims to consider all aspects of US energy markets.

To this end, Chapter 2 investigates the ability of Log Periodic Power Law models to identify bubble (s) and its corresponding termination point (s). Our findings reinforce the fact that energy market prices during bubble periods oscillate with decreasing amplitude around a faster-than-exponential growth. Beside that our results retell that the application of the LPPL model in US energy market are error-free where the actual regime shift date encloses by the confidence interval of termination point.

Chapter 3 targets at the automatic specification of the optimal number of clusters for an unlabeled energy data set through five classes of various evolutionary techniques with utilizing a new representation scheme for the search variables to assign the optimal number of clusters. Our results show that adopting the proposed method to label energy market, reveals two different classes, while generalization of this approach using the iterative algorithm, detect the correct interval for the crude oil bubble at 2008.

Application of some nonlinear methods in electricity forecasting are covered in Chapter 4. Electricity cannot be saved while power system stability depends upon a constant balance between production and consumption. These items give rise to severe price volatility and sudden changes known as spikes. Therefore good forecast needs good model to capture complex behavior in US electricity market. We developed a new machine learning methodology in electricity forecasting namely as GMDH. We found that this class outperformed by the other counterparts in all forecast horizons.

Chapter 5 attempts to contribute to the literature on forecasting inflation by evaluating the performance of a suite of non-linear models. As opined by economists, over a short period and average duration, oil prices are capable of driving some variation in inflation. International factors which include commodity prices like oil could have a significant hit on inflation. Hence, the central bank should be informed about the expected future inflation in association with publicized inflation target to set its policy actions related to the inflation targeting framework. Therefore proper forecasting technique needs to be adapted for reliable forecasts. We introduced a new type of neuro-fuzzy model called as a LoLiMoT on forecasting inflation. We found that, model outdoes that of the other models across all forecast horizons.

REFERENCES

Acemoglu, D., & Robinson, J. A. (2013). Why nations fail*: The origins of power, prosperity, and poverty*: Crown Business.

Ahmadi, M. H., Ahmadi, M.-A., Mehrpooya, M., & Rosen, M. A. (2015). Using GMDH neural networks to model the power and torque of a stirling engine. *Sustainability, 7*(2), 2243-2255.

Ascari, G., & Marrocu, E. (2003). Forecasting inflation: a comparison of linear Phillips curve models and nonlinear time series models.

Axelrod, R. (2010). Governing the Commons: The Evolution of Institutions for Collective Action. *Perspectives on Politics, 8*(2), 580.

Bäck, T., Fogel, D. B., & Michalewicz, Z. (1997). Handbook of evolutionary computation. *New York: Oxford*.

Balcilar, M., Gupta, R., & Miller, S. M. (2015). The out-of-sample forecasting performance of nonlinear models of regional housing prices in the US. *Applied Economics, 47*(22), 2259-2277.

Balcilar, M., Ozdemir, Z. A., & Yetkiner, H. (2014). Are there really bubbles in oil prices? *Physica A: Statistical Mechanics and its Applications, 416*, 631-638.

Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M., & Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. *Applied Economics, 37*(6), 665-680.

Bishop, C. M. (1995). *Neural networks for pattern recognition*: Oxford university press.

Calhoun, G., & Elliott, G. (2012). *Why Do Nonlinear Models Provide Poor Macroeconomic Forecasts?* Paper presented at the Seventh ECB Workshop on Forecasting Techniques-New Directions for Forecasting, Frankfurt am Main, Germany.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods, 3*(1), 1-27.

Chauvet, M., Lima, E. C., & Vasquez, B. (2002). Forecasting Brazilian output in the presence of breaks: A comparison of linear and nonlinear models.

Chou, C.-H., Su, M.-C., & Lai, E. (2004). A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications, 7*(2), 205-220.

Comtois, C., & Slack, B. (2009). *The geography of transport systems*: Routledge.

Croce, E., & Khan, M. S. (2000). Monetary regimes and inflation targeting. *Finance and development, 37*(3), 48.

Das, S., Abraham, A., & Konar, A. (2008). Automatic clustering using an improved differential evolution algorithm. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 38*(1), 218-237.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on pattern analysis and machine intelligence*(2), 224-227.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics, 20*(1), 134-144.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics, 4*(1), 95-104.

