Forecasting Realized Volatility of the Oil Future Prices via Machine Learning

2024.11.29.

Myung Jun Kim

POSTECH IME

Joint Work with Taeyoon Kim, Byung-June Kim, and Bong-Gyu Jang

Prepared for 2024년 한국금융공학회 추계학술대회

Contents

Abstract

Abstract

- This paper explores the potential use of machine learning models in crude oil realized volatility forecasting through a variety of empirical analyses and robustness checks
- Although the conventional Heterogeneous Autoregressive (HAR) model is widely accepted, the machine learning models with the HAR factors can significantly improve its forecasting performance
- We also found that macroeconomic variables such as supply factors, implied volatility indices and uncertainty factors can be useful in forecasting oil volatility

1 Introduction

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Introduction

- COVID-19 has created a big lurch in the commodity market
  - Stumbling demand pushes oil future prices down to even the negative territory, -$36.98, in May 2020

- Even before the global impact of COVID-19, oil price uncertainty has steadily increased over recent decades, which can be a significant threat to the global economy
- In addition to economic decision-making, *Volatility* is one of the key variables for trading strategies, asset allocation, and risk management

Introduction

- This paper explores the possibility of the potential usage of machine learning models in the field of volatility forecasting
  - Comparison of various forecasting models with a rich set of data and various forecasting horizons guide us to the fact that a combination of conventional models and ML techniques can improve the forecasting performance in out-of-sample

- We show that the ML models with factors in the HAR model as input variables have the potential to enhance the out-of-sample forecasting performance with one-week, bi-week, and month ahead horizon
- We examine the out-of-sample forecasting performance of ten different models as well as their robustness with five different types of performance metrics
- We consider the period from 2002.04. to 2024.04. to include the Great Recession and COVID-19
- In addition, we construct a rich set of data including uncertainty indices, following Miao et al. (2017), Wei et al. (2017), and Ma et al. (2018)

2

Data

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Data: Realized Volatility

- We construct the realized volatility of oil future contracts on the WTI crude oil as follows:
- $RV_t = \sqrt{\sum_{j=1}^{N_t} r_{t,j}^2}\,, for\ r_{t,j} = 100 \times \log\left(\frac{p_{t,j}}{p_{t,j-1}}\right),$
- N_t: # of business days in the t-th week/bi-week/month
- $p_{t,j}$: Daily WTI future prices on j-th business day of the t-th week/bi-week/month

Data: Explanatory Variables

- 6 Categories
  - Prices
  - Supply Factors
  - Demand Factors
  - Financial Factors
  - Implied Volatility Indices
  - Uncertainty Factors

