Forecasting Realized Volatility of Crude Oil Futures Prices Based on Machine Learning

Jiawen Luo, Ph.D.
School of Business Administration, South China University of Technology, China

Tony Klein, Ph.D.
Faculty of Business and Economics, Technische Universität Chemnitz, Germany

Thomas Walther, Ph.D.
Utrecht School of Economics, Utrecht University, The Netherlands; and Editorial Board Member, *Commodity Insights Digest*

Qiang Ji, Ph.D.
Institutes of Science and Development, Chinese Academy of Sciences, China

Published in *Journal of Forecasting* and Available at: https://onlinelibrary.wiley.com/doi/10.1002/for.3077

Extending the popular heterogeneous autoregressive (HAR) model with additional information channels to forecast realized volatility of WTI futures prices, we show that machine learning-generated forecasts provide better forecasting quality and that portfolios that are constructed with these forecasts outperform their competing models resulting in economic gains. Analyzing the selection process, we show that information channels vary across forecasting horizons. Variable selection produces clusters and provides evidence that there are structural changes with regard to the significance of information channels.

Introduction

Modelling and forecasting volatility is a crucial task in risk management to quantify *e.g.,* commodity market risk exposure. One of the tools to predict asset price fluctuations is the heterogeneous autoregressive (HAR) model (Corsi, 2009). Structurally a simple, linear model of three different horizons of realized volatility, it has been proven hard to beat by more complex models in out-of-sample settings (Degiannakis *et al*., 2022). In machine learning (ML) terminology, the more complex models typically rely on features based on returns, *e.g.,* asymmetry between positive and negative returns or jump processes. In this study, we extend the baseline HAR model with exogenous information. In particular, we include equity, sentiment, other commodities, attention, and foreign exchange information channels.

In addition, we use ML techniques to improve the HAR model and better select relevant predictors. In this way, the inclusion of certain information channels is purely data-driven and not at the discretion of the forecaster. We show that the ML variants can outperform the baseline model both in accuracy of the predictions and economic gains. Building on the time-varying selection of predictors, we further analyze the importance of the information from different channels. Particularly, foreign exchange information adds value beyond realized volatility.

The main contribution of this paper is to provide new evidence on the role of diverse exogenous information channels, including equity, commodity, FX, sentiment, and Google Trends, in explaining the realized volatility of WTI crude oil futures. Extending the baseline HAR model with 84 predictors, we apply

and compare multiple ML-based variable selection methods (LASSO, BMA, BAG, and SSVS) to address parameter proliferation and multicollinearity. We show that incorporating selected exogenous variables improves both forecasting accuracy and economic value, with FX markets emerging as particularly important. Our results further reveal that the number of relevant channels increases with the forecasting horizon, contributing to the literature on market connectedness and HAR model forecasting.

Methodology and Data

We model realized volatility using high-frequency intraday data and extend the heterogeneous autoregressive (HAR) framework to incorporate a broad range of exogenous information channels. Following Corsi (2009), we compute daily realized volatility (RV) as the sum of squared intraday returns sampled at 5-minute intervals:

$$RV_t = \sum_{i=1}^{m} r_{t,i}^2$$

where $r_{t,i} = 100 \cdot (\log(P_{t,i}) - \log P_{t,i-1})$. Our baseline model is the standard HAR specification, in which the forecast of log realized volatility over a given horizon *h* depends on daily, weekly (5-day), and monthly (22-day) averages of past log realized volatility. We estimate separate regressions for each forecasting horizon rather than imposing fixed weights across horizons, allowing for greater flexibility in capturing horizon-specific dynamics:

$$\log RV_{t+h}^{(h)} = \beta_0 + \beta_1 \log RV_t + \beta_2 \log RV_t^{(5)} + \beta_3 \log RV_t^{(22)} + u_{t+h}^{(h)},$$

