Energy Economics 32 (2010) 1507–1519

Contents lists available at ScienceDirect

Energy Economics

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

Forecasting oil price trends using wavelets and hidden Markov models

Edmundo G. de Souza e Silva [a,*], Luiz F.L. Legey [a], Edmundo A. de Souza e Silva [b]

[a] Energy Planing Department, Federal University of Rio de Janeiro/COPPE P.O. Box 68565, 21941-972, Rio de Janeiro, RJ, Brazil
[b] System Engineering and Computer Science Department, Federal University of Rio de Janeiro/COPPE, P.O. Box 68511, 21941-972, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Article history:
Received 9 January 2010
Received in revised form 23 June 2010
Accepted 18 August 2010
Available online 25 August 2010

Keywords:
Oil price trends
Forecasting
Hidden Markov model
Wavelets

ABSTRACT

The crude oil price is influenced by a great number of factors, most of which interact in very complex ways. For this reason, forecasting it through a fundamentalist approach is a difficult task. An alternative is to use time series methodologies, with which the price's past behavior is conveniently analyzed, and used to predict future movements. In this paper, we investigate the usefulness of a nonlinear time series model, known as hidden Markov model (HMM), to predict future crude oil price movements. Using an HMM, we develop a forecasting methodology that consists of, basically, three steps. First, we employ wavelet analysis to remove high frequency price movements, which can be assumed as noise. Then, the HMM is used to forecast the probability distribution of the price return accumulated over the next F days. Finally, from this distribution, we infer future price trends. Our results indicate that the proposed methodology might be a useful decision support tool for agents participating in the crude oil market.

1. Introduction

The crude oil price is determined by its supply and demand balance. Despite being apparently simple, this equilibrium is quite complex, due to the great number, and complexity, of interacting factors that can affect it. Wars and geopolitical tensions; economic growth; discovery of new oil reserves; development of new energy sources; weather conditions, are just a few examples of the elements that can alter this balance.

Since a great number of these elements are random in nature and, usually, are very hard to predict, it is difficult for market participants to determine an adequate price for the crude oil. This, in turn, leads to the big price fluctuations frequently observed in both short and medium terms. For example, from July to August of 2004, the price of the West Texas Intermediate (WTI) barrel rose from US$40, to nearly US$45, which represents, approximately, a 10% rise in less than one month.

This large volatility, intrinsic to the crude oil price, is very harmful for both producers and consumers, because investment decisions are, consequently, much harder to make. Therefore, the development of good forecasting models for oil prices is of great importance. There are, essentially, two ways through which this can be accomplished. The first, known as prediction by fundamentals, consists in identifying the main factors that influence prices, determining how the variation of each one affects them, and, finally, building a cause and effect model, such as a regression model. As mentioned before, the main problem with this approach is the large number of elements to be considered, and the difficulty to establish and understand the relationships between them.

The most commonly used explanatory variables in fundamentalist regression models usually attempt to describe market demand and the role of OPEC in the oil market, by measuring OPEC's supply. In (Ye et al., 2002, 2005, 2006), for example, the authors suggest that the petroleum inventory levels maintained by OECD countries may be used to predict future crude oil prices. Along those lines, they build a regression model, which describes the inventory–price relation, and use it to forecast WTI prices. The main drawback of this model is that it relies on accurate inventory-level forecasts to make its predictions. Also, Merino and Ortiz (2005) have found that this inventory–price correlation has recently weakened, as the premium sustained by real prices, against those forecasted by the model, shows. In Kaufman et al. (2004) the authors associate oil prices to OPEC's production policies (such as production quotas) and to OECD's stocks and demand. Results confirm that OPEC decisions are indeed important price drivers. However, the main problem in using such a model in oil price forecasting is the need to predict OPEC's conduct accurately, as well as future oil supply and demand, which can be very difficult. A more thorough discussion on these and other fundamentalist models can be found on Frey et al. (2009).

An alternative way of forecasting crude oil prices is to use time series models, in which future price behavior is inferred from its own historical data. These models are mostly employed when *(i)* the data exhibits some systematic pattern, such as autocorrelation; *(ii)* the overwhelming number of possible explanatory variables, and their interactions, render a structural model which is very difficult to implement; or *(iii)* the forecast depends on the prediction of the

* Corresponding author. Tel.: +55 21 2562 8788; fax: +55 21 2562 8789.
E-mail addresses: edmundosilva@gmail.com (E.G. de Souza e Silva), legey@ppe.ufrj.br (L.F.L. Legey), edmundo@land.ufrj.br (E.A. de Souza e Silva).