Factor Group	Individual Variable	Code	Frequency	Source
Prices	WTI (West Texas Intermediate) future prices	-	Daily	Energy Information Administration
	WTI spot prices	P1	Daily	Energy Information Administration
	Brent oil spot prices	P2	Daily	Energy Information Administration
	NGL (Natural Gas Liquids) future prices	P3	Daily	Energy Information Administration
	NGL spot prices	P4	Daily	Energy Information Administration
Supply Factors	Global crude oil production	S1	Monthly	JODI-Oil Database
	Global crude oil stock	S2	Monthly	JODI-Oil Database
	Global crude oil export	S3	Monthly	JODI-Oil Database
	Total OPEC production capacity	S4	Monthly	Energy Information Administration
	Capacity utilization rate	S5	Weekly	Energy Information Administration
Demand Factors	Global crude oil import	D1	Monthly	JODI-Oil Database
	Liquid fuels consumption in World	D2	Monthly	Energy Information Administration
	PPI in China	D3	Monthly	Federal Reserve Bank of St. Louis Economic Database
	PPI in US	D4	Monthly	Federal Reserve Bank of St. Louis Economic Database
	PPI in EU	D5	Monthly	Federal Reserve Bank of St. Louis Economic Database
Financial Factors	S&P 500 Adjusted Close	F1	Daily	Yahoo Finance
	Japan / US Foreign Exchange Rate	F2	Daily	Federal Reserve Bank of St. Louis Economic Database
	US / Euro Foreign Exchange Rate	F3	Daily	Federal Reserve Bank of St. Louis Economic Database
	US / UK Foreign Exchange Rate	F4	Daily	Federal Reserve Bank of St. Louis Economic Database
	China / US Foreign Exchange Rate	F5	Daily	Federal Reserve Bank of St. Louis Economic Database
	Federal Funds Rate	F6	Monthly	Federal Reserve Bank of St. Louis Economic Database
	MSCI World Standard (Large+Mid Cap)	F7	Monthly	MSCI
Implied Volatility Indices	CBOE Volatility Index	V1	Daily	Yahoo Finance
	CBOE Crude Oil Volatility Index	V2	Daily	Yahoo Finance
	CBOE DJIA Volatility Index	V3	Daily	Yahoo Finance
	CBOE Gold Volatility Index	V4	Daily	Yahoo Finance
Uncertainty Factors	Global Economic Policy Uncertainty (current)	U1	Monthly	Davis (2016), https://www.policyuncertainty.com
	Global Economic Policy Uncertainty (ppp)	U2	Monthly	Davis (2016), https://www.policyuncertainty.com
	Daily Infectious Disease Equity Market Volatility Tracker	U3	Daily	Baker et al. (2019) and Baker et al. (2020) https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Economic Policy Uncertainty	U4	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Monetary Policy	U5	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Fiscal Policy (Taxes or spending)	U6	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Taxes	U7	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Government Spending	U8	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Healthcare	U9	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in National Security	U10	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Entitlement Programs	U11	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Regulation	U12	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Financial Regulation	U13	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Trade Policy	U14	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	US Economic Policy Uncertainty in Sovereign Debt and currency crises	U15	Monthly	Baker et al. (2016), https://www.policyuncertainty.com
	World Uncertainty Index	U16	Quarterly	Ahir et al. (2018), https://worlduncertaintyindex.com/
	World Pandemic Uncertainty Index	U17	Quarterly	Ahir et al. (2018), https://worlduncertaintyindex.com/

Table 1. There are six groups of data - prices, supply factors, demand factors, financial factors, implied volatility indices, and uncertainty factors. The dependent variable is WTI future price, and the rest of the data are explanatory variables used in the models.

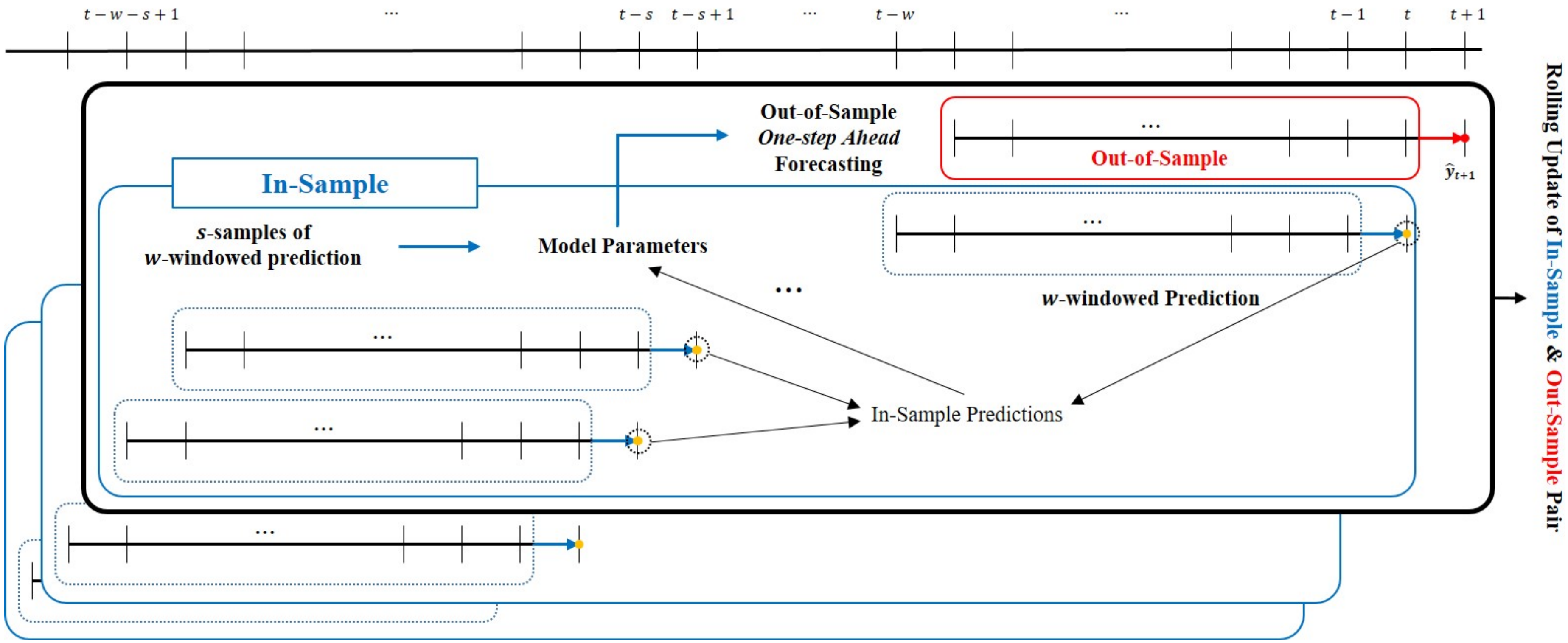
POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

