COMMODITY INSIGHTS DIGEST

SUMMER 2023

RESEARCH DIGEST ARTICLES

"THE LEAD-LAG RELATIONS IN THE COMMODITY FUTURES RETURNS: A MACHINE LEARNING APPROACH"

BY YUFENG HAN, Ph.D., BELK COLLEGE OF BUSINESS, UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE; and

LINGFEI KONG, Ph.D., OLIN SCHOOL OF BUSINESS, WASHINGTON UNIVERSITY IN ST LOUIS

The Lead-lag Relations in the Commodity Futures Returns: A Machine Learning Approach

Yufeng Han, Ph.D.
Professor of Finance, Belk College of Business, University of North Carolina at Charlotte

Lingfei Kong, Ph.D.
Postdoctoral Research Associate in Finance, Olin School of Business, Washington University in St. Louis

Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3536046

This paper uses machine learning tools to study the lead-lag relations in commodity futures returns. We use LASSO (Least Absolute Shrinkage and Selection Operator) to select the predictors because the number of predictors is large relative to the number of observations. We find significant full-sample and out-of-sample predictability. In the full sample, we find that LASSO can identify a sparse set of predictors that either come from economically linked commodities or are likely driven by excessive speculative trading. The out-of-sample forecasts based on LASSO generate statistically and economically large gains. When we use more complex machine learning models such as regression trees and neural networks to forecast commodity futures returns, the out-of-sample performance is worse than LASSO portfolios, suggesting that nonlinearities and interactions do not appear substantial in the data. Finally, we find that index trading due to financialization has been associated with excess comovement among commodity futures.

Introduction

Commodity futures contracts are agreements to buy or sell a predetermined quantity of a commodity at a specified price on a particular date in the future. Historically, commodity futures were primarily used by farmers, producers, and commodity merchandisers to lock in the price and reduce the risk of financial losses from price changes. Over the past two decades, financial institutions such as hedge funds, swap dealers, and mutual funds have dramatically increased their exposure to commodities. The advent of commodity futures Exchange-Traded Funds (ETFs) and Exchange-Traded Notes (ETNs) also gives individual investors easy access to commodity futures. According to BarclayHedge, the assets under management for managed futures have grown from $95.7 billion to $318.4 billion from 2003 to 2019; managed futures programs, in turn, include the trading of both financial futures and physical commodity futures (Gupta and Wilkens, 2007). The phenomenon of investor participation in physical commodity futures trading is called “financialization” by researchers (*e.g.*, Tang and Xiong, 2012; Basak and Pavlova, 2016), and has received extensive attention from researchers and practitioners.

This paper examines the predictability of commodity futures returns. Specifically, we examine the lead-lag relations in the commodity futures returns by predicting commodity futures returns with the commodity futures’ own lagged returns as well as other commodity futures’ lagged returns. We focus on the post-financialization period (2004-2019) when commodity futures become a popular investment asset class, and futures markets become more liquid than before. The inclusion of the broad set of lagged returns is motivated by the following arguments. First, many futures have economic links that can drive the lead-lag relations among futures returns. For instance, (a) heating oil is refined from crude oil, so via the crack spreads, the lagged returns of crude oil futures would be tied to the returns of heating oil futures; (b) corn and soybean meal are the primary feeds for livestock such as pigs and cows, thus via the feed margins, the returns of corn and soybean futures would impact the returns of lean hogs and live cattle futures at some stage of the production; and (c) crops such as corn and soybeans are used to produce biofuels, which are substitutes for fossil fuels such as crude oil, so crude oil futures prices would be expected to be related to grain futures prices. Mensi *et al*. (2014) find significant spillovers among grain and energy commodity prices by studying the lead-lag relations of the returns and volatilities of these commodities. Second, the increased trading activities by financial institutions and retail investors can either weaken or strengthen these effects, or even introduce new relations that are not driven by any economic links; for example, Till (2008) describes how seemingly unrelated commodities and financial instruments became (temporarily) very correlated during "risk off" episodes. Unlike producers/merchants who use futures contracts to lock in the price of their products or inputs, these traders generally do not have physical exposure to the commodities but use commodity futures as investment tools, and thus are non-commercial traders.

