Journal of Natural Gas Science and Engineering 14 (2013) 238–249

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: www.elsevier.com/locate/jngse

Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis

Narges Salehnia [a,*], Mohammad Ali Falahi [b,1], Ahmad Seifi [b,1], Mohammad Hossein Mahdavi Adeli [b,1]

[a] Energy Economics, Ferdowsi University of Mashhad (FUM), #140, No. 11 Soroosh St., Vakil Abad Avenue, Mashhad, Khorasan Razavi 9188875637, Iran
[b] Economics Department, Ferdowsi University of Mashhad (FUM), Iran

A R T I C L E I N F O

Article history:
Received 4 January 2013
Received in revised form
8 July 2013
Accepted 11 July 2013
Available online 9 August 2013

JEL classification:
Q42
Q47
C14
C45

Keywords:
Natural gas
Spot price forecasting
Gamma test
Nonparametric nonlinear model

A B S T R A C T

Developing models for accurate natural gas spot price forecasting is critical because these forecasts are useful in determining a whole range of regulatory decisions covering both supply and demand of natural gas or for market participants. A price forecasting modeler needs to use trial and error to build mathematical models (such as ANN) for different input combinations. This is very time consuming since the modeler needs to calibrate and test different model structures with all the likely input combinations. In addition, there is no guidance about how many data points should be used in the calibration and what accuracy the best model is able to achieve. In this study, the Gamma test has been used for the first time as a mathematically nonparametric nonlinear smooth modeling tool to choose the best input combination before calibrating and testing models. Then, several nonlinear models have been developed efficiently with the aid of the Gamma test, including regression models; Local Linear Regression (LLR), Dynamic Local Linear Regression (DLLR) and Artificial Neural Networks (ANN) models. We used daily, weekly and monthly spot prices in Henry Hub from Jan 7, 1997 to Mar 20, 2012 for modeling and forecasting. Comparison of the results of regression models show that DLLR model yields higher correlation coefficient and lower MSError than LLR and will make steadily better predictions. The calibrated ANN models show the shorter the period of forecasting, the more accurate results will be. Therefore, the forecasting model of daily spot prices with ANN can provide an accurate view. Moreover, the ANN models have superior performance compared with LLR and DLLR. Although ANN models present a close up view and a high accuracy of natural gas spot price trend forecasting in different timescales, their ability in forecasting price shocks of the market is not notable.

1. Introduction

Natural gas is the cleanest-burning fossil fuel, producing significantly lower carbon emissions than coal or oil, as well as lower levels of other pollutants. Around one-fifth of the globe's energy needs is met from natural gas, compared with one-third from oil and one-fourth from coal. Therefore, developing models for accurate natural gas price forecasting and direction of price changes is critical because these forecasts are commonly used in determining a range of regulatory decisions covering both supply and demand of natural gas or for market participants; they are also often a crucial variable in electric generation capacity planning and in analyzing the benefit-cost relationship for demand-side and energy-efficiency programs.

In general, forecasting has been very important in decision making at all levels and sectors of the economy. In energy sector, where the decision environment is characterized by risks and uncertainty, decision makers require some information about the possible future outcomes.

Having accurate forecasts of spot prices one, three, five, or more months into the future is vital for profitable production decisions, purchases, and planning; there are many studies and multiple methods have been conducted so far in connection with the price forecasting. An appropriate method may be chosen based on the nature of available data, desired nature, and level of details in the estimation. A few studies worked in this area, including natural gas demand estimation (Azadeh et al., 2010), natural gas price in domestic sector (Dudek et al., 2006), and the trend of long-term price of

* Corresponding author. Tel./fax: +98 5116058095, +98 9151170016.
E-mail addresses: salehnian@gmail.com, salehnia_n@yahoo.com (N. Salehnia), falahi@um.ac.ir (M.A. Falahi), spring05@um.ac.ir (A. Seifi), mh-mahdavi@um.ac.ir (M.H. Mahdavi Adeli).
[1] Tel./fax: +98 5118813090.

1875-5100/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jngse.2013.07.002

gas (MacAvoy and Moshkin, 2000). Woo et al. (2006) evaluated the natural gas price trend in California utilizing a partial adjustment regression model and found a relation. Also, Serletis and Rangel-Ruiz (2004) examined the interconnectedness of North American natural gas markets using only two daily spot markets prices, U.S. Henry Hub and AECO Alberta. They conclude that since deregulation, North American natural gas prices are largely defined by Henry Hub price trends. Notably, several studies have reported that the NYMEX natural gas futures prices are downward biased in forecasting the subsequently realized spot price and have interpreted this bias as representing a risk premium (Walls, 1995; Modjtahedi and Movassagh, 2005; Movassagh and Modjtahedi, 2005). Moreover, Al-Fattah and Startzman (2001) used ANN model to forecast US natural gas supply to the year 2020. The input parameters included data of gas exploratory wells, oil/gas exploratory wells, oil exploratory wells, gas depletion rate, proved reserves, gas wellhead prices, and growth rate of gross domestic product. Evaluation of parameters' influence on the US natural gas production was performed using fuzzy combinatorial analysis by Garcia and Mohaghegh (2004). Their input variables for the neural model were US natural gas production from previous year, GDP, population, average depth of oil and gas wells, and annual gas depletion rate. They showed that US gas production is not a crisp number for each year rather it is a range that includes a minimum, a maximum and a most likely value.

