Prague Economic Papers
2024, Volume 33, Issue 6
https://doi.org/10.18267/j.pep.879

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

The Second RP-PCA Factor and Crude Oil Price Predictability

Qi Shi

Department of Finance, Zhe Jiang University of Finance and Economics, Hang Zhou, China, email: qishi@zufe.edu.cn

Abstract

Although it is notoriously difficult to utilize financial ratios to forecast the crude oil market prices, our study challenges this perception and reveals that the second risk premium principal component analysis (RP-PCA) factor may contain statistically significant information for both in-sample and out-of-sample forecasts of future crude oil prices. Our evidence illustrates that the second RP-PCA factor substantially outperforms many other popular predictors (approximately 30 conventional predictors) in forecasting crude oil prices and generating adequate higher values of economic profits. We conduct a range of informative tests, including bootstrap simulation, success ratio tests, alternative out-of-sample evaluation periods, and structure break tests. Furthermore, we illustrate that the forecasting ability of the second RP-PCA factor may stem from its ability to forecast oil market sentiment. Our study presents a novel and indicatable financial instrument for policymakers to predict crude oil prices robustly. The theoretical motivation of this study links to Cochrane's (2005) framework for general candidate factors in asset pricing.

Keywords: RP-PCA factor, forecasting, crude oil prices, economic profits, oil market sentiment, policymakers

JEL Classification: C53, G12, G17, Q47

1. Introduction

The oil industry is sometimes known as the "lifeblood industry" since oil plays a crucial role in key economic activities. The predictability of crude oil price is one of the most popular topics in economics, where it is often[1] forecast via conventional economic/financial variables. Despite

1 For example, our study scans more than 30 reasonable economic/financial variables, but the overall results are disappointing in that most of them fail to properly forecast crude oil prices.

the difficulty of forecasting, numerous studies explore this research gap with fruitful outcomes, such as the technique factor (Yin and Yang, 2016), stock prices (Chen, 2014), the dollar exchange rate (Chen and Chen, 2007; Brahmasrene et al., 2014), monetary policy (Hamilton and Herrera, 2004; Herrera and Pesavento, 2009), oil shocks (Kilian, 2009; Aastveit et al., 2015), and bond yield (Dai and Kang, 2021). Instead, our study is the first to use updated actual financial risk ratios to forecast oil prices. Previous studies on asset pricing suggest that a theoretically sound pricing factor in the framework of the pricing kernel[2] may forecast future asset returns and future economic activity to a certain extent (see Rapach et al., 2005; and Cochrane, 2005). Cochrane (2005, 2007) suggested that predictors can more plausibly forecast equity premiums if these predictors can also forecast economic activity[3]. Financial ratios in asset pricing models are widely popular for forecasting excess stock returns or economic activity (Lettau and Ludvigson, 2001; Maio and Santa-Clara, 2012; Boons, 2016; Cooper and Maio, 2019; and others); however, few studies have reported their valid role in predicting oil prices. For example, there is no empirical link between the predictability of oil prices and RP-PCA factors. The predictability of Lettau and Pelger's (2020) RP-PCA factors is successfully proven in the financial stock market (Shi, 2023). Instead, our study aims to fill this gap, in which we further extend Shi (2023) to explore whether predictive RP-PCA factors may also forecast crude oil prices.

In the past literature, conventional economic/financial variables generally have difficulty forecasting crude oil prices. Hence, the prevailing research questions are as follows: can crude oil prices be predicted, and by which sound financial variables? And why? In contrast to previous financial ratios, the RP-PCA methods extract five latent factors from large financial datasets (i.e., a large cross-section of portfolios sorted into 37 well-known anomalies). RP-PCA factors refine the efficient information from a large cross-section of financial portfolios sorted into 37 anomalies. Cochrane (2005) noted that any sound pricing factor in the linear asset pricing model may forecast asset return and economic activity. Since the RP-PCA factors fall within the theoretical framework of APT and adequately fit the cross-section and time series of a large panel of equity returns, we suggest that the RP-PCA factors may contain much more additional information than any individual macro/financial factor in forecasting economic activity (e.g., the crude oil market prices).

2 In our case, Lettau and Pleger (2020) note that RP-PCA factors should be incorporated into Ross's (1976) APT model. In other words, RP-PCA factors should go into the pricing kernel of Lettau and Pleger's (2020) multifactor APT model. RP-PCA refers to risk-premium principal component analysis. Lettau and Pelger (2020) clearly state the details of RP-PCA factors that are extracted by the RP-PCA method.

