Crude oil Prediction

Literature review

We divide crude oil price forecasting approaches into three categories: 1) heuristic approaches; 2) econometric models; and 3) machine learning techniques. Heuristic approaches for oil price prediction include professional and survey forecasts, which are mainly based on professional knowledge, judgments, opinion and intuition. Another heuristic approach, the so-called no-change forecast, uses the current price of oil as the best prediction of future oil prices. Despite its simplicity, the no-change forecast appeared to be a good baseline approach for oil price prediction and was better than other heuristic judgmental approaches (Alquist et al., 2013).

Econometric models are the most widely used approaches for oil price prediction, which include autoregressive moving average (ARMA) models and vector autoregressive (VAR) models, with possibly different input variables (Pindyck, 1999; Frey et al., 2009). These econometric models provide more accurate prediction than the no-change model at least at some horizons (Alguist et al., 2013; Baumeister and Kilian, 2015). Recently, a forecast combination approach was proposed by Baumeister and Kilian (2015), which combines 6 different oil price prediction models including both econometric models (such as the VAR model) and the no-change model. It should be noted that most of the econometric models are linear models and are not be able to capture the nonlinearity of oil prices.

Several machine learning techniques were proposed for oil price prediction, such as artificial neural networks (ANN) (Yu et al., 2008; Kulkarni and Haidar, 2009), and support vector machine (SVM) (Xie et al., 2006). These are nonlinear models which may produce more accurate predictions if the oil price data are strongly nonlinear (Behmiri and PiresManso, 2013). However, these machine learning techniques, like other traditional machine learning techniques, rely on a fixed set of training data to train a machine learning model and then apply the model to a test set. Such an approach works well if the training data and the test data are generated from a stationary process, but may not be effective for non-stationary time series data a such as oil price data.