Elliott, G., & Timmermann, A. (2013). *Handbook of economic forecasting*: Elsevier.

Evangelou, I. E., Hadjimitsis, D. G., Lazakidou, A. A., & Clayton, C. (2001). *Data mining and knowledge discovery in complex image data using artificial neural networks*. Paper presented at the in Proc. Workshop Complex Reason. Geogr. Data, Paphos.

Evans, C. L., & Fisher, J. D. (2011). What are the implications of rising commodity prices for inflation and monetary policy? *Chicago Fed Letter*(May).

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business, 38*(1), 34-105.

Farlow, S. J. (1984). *Self-organizing methods in modeling: GMDH type algorithms* (Vol. 54): CrC Press.

Fawley, B., & Juvenal, L. (2011). Commodity price gains: speculation vs. fundamentals. *The regional economist, 19*(3), 4-9.

Filimonov, V., & Sornette, D. (2013). A stable and robust calibration scheme of the log-periodic power law model. *Physica A: Statistical Mechanics and its Applications, 392*(17), 3698-3707.

Geem, Z. W., Kim, J. H., & Loganathan, G. (2001). A new heuristic optimization algorithm: harmony search. *Simulation, 76*(2), 60-68.

Geraskin, P., & Fantazzini, D. (2013). Everything you always wanted to know about log-periodic power laws for bubble modeling but were afraid to ask. *The European Journal of Finance, 19*(5), 366-391.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning 'addison-wesley, 1989. *Reading, MA*.

Green, J. H. (1996). Inflation targeting: theory and policy implications. *Staff Papers, 43*(4), 779-795.

Gustavsson, M., & Levén, D. (2015). The Predictability of Speculative Bubbles: An examination of the log-periodic power law model.

Hansen, J. V., McDonald, J. B., & Nelson, R. D. (1999). Time Series Prediction With Genetic-Algorithm Designed Neural Networks: An Empirical Comparison With Modern Statistical Models. *Computational Intelligence, 15*(3), 171-184.

Hillman, R. (1998). Forecasting Inflation with a Non-linear Output Gap Model. *WP99-13*.

Holland, J. (1975). Adaption in natural and artificial systems. *Ann Arbor MI: The University of Michigan Press*.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT press.

Irwin, S. H., & Sanders, D. R. (2010). The impact of index and swap funds on commodity futures markets.

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE transactions on Systems, Man, and Cybernetics, 1*(4), 364-378.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR), 31*(3), 264-323.

Johansen, A., Ledoit, O., & Sornette, D. (2000). Crashes as critical points. *International Journal of Theoretical and Applied Finance, 3*(02), 219-255.

Johansen, A., & Sornette, D. (1999). Predicting financial crashes using discrete scale invariance.

Johansen, A., & Sornette, D. (2002). Endogenous versus exogenous crashes in financial markets.

Kanda, P. T., Balcilar, M., Bahramian, P., & Gupta, R. (2016). Forecasting South African inflation using non-linearmodels: a weighted loss-based evaluation. *Applied Economics, 48*(26), 2412-2427.

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization*. Retrieved from

Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization*. Retrieved from

Karabulut, H., Çınar, C., Oztürk, E., & Yücesu, H. (2010). Torque and power characteristics of a helium charged Stirling engine with a lever controlled displacer driving mechanism. *Renewable Energy, 35*(1), 138-143.

Kelleher, D., Hall, S., & Bradley, K. (2012). The Cost of the Wall Street-Caused Financial Collapse and Ongoing Economic Crisis is More Than $12.8 Trillion. *report by Better Markets, Inc.,(September 15)*.

Kennedy, J. (2011). Particle swarm optimization *Encyclopedia of machine learning* (pp. 760-766): Springer.

Kennedy, J. F., Kennedy, J., Eberhart, R. C., & Shi, Y. (2001). *Swarm intelligence*: Morgan Kaufmann.

Kirschen, D. S., & Strbac, G. (2004). *Fundamentals of power system economics*: John Wiley & Sons.