with

$$\log RV_{t+h}^{(h)} = \frac{1}{h}\sum_{j=1}^{h} \log RV_{t+h-j+1}.$$

In this paper, we forecast the volatility of the WTI crude oil futures. For forecasts beyond 1-step ahead, our models predict the average realized volatility over the forecasting horizon, which we use as a proxy for the realized volatility h-steps ahead. To enrich the information set, we extend the baseline HAR by including variables from five exogenous predictor categories: (i) equity market volatilities from major global indices: Dow Jones Industrial Average, CAC40, FTSE, DAX30, HSI, NIKKEI225, S&P 500, SSE, and EUROSTOXX; (ii) sentiment and uncertainty indicators: the Global Financial Stress Index, VIX, Oil Volatility Index OVX, and Economic Policy Uncertainty; (iii) commodity futures volatilities, including Natural Gas, Corn, Soybeans, and Gold; (iv) foreign exchange market volatilities: U.S. weighted currency basket, EUR/USD, Yen/USD, Yuan/USD, Ruble/USD, and GBP/USD; and (v) Google Trends search volumes to gauge the attention to “Oil Production”, “Financial Crisis”, “Oil Demand”, “Oil Price”, and “OPEC Conference”. Given the five predictor sets, we construct HAR-Equity, HAR-Sentiment, HAR-Commodity, HAR-FX and HAR-Google models respectively. For Equity, we use realized volatilities sampled at a matching 5-min interval. For the commodity futures and FX, we proxy volatilities by using squared daily returns. For each variable, we construct daily, weekly, and monthly averages to match the HAR structure, which results in a

total of 84 exogenous predictors when all categories are combined. We estimate category-specific HAR extensions as well as two "kitchen-sink" approaches: a comprehensive "HAR-All" specification that simultaneously incorporates all predictors; and a specification which first runs a Principal Component Analysis and only uses the first principal component of each information channel and each horizon and we define it as HAR-factor model. Lastly, we also add an Autoregressive (AR) model of order 1 to the set of models to have a very simple benchmark.

The large number of predictors in these extended models raises issues of over-parameterization and multicollinearity, both within and across information channels. To address this, we employ several ML–based techniques that are well-suited to high-dimensional forecasting problems. Specifically, we use the Least Absolute Shrinkage and Selection Operator (LASSO), Bayesian Model Averaging (BMA), Bagging, and Stochastic Search Variable Selection (SSVS). While LASSO and BAG reduce dimensionality, the BMA and SSVS use a weighting of predictors to emphasize the most important variables.

Results

We use the above models to forecast oil futures volatility (WTI) for the next day, the next week (5 days), and the next month (22 days). In the first part of the results, we analyze the predictive performance in terms of accuracy, *i.e.,* we compare the mean squared and mean absolute errors of the different model configurations. In order to understand which models actually perform better than others (in statistical terms), we use the Model Confidence Set (MCS, Hansen *et al.*, 2011). Given a confidence level, the MCS is a subset of models with equal predictive performance.

In the first set of our results, we find that most model-predictor combinations have equal predictive performance when it comes to the mean squared error. Only for mean absolute errors, we find a more exclusive set if we lower our confidence (we use confidence sets of 10% and 25%, with 25% being more exclusive but less confident). Especially for the 1-day ahead forecast, the MCS only contains four models from the set of 42 models: BMA- and BAG-HAR-Google, LASSO- and BAG-HAR-All.

For the 1-week ahead forecast (5 days), we find mostly LASSO models to be included in the 25% MCS of both MSE and MAE. While the LASSO-HAR-FX combination has the lowest predictive errors, its performance is not statistically different from the LASSO-HAR, *i.e.*, the additional information from the FX channel only plays a marginal role.

For the 22-day ahead (1-month) forecasts, statistical accuracy is very similar across models. For the MSE, even the more restrictive 25% confidence set contains all HAR variants. The MAE 25% set mainly includes the LASSO-based models together with a few others. For longer horizons, the Google-augmented variants appear to be the best-performing models.