Some researchers have concluded that the proliferation of non-commercial trading may have caused price distortion and excessive price comovement, depending on the frequency of observation. For example, Tang and Xiong (2012) find that since 2004 when financial institutions significantly increased their positions on commodity futures, the prices of non-energy commodity futures have become increasingly correlated with oil futures, and the correlations are stronger for commodities that are included in the broadly traded commodity indices than for non-indexed commodities.

To provide an even balance of views, one must also note that there may be other explanations besides financialization for the increased correlations of oil futures with non-energy commodity futures, which would include a broad-based increase in commodity demand from emerging economies (Cevik and Sedik, 2011). Further, IEA (2011) states that "[a] … comparison of the non-exchange-traded commodity price index, as well as [the] crude oil price series, supports the notion that, starting in 2003 and more strongly after 2004, a demand shock pushed upward the price of most commodities." In addition, based on an examination of crude oil, natural gas, and corn futures markets, Brunetti *et al.* (2015) "find little evidence that speculators destabilize … markets."

Relevance of the Research Question

This paper contributes to the existing literature in the following ways. First, the majority of the literature, after financialization, studies the comovement among commodity futures in the context of contemporaneous relations rather than lead-lag relations. Moreover, most papers regarding return predictability in commodity futures focus on the futures' own past returns or other characteristics but ignore the cross-serial correlations. As far as we can tell, this paper is the first to analyze the predictability of the lagged futures returns that allows each individual commodity futures' return to respond to the lagged returns for all commodity futures, thereby accommodating a large dimension of commodity links, both direct and indirect. Second, we use machine learning techniques to overcome the potential overfitting problem. With a large number of predictors, OLS (ordinary least squares) estimation is subject to overfitting. Third, we find that incorporating the lagged returns of commodity futures can help to forecast the returns for individual commodity futures and construct profitable trading strategies. The performance of the long-only futures indices has been lackluster in the last decade. By contrast, the actively constructed single-sort or timing strategies based on LASSO generate both statistically and economically large out-of-sample returns.

Data and Methodology

We study 27 commodity futures that are traded actively in the U.S. The data are collected from Bloomberg. The 27 commodity futures cover five main sectors, namely, grains, softs, livestock, energy, and metals. There are 8 grains futures, 7 softs futures, 3 livestock futures, 4 energy futures, and 5 metals futures in the sample. The sample period is from January 2004 to December 2019.

The main methodology used to forecast commodity futures returns is LASSO (Least Absolute Shrinkage and Selection Operator), as proposed by Tibshirani (1996). The reason why we use LASSO is that the number of predictors is large compared with the number of observations used to fit the model. With a large number of predictors, OLS model causes overfitting, which means it reduces the in-sample mean squared error but often results in a large out-of-sample mean squared error. LASSO selects a subset of predictors and thus reduces overfitting. We use the corrected Akaike information criterion (AICC) to select the shrinkage parameter in LASSO. We use OLS to estimate the coefficients for the predictors selected by LASSO (OLS post-LASSO) since LASSO can over-shrink the magnitudes of the coefficients for the selected predictors (Efron *et al*., 2004). Our main benchmark forecast models are OLS that include all the lagged returns as predictors and prevailing mean, which is the moving average of the commodity futures returns during the estimation window.

In the in-sample analysis, we use the full sample to estimate the OLS post-LASSO regression. In the out-of-sample analysis, we run a 60-month rolling LASSO regression and calculate the one-step ahead out-of-sample return forecast. To evaluate the out-of-sample forecast performance, we construct two trading strategies, one is a single-sort portfolio, in which we form equally-weighted portfolios using the five commodities with the highest forecasted returns and the five commodity futures with the lowest forecasted returns. We then calculate the return differences between the two portfolios. The other strategy is a moving-average timing portfolio; that is, we establish a long position if its forecasted return is positive and a short position if its forecasted return is negative, then form an equally-weighted portfolio of all the positions. There are two versions of LASSO models in the out-of-sample analyses: they are LASSO that excludes any futures that do not have any lagged returns selected by the LASSO, and LASSO (All) that includes all contracts including those without any predictors.