3 Shi and Li (2022) provide such a study on conventional financial ratios in forecasting economic activity.

We illustrate our key contributions as follows. First, under the theoretical framework of Cochrane (2005), we comprehensively scan the link between recent popular economic/financial ratios and future crude oil prices, and we highlight the significant and distinctive role of the second RP-PCA factor in both in- and out-of-sample forecasting of crude oil prices. Second, our rich array of evidence further illustrates that both the second RP-PCA factor and the corporate bond yield (see Dai and Kang, 2021) substantially outperform 30[4] other popular potential economic or financial predictors with the highest explanatory power. Consistently, in terms of the economically significant asset allocation estimates, the second RP-PCA factor can generate positive and the highest utility and Sharpe ratio gains among many benchmark predictors. Finally, we reveal that the second RP-PCA factor has a significant link with oil market sentiment, which may drive crude oil prices.

We use rigorous forecasting tests and report significant and informative outcomes as follows. Following Lettau and Ludvigson (2005), Shi (2023) and many other studies, we employ univariate in-sample analysis with a one-sided bootstrap p-value estimation to forecast crude oil prices. Our evidence shows that the second RP-PCA factor, rather than other RP-PCA factors, can consistently and robustly predict the WTI and Brent oil prices at all forecasting horizons, with a sound *R*-squared value and a significant t-statistic. Moreover, our out-of-sample analysis confirms that the second RP-PCA factor can reasonably capture the forecasting power of crude oil prices for one-month and one-quarter (3-month) horizons, whereas the other RP-PCA factors and most other popular economic/financial predictors (approximately 30 popular predictors) appear not to exhibit forecasting power for crude oil prices. Furthermore, the success ratio tests also do not reject the second RP-PCA factor forecast with the correct signs. The out-of-sample evidence is robust to different initial estimation windows and different benchmark models (including historical mean forecasts, no change forecast models, and autoregressive distributed lag forecast models). Following Neely et al. (2014) and Huang et al. (2015), we also estimate the utility and Sharpe ratio gains to account for economic performance. The second RP-PCA factor provides positive and large values of utility and Sharpe ratio gains, which consistently beat most of the other popular economic/financial predictors except for the corporate bond yield proposed by Dai and Kang (2021). Finally, we reveal that the predictive power of the second RP-PCA factor for crude oil prices may stem from its ability to forecast oil market sentiment.

4 Specifically, a total of approximately 30 predictors includes RP-PCA factors, 14 Welch and Goyal (2008) economic factors, 6 other benchmark factors, and 11 recent popular financial ratios.

Shi (2023)[5] revealed that RP-PCA factors can robustly forecast stock market excess returns via aligned approach. Since valid forecasts of crude oil prices benefit many economists and decision makers (Alquist et al., 2013; Baumeister and Kilian, 2014), our study further confirms that forecasting oil via some economic tools is a profitable strategy for policymakers and investors. In particular, policymakers and investors can alternatively notice that new financial instruments (e.g., the second RP-PCA factor) should contain sound predictive information for forecasting crude oil prices and generate substantially higher values of utility gain and Sharpe ratio gain.

Our paper is organized as follows. Section 2 presents our theoretical motivation. Section 3 introduces the data. Sections 4, 5, and 6 present the research methods and analyze the empirical work, and Section 7 summarizes the conclusions.

2. Theoretical Motivation

Cochrane (2005)[6] proposed the pricing kernel of a linear multifactor pricing model via the following euqation:

$$m_{t+1} = a + b' f_{t+1} \tag{1}$$

where *f* are RP-PCA factors in our study, *a* and *b* denote free parameters, and *m* is the pricing kernel. Cochrane (2005) provides in-depth insights into any candidate factors *f* in Equation (1): can financial ratios forecast proxies for good or bad states of future economic activity, such as the investment opportunity set or representative macroeconomic activity? Alternatively, all the factors of factor pricing models may have valid forecasting abilities for future economic circumstances once they are considered as candidate factors according to the theoretical APT[7] framework.

RP-PCA methods extract five latent factors from large financial datasets (i.e., a large cross-section of portfolios sorted into 37 well-known anomalies). Lettau and Pelger's (2020) RP-PCA factors refine the efficient information from a large cross-section of financial portfolios sorted into 37 anomalies. Since the RP-PCA factors fall within the theoretical framework of Ross's (1976) APT[8] and adequately fit the cross-section and time series of a large panel of equity returns, we suggest that the RP-PCA factors may contain much more additional information

5 Similarly, Dong et al. (2022) suggest that a variety of shrinkage methods helps to extract efficient information from 100 representative portfolios, which predict excess stock return significantly.