After looking at the statistical accuracy, we are interested in the economic value of the forecasts. Particularly, we compare the values to the standard AR with and without information from either Google attention or other asset classes. In this way, we can assess both the addition of using ML techniques and the addition of extra information over a very naïve benchmark. We assume an investor with a quadratic

utility function with either mild ($\gamma = 1$) or very strong ($\gamma = 10$) risk aversion and compare the models over the forecasting horizons of one month. For each model-predictor combination, we forecast the next step's volatility. In order to predict the next period's return, we use a simple 1-year moving average for all combinations. Then the allocation to the oil futures (the risky asset) is $\widehat{\omega}_t = 1\backslash\gamma \left(\frac{\widehat{r_{t+1}}}{\widehat{\sigma}^2_{t+1}}\right)$. The portfolio return comprises the realized return from the investment in the oil futures plus the investment of the remainder into the 5-year U.S. Treasury Note.

Our findings show that while ML-augmented HAR models outperform standard HAR models in terms of statistical accuracy, this advantage does not always carry over to economic value. In portfolio applications, the BMA-based models achieve the greatest economic gains. The BMA-HAR-FX yields the highest annualized return and economic value compared to both AR(1) and standard HAR benchmarks. Risk-adjusted performance, however, is strongest for BAG models, with the BAG-HAR-All achieving the highest Sharpe ratio. LASSO models, on the other hand, do not translate statistical accuracy into higher utility.

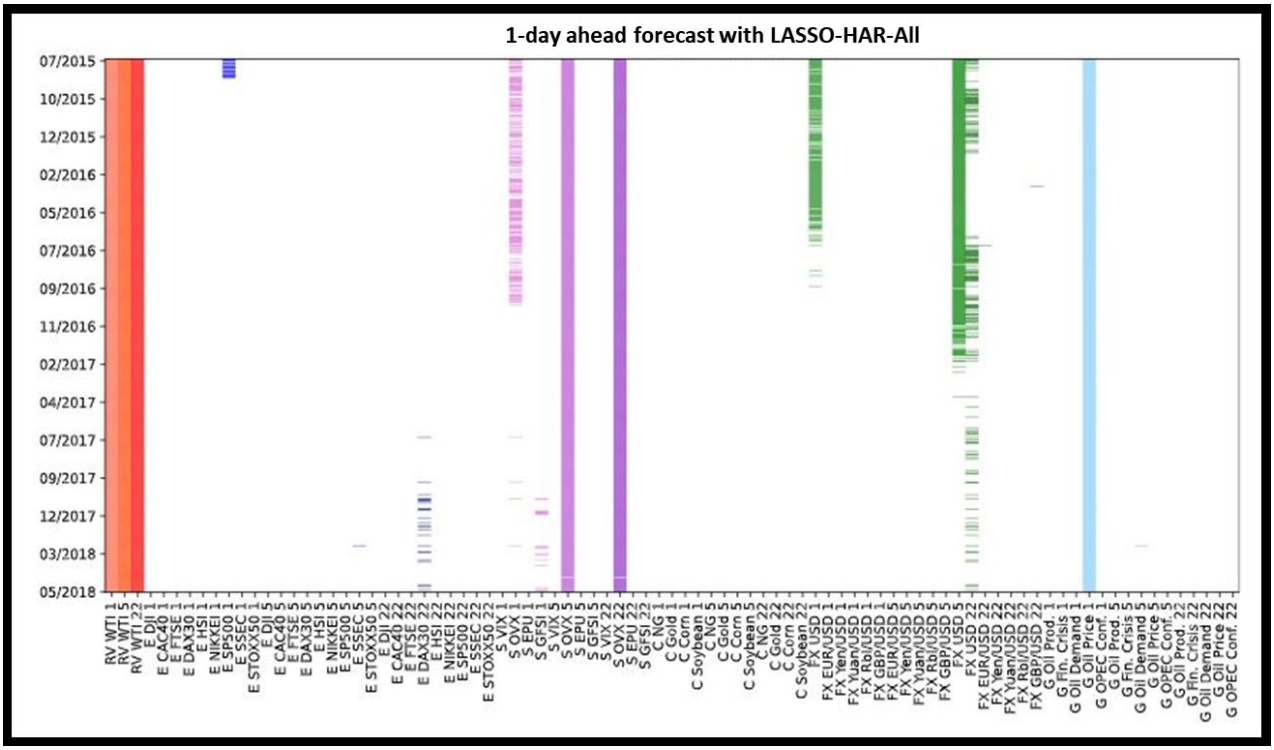
We observe two general trends in our results. Using additional information channels (especially from sentiment, commodity markets, or FX) lead to higher economic gains. ML-models can then further improve the forecasts, but only marginally.

