Predictability of Commodity Futures Returns with Machine Learning Models

Shirui Wang, Ph.D.
Independent Researcher, China

Tianyang Zhang, Ph.D.
Wenlan School of Business, Zhongnan University of Economics and Law, China

Published in *Journal of Futures Markets* and Available at: https://doi.org/10.1002/fut.22471

We use prevailing machine learning models to investigate the predictability of futures returns in 22 commodities with commodity-specific and macroeconomic factors as predictors. Out-of-sample prediction errors for the majority of futures contracts are lowered compared with those obtained by the baseline models of AR(1) and forecast combinations. Using Shapley values to explain feature importance, we identify dominant predictors for each commodity. A long–short portfolio strategy based on monthly Light Gradient-boosting Machine predictions outperforms the benchmark linear models in terms of annual return, Sharpe ratio, and maximum drawdown.

Introduction

The predictability of futures returns is one of the main questions of interest in the commodity markets in recent decades. As an alternative asset class from traditional financial assets such as equities, the factors that price the traditional asset classes are not appropriate for commodities (Giampietro *et al*., 2018; Gorton and Rouwenhorst, 2006). Besides commodity-specific factors, commodity markets are also well known to be affected by macroeconomic factors. There is still debate about the predictability of risk factors since previous studies have shown mixed results. Examples include Daskalaki *et al*. (2014), Fernandez-Perez *et al*. (2018), and others.

Most studies in the commodity markets mainly use the traditional asset pricing models, such as linear regression models. However, the underlying data-generating process of asset returns is likely nonlinear, thus predictions are typically improved by adding nonlinearities while accounting for overfitting. We employ machine learning methods that allow for nonlinear forms of predictors that are not captured by traditional linear pricing models.

Our paper has several contributions. First, we apply modern machine learning techniques to select and interpret factors, and find that the significant predictive factors of each commodity are distinct, which explains why there is no agreement on the common predictive factors in the commodity markets based on linear relationships. This indicates that it is critical to focus on each contract instead of doing a cross-sectional test as in most previous research in the literature. Second, we implement a comprehensive comparison of models and contracts, and find that Light Gradient-boosting Machine (LightGBM) performs the best in predicting futures returns. Third, we construct a practical long–short portfolio strategy based on return predictions that outperform strategies with AR(1) and forecast combinations.

Data and Methodology

The data include futures contracts for 22 commodities, covering the period from May 1997 to December 2018. For each commodity, we construct continuous time series monthly futures returns. Following Gu *et al*. (2020) and Leippold *et al*. (2022), we adopt an additive prediction error model to study the predictability of futures return of commodity i at time $t+1$ ($r_{i,t+1}$) when all available risk factors are included:

$$r_{i,t+1} = f(Z_{i,t}) + \varepsilon_{i,t+1} \quad (1)$$

where $Z_{i,t}$ is a vector of predictors at time t. The specific formula of $f(\cdot)$ is undefined and dependent on the model of choice.

The predictors of futures return are sorted into two groups: commodity-specific and macroeconomic factors. Commodity-specific factors are constructed for each commodity, while the macroeconomic factors are common for all commodities. On the basis of the findings in the previous studies, 15 commodity-specific factors used in this study are lagged return, liquidity factor, futures basis, hedging pressure, moving average of hedging pressure, call-to-put volume ratio, skewness, past 6-month futures returns, past 12-month futures returns, the difference between lagged return and past 6-month futures returns, the difference between lagged return and past 12-month futures returns, the logarithm value of futures open interest, the logarithm value of futures trading volume, the average lagged return in the sector, and the average lagged return in the whole market. To capture seasonal information, we calculate the rolling mean, standard deviation, minimum, and maximum in the same month with past values for the commodity-specific factors. In addition, we construct the interaction terms that are formed by two commodity-specific factors. To control for the effect of market volatility, we include the squared of market return. As a result, the total number of factors that are related to commodity markets is 181.

Monthly macroeconomic factors are combined from the updated data of Welch and Goyal (2008) and McCracken and Ng (2016). If there are common factors in these two databases, we only keep the ones from the database in McCracken and Ng (2016). As a result, there are 136 macroeconomic factors included in our study. In the appendix of the published paper, we provide the explanations of the factors selected by the model for prediction. Details of other macroeconomic factors are referred to Welch and Goyal (2008) and McCracken and Ng (2016). We also include three seasonal dummy variables. Thus, the total number of risk factors in this study is 320.

We choose five prevailing machine learning models: Elastic net (ENET), principal component regression (PCR), support vector regression (SVR), LightGBM (LGB), and Neural Network (NN). Modern machine learning methods are particularly good at feature selection and extraction of nonlinear relationships among features. To this end, we follow the standard practitioner's procedure by first creating statistics and interactions terms to create a rich set of features, and then allow models to select features based on prediction errors of validation sets. Feature selection is essential to balance the bias-variance trade-off and avoid overfitting.