6 See Cochrane (2005) for further details.

7 Rapach et al. (2005) provides similar motivation for the predictive role of financial ratios/macro ratios.

8 In Lettau and Pelger's (2020) study, RP-PCA factors can be proxied by *f* the pricing kernel function of an APT model (Equation 1).

than any individual macro/financial factor in forecasting aggregate stock returns (Shi, 2023) and economic activity (e.g., crude oil prices) under the framework of Cochrane's (2005) linear asset pricing model.

The oil industry is a vital economic activity that is well known as the "lifeblood industry". Forecasting this key economic activity is of considerable interest to academics. Most conventional financial ratios and economic factors cannot consistently predict this macroeconomic activity. One of our key research questions is whether crude oil prices can be predicted by any financial variables. Our study extracts the key predictive role of the second RP-PCA factor, which answers the posed research question.

3. Data

Following a large body of research on oil (Baumeister and Kilian, 2012, 2014; Zhang et al., 2018; Dai and Kang, 2021; among others), we use spot prices of West Texas Intermediate crude oil (WTI) and North Sea Brent crude oil (Brent) as proxies for crude oil prices, where the corresponding data are sourced from the U.S. Energy Information Administration. Following the past literature, we deflated[9] the original data (nominal oil prices) by the U.S. consumer price index[10] (CPI); thus, we derived real oil prices. The WTI and Brent data start at the beginning of June 1987.

We collect the data on RP-PCA factors from Markus Pelgerthe website, where the five RP-PCA data series are available until December 2017. In line with Huang et al. (2015) and Shi (2023), we employ a six-month moving average method to rule out extreme outliers. Hence, our sample data are monthly, beginning in June 1987 and ending in December 2017, with a total of 367 observations.

In addition to RP-PCA factors, we consider two sets of benchmark comparative predictors. The first dataset is Welch and Goyal's (2008)[11] variable set, where DP = log dividend–price ratio; DY = log dividend yield; EP = the log earnings–price ratio; DE = the log dividend–payout ratio; RVOL = the volatility of excess stock returns; BM = the book–to–market value ratio for the Dow Jones Industrial Average; NTIS = net equity expansion; TBL = the interest rate of a three-month Treasury bill; LTY = the long-term government bond yield; LTR = the return of long-term government bonds; TMS = the long-term government bond yield minus the Treasury

9 See Zhang et al. (2019), He et al. (2021), and Dai and Kang (2021) for details about the deflation of nominal crude oil prices.

10 CPI data are sourced from Federal Reserve Economic Data (FRED).

11 See Welch and Goyal's (2008) detailed information about the setup of these economic variables.

bill rate; DFY = the difference between Moody's BAA and AAA rated corporate bond yields; DFR = the long-term corporate bond return minus the long-term government bond return; and INFL = inflation calculated from the CPI for all urban consumers.

The second set of variables is another set of popular oil price predictors[12], where EPU = economic policy uncertainty; MS = money supply; IPI = industry production index; COP = growth of crude oil production; OI = growth of returns and excess returns on oil company stocks; and KI = growth of Kilian's (2009) index.

Table 1: Summary statistics

Panel A: Summary statistics							
	Observations	**Mean**	**Standard deviation**	**Minimum**	**Maximum**	***ρ*(1)**	***ADF***
RPPCA1	367	5.706	19.196	−83.319	69.479	0.844	−5.589***
RPPCA2	367	−1.854	4.336	−23.999	17.258	0.867	−5.108***
RPPCA3	367	−0.147	4.948	−13.018	21.448	0.878	−4.863***
RPPCA4	367	−0.463	3.248	−11.642	26.163	0.808	−6.220***
RPPCA5	367	0.411	2.368	−9.743	8.446	0.842	−5.614***

	Panel B: Correlation Matrix			
	RPPCA1	***RPPCA2***	***RPPCA3***	***RPPCA4***
RPPCA2	0.481			
RPPCA3	0.103	−0.047		
RPPCA4	0.124	0.104	−0.093	
RPPCA5	−0.334	0.314	−0.203	0.081

Notes: the independent variable is defined as follows: *RPPCA*1–*RPPCA*5 = five RP-PCA factors proposed by Lettau and Pelger (2020). ρ(1) represents the coefficient of first-order autocorrelation. *** represents a statistical significance level of 1%. ADF refers to augmented Dickey–Fuller (ADF) tests proposed by Dickey and Fuller (1979).

Source: author's calculations

12 Yin and Yang (2016) and Zhang et al. (2019) provide detailed information about the second list of popular real oil price predictors.