Our last set of results looks at the time-varying drivers of the forecasts. For this exercise, we use the LASSO-HAR-All specification which contains every information channel at three different horizons. This allows us at every forecasting step to assess which of the predictors the LASSO selected over our out-of-sample period from 2015 to 2018. In this way, we can answer which are the most important variables, how sticky they are over time, and if additional information channels are able to replace the information from the previous realized volatility in the HAR model. See the three plots in Figure 1 below.

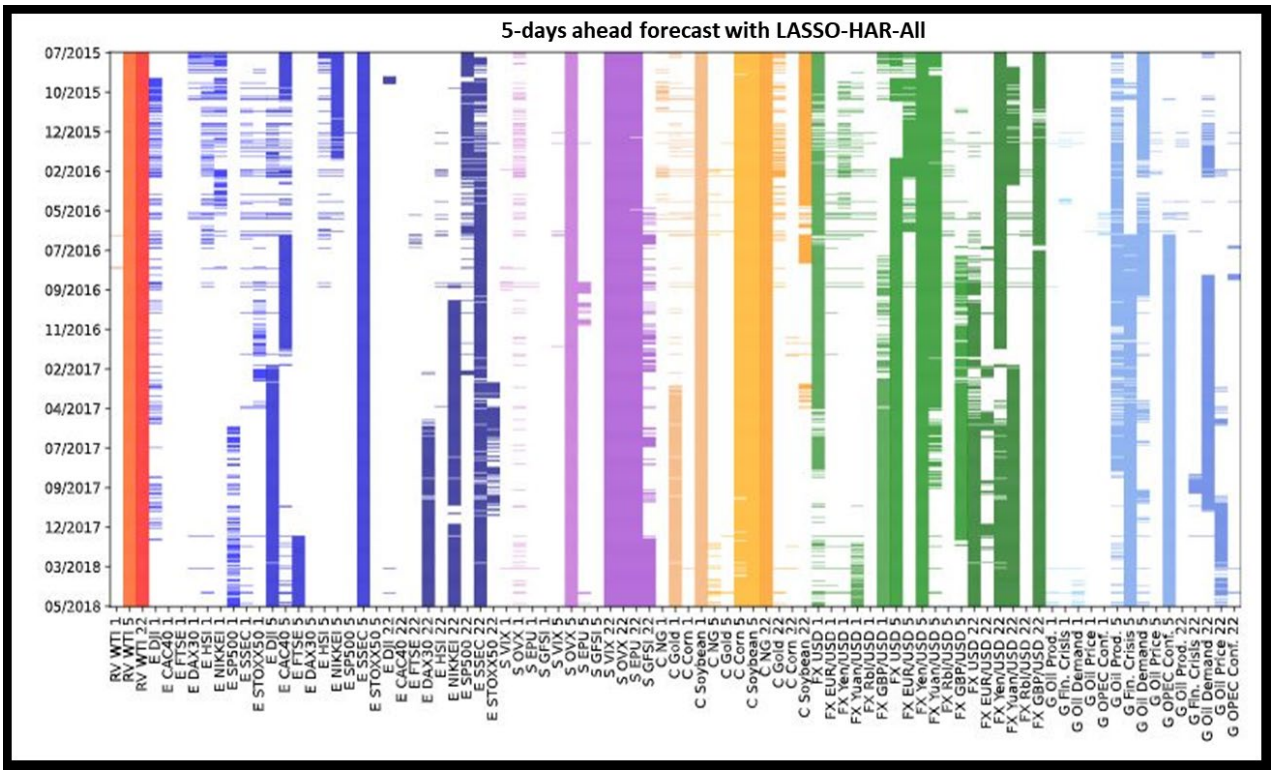
Three main observations emerge: (1) the information from the previous volatility remains important and is not replaced by any drivers. Only in the 5-day ahead model, we find that the last day's realized volatility is not important for the model. (2) While only a few predictors are selected for the 1-day ahead forecasts, we find that the longer the forecasting horizons, the more predictors are selected by the LASSO. (3) Apart from a few very persistent variables over time (*e.g.,* OVX and attention to Oil Price for 1-day; SSEC, OVX for 5-days; and NIKKEI and OVX for 22-days), the drivers vary a lot over time. We even observe clustering of some variables, meaning they are persistent for a short period of time before falling out of the selection again. For example, in the 22-days forecast, short-term information from the stock markets were selected until 2016 and fell out almost entirely afterwards.