3 Forecasting Models

Cross Validation

- In the time-series model, preventing the use of future information is a pertinent first step
- Similar to Gu et al. (2020), we construct the out-of-sample data and design cross-validation as shown below



Forecasting Models

- Heterogeneous Autoregressive with Exogenous Variables (HAR-X)
  - Contains two explanatory variables which represent short-term and long-term memories
  - $$\begin{cases} \widehat{RV_{t+1}^{W}} = c_W + \beta_{W,w} RV_t^W + \beta_{W,m} RV_{t-4:t}^W + \beta_{W,q} RV_{t-13:t}^W + \sum_{k=0}^{w-1} \sum_{m=1}^{n} \eta_{W,m,t-k} x_{m,t-k} + \epsilon_{W,t+1}, \\ \widehat{RV_{t+1}^{2W}} = c_{2W} + \beta_{2W,w} RV_t^{2W} + \beta_{2W,q} RV_{t-6:t}^{2W} + \beta_{2W,y} RV_{t-26:t}^{2W} + \sum_{k=0}^{w-1} \sum_{m=1}^{n} \eta_{2W,m,t-k} x_{m,t-k} + \epsilon_{2W,t+1}, \\ \widehat{RV_{t+1}^{M}} = c_M + \beta_{M,w} RV_t^M + \beta_{M,q} RV_{t-3:t}^M + \beta_{M,y} RV_{t-12:t}^M + \sum_{k=0}^{w-1} \sum_{m=1}^{n} \eta_{M,m,t-k} x_{m,t-k} + \epsilon_{M,t+1}, \end{cases}$$
- Time-Varying Parameter Heterogeneous Autoregressive (TVP-HAR)
  - Extension of HAR, considering the parameter changes over time
  - Commonly used to enhance the power of HAR

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Forecasting Models

- **Least Absolute Shrinkage and Selection Operator (LASSO)**
  - Reduces the risk of overfitting by adding a penalty term to the cost function
  - $\min_{\phi,\eta}\left[\left(RV_{t+1}-\left(\sum_{i=0}^{w-1}\phi_i RV_{t-i}+\sum_{k=0}^{w-1}\left\{\sum_{m=1}^{n}\eta_{m,t-k}x_{m,t-k}\right)\right)^2\right\}\right]$, subject to $\sum|\phi|+\sum|\eta|<c$(constant)

- **Elastic Net**
  - Combinatorial extension of the Ridge and the LASSO
  - Considers both L1 and L2 norms

Forecasting Models

- Decision Tree Regression (DTR)
  - Builds models in the form of a tree structure by dividing the data set into smaller subsets while gradually elaborating related decision trees

- Random Forest Regression (RFR)
  - Learning algorithm that uses an ensemble learning method for several decision trees
  - Generates multiple decision trees and averages them to reduce noise and the risk of overfitting.