The reasons for using these two sets of popular predictors as benchmark comparative variables are as follows. Following the procedures of Ma et al. (2018) and Lv and Qi (2022), Welch and Goyal's (2008) 14 macroeconomic factors all contain predictive information for forecasts in the financial market. Additionally, following Dai and Kang (2021), Yin and Yang (2016), and Zhang et al. (2019), another six popular predictors (EPU, MS, IPI COP, OI, and KI) denote fluctuations in demand and supply pressures affecting oil markets. However, the oil market is also a vital economic market. Since all these comparative predictors contain some valuable predictive economic content and most of them are proven to forecast the financial market, they all may forecast crude oil prices relatively well.

We display the summary statistics for the RP-PCA financial data[13] in Table1. Unsurprisingly, all the RP-PCA financial data have a highly persistent value, implying that our sample may suffer from the well-known Stambaugh (1999) small-sample bias, where our in-sample estimations may particularly produce distorted t-ratios in our finite sample. Following the approaches of Neely et al. (2014) and Rapach et al. (2016), we address this problem by employing asymptotic t-ratios (Newey and West, 1987) with certain lags and conducting a wild bootstrap procedure. Consistent with Shi (2023), the correlations between each pair of RP-PCA predictive factors are relatively low, indicating that the RP-PCA factors do not share the same predictive information in forecasting economic conditions (e.g., real oil price). All the RP-PCA factors are stationary based on augmented Dickey–Fuller (ADF) tests.

4. In-sample Analysis

Our in-sample univariate regression follows the following conventional regression model:

$$R_{t+h} = b_0 + b_1 Factor_t + \varepsilon_{t+h} \tag{5}$$

where R_{t+h} denotes the cumulative log return of WTI or Brent for the h-month forecasting horizon[14] and where $Factor_t$ denotes the lagged RP-PCA factors.

13 In addition, we standardize the RP-PCA factors before our empirical analysis.

14 Note the formulation:.

Table 2: In-sample forecasting of the WTI spot price

H **=**	**1**	**3**	**6**	**12**
RPPCA1	**0.011**	0.007	0.003	0.002
t-stat (NW)	**1.658**	1.187	0.718	0.564
Boots	**0.058**	0.165	0.282	0.332
R^2	**0.020**	0.015	0.006	0.004
RPPCA2	**0.014**	**0.013**	**0.010**	**0.005**
t-stat (NW)	**2.864**	**3.182**	**3.241**	**2.776**
Boots	**0.003**	**0.004**	**0.007**	**0.030**
R^2	**0.028**	**0.053**	**0.065**	**0.035**
RPPCA3	0.001	–0.002	–0.002	0.001
t-stat (NW)	0.149	–0.373	–0.660	0.382
Boots	0.438	0.378	0.292	0.378
R^2	0.000	0.001	0.003	0.002
RPPCA4	0.002	0.004	0.004	0.002
t-stat (NW)	0.602	1.144	1.320	1.018
Boots	0.284	0.177	0.155	0.219
R^2	0.001	0.004	0.009	0.007
RPPCA5	0.002	0.006	0.004	0.004
t-stat (NW)	0.381	1.404	1.155	1.491
Boots	0.351	0.116	0.177	0.125
R^2	0.001	0.013	0.011	0.019

Notes: In each model, we report the OLS coefficient estimations in the first row, the Newey and West (1987) *t*-statistics in the second row, the *p*-values of one-sided wild bootstrap in the third row, and the R^2 in the fourth row. The predictive variables are each of the RP-PCA factors. *H* denotes the monthly forecasting horizons. We highlight any significant point for in-sample forecasting.

Source: author's calculations

Tables 2 and 3 illustrate the in-sample forecasting of RP-PCA factors in predicting WTI and Brent, respectively[15]. The evidence indicates that only the second RP-PCA factor can

15 We also investigate the structural breaks of the RP-PCA indices in forecasting crude oil prices in the Appendix Table 1. The results of Elliott and Müller's (2006) qLL significant statistics reject the null hypothesis that the model is unstable for RP-PCA predictors, suggesting that there is no structural problem in our in-sample analysis.

forecast both WTI and Brent robustly in horizons from one month to one year owing to the high *R*-squared values and significant p-values of the bootstrap statistics; the first RP-PCA factor appears to forecast the crude oil prices in the 1-month forecasting horizon to some degree. Other RP-PCA factors fail to robustly forecast the oil prices. The above evidence implies that the second RP-PCA factor, rather than the other four RP-PCA indices, contains vital information for in-sample forecasting of real oil prices through all horizons.

