OIL SPOT PRICE FORECASTING WITH VARIABLE UNCERTAINTY: A REVIEW OF SOME NOVEL METHODS

Krzysztof Drachal, Faculty of Economic Sciences, University of Warsaw, +48 22 55 49 111, kdrachal@wne.uw.edu.pl

Overview

This research is focused on applying various forecasting econometric models to crude oil spot prices. In particular, the applied models are those able to deal with variable (feature) selection issues. For instance, the novel approach to symbolic regression with Bayesian symbolic trees (instead of conventionally used genetic programming methods) is examined (Jin et al., 2019; Koza, 1992). Indeed, numerous studies report that there can be numerous important crude oil price drivers. Moreover, in different time periods different drivers can play the main role. In particular, suppy and demand factors, stocks quotas, stock market indices, a market stress index, interest rates, the Kilian global economic activity index and exchange rates are used as potentially important oil price drivers (Manickavasagam, 2020; Su et al., 2020; Miao et al, 2017). Monthly data beginning in 1989 and ending in 2021 are used. Several models dealing with variable uncertainty are estimated: LASSO, RIDGE, elastic net, least-angle regression, Dynamic Model Averaging, Bayesian Model Averaging, etc. Additionally, time-varying parameters regression, ARIMA and no-change forecasts are estimated. Forecasts accuracies are measured by Root Mean Square Error, Mean Absolute Error and Mean Absolute Scaled Error (Hyndman and Koehler, 2006). Forecasts performance is additionally examined with the Diebold-Mariano test (Diebold and Mariano, 1995) and the Model Confidence Set testing procedure (Hansen et al., 2011). There is a strong evidence in favour of treating Dynamic Model Averaging and ARIMA as superior forecasting models. However, the novel Bayesian version of symbolic regression generates at least not less accurate forecasts than those generated by the other models.

Methods

Monthly data beginning in 1989 and ending in 2021 are analysed. Oil spot prices are measured separately by WTI, Brent and Dubai prices for robustness of the results. The possible oil price drivers considered are: the world production of crude oil including lead condensate, OECD refined petroleum products consumption, U.S. ending stocks excluding Strategic Petroleum Reserves of crude oil and petroleum products, MSCI World stock market index for developed markets, VXO index, Kilian index of global economic activity and real narrow effective exchange rate for U.S. The index of Chinese stock market is constructed by glueing Hang Seng and SSE Composite index in December 1991 (CBOE, 2021; EIA, 2021; FRED, 2021; MSCI, 2021; Stooq, 2021; The World Bank, 2021; Kilian, 2009). The usual transformations of time-series before inserting them into the models are performed. In particular: logarithmic 1st differences, logarithmic 12th differences and standardization. Means and standard deviations for standardization are computed on the basis of the first 100 observations in order to omit the forward looking bias in forecasting. However, forecasts accuracies and testing procedures are performed after backward transformions from differences to raw time-series levels. Except the quite well explored models, like, penalized regressions and certain Bayesian model combination schemes, the novel method, i.e., Bayesian symbolic regression, is more deeply analysed. The original method is improved with some model averaging schemes (Steel, 2020; Wang et al., 2017). The very simple set of operators is considered for symbolic regression, because the aim is stressed on variable uncertainty over model uncertainty (simple functions are ad hoc preferred herein). All of the obtained forecasts are compared between themselves with the Model Confidence Set testing procedure. Several models are estimated: symbolic regression with genetic programming, symbolic regression with Bayesian symbolic trees, LASSO, RIDGE, elastic net, least-angle regression, Dynamic Model Averaging, Bayesian Model Averaging, time-varying parameters regression, ARIMA and no-change method (Jin, 2021; Stephens, 2021; Gramacy, 2019; Onorante and Raftery, 2016; Hastie and Efron, 2013; Friedman et al., 2010; Raftery et al., 2010; Hyndman and Khandakar, 2008). Some of these models are improved and modified by implementing certain model combination schemes (averaging and selection ones). Forecasts accuracies are measured by Root Mean Square Error, Mean Absolute Error and Mean Absolute Scaled Error. Forecasts performance is tested with the Diebold-Mariano test and the Model Confidence Set procedure. Analysis with the Diebold-Mariano test is improved by the Giacomini-Rossi fluctuation test (Giacomini and Rossi, 2010).