- Gradient Boosting Regression (GBR)
  - Produces a model from an ensemble of weak predictive models (trees)
  - Updates its predictor by calculating the negative gradient of the loss criteria, then regress a tree to those residuals

Forecasting Models

- **Artificial Neural Network (ANN)**
  - Computational model inspired by the human brain
  - Designed to recognize the patterns and make decisions
  - It consists of interconnected layers of nodes (neurons), including an input layer, one or more hidden layers, and an output layer

- **Recurrent Neural Network (RNN)**
  - Specific type of ANN, designed to process sequential data
  - Have connections from directed cycles, allowing them to retain information from earlier input sequences

4 Empirical Analysis

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Empirical Results and Potential Explanation

- Overall Performance

Forecasting Model	Number of Feature Selection	$\hat{R}^2_{oos}$ With Uncertainty M	With Uncertainty 2W	With Uncertainty W	Without Uncertainty M	Without Uncertainty 2W	Without Uncertainty W
HAR	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		(0.000E+00)	(0.000E+00)	(0.000E+00)	(0.000E+00)	(0.000E+00)	(0.000E+00)
TVP-HAR	-	-0.0203	0.2508	0.2392*	-0.0203	0.2508	0.2392*
		(2.079E-13)	(2.648E-19)	(1.421E-19)	(2.079E-13)	(2.648E-19)	(1.421E-19)
HAR-X	0	-0.0203	0.2508	0.2392*	-0.0203	0.2508	0.2392*
		(2.079E-13)	(2.648E-19)	(1.421E-19)	(2.079E-13)	(2.648E-19)	(1.421E-19)
	10	0.0958	0.3445	0.2528*	-0.2564	0.2779	0.1377*
		(4.276E-11)	(8.056E-16)	(1.967E-17)	(2.145E-07)	(1.239E-07)	(3.584E-14)
	20	0.1260	0.3085	0.2505*	0.0739	0.2377	0.0311
		(4.334E-12)	(2.286E-15)	(5.876E-17)	(1.379E-15)	(3.635E-17)	(4.732E-15)
	all	**0.3708***	**0.4154***	**0.3236***	0.0125	0.2106	-1.2143
		(4.671E-11)	(3.443E-13)	(3.356E-12)	(6.184E-17)	(1.366E-16)	(6.322E-06)
LASSO	0	0.0599	0.3018	0.2756*	0.0599	0.3018*	0.2756*
		(5.319E-15)	(4.384E-20)	(1.552E-16)	(5.319E-15)	(4.384E-20)	(1.552E-16)
	10	0.1964	0.3622*	0.2766*	0.0452	0.2335	0.2605*
		(1.983E-14)	(1.719E-19)	(1.717E-16)	(1.151E-16)	(2.426E-15)	(2.258E-17)
	20	0.2110	0.3623*	0.2766*	0.0550	0.2458	0.2603*
		(1.528E-14)	(1.723E-19)	(1.718E-16)	(6.858E-17)	(3.340E-15)	(2.308E-17)
	all	**0.4007**	**0.4512***	**0.3547***	0.2554*	0.3980*	0.2439*
		(2.232E-17)	(8.323E-14)	(5.456E-14)	(7.102E-17)	(5.951E-16)	(1.809E-14)
ElasticNet	0	0.0613	0.2985	0.2698*	0.0613	0.2985*	0.2698*
		(6.682E-15)	(3.345E-20)	(1.053E-17)	(6.682E-15)	(3.345E-20)	(1.053E-17)
	10	0.2137	0.3849*	0.2969*	0.1537*	0.3245*	0.2714*
		(2.232E-14)	(4.921E-19)	(2.775E-17)	(3.724E-15)	(3.286E-14)	(3.825E-18)
	20	0.2237	0.3807*	0.2971*	0.2023*	0.3319*	0.2726*
		(2.426E-14)	(3.516E-19)	(3.317E-17)	(1.340E-20)	(1.301E-20)	(5.200E-18)
	all	**0.4043***	**0.4397***	**0.3533***	0.2396*	0.3884*	0.2007*
		(3.434E-17)	(3.443E-14)	(5.409E-14)	(6.246E-17)	(8.830E-17)	(1.019E-13)
DTR	0	0.0513	-0.0220	-0.2066	0.0240	-0.1064	-0.1085