Table 3: In-sample forecasting of the Brent spot price

***H* =**	**1**	**3**	**6**	**12**
RPPCA1	**0.012**	0.008	0.004	0.002
t-stat (NW)	**1.940**	1.402	0.950	0.647
Boots	**0.029**	0.124	0.221	0.316
R^2	**0.019**	0.017	0.010	0.005
RPPCA2	**0.014**	**0.013**	**0.010**	**0.005**
t-stat (NW)	**2.677**	**2.832**	**2.970**	**2.856**
Boots	**0.007**	**0.012**	**0.016**	**0.024**
R^2	**0.026**	**0.049**	**0.061**	**0.034**
RPPCA3	0.001	–0.001	–0.002	0.002
t-stat (NW)	0.179	–0.301	–0.621	0.445
Boots	0.425	0.400	0.299	0.366
R^2	0.000	0.001	0.003	0.003
RPPCA4	0.002	0.004	0.005	0.002
t-stat (NW)	0.371	1.260	1.572	0.969
Boots	0.365	0.167	0.128	0.252
R^2	0.000	0.005	0.012	0.006
RPPCA5	0.003	0.006	0.004	0.003
t-stat (NW)	0.480	1.367	0.982	1.393
Boots	0.322	0.128	0.215	0.145
R^2	0.001	0.011	0.007	0.015

Source: author's calculations

5. Out-of-sample Findings

To investigate the statistical out-of-sample performance, we follow Campbell and Thompson's (2008) approach for the application of the statistic:

$$R_{os}^2 = 1 - \frac{MSPE_{model}}{MSPE_{HA}} \tag{6}$$

where $MSPE_{model} = \frac{1}{T-P}\sum_{t=p+1}^{T}(r_t - \hat{r}_t)^2$ and $MSPE_{HA} = \frac{1}{T-P}\sum_{t=p+1}^{T}(r_t - \overline{r}_t)^2$. In other words, we compute the R_{os}^2 statistic as the reduction in the mean squared predictive error (MSPE) for a competing forecast (a model including any of the useful economic predictors) relative to the benchmark historical average (HA) model. Next, for the corresponding positive $R_{os}^2 > 0$, the competing model forecasts the crude oil price more accurately than does the HA forecast approach. Hence, the R_{os}^2 statistic provides a criterion for evaluating competing predictive performance relatively to an HA strategy.

Furthermore, we employ Clark and West's (2007) adjusted MSFE statistic to assess the null hypothesis, which states that the MSFE of the HA forecast method is greater than or equal to the MSFE of the competing forecast (a model including any of the useful economic predictors), corresponding to H_0: $R_{os}^2 \leq 0$: against H_{A}: $R_{os}^2 > 0$:

Given the considerably reduced power of the recursive regressions when the forecasting horizons increase, particularly in forecasting crude oil prices (see, e.g., Zhang et al., 2019; Yin and Yang, 2016; and Dai and Kang 2021), we report forecasting horizons only up to 1 quarter (3 months) ahead.

We present our out-of-sample analysis with a large range of economic variables[16] in Table 4, where the initial estimation period is 60 months[17]. As expected by Dai and Kang (2021), two bond yields (LTR and DFR) consistently achieve significant Clark and West (2007) MSFE statistics and positive and high values of R_{os}^2.

16 We also report out-of-sample performance for recent popular financial risk ratios in Table A2. Most financial risk ratios demonstrate insignificant forecasting power in predicting crude oil prices. Only MGMT provides some predictive ability; however, the predictive power is relatively low with an approximately 0.01 magnitude of *R*-squared at the one month-forecasting horizon; the predictive power is no longer valid when the forecasting horizon is one quarter or longer. In other words, the second RP-PCA should quantitively outperform MGMT. Chib et al. (2024) provide a detailed description of the majority of popular financial risk factors adopted in Table A2. Since these financial data are all quite popular and available, one can obtain these data from corresponding authors' websites.