Prediction Results

In this study, we employ AR(1) and forecast combinations models as two benchmark models. The out-of-sample mean squared predictions error (MSE) was used to assess the predictive performance of AR(1) and forecast combinations models, along with five machine learning models. For commodity i, this is defined as

$$MSE(i) = \frac{1}{N}\sum_{(i,t)\in\Theta}(r_{i,t} - \hat{r}_{i,t+1}) \quad (2)$$

where $r_{i,t}$ and $\hat{r}_{i,t+1}$ are observed and predicted futures return, N is the number of observations in the out-of-sample test and Θ indicates the fits that are only assessed on the testing subsample. For each commodity, we make a 70%/20%/10% time split for the training, validation, and test sets.

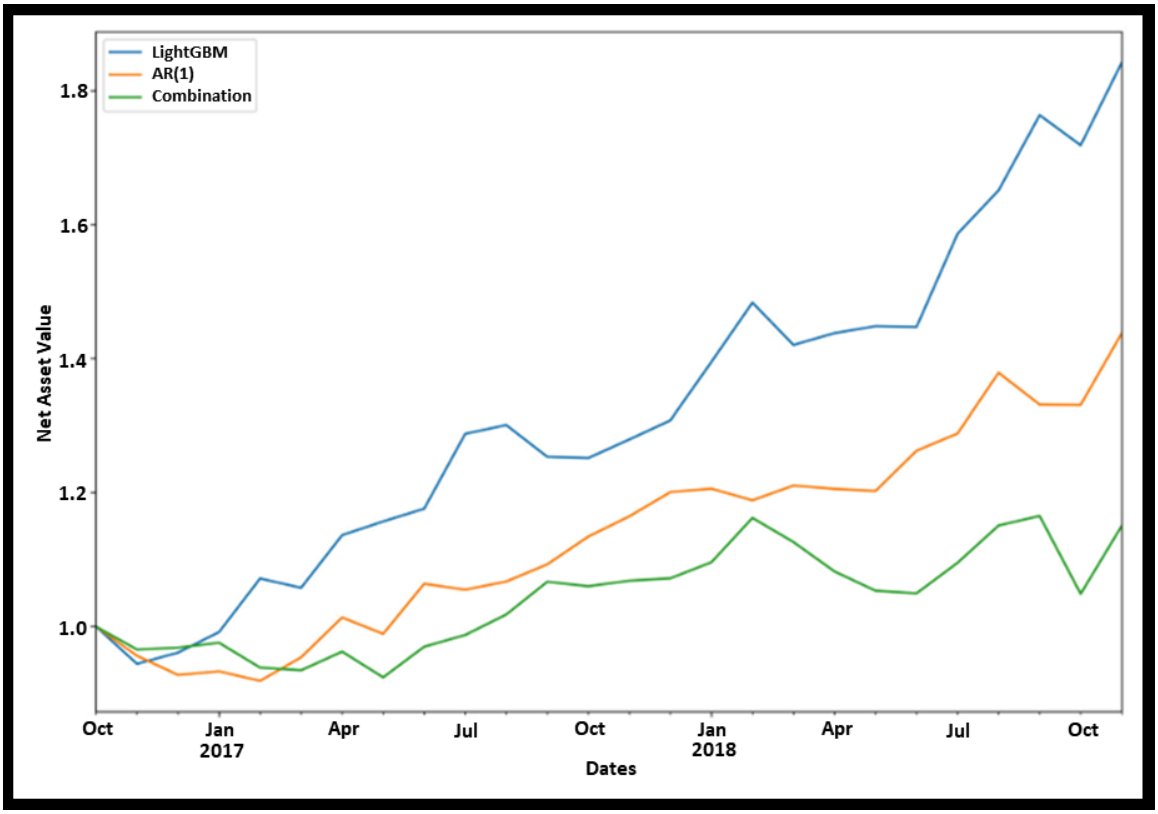
All predictive variables are included when comparing different models, although which factors are fed into the final models depends on whether they survive feature selections that are distinct for each model. The results show that the lowest out-of-sample MSE values differ a lot across different commodities. For the average MSEs among all commodities, LightGBM is the lowest with 0.438, followed by AR(1), ENET, NN, PCR, forecast combinations, and SVR. The results of the CW-test also show that LightGBM has significantly better predictive performance than AR(1) and forecast combinations.