		(1.209E-12)	(7.737E-08)	(1.258E-17)	(2.446E-12)	(2.973E-08)	(7.020E-16)
	10	0.3294*	**0.3065**	-0.2111	0.2264*	0.0022	-0.2102
		(4.235E-14)	(3.638E-12)	(7.734E-15)	(7.963E-16)	(4.163E-15)	(1.461E-19)
	20	**0.3452***	0.2293	-0.2178	0.2755*	-0.0087	**0.0697**
		(2.868E-15)	(5.137E-13)	(4.559E-18)	(3.774E-14)	(1.436E-13)	(1.118E-12)
	all	0.3308*	0.1281	-0.2805	0.0503	0.1752	-0.0473
		(1.535E-15)	(1.858E-11)	(7.418E-19)	(5.726E-08)	(2.422E-18)	(3.920E-13)
RFR	0	0.2827*	0.3737*	0.2297	0.2823*	0.3720*	0.2428*
		(1.293E-12)	(2.818E-16)	(1.783E-17)	(1.240E-12)	(2.061E-16)	(3.330E-17)
	10	0.4623*	**0.4884***	0.3434*	0.4228*	0.4060*	0.2760*
		(3.638E-16)	(1.350E-15)	(3.149E-16)	(1.145E-13)	(2.940E-18)	(2.131E-15)
	20	**0.4909***	0.4678*	0.3743*	0.4216*	0.4604*	0.3352*
		(1.028E-15)	(2.964E-17)	(2.684E-17)	(1.606E-13)	(3.747E-17)	(2.574E-16)
	all	0.4843*	0.4668*	**0.3751***	0.4175*	0.4470*	0.3328*
GBR	0	0.2160	0.3269	0.2092	0.2296*	0.3205*	0.2233*
		(5.184E-13)	(3.493E-17)	(6.703E-16)	(6.920E-13)	(2.413E-17)	(3.080E-16)
	10	0.4292*	**0.4891***	0.3221*	0.3102*	0.3305*	0.0601
		(1.710E-15)	(1.786E-13)	(1.266E-15)	(1.699E-06)	(1.026E-16)	(2.916E-19)
	20	**0.4783***	0.4644*	**0.3601***	0.3382*	0.3751*	0.2670*
		(2.345E-15)	(1.528E-16)	(3.625E-16)	(2.420E-16)	(4.168E-18)	(2.752E-19)
	all	0.4641*	0.4709*	0.3591*	0.3575*	0.3919*	0.2863*
		(1.117E-15)	(1.059E-16)	(1.366E-16)	(3.946E-17)	(8.140E-19)	(2.269E-17)
ANN	0	-0.0483	0.2008	0.2090	-0.0554	0.2214	0.1738*
		(1.238E-13)	(8.166E-20)	(6.319E-18)	(4.572E-14)	(1.274E-19)	(4.084E-18)
	10	0.2757*	0.4146*	**0.3610***	0.1210	0.4633*	0.2837*
		(9.829E-11)	(5.968E-13)	(3.832E-13)	(4.540E-14)	(1.464E-13)	(8.791E-13)
	20	0.2485	0.3708*	0.2801*	0.2626*	**0.5103***	0.0482
		(2.381E-12)	(2.239E-15)	(4.347E-14)	(1.589E-13)	(1.832E-12)	(4.429E-13)
	all	**0.2923***	0.3285	0.2618*	-2.2483	-0.0852	-0.7148
		(5.360E-12)	(4.946E-16)	(1.763E-13)	(2.300E-18)	(3.332E-14)	(1.649E-11)
RNN	0	0.4398*	0.3868*	0.3121*	0.4372*	0.3878*	0.3132*
		(5.641E-10)	(2.328E-15)	(8.896E-15)	(6.683E-10)	(8.813E-16)	(1.264E-14)
	10	0.4711*	**0.4777***	0.3956*	0.4800*	0.4696*	**0.4070***
		(1.647E-11)	(1.101E-13)	(7.267E-14)	(5.836E-11)	(1.699E-12)	(6.483E-15)
	20	**0.4906***	0.4549*	0.3841*	0.4803*	0.4457*	0.3488*
		(4.541E-10)	(3.328E-15)	(1.998E-15)	(3.066E-10)	(2.522E-07)	(3.794E-14)
	all	0.4813*	0.4316*	0.3347*	0.4742*	0.4135*	0.3280*
		(6.191E-10)	(2.055E-13)	(4.252E-15)	(8.149E-10)	(5.288E-14)	(4.572E-15)

Table 2. Forecasting Performance Compared to the Benchmark HAR. $\hat{R}^2_{oos}$ is the performance measurement compared to the benchmark HAR. For each model and each forecasting horizon, bold numbers are the best performance measurements with feature selection and data set. For each time horizon, models that pass the MCS test with a significant level of 0.05 have an asterisk with each performance measure. The numbers in parentheses indicate the p-values of residual stationarity.