Figure 1

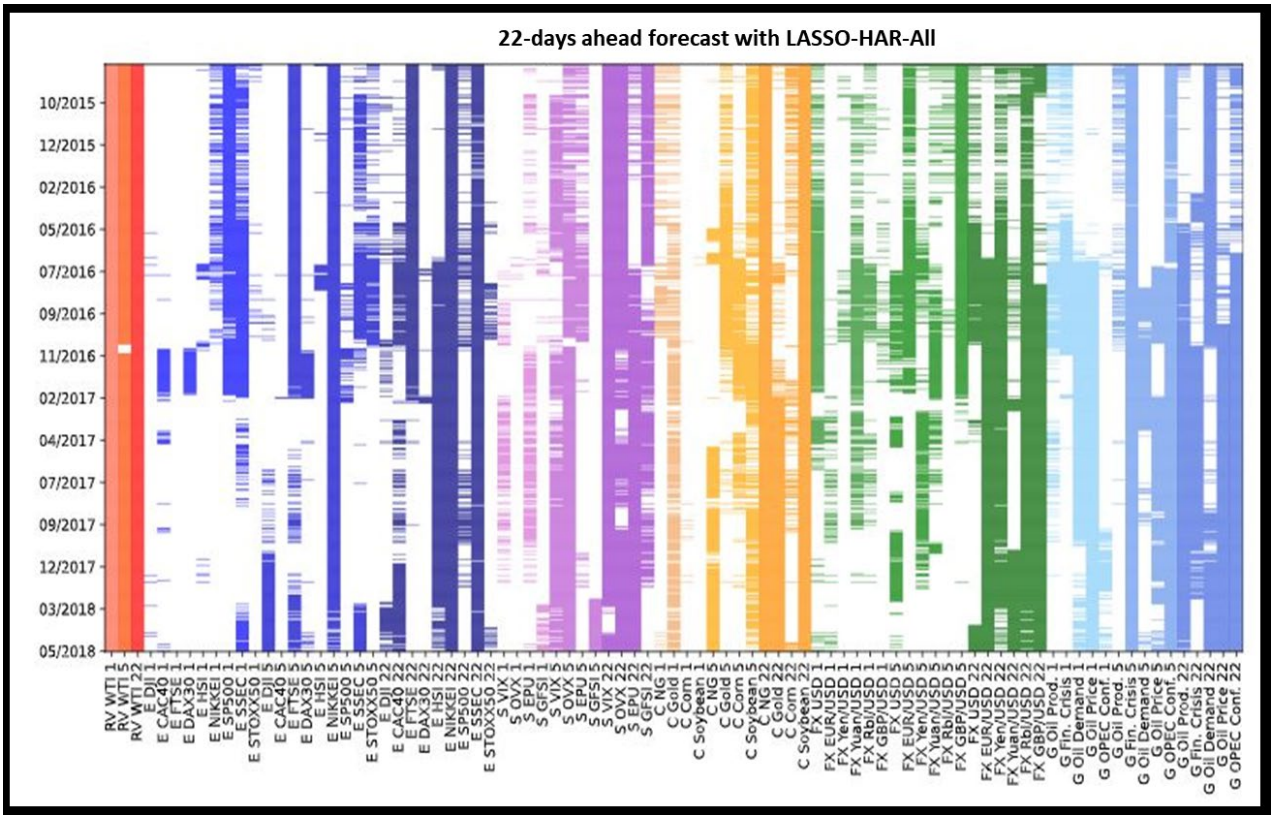
(a) Variable Selection Over Time with the LASSO-HAR-All (1-day ahead forecast)