17 Following Dai and Kang (2021), we use 60 months as the initial estimation window.

Table 4: Estimations of out-of-sample predictability

	WTI spot price				Brent spot price			
	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	0.374	0.171	−3.037	0.310	0.648	0.106	−2.749	0.224
RPPCA2	**2.090**	**0.007**	**2.507**	**0.000**	**1.695**	**0.010**	**0.372**	**0.001**
RPPCA3	−0.581	0.688	−1.383	0.773	−0.686	0.807	−1.561	0.918
RPPCA4	−0.645	0.529	−1.297	0.383	−0.918	0.731	−1.661	0.252
RPPCA5	−0.930	0.377	−0.553	0.044	−0.889	0.310	−0.993	0.037
Welch and Goyal's (2008) benchmark factors								
DP	−0.796	0.566	−3.789	0.745	−0.884	0.570	−4.976	0.743
DY	−1.833	0.596	−3.627	0.809	−2.158	0.611	−5.016	0.794
EP	−0.384	0.455	−1.284	0.525	−0.331	0.422	−1.432	0.530
DE	−2.058	0.534	−8.079	0.756	−1.611	0.532	−7.218	0.758
RVOL	−2.540	0.757	−4.467	0.895	−2.820	0.773	−5.075	0.883
BM	−0.838	0.674	−3.437	0.876	−0.904	0.690	−4.216	0.850
NTIS	−0.774	0.716	−4.741	0.942	−0.649	0.705	−4.321	0.924
TBL	−1.399	0.809	−3.643	0.893	−1.295	0.821	−3.361	0.862
LTY	−1.195	0.838	−4.011	0.922	−1.203	0.820	−3.894	0.897
LTR	**2.045**	**0.006**	**0.384**	**0.076**	**1.218**	**0.008**	**0.381**	**0.036**
TMS	−0.950	0.791	−2.161	0.748	−0.962	0.830	−2.251	0.742
DFY	−5.520	0.518	−17.090	0.973	−5.532	0.550	−16.497	0.981
DFR	**3.161**	**0.011**	−0.512	0.076	**4.708**	**0.004**	**0.699**	**0.048**
INFL	−0.610	0.552	−3.001	0.761	−1.125	0.891	−2.755	0.756
Other popular benchmark factors								
EPU	−0.352	0.129	−3.934	0.380	−0.700	0.123	−4.005	0.340
MS	**0.675**	**0.078**	−0.806	0.706	−0.785	0.553	−1.079	0.943
IPI	−0.012	0.309	−0.544	0.629	−0.175	0.470	−0.593	0.449
COP	−1.364	0.961	−0.106	0.463	−1.298	0.976	−0.298	0.565
OI	−3.494	0.903	0.297	0.285	−3.134	0.876	0.329	0.262
KI	1.249	0.116	−4.827	0.465	−1.104	0.268	−6.602	0.825

Notes: we present the out-of-sample predictive forecast for the WTI spot price. We report the R^2 statistic, which denotes the out-of-sample coefficient of determination (in %) in each column of *H*=1 and 3, where *H* represents the forecasting horizon. We also report the statistics of Clark and West's (2007) MSFE adjusted p-value. Forecasts begin 5 years after the start of the sample. The values in bold are significant at the 10% level. The benchmark is historical mean forecasts.

Source: author's calculations

Table 5: Estimations of out-of-sample predictability

	WTI spot price				Brent spot price			
RPPCA1	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	0.959	0.145	−2.737	0.304	**1.250**	**0.084**	−2.258	0.214
RPPCA2	**2.431**	**0.006**	**1.576**	**0.001**	**1.990**	**0.009**	−1.008	0.002
RPPCA3	−0.560	0.645	−1.221	0.709	−0.682	0.779	−1.373	0.871
RPPCA4	−0.719	0.587	−0.819	0.223	−0.940	0.749	−1.095	0.139
RPPCA5	−1.399	0.591	−1.953	0.155	−1.398	0.508	−2.513	0.167
180months	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	−0.024	0.245	−5.950	0.548	0.536	0.164	−5.270	0.475
RPPCA2	**2.729**	**0.035**	**2.512**	**0.039**	**2.596**	**0.034**	**2.668**	**0.037**
RPPCA3	−0.630	0.799	−1.125	0.770	−0.505	0.858	−0.888	0.796
RPPCA4	−0.126	0.643	−0.441	0.607	−0.142	0.734	−0.675	0.619
RPPCA5	−1.425	0.797	−5.095	0.758	−1.735	0.807	−5.846	0.818
240months	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	0.761	0.239	−5.541	0.567	1.420	0.167	−5.135	0.524
RPPCA2	**4.040**	**0.027**	**4.936**	**0.028**	**4.191**	**0.024**	**5.304**	**0.025**
RPPCA3	−0.568	0.724	−0.906	0.683	−0.467	0.768	−0.732	0.708
RPPCA4	−0.152	0.694	−0.513	0.672	−0.180	0.787	−0.846	0.710
RPPCA5	−0.979	0.940	−2.044	0.687	−0.873	0.867	−1.692	0.646

Notes: we present the out-of-sample predictive forecast for the WTI spot price. We report the R^2 statistic, which denotes the out-of-sample coefficient of determination (in %) in each column of *H*=1 and 3, where *H* represents the forecasting horizons. We also report the statistics of Clark and West's (2007) MSFE adjusted p-value. Forecasts begin 10, 15, and 20 years after the start of the sample. Significance at the 10% level is in bold. The benchmark is historical mean forecast.

Source: author's calculations

In addition to the findings of Dai and Kang (2021), our study further reveals that only the second RP-PCA factor[18] outperforms most of the benchmark economic variables and displays the highest positive value of R^2_{oos} with a significant Clark and West (2007) MSFE statistic.