Knutsen, S., & Sjögren, H. (2009). Institutional Clash and Financial Fragility. An Evolutionary Model of Banking Crises.

Leiderman, L., & Svensson, L. E. (1995). *Inflation targets*: Centre for Economic Policy Research London.

Lemke, F., & Mueller, J.-A. (2003). Medical data analysis using self-organizing data mining technologies. *Systems Analysis Modelling Simulation, 43*(10), 1399-1408.

Leung, Y., Zhang, J.-S., & Xu, Z.-B. (2000). Clustering by scale-space filtering. *IEEE Transactions on pattern analysis and machine intelligence, 22*(12), 1396-1410.

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Paper presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

Marcellino, M. (2007). A comparison of time series models for forecasting GDP growth and inflation. *Bocconi University, Italia*.

Milani, F. (2012). The modeling of expectations in empirical DSGE models: A survey *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments* (pp. 3-38): Emerald Group Publishing Limited.

Minsky, H. (1974). The modeling of financial instability: An introduction Modeling and Simulation Proceedings of the Fifth Annual Pittsburgh Conference 5.

Nariman-Zadeh, N., Darvizeh, A., Felezi, M., & Gharababaei, H. (2002). Polynomial modelling of explosive compaction process of metallic powders using GMDH-type neural networks and singular value decomposition. *Modelling and Simulation in Materials Science and Engineering, 10*(6), 727.

Neely, C. J., & Rapach, D. E. (2011). International comovements in inflation rates and country characteristics. *Journal of International Money and Finance, 30*(7), 1471-1490.

North, D. (2005). Understanding the process of economic change. *SPE, Storia del pensiero economico. Fascicolo 2, 2005*(2), 1000-1004.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies, 23*(2), 821-862.

Rocca, P., Oliveri, G., & Massa, A. (2011). Differential evolution as applied to electromagnetics. *IEEE Antennas and Propagation Magazine, 53*(1), 38-49.

Sek, S. K., Teo, X. Q., & Wong, Y. N. (2015). A Comparative Study on the Effects of Oil Price Changes on Inflation. *Procedia Economics and Finance, 26*, 630-636.

Selim, S. Z., & Alsultan, K. (1991). A simulated annealing algorithm for the clustering problem. *Pattern recognition, 24*(10), 1003-1008.

Shiller, R. J. (1987). Investor behavior in the October 1987 stock market crash: Survey evidence: National Bureau of Economic Research Cambridge, Mass., USA.

Smith, V. L., King, R., Williams, A. W., & Van Boening, M. (1993). The robustness of bubbles and crashes in experimental stock markets.

Sornette, D., Johansen, A., & Bouchaud, J.-P. (1996). Stock market crashes, precursors and replicas. *Journal de Physique I, 6*(1), 167-175.

Sornette, D., Woodard, R., Yan, W., & Zhou, W.-X. (2013). Clarifications to questions and criticisms on the Johansen–Ledoit–Sornette financial bubble model. *Physica A: Statistical Mechanics and its Applications, 392*(19), 4417-4428.

Sornette, D., Woodard, R., & Zhou, W.-X. (2009). The 2006–2008 oil bubble: Evidence of speculation, and prediction. *Physica A: Statistical Mechanics and its Applications, 388*(8), 1571-1576.

Stiglitz, J. E. (2010). *Freefall: America, free markets, and the sinking of the world economy*: WW Norton & Company.

Stock, J. H., & Watson, M. W. (1998). *A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series*. Retrieved from

Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics, 44*(2), 293-335.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature, 41*(3), 788-829.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*(6), 405-430.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization, 11*(4), 341-359.

Teräsvirta, T. (2006). Forecasting economic variables with nonlinear models. *Handbook of economic forecasting, 1*, 413-457.

Teräsvirta, T., Van Dijk, D., & Medeiros, M. C. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting, 21*(4), 755-774.

Trehan, B. (2005). Oil price shocks and inflation. *FRBSF Economic Letter*.

Vasechkina, E., & Yarin, V. (2001). Evolving polynomial neural network by means of genetic algorithm: some application examples. *Complexity international, 9*, 1-13.

White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns.

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association, 84*(408), 1003-1013.