In energy markets, a wide range of bottom-up models that include supply/demand fundamentals is also available (see, e.g., Fleten and Lemming, 2003; Kumbarouglu and Madlener, 2003; Martinsen et al., 2003). While these models may be used more by practitioners, time series models require access only to market prices, which are more readily available than bottom-up data. Forecasting inter-related energy product prices using both linear and nonlinear techniques are presented in Malliarisa and Malliarisb (2008). Results show that, for crude oil, heating oil, gasoline and natural gas, the nonlinear forecasts were best, while for propane, the linear model gave the lowest error.

So, there have been abundant studies on analysis and forecasting of energy commodity prices. The approaches can be grouped into two categories: structural models and data-driven methods. Standard structural models outline the world oil market and then analyze the oil price volatility in terms of a supply-demand equilibrium schedule (e.g. Bacon, 1991; Al Faris, 1991; Huntington, 1994; Zamani, 2004; Yang et al., 2002). Data-driven models include linear models such as Autoregressive Moving Average (ARMA), Autoregressive Conditional Heteroscedasticity (ARCH) type models (e.g. Sadorsky, 2002; Morana, 2001; Buchananan et al., 2001 and etc.), and nonlinear models such as Artificial Neural Network (e.g. Mirmirani and Li, 2004; Moshiri, 2004; Nelson et al., 1994; Yu et al., 2006), Support Vector Regression (Xie et al., 2006), etc.

Despite an abundance of studies on prediction and modeling of forecasting commodity prices using classical time series models, financial models and nonlinear techniques like artificial neural networks, there are still many questions that need to be answered. For example, how many data points are required to make a prediction with a best possible accuracy? Which inputs are relevant in making the prediction and which are irrelevant?

Moreover, the literature on natural gas price forecasting has focused on two main classes of linear, single-equation, reduced-form econometric models. The first group ("financial" models) includes models which are directly inspired by financial economic theory and are based on the market efficiency hypothesis (MEH), while models belonging to the second class ("structural" models) consider the effects of natural gas market agents and real variables on natural gas prices. Both financial and structural models often use pure time series specifications for benchmarking. However, in this paper, due to the advancement of modern computing technology and a recent algorithm from the computing science community we want to present the third group which is a nonparametric method called the Gamma test (GT) (Stefánsson et al., 1997; Koncar, 1997; Tsui, 1999; Guedes de Oliveira, 1999).

Nonparametric methods for forecasting in time series can be viewed, up to a certain extent, as a particular case of nonparametric regression estimation under dependence. Some significant papers in this field are those by Gyorfi et al. (1989),Hardle and Vieu (1992), Hart (1991), Masry and Tjostheim (1995), Hart (1996), Hardle et al. (1997, 1998) and Bosq (1998), among many others.

A formal proof for the Gamma test can be found in Evans (2002), Remesan et al. (2008), and Kemp (2006). This novel technique would help modelers to determine the best model input combinations of data time intervals to achieve a particular target output and efficiently calculate directly from the data an estimate, the Gamma statistic, of the best mean-squared error on an output with a smooth model. By examining the Gamma statistic for different selections of natural gas price data time intervals, we choose that selection which minimizes the expected mean squared error. It is this application that the present paper is designed to illustrate using time series data. Moreover, The Gamma test is also designed to solve the overtraining problem associated with almost all nonlinear modeling techniques including ANN, efficiently by giving an estimate of how closely any smooth model could fit the unseen data (Remesan et al., 2008).

To model time series data, we need to construct the model by choosing the past values of natural gas spot prices, up to some number m (often called the embedding dimension) to specify the inputs of the model. The output is then the current value of the time series. Thus, an embedding of a time series is a selection of past values which are used to predict the current value via a model constructed from the data. A regular embedding takes all past values up to m.

Regular embedding was first applied for dynamical system modeling by Takens (1981). An irregular embedding chooses some subset of the m past values, and there are $2^m - 1$ possible irregular embeddings once m is chosen. It was suggested by Judd and Mees (1998) that irregular embeddings may often provide a better model. Therefore, the choice of a suitable irregular embedding can be critical for model-based time series prediction. Clearly, as mentioned, one of the key problems in forming a smooth model from time series data is the determination of which lagged variables are relevant in predicting a given output. Tsui et al. (2002) showed the best input combination of lagged variables from a smooth model of an output with Gamma test capability as a nonparametric method. They conclude that the Gamma test is an effective tool in the determination of irregular time series embeddings. Moreover, forecasting a seasonal time series for El-Nino from 1950 to 2008 with a nonparametric method is presented by Shang and Hyndman (2009). They propose four dynamic updating methods to improve point forecast accuracy.

In this paper, we illustrate the use of the Gamma test in selecting irregular embeddings for time series data and the goal is to form a good predictive model for natural gas spot price time series. We consider here three modeling techniques, local linear regression (LLR), dynamic local linear regression (DLLR) and artificial neural networks (ANNs) trained using Matlab Software because it is of a high flexibility in neurons and layers variations.

2. Methods and materials

2.1. Study area and data used

Well-developed day ahead, weekly and monthly spot markets for natural gas have thrived since the early 1990s. The U.S. has