(b) Variable Selection Over Time with the LASSO-HAR-All (5-days ahead forecast)



(c) Variable Selection Over Time with the LASSO-HAR-All (22-days ahead forecast)



Lastly, we also compare the variable importance. In short, the variable importance is the absolute, standardized coefficient for a predictor over the sum of all absolute, standardized coefficients from a LASSO regression. We then take the average for each predictor over the entire out-of-sample period. See Figure 2, which is further discussed on the next page.

Figure 2

Ranked Variable Importance of Individual Variables (Top 10) Across the Applied Forecasting Horizons

Rank	1 day	VI	5 days	VI	22 days	VI
1	RV WTI 22	47.8%	RV WTI 22	53.6%	RV WTI 22	66.4%
2	RV WTI 5	30.9%	FX Yuan/USD 22	9.4%	FX EUR/USD 22	5.6%
3	RV WTI 1	8.9%	RV WTI 5	7.4%	FX Yuan/USD 22	4.1%
4	FX USD 5	5.1%	S OVX 22	5.2%	FX Yen/USD 22	2.8%
5	S OVX 22	3.5%	FX Yen/USD 5	4.6%	E NIKKEI 22	2.7%
6	S VIX 5	2.0%	FX Yuan/USD 5	4.3%	E HSI 22	2.6%
7	FX USD 22	0.7%	FX GBP/USD 22	3.0%	S GFSI 22	1.6%
8	FX USD 1	0.6%	FX USD 5	2.7%	RV WTI 5	1.5%
9	G OiL Price 1	0.5%	FX USD 22	2.2%	FX GBP/USD 22	1.3%
10	S OVX 1	<0.1%	FX Yen/USD 22	2.0%	C Soybean 22	1.3%

We first look at all individual predictor-horizon combinations and find the 22-day realized volatility to be the most important variable over all three forecasting horizons. It makes up 47.8% in the 1-day ahead case up to 66.4% in the 22-day ahead forecast. For the 1-day ahead prediction, the order is 22-day RV, 5-day RV, and 1-day RV. For the longer forecasting horizons, the previous day's RV falls entirely out of the top 10 predictors. Predictors from other channels contribute more to the forecasts. Within the top 10 predictors, we find FX and sentiment predictors over all forecasting horizons. Other information channels appear also, but only with incremental contribution (*e.g.,* Google attention to Oil Price for 1-day ahead, equity market for 22-days ahead).

Finally, we condense the information channels regardless of their horizon and look at their importance over the entire period. As stated above, the past realized volatility is by far the strongest contributor to the prediction. Interestingly, FX is the runner-up for all horizons with contributions between 6.3% (1-day) to even 28.8% (5-days). Information from other commodities and Google attention play a minor role for the oil price volatility forecast. In the 1-day ahead case, *e.g.*, they are never selected.

Conclusions

This paper shows that extending HAR models with exogenous information and recent ML techniques improves both forecasting accuracy and portfolio performance. ML-based variable selection, particularly through LASSO, improves statistical accuracy, while BMA models provide the greatest economic gains. We find that the relevance of predictors is horizon-dependent: short-term forecasts are dominated by past volatility and use only a few additional information channels. For longer horizons, sentiment, equity, commodity, and especially FX markets have more importance without replacing realized volatility. Google-based indicators play a role but remain of limited importance compared to traditional financial channels. Overall, our results highlight that while past realized volatility remains the most important predictor, incorporating broader information channels and ML-driven selection yields meaningful improvements in both predictive and economic value.