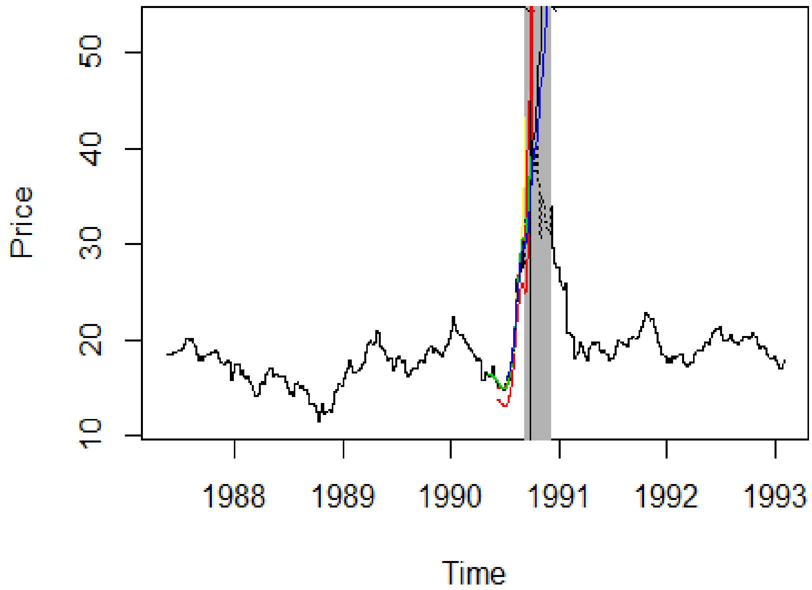
Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE, 87*(9), 1423-1447.

Zhou, W.-X., & Sornette, D. (2002). Generalized q analysis of log-periodicity: Applications to critical ruptures. *Physical Review E, 66*(4), 046111.