Results

Dynamic Model Averaging and ARIMA models are the superior models in a sense of forecast accuracy according to the Model Confidence Set. The Diebold-Mariano test confirms that forcasts generated by Dynamic

Model Averaging are statistically significantly more accurate than forecasts generated by the other models. Nevertheless, forecasts generated by symbolic regression with Bayesian symbolic trees are at least not less accurate than the ones generated by the competing models. Performing recursive computations (expanding in-sample period) tentds to slightly improve forecast accuracy comparing with fixed estimations.

Conclusions

The newly proposed symbolic regression with Bayesian symbolic trees has a good forecasting potential. However, (also quite a novel) Dynamic Model Averaging happens to be the most accurate forecasting method. Indeed, this method has recently gain a very high interest from researchers and pracitioners (Nonejad, 2021). Neverthelss, this research still can be deepened and explored more thoroughly.

Acknowledgements

Research funded by the grant of the National Science Centre, Poland, under the contract number DEC-2018/31/B/HS4/02021.

References

CBOE, 2021, VIX Historical Price Data, https://www.cboe.com/tradable_products/vix/vix_historical_data

Diebold, F. X., Mariano, R. S., 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253-263.

EIA, 2021, U.S. Energy Information Administration, https://www.eia.gov

FRED, 2021, Economic Data, https://fred.stlouisfed.org

Friedman, J., Hastie, T., Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software 33, 1-22.

Giacomini, R., Rossi, B., 2010, Forecast comparisons in unstable environments, Journal of Applied Econometrics 25, 595-620.

Gramacy, R. B., 2019, monomvn: Estimation for MVN and Student-t Data with Monotone Missingness, https://CRAN.R-project.org/package=monomvn

Hansen, P. R., Lunde, A., Nason, J. M., 2011, The Model Confidence Set, Econometrica 79, 453-497.

Hastie, T., Efron, B., 2013, lars: Least Angle Regression, Lasso and Forward Stagewise, https://CRAN.R-project.org/package=lars

Hyndman, R. J., Khandakar, Y., 2008, Automatic time series forecasting: the forecast package for R, Journal of Statistical Software 26, 1-22.

Hyndman, R. J., Koehler, A. B., 2006, Another look at measures of forecast accuracy, International Journal of Forecasting 22, 679-688.

Jin, Y., 2021, A Bayesian MCMC Based Symbolic Regression Algorithm, https://github.com/ying531/MCMC-SymReg

Jin, Y., Fu, W., Kang, J., Guo, J., Guo, J., 2019, Bayesian symbolic regression, https://arxiv.org/abs/1910.08892

Kilian, L., 2009, Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, American Economic Review 99, 1053-1069.

Koza, J. R., 1992, Genetic Programming, Cambridge, MA: MIT Press.

Manickavasagam, J., 2020, Drivers of global crude oil price: A review, IMI Konnect 9, 69-75.

Miao, H., Ramchander, S., Wang, T., Yang, D., 2017, Influential factors in crude oil price forecasting, Energy Economics 68, 77-88.

MSCI, 2021, End of Day Index Data Search, https://www.msci.com/end-of-day-data-search

Nonejad, N., 2021, An overview of Dynamic Model Averaging techniques in time-series econometrics, Journal of Ecnomic Surveys 35, 566-614.

Onorante, L., Raftery, A. E., 2016, Dynamic model averaging in large model spaces using dynamic Occam's window, European Economic Review 81, 2-14.

Raftery, A. E., Karny, M., Ettler, P., 2010, Online prediction under model uncertainty via Dynamic Model Averaging: Application to a cold rolling mill, Technometrics 52, 52-66.

Steel, M. F., 2020, Model averaging and its use in economics, Journal of Economic Literature 58, 644-719.

Stephens, T., 2021, Genetic Programming in Python, with a Scikit-learn Inspired API: gplearn, https://github.com/trevorstephens/gplearn

Stooq, 2021, Quotes, https://stooq.com

Su, C. W., Qin, M., Tao, R., Moldovan, N. C., Lobont, O. R., 2020, Factors driving oil price - From the perspective of United States. Energy 197, 117219.

The World Bank, 2021, Commodities Markets, https://www.worldbank.org/en/research/commodity-markets

Wang, Y., Liu, L., Wu, C., 2017, Forecasting the real prices of crude oil using forecast combinations, Energy Economics 66, 337-348.