To test whether financialization has been associated with the excess comovement among commodity futures, we have both in-sample tests and out-of-sample tests. In the in-sample test, we create a dummy variable that is equal to 1 when the commodity futures are members of indices and when the date is after the inception of three ETF or ETN products (September 2006) that track the returns of the Bloomberg Commodity Index (BCOM) or the S&P Goldman Sachs Commodity Index (GSCI); otherwise, the dummy variable is set to 0. We then add the dummy as well as its interaction with all the lagged returns as the additional predictors and run the OLS post-LASSO regression using the full sample. In the out-of-sample test, we classify the commodity futures included in both BCOM and GSCI as indexed futures, and those included neither in BCOM nor GSCI index as non-indexed futures. There are 16 indexed futures and 7 non-indexed futures. We then reconstruct the single sort and timing portfolios using the 16 indexed futures or the 7 non-indexed futures. We also test the lead-lag relations and construct single-sort and timing portfolios using data during 1970-2003, which is before the financialization period.

Main Results

The LASSO single-sort portfolio has an annualized mean return of 15.15%, which is more than eight times the mean return of the prevailing mean portfolio (1.78%). The annualized Sharpe ratio of this LASSO portfolio is also much higher than those of the two benchmark portfolios, 0.93 versus 0.10 (prevailing mean) and -0.21 (OLS). The LASSO timing portfolio yields an average excess return of 6.15% per year and a Sharpe ratio of 0.72. In contrast, the prevailing mean timing portfolio generates an annual return of only 1.80%, the OLS timing portfolio generates an even smaller average return that is less than 1%. Overall, the results show that lagged returns contain predictability for the cross-section of commodity futures returns and LASSO helps to identify important predictors. The above results are robust to transaction costs, an alternative number of lagged returns, seasonality, and different machine learning models such as elastic net.

When we interact the ETF inception dummy with lagged returns for the whole sample, at least one lagged commodity futures return is selected for 14 of the 16 indexed futures. In addition, 10 out of the 14 futures have interactions of the ETF inception dummy with lagged returns as selected effects, and the hypothesis that the interactions are all equal to 0 is rejected for all the 10 futures. When we use indexed-only futures to run the lead-lag predictive regressions and construct portfolios, the LASSO portfolio generates economically and statistically significant out-of-sample returns in both the single-sort and timing strategies. However, when we use the non-indexed futures to run the lead-lag predictive regressions, the LASSO portfolio no longer generates any significant out-of-sample returns. In addition, when we examine the lead lag relations from 1970-2003, the two LASSO portfolios do not generate significant out-of-sample returns.

Conclusions

In this paper, we exploit the lead-lag relations in the commodity futures returns via machine learning tools (mainly LASSO) to analyze return predictability in commodity futures markets. The portfolio analysis reveals that the LASSO strategies outperform the well-known OLS and historical moving average by generating statistically and economically larger excess returns. We find that LASSO strategies perform well after the financialization period but not before financialization. When we separate the indexed futures from the non-indexed futures, we find that LASSO strategies only work in indexed futures. Moreover, the lead-lag relations are stronger after the advent of ETFs or ETNs that track the broad futures indices such as GSCI and BCOM indices. Overall, our results suggest that lagged returns contain valuable information to predict futures returns and that machine learning is a useful and effective tool to increase out-of-sample predictability.

Endnote

Dr. Kong presented on this topic at the J.P. Morgan Center for Commodities' 5th Annual International Symposium at the University of Colorado Denver Business School on August 15, 2022.

References

Basak, S. and A. Pavlova, 2016, "A Model of Financialization of Commodities," *The Journal of Finance*, Vol. 71, No. 4, August, pp. 1511-1556.

Brunetti, C., Büyükşahin, B. and J. Harris, 2015, "Speculators, Prices and Market Volatility, Bank of Canada Working Paper, November.

Cevik, S. and T. Sedik, 2011, "A Barrel of Oil or a Bottle of Wine: How Do Global Growth Dynamics Affect Commodity Prices?", International Monetary Fund Working Paper, January.

Efron, B., Hastie, T., Johnstone, I. and R. Tibshirani, 2004, "Least Angle Regression," *The Annals of Statistics*, Vol. 32, No. 2, April, pp. 407-499.