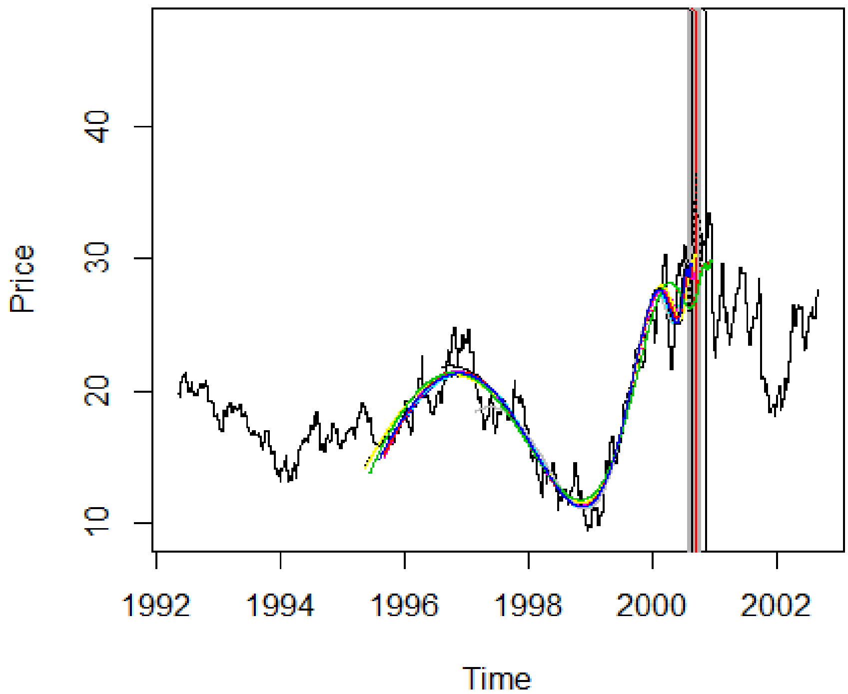
APPENDICES

Appendix A

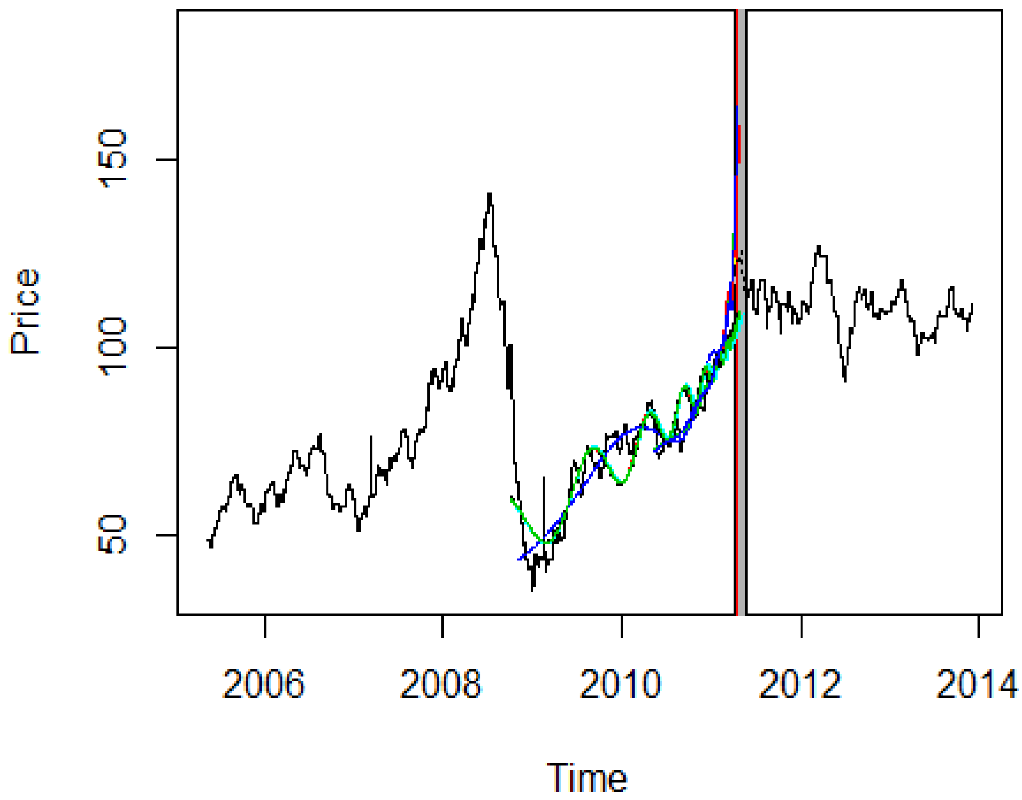
The LPPL estimation's Plots for the US crude oil market.



LPPL Estimation for Crude Oil in 1999.



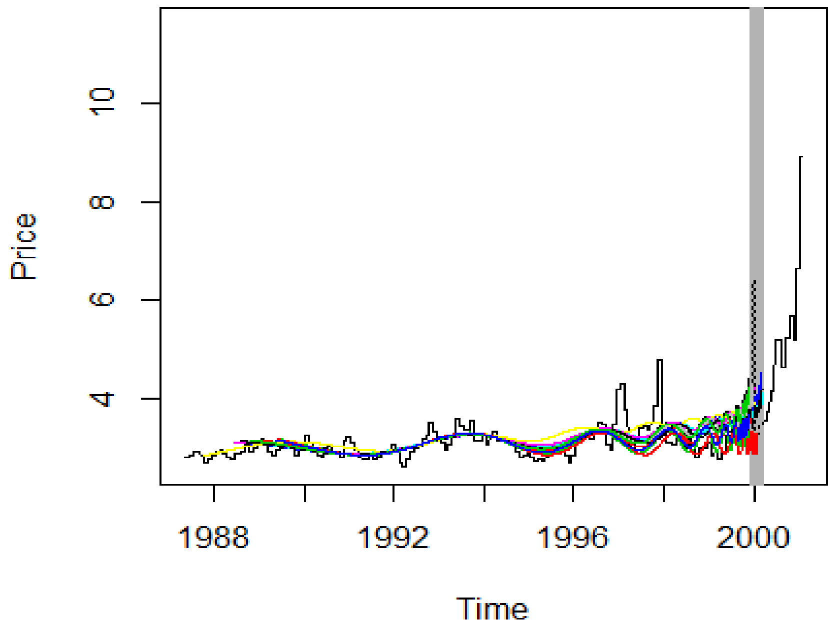
LPPL Estimation for Crude Oil in 2000.