0140-9883/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.eneco.2010.08.006

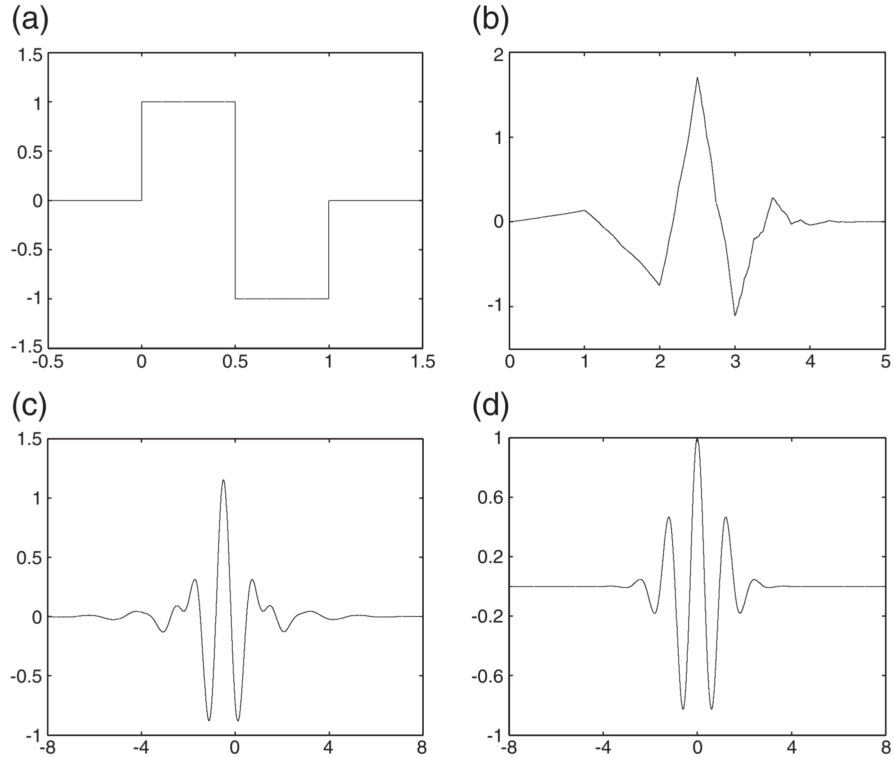

Fig. 1. Some wavelet examples: (a) Haar wavelet, (b) Daubechie wavelet (of order 3), (c) Meyer wavelet, and (d) Morlet wavelet.

explanatory variables, which may be even more difficult than forecasting the variable itself. This seems to be the case of oil prices. Actually, the number and nature of the variables influencing them — which encompasses economical, political and physical aspects — indicate that it may be suitable to check for any systematic pattern in the data series, and try to use it to predict the series' future values, rather than attempting to build a structural model.

In fact, several authors have used time series models to forecast oil prices. For example, Xie et al. (2006) experimented with the linear time series model ARIMA to forecast WTI prices, but concluded against it, arguing that oil prices exhibit nonlinear behavior that cannot be captured by a linear model. Other attempts to use linear time series models have come to controversial results, as can be seen in Frey et al. (2009).

More recently, artificial intelligence techniques, such as artificial neural networks (ANN) and nonlinear time series models, such as hidden Markov models, have been applied to financial data forecasting. In (Wang et al., 2004; Xie et al., 2006; Shambora and Rossiter, 2007) ANNs were applied to the oil price forecasting problem. Their results indicate that these models may have good predictive capabilities. However, they mostly focus on the model's forecasting success rate (which is defined by the number of times the model correctly predicts price direction movements, over the total number of predictions made), and not on the investor's financial return. The later constitutes a particularly important measure, because it might well be that the model is only correctly forecasting small price movements, but missing important, bigger and more profitable ones. Additionally, their results are based on forecasting exercises done over a small period of time — usually from 3 to 6 years — thus making it hard to evaluate whether the models would perform equally well over longer time periods. Finally, they all forecast only a single price value, which makes it impossible analyze the risk of any decision based on the model's forecast. In (Shi and Weigend, 1997; Zhang, 2004), HMMs were used to predict the Down Jones Industrial Average stock index, with results showing an even greater financial data predictive capability then ANNs.

In this paper, we investigate the performance of an HMM in predicting future crude oil price movements. Unlike the previous approaches presented in the aforementioned papers, we aim at forecasting medium-term price movements (roughly 20 to 30 days) instead of short-term (1 day) movements, because investors, producers and consumers are usually more concerned with (and affected by) medium and long-term oil price trends, than their short-term daily variations. Additionally, and more important, our model outputs a complete distribution, which allows investors to, for example, analyze the risk of using a particular forecasted value, and create different strategies which suit his/her needs or profile. To this end, we develop a forecasting methodology that consists of, basically, three steps. First, we employ wavelet analysis to remove high frequency price movements, which can be assumed as noise. Then, the HMM we propose is used to forecast the probability distribution of the price return accumulated over the next F days. Finally, from this distribution, we infer future price trends. Our results indicate that the proposed methodology might be a useful decision support tool for agents participating in the crude oil market.

The remainder of this paper is organized as follows. In Section 2 we introduce the necessary mathematical background. In Section 3, we describe our forecasting methodology; and in Section 4 we show selected results. Finally, our concluding remarks are presented in Section 5.

2. Mathematical background

2.1. Wavelets

Wavelets are undulatory mathematical functions, with limited duration, whose time integral equals to zero. Fig. 1 shows some examples of commonly used wavelets. Like the sine and cosine functions in the Fourier analysis, wavelets are used as a base to represent functions in $\mathbf{L}^2(\mathbb{R})$[1]. Today, wavelets are applied in several different fields, such as data compression; image and signal

[1] $\mathbf{L}^2(\mathbb{R})$ denotes the vector space of measurable, square-integrable one-dimensional functions $f(x)$. A function is said to be square-integrable if $\int_{-\infty}^{\infty}|f(x)|^2 dx$ is finite.