Global Oil Export Destination Prediction: A Machine Learning Approach

Haiying Jia,[a] Roar Adland,[b] and Yuchen Wang[c]

The geographical separation of oil consuming and producing nations means that oil needs to be transported at great distances, with forty percent of the annual global oil production transported via the oceans in specialized oil tankers. The objective of this paper is to predict the destination of oil exports at the micro level in a data-driven framework by utilizing actual oil shipment information and training machine learning algorithms based on supervised classification techniques. We contribute to the academic literature by providing the first machine learning application to oil shipment data, and by providing new knowledge on the determinants of global crude oil flows.

Based on crude oil shipment data for the period January 2013 through mid-March 2016, we investigate how destinations are determined based on four attribute clusters: cargo information (such as sellers' identity, cargo grade and cargo size), vessel information (such as vessel identity and its technical specifications), geographic information (load terminals and ports), and macroeconomic data (e.g. regional oil prices and crack spreads). We train the machine learning algorithm based on historical data and demonstrate the models' out-of-sample accuracy.

The results show that micro-level attributes of the oil shipment such as quality and cargo size dominate in the destination prediction. The machine-learning models used to predict the importing country can reach an accuracy of above 71% for the major oil exporting countries based on out-of-sample tests and outperform both naïve models and discrete regression models.

Our research is an important building block in commercial applications that deal with oil and freight market analysis. For instance, the public destination information in ship tracking data is known to be of low quality and can be easily manipulated. Accordingly, analysts that want to track cargoes as a proxy for economic activity or to estimate short-term regional supply of crude oil need a tool to benchmark such information against the likely outcome predicted from past trading patterns and micro data. Importantly, our work suggests that micro data is substantially more valuable for predictive oil trade models than observable macroeconomic data such as crack spreads and oil prices.

a Center for Applied Research (SNF) and Department of Business and Management Science, Norwegian School of Economics, Bergen, Norway. E-mail: haiying.jia@nhh.no

b Corresponding author. Department of Business and Management Science, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway. E-mail: roar.adland@nhh.no

c Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, USA. E-mail: yuchenw@mit.edu

The Energy Journal, Vol. 42, No. 4