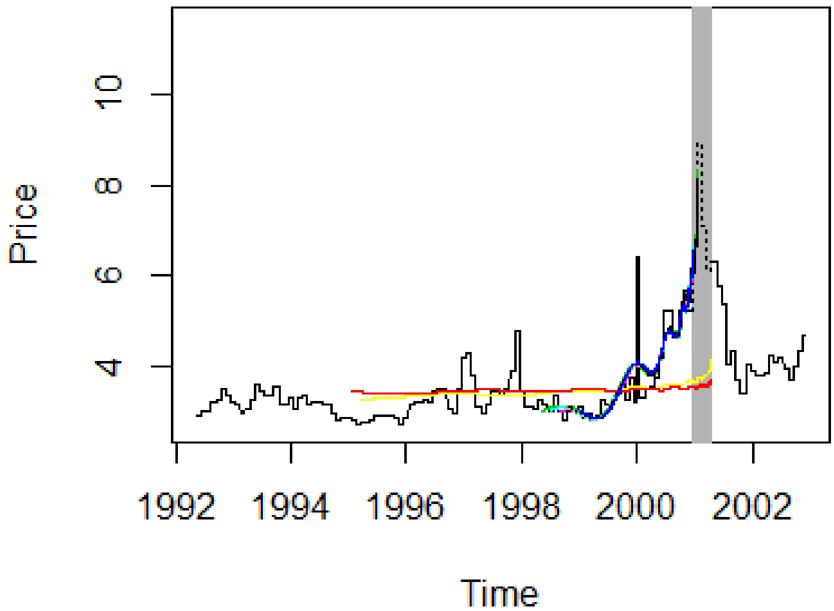
LPPL Estimation for Crude Oil in 2011.

Appendix B

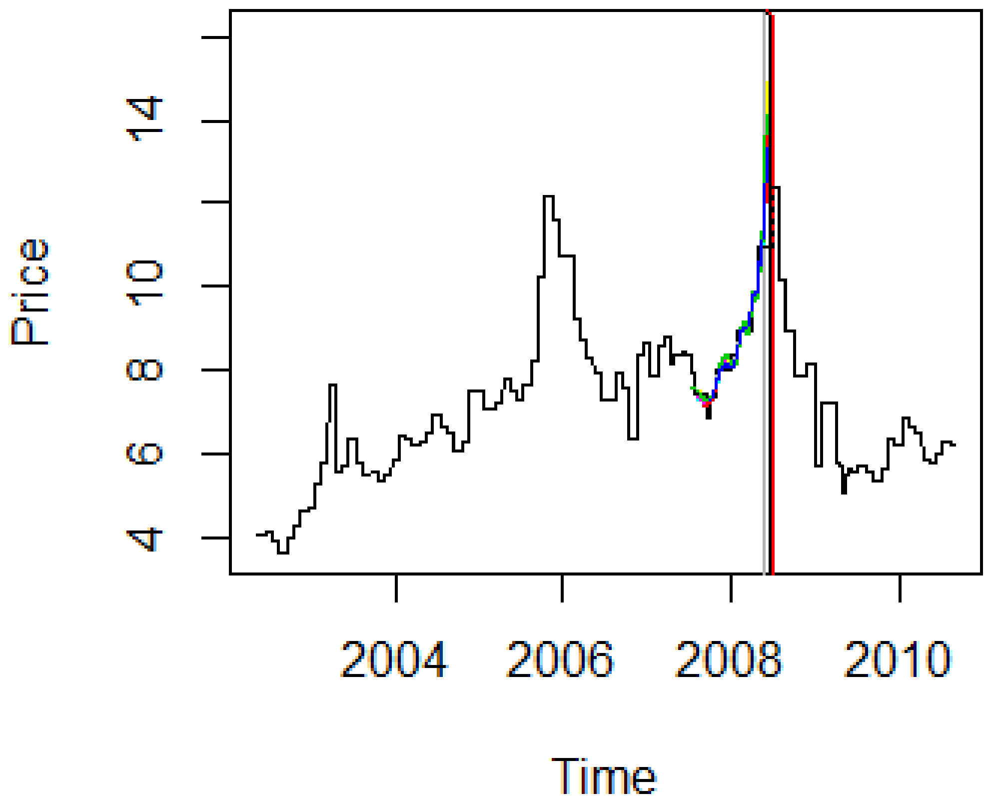
The LPPL estimation’s Plots for the US Natural Gas market.