References

Corsi, F., 2009, "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, Vol. 7, No. 2, Spring, pp. 174-196. https://doi.org/10.1093/jjfinec/nbp001.

Degiannakis, S., Filis, G., Klein, T. and T. Walther, 2022, "Forecasting Realized Volatility of Agricultural Commodities," *International Journal of Forecasting*, Vol. 38, No. 1, January-March pp. 74-96. https://doi.org/10.1016/j.ijforecast.2019.08.011.

Hansen, P. R., Lunde, A. and J. M. Nason, 2011, "The Model Confidence Set," *Econometrica*, Vol. 79, No. 2, March, pp. 453-497. https://doi.org/10.3982/ECTA5771.

Keywords

Forecasting, crude oil, realized volatility, exogenous predictors, machine learning.

Author Biographies

JIAWEN LUO, Ph.D.
School of Business Administration, South China University of Technology, China

Jiawen Luo is an Associate Professor at the South China University of Technology. Her research interests include financial economics, financial risk analysis, and energy finance. Her research is published in journals such as *International Journal of Forecasting*, *Journal of Forecasting*, *Journal of Futures Markets*, *Energy Economics*, and *European Journal of Finance*.

TONY KLEIN, Ph.D.
Faculty of Business and Economics, Technische Universität Chemnitz, Germany

Tony Klein is a Professor of Finance and holds the Chair of Finance and Banking at Technische Universität Chemnitz, Germany. He previously held positions at Queen's University Belfast, U.K., and Universitat de Barcelona, Spain. His research interests include econometric modelling and machine learning within the fields of finance and economics, energy and carbon economics, and the financialization of commodity markets and connectedness to financial markets.

THOMAS WALTHER, Ph.D.
Utrecht School of Economics, Utrecht University, The Netherlands and Technische Universität Dresden, Germany

Thomas Walther is an Associate Professor of Finance with the Utrecht School of Economics, Utrecht University, The Netherlands. Formerly, he was an Assistant Professor for Energy Finance (Non-Tenure Track) at the School of Finance, University of St. Gallen, Switzerland. He is also a Research Fellow for Finance at Technische Universität Dresden, Germany from which he received a B.Sc. (2010) and M.Sc. (2013) in Industrial Engineering and Management and a Ph.D. (2017) in Business and Economics. His research focuses on commodity/energy/climate finance. He initiated the European Sustainable Finance Ph.D. Workshop in 2023 in Utrecht (with more information about the latest workshop available here: http://www.esfpw.eu). He is also the co-founder of the Dutch Sustainable Finance Network (http://www.dsfn.eu), which has organized an annual workshop since 2024.

QIANG JI, Ph.D.
Institutes of Science and Development, Chinese Academy of Sciences, China

Qiang Ji is the executive director of the support center for discipline research in the Institutes of Science and Development, Chinese Academy of Sciences (ISDCAS). He presides over a program supported by the Excellent Young Scientists Fund of the National Natural Science Foundation of China and is granted the Excellent Member by the Youth Innovation Promotion Association of CAS. He has been engaged in research work in the fields of energy strategic management, energy and climate finance, energy forecasting, and risk management. He is the co-founder of the China Energy Finance Network and the vice president of the International Society for Energy Transition Studies and Society for the Studies of Climate Finance. He is a founding editor of the *Journal of Climate Finance* and serves as senior editor and associate editor for over 10 energy and financial journals, such as *Energy Economics*, *Finance Research Letters*, and *International Review of Financial Analysis*. He has published more than 150 SCI/SSCI indexed articles in peer-reviewed journals and has been named in the global "Highly Cited Researchers" lists in the year 2021–2023 released by Clarivate Analytics and "World's Top 2% Scientists" lists released by Stanford University in the year 2020–2023.