In addition to analyzing the out-of-sample MSE values of each model, we use Shapley values obtained from LightGBM to quantify the importance of each factor in affecting futures returns and rank all factors based on the predictive performance for each commodity. All commodities are categorized into four broad sectors: agriculture (13 commodities), energy (three commodities), livestock (three commodities), and metals (three commodities). In summary, there are no common factors in predicting futures returns since the top performance factors are distinct for each commodity.

The empirical results show that the factors related to hedging pressure and momentum perform best among all the factors associated with commodity markets. The fact that the hedging pressure effects only exist in some commodities is consistent with the finding that there is not an agreement on the hedging pressure effect in commodity markets. Regarding macroeconomic factors, the information mainly from labor markets, industrial production, credit markets, interest rate markets, and stock markets has predictive power for certain commodities.

Next, we take advantage of the machine learning methods in predicting expected returns and evaluate trading strategies based on the predictions. A market-neutral strategy is built by going long the contracts with the highest predicted returns and shorting the losers. Figure 1 on the next page shows the net asset values of the three strategies. The LightGBM strategy achieves a 35% annualized return and 2.07 Sharpe ratio in the testing sample, compared to the 20% (8%) and 1.52 (0.54) Sharpe ratio of the AR(1) (forecast combinations) strategy in the same period. The LightGBM strategy's performance is consistently best over time, except for the period in the first two months of the testing sample.

Figure 1
Net Asset Values of LightGBM, AR(1), and Forecast Combination Strategies, October 2016-November 2018



Conclusion

This study investigates the performance of several machine learning models in predicting commodity futures returns for 22 commodities. All the risk factors are sorted into two groups that are commodity-specific and macroeconomic factors. Compared with benchmark AR(1) and forecast combinations models, LightGBM performs the best among other machine learning models in terms of out-of-sample MSE overall. The top predictive factors based on the Shapley values are provided for each commodity. Interestingly, we do not find common predictors for all commodities, which indicates the importance of analyzing commodities individually. Overall, the performance of factors that are associated with momentum and hedging pressure are better than the other factors related to commodity markets. In addition, the macroeconomic factors demonstrating predictive power for futures returns belong to multiple groups, with labor and stock markets serving as two examples. A long–short strategy based on predicted returns with LightGBM is shown to be profitable and superior to the benchmark, with a 35% annualized return and a Sharpe ratio of 2.07. The downside risk control is also practically desirable.

Endnote

All errors are our own. The authors declare no conflict of interest.

References

Daskalaki, C., Kostakis, A. and G. Skiadopoulos, 2014, "Are There Common Factors in Individual Commodity Futures Returns?", *Journal of Banking & Finance*, Vol. 40, March, pp. 346-363.

Fernandez-Perez, A., Frijns, B., Fuertes, A.-M. and J. Miffre, 2018, "The Skewness of Commodity Futures Returns," *Journal of Banking & Finance*, Vol. 86, January, pp. 143-158.

Giampietro, M., Guidolin, M. and M. Pedio, 2018, "Estimating Stochastic Discount Factor Models with Hidden Regimes: Applications to Commodity Pricing," *European Journal of Operational Research*, Vol. 265, No. 2, March, pp. 685-702.

Gorton, G. and K. G. Rouwenhorst, 2006, "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal*, Vol. 62, No. 2, March-April, pp. 47-68.

Gu, S., Kelly, B. and D. Xiu, 2020, "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, Vol. 33, No. 5, May, pp. 2223-2273.

Leippold, M., Wang, Q. and W. Zhou, 2022, "Machine Learning in the Chinese Stock Market," *Journal of Financial Economics*, Vol. 145, No. 2, August, pp. 64-82.

McCracken, M. W. and S. Ng, 2016, "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, Vol. 34, No. 4, October, pp. 57-589.

Welch, I. and A. Goyal, 2008, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *The Review of Financial Studies*, Vol. 21, No. 4, July, pp. 1455-1508.

Keywords

Commodity futures, machine learning, LightGBM.

Author Biographies

SHIRUI WANG, Ph.D.
Independent Researcher, China

Shirui Wang is an independent researcher and crypto trader. He received his Ph.D. in Economics from Iowa State University in 2020. His research focuses on commodity and crypto markets. He has published papers in *Journal of Futures Markets,* and *Economic Theory and Business Management (Chinese).*

TIANYANG ZHANG, Ph.D.
Wenlan School of Business, Zhongnan University of Economics and Law, China

Tianyang Zhang is an Assistant Professor in the Wenlan School of Business, Zhongnan University of Economics and Law. He received his Ph.D. in Economics from Iowa State University in 2020. His fields of interest are within commodity markets and cryptocurrency markets. He has published papers in *Journal of Futures Markets*, *The North American Journal of Economics and Finance*, and *Finance Research Letters*.