LPPL Estimation for Natural Gas in 1999.



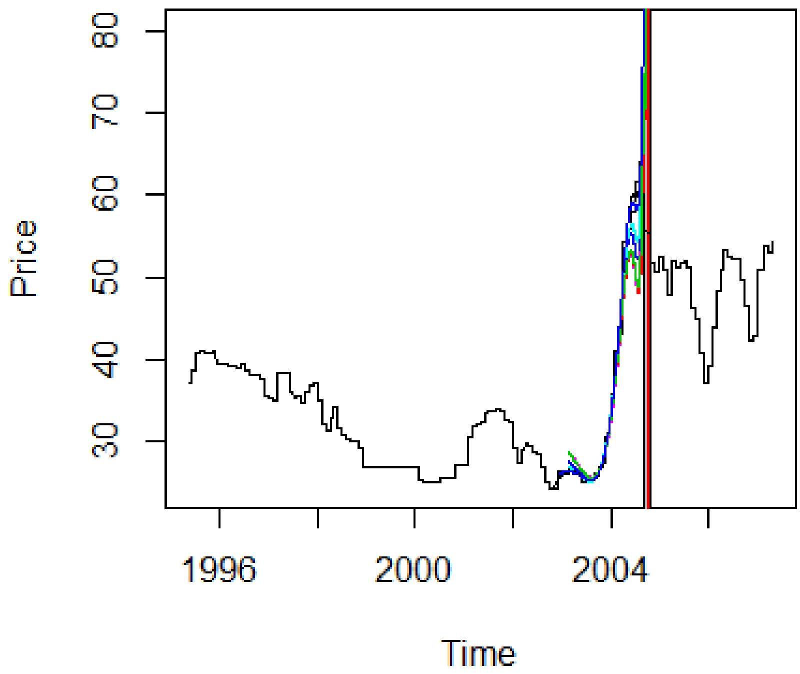
LPPL Estimation for Natural Gas in 2001



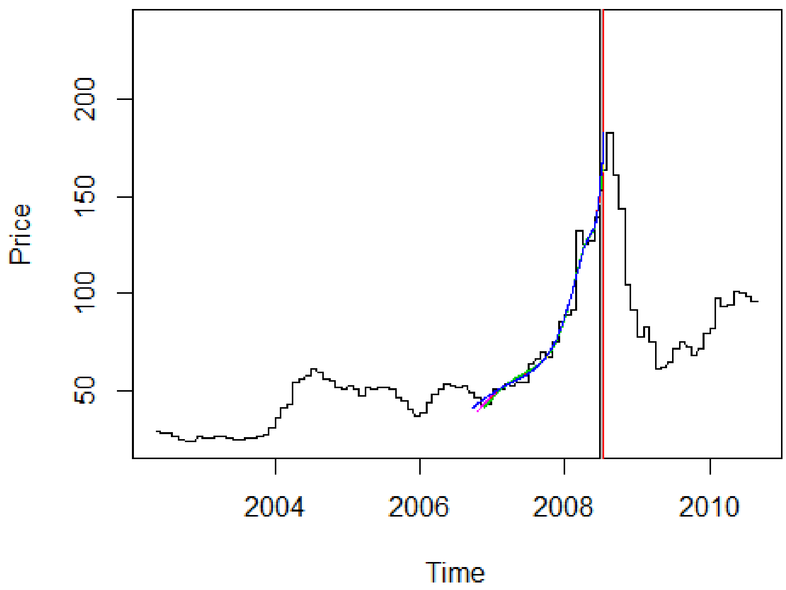
LPPL Estimation for Natural Gas in 2008.

Appendix C

The LPPL estimation's Plots for the US Coal market.



LPPL Estimation for Coal in 2004.



LPPL Estimation for Natural Gas in 2008.