Gupta, B. and K. Wilkens, 2007, "An Evaluation of the *Commodity* Exposure in Commodity Trading Advisers," in H. Till and J. Eagleeye (eds) Intelligent Commodity Investing, London: Risk Books, pp. 179-186.

IEA [International Energy Agency], 2011, "Oil Market Report," March 15, pp. 44-45.

Mensi, W., Hammoudeh, S., Nguyen, D.K. and S. M. Yoon, 2014, "Dynamic Spillovers Among Major Energy and Cereal Commodity Prices," *Energy Economics*, Vol. 43, May, pp. 225-243.

Tang, K. and W. Xiong, 2012, "Index Investment and the Financialization of Commodities," *Financial Analysts Journal*, Vol. 68, No. 6, November/December, pp. 54-74.

Tibshirani, R., 1996, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 1, pp. 267-288.

Till, H., 2008, "Case Studies and Risk Management in Commodity Derivatives Trading," in H. Geman (ed) Risk Management in Commodity Markets, Chichester (U.K.): John Wiley & Sons Ltd., pp. 255-291.

Keywords

Commodity futures, LASSO, machine learning, predictability, lead-lag relations, financialization, comovement.

Author Biographies

YUFENG HAN, Ph.D.
Professor of Finance, Belk College of Business, University of North Carolina at Charlotte

Dr. Yufeng Han joined the Belk College faculty as an Associate Professor of Finance in 2016. Prior to joining UNC Charlotte, he served as a faculty member at the University of Colorado Denver and Tulane University. Dr. Han's primary research interests are empirical asset pricing, investment, mutual funds, and econometrics. He has published in the *Review of Financial Studies, Journal of Financial Economics, Journal of Financial and Quantitative Analysis, Journal of Banking and Finance, Real Estate Economics* and many other finance and economics journals. Dr. Han received a Bachelor's degree from Tsinghua University (China) and a Ph.D. from Washington University in St. Louis.

The Lead-lag Relations in the Commodity Futures Returns: A Machine Learning Approach

LINGFEI KONG, Ph.D.
Postdoctoral Research Associate in Finance, Olin School of Business, Washington University in St. Louis

Dr. Lingfei Kong is a Postdoctoral Research Associate in Finance at the Olin School of Business, Washington University in St. Louis. Dr. Kong received her Ph.D. in Finance from the University of North Carolina at Charlotte. Her research focuses on empirical asset pricing, return predictability, big data analytics, commodity futures, corporate finance, and product innovation.

COMMODITY INSIGHTS
DIGEST

Commodity Insights Digest (CID) is a publication of Bayes Business School – City, University of London (U.K.), in association with Chicago-based Premia Research LLC (U.S.A.). The digest is co-edited by Ana-Maria Fuertes, Ph.D. in International Finance, and Hilary Till, M.Sc. in Statistics.

CID seeks to foster knowledge transfer between academics and practitioners by publishing:

- Scholarly research digest articles;
- Practitioner insights and interviews;
- Articles on economic history; and
- Book reviews.

Contributors to *CID* include (a) scholars seeking to enhance the impact of their research by delving into the practical implications of their theoretical or empirical studies; (b) consultants reporting on challenges faced by commodity market participants; and (c) industry economists providing analyses on their areas of market expertise.

Bayes Business School – City, University of London is a world-leading provider of education and cutting-edge research. Bayes' unique location enables strong links with industry in the City of London.

The Summer 2023 *CID* cover image is cropped from the following artwork, whose complete image is above: Poster for the London Midland & Scottish (LMS) Railway from the 1930s: The Mersey from Runcorn Bridge – Euston/Liverpool Line. Artist: Norman Wilkinson, R.I. Printed by McCorquodale & Co Ltd, Glasgow & London. Other ids: From the LMS Carriage Window Series (1). 1991–7161 Science Museum Group Collection Online. Copyright © National Railway Museum / Science & Society Picture Library. All rights reserved.

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Physical Address
Bayes Business School
106 Bunhill Row
London, EC1Y 8TZ
United Kingdom

Website
bayes-cid.com

Contact
To submit articles, authors can contact the editors via editors@bayes-cid.com

City is a member institution of the University of London

UNIVERSITY OF LONDON