18 Indeed, the out-of-sample forecasting power for the crude oil prices generally weakens as the forecasting horizon becomes longer (e.g., half a year (six months) or one year (twelve years)), and this outcome appears to be consistent with Dai and Kang's (2021) findings.

In Table 5, we further investigate the out-of-sample performance in a robustness test when the initial estimation period is 10 years, 15 years, or 20 years. The second RP-PCA factor, but not the other RP-PCA factors, could broadly forecast WTI and Brent quite well with a sound R_{os}^2 and a significant MSFE statistic (Clark and West, 2007).

We also evaluate the out-of-sample estimations for predicting the WTI spot price[19] via the other two most common benchmark models (i.e., the no-change forecast model and autoregressive distributed lag model[20]) in Tables A3 and A4, where the initial estimation period is either five years or fifteen years. We reveal that among a long list of economic predictors, the second RP-PCA factor stably[21] demonstrates the second-highest forecasting power at one-month forecasting horizon, while it achieves the highest predictive power at one-quarter forecasting horizon. The presented evidence strongly suggests that there is a solid link between the second RP-PCA factor and crude oil prices.

Leitch and Tanner (1991) suggest a more meaningful view by correctly judging the market direction in investment activity. Following their approach, we compute the success ratio of each competing model, which indicates that the predictive sign of the crude oil return is correct and the test p-value of Pesaran and Timmermann's (1992) success ratio statistic.

With respect to the success ratio, all the outcomes of the second RP-PCA ratio surpass 0.5[22], with a corresponding significant p-value under the 0.05 level of significance, suggesting that the RP-PCA factor can correctly forecast the sign of future crude oil prices. Nevertheless, none of the other RP-PCA factors pass the p-value of Pesaran and Timmermann's (1992) success ratio at the 0.05 level of significance, indicating that these RP-PCA factors may make incorrect forecasts.

19 We obtain similar out-of-sample forecasting evidence for Brent crude oil prices particularly when initial estimation periods are somewhat longer (e.g., 15 years or 20 years). These results are available upon request.

20 We follow Rapach et al.'s (2010) ARDL model to provide another robust test. Then ARDL forecasting model is: $R_{t+h} = b_0 + b_1 R_t + b_2 Factor_t + \varepsilon_{t+h}$, where denotes the cumulative log returns of WTI or Brent for the h-month forecasting horizon (i.e.,), $R_{t \to t+H} = R_{t+1} + R_{t+2} + \cdots + R_{t+H}$ denotes the first-order lag of log returns of WTI or Brent and $Factor_t$ denotes any lagged predictor.

21 Notably, when the initial estimation period varies (changes from 5-year to 15-year), the return on long-term government bonds (LTR) cannot stably forecast WTI spot price in out-of-sample estimations.

22 Yin and Yang (2016) suggest that if the value of the success ratio is higher than 0.5, it implies that the predictor in the predictive regression model outperforms the no-change forecast.

Table 6: Success ratio tests

	WTI spot price				Brent spot price			
P value	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	0.508	0.501	0.554	0.052	0.518	0.364	0.554	0.058
RPPCA2	**0.567**	**0.017**	**0.636**	**0.000**	**0.580**	**0.005**	**0.636**	**0.000**
RPPCA3	0.495	0.585	0.485	0.700	0.505	0.477	0.479	0.800
RPPCA4	0.518	0.456	0.515	0.385	0.534	0.246	0.521	0.284
RPPCA5	0.511	0.310	0.528	0.095	0.518	0.259	0.531	0.082

Notes: we present Pesaran and Timmermann's (1992) success ratio statistic for the out-of-sample predictive forecasts of the WTI spot price and the Brent spot price, respectively. Forecasts begin 5 years after the start of the sample. *H* represents the forecasting horizon. Significance at the 5% level is in bold.