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Empirical Results and Potential Explanation

- **Overall Performance**
  - HAR-X and TVP-HAR demonstrate a less impressive performance compared to the HAR
  - Linear or simple models demonstrate inconsistent performance
    - Superior performance when forecasting past periods
  - # of feature selection plays important role
    - Linear or Simple: Better with large number
    - Nonlinear: Too much is as bad as too little
  - Nonlinear ML models with ensemble shows great performance

Empirical Results and Potential Explanation

- **Forecasting Horizon**
  - The longer the forecast horizon, the better the performance
  - However, BM model decreases as the forecast horizon increases
  - Short-term and long-term variables of HAR do not fully reflect the momentum of the volatility
  - RFR and RNN models tend to make robust predictions over the period

Empirical Results and Potential Explanation

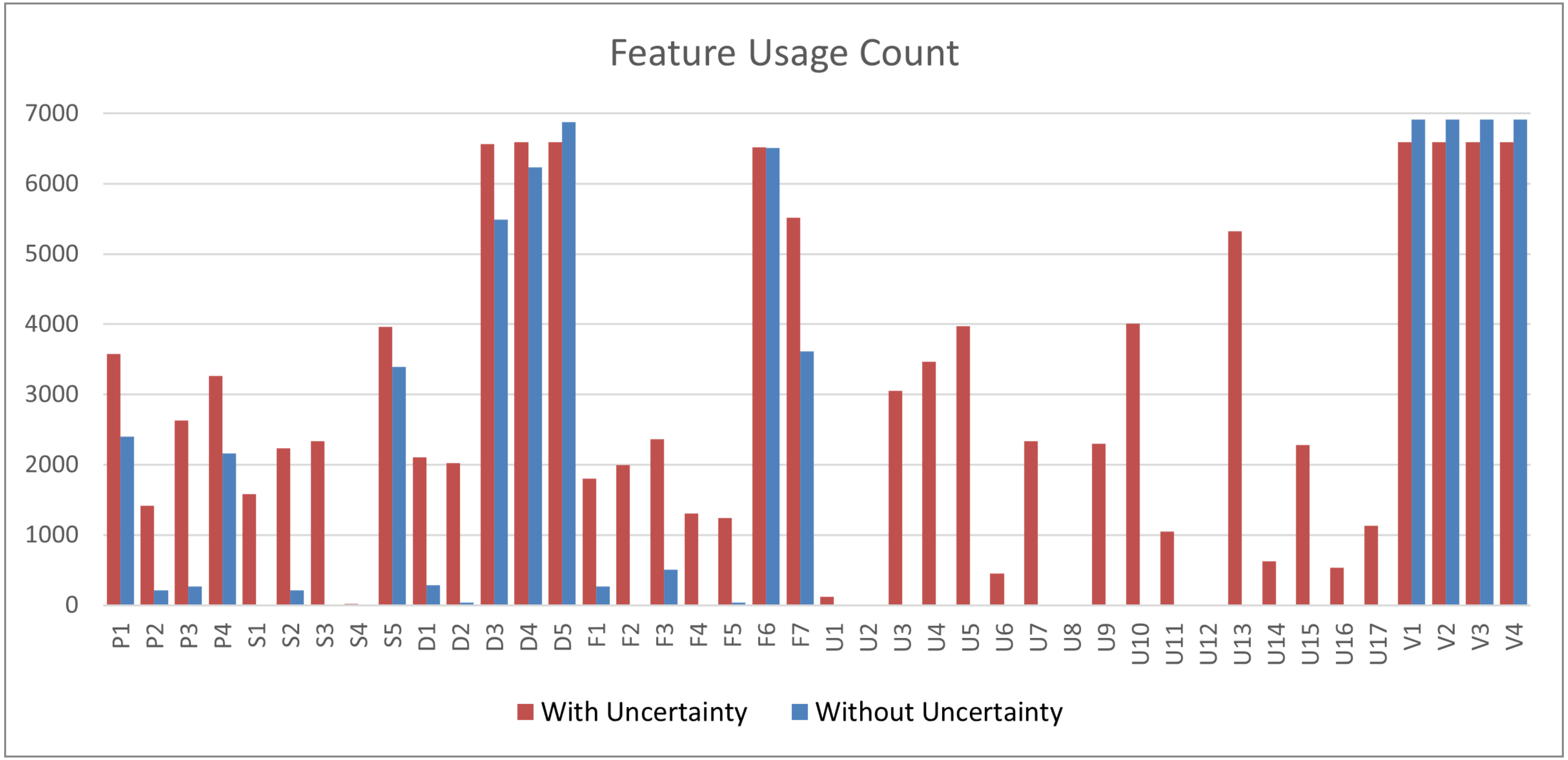
- **Robustness to sub-period**
  - LASSO, Elastic Net: Great performance for longer sub-period
    - However, the ranking of the model did not fall in recent period
    - Oil price have been very volatile recently
  - RFR, RNN: Stable even during the COVID-19

Empirical Results and Potential Explanation

- **Uncertainty Indices**
  - Can push the accuracy of models even higher
  - Results vary slightly by model
    - Elastic Net, DTR, and RNN: About half of the cases
    - LASSO, RFR, and ANN: Most case
  - Models with shorter forecast horizon, the better the use of uncertainty indices
    - Can reduce overfitting and are more influential in dealing with noise in short-term

Empirical Results and Potential Explanation

- Feature Selection



Empirical Results and Potential Explanation

- **Highly Selected Features**
  - Volatility indices, PPI in CN, US, EU, Federal funds rate, MSCI World Index
  - WTI, NGL Spot prices, Capacity utilization rate
  - US economic policy uncertainty indices (Financial regulation, Monetary policy, National security, economic policy regulation)

- **Lowly Selected Features**
  - OPEC production capacity, Brent oil Spot price, US/UK, CN/UK FX rate
  - US economic policy uncertainty indices (Fiscal policy, Government spending, regulation, trade policy), WUI

Empirical Results and Potential Explanation

- **Selected in the Models with Uncertainty Indices**
  - Price features, Global crude oil production, stock and export
  - MSCI World Standard Index
  - Global crude oil import and Liquid fuels consumption

Potential Explanation

- **Why the performance orders are different for each forecast horizon?**
  - HAR uses only the lagged values, reflecting the autocorrelation
  - Different models are exposed to the different levels of AR effect, therefore the performance orders might change
  - Features selected varies depending on the forecast horizon, suggesting that the nature of forecasting short-term and long-term horizon is distinct

Potential Explanation

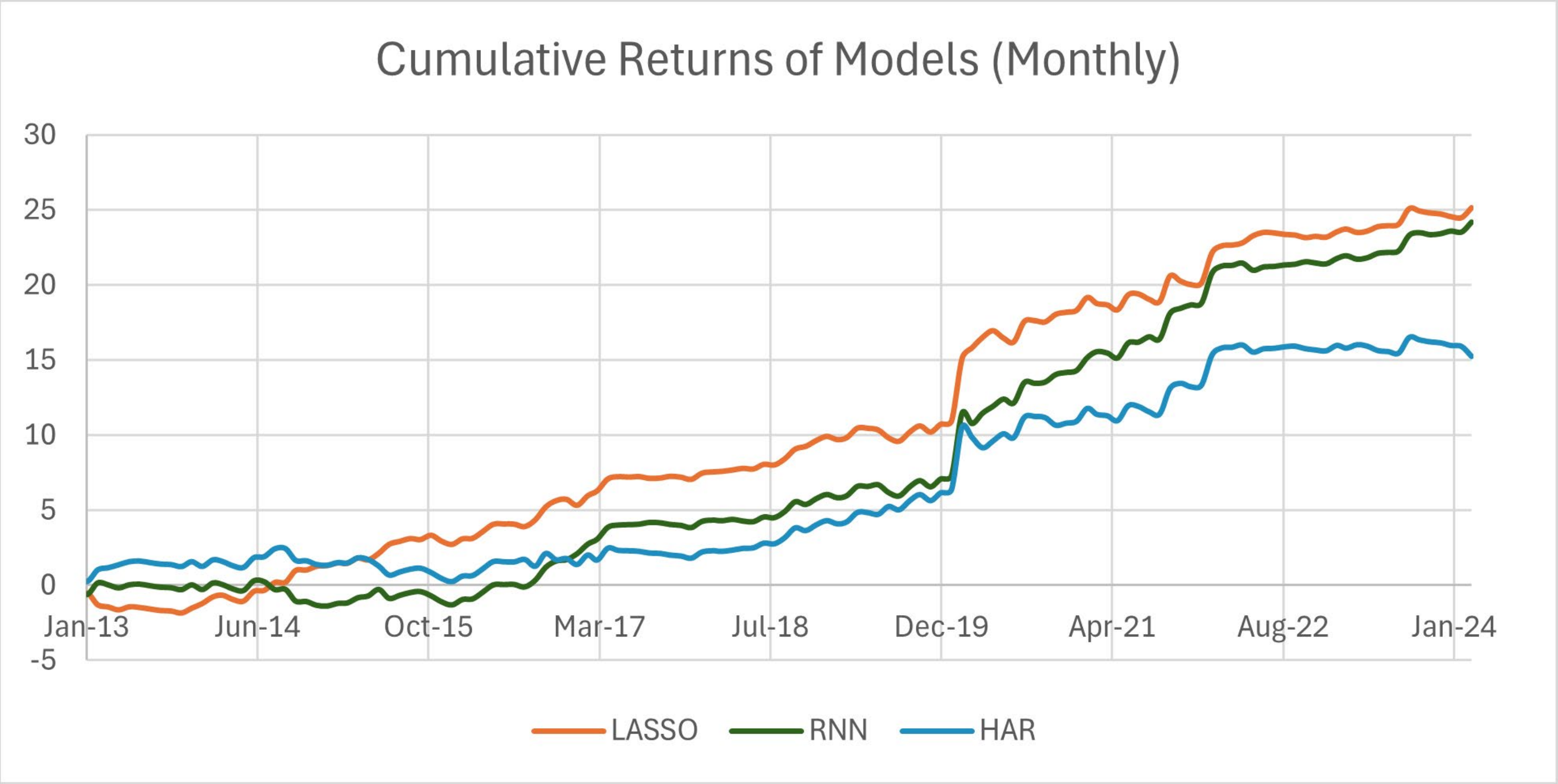
- **How do some ML models outperform linear model?**
  - Assumption of Gaussian distribution seems most restricted to the linear models
    - Fat-tailed data
  - Overfitting
    - Some ML models are in-sample over-fitted
  - Convergence problems
    - Some ML models might not converge well

Discussion

- **Volatility Trading Backtest**
  - Invest in OVX(ETF)
  - Simple Strategy
    - Long($\boldsymbol{d} = +1$) when the prediction is higher than current
    - Short($\boldsymbol{d} = -1$) when the prediction is lower than current
    - $\mathrm{r}^{\mathrm{i}} = \sum_{t=1}^{t_{N-s}} \frac{(\mathrm{RV}_{t+s} - \mathrm{RV}_t)\mathrm{d}_t^{\mathrm{i}}}{\mathrm{RV}_t}$

Discussion

- Volatility Trading Backtest
- Return: 2,448%, 2,355%, 1,589%



5 Conclusion

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM

Conclusion

- This paper compares and analyzes the predictability of the realized volatility of the crude oil future prices
  - 10 Various forecasting models includes both linear and ML models
  - From 2002 April to 2024 April to include the Great Recession and COVID-19
  - 6 different categories(Prices, Supply, Demand, Financial, Implied Volatility Indices, Uncertainty) of Features

- HAR is popular model however, some ML models outperformed, such as RFR and RNN
- Uncertainty features can enhance the model performances
- Implied volatility indices and PPI, federal funds rate, MSCI World Index were most selected features

End of Document

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

FIRM