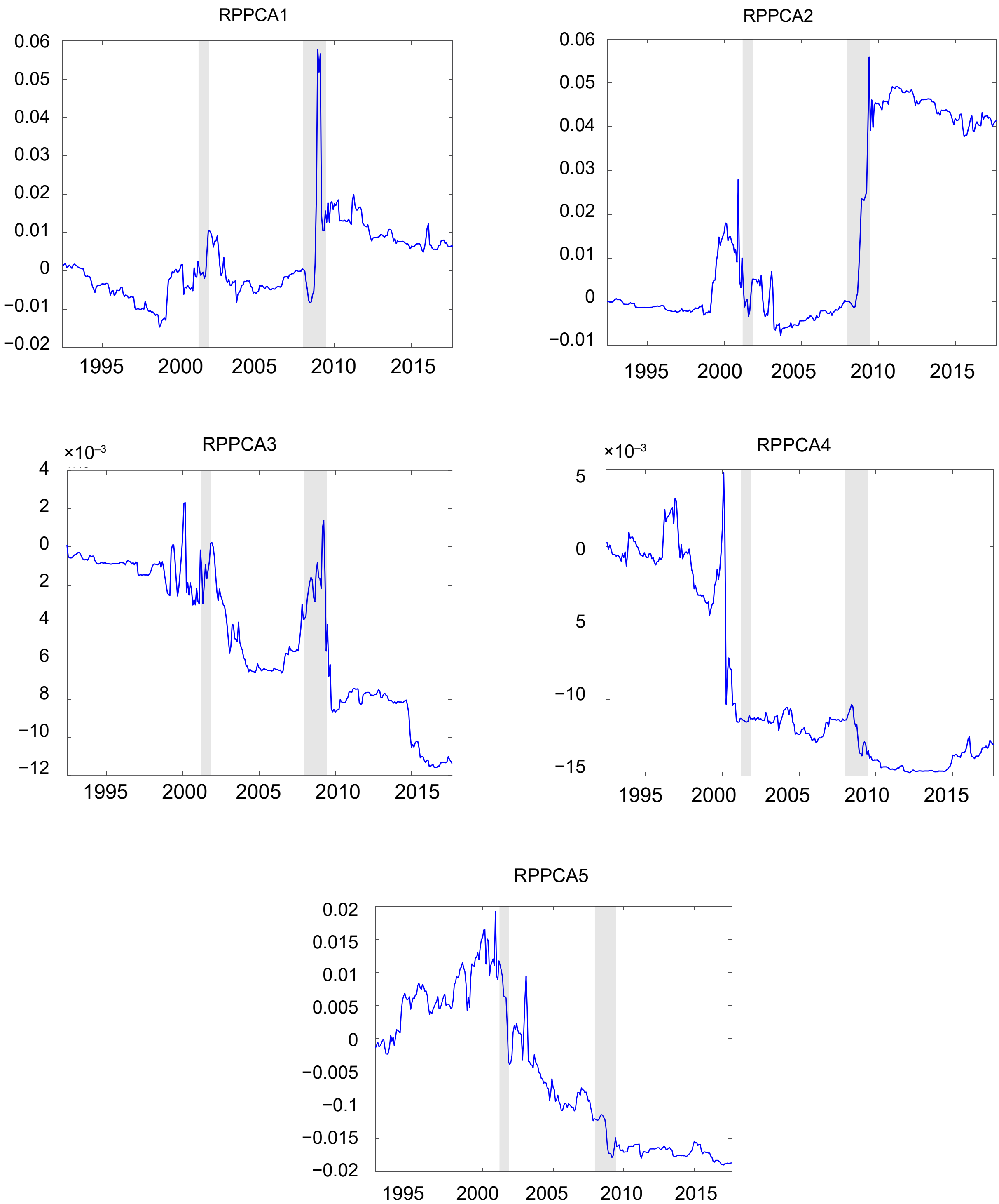
Source: author's calculations

To offer a clearer view of our study, the out-of-sample analysis is presented in Figures 1 and 2. We concentrate on presenting graphs for the RP-PCA roles in predicting WTI prices at one-month-ahead and one-quarter-ahead horizons. Figures 1 and 2 clearly show that only the second RP-PCA factor performs quite well[23] in out-of-sample forecasting of WTI, as most of the values of the second RP-PCA factors are above zero.

Most studies attempt to build a forecasting link between different macroeconomic variables and crude oil prices. Instead, our study explores financial factors in an asset pricing model that may predict economic activity (e.g., the crude oil market prices). We present 11 popular financial ratios in Table A2; we find that the second RP-PCA factor provides additional useful information in predicting future crude oil prices and outperforms a large body of popular economic predictors. Lettau and Pelger (2020) conclude that the first, second, and fifth RP-PCA factors capture most of the cross-sectional variation in equity returns. Hence, following Cochrane's (2005) framework regarding any sound pricing factors in a linear asset pricing model (e.g., Ross APT model), only these three RP-PCA factors are most likely to contain additional information in forecasting economic activity, such as future crude oil prices. Consistent with that conjecture, our empirical study further reveals that the second RP-PCA provides such robust forecasting ability.

23 The blue color line represents cumulative squared prediction error. When the blue color curve is broadly above zero in the figures, it implies that the out-of-sample forecast of the WTI crude oil price performs quite well. Notably, unlike the graphs for the other RP-PCA factors, the values of the second RP-PCA factor are highly positive, suggesting that only the second RP-PCA index can predict crude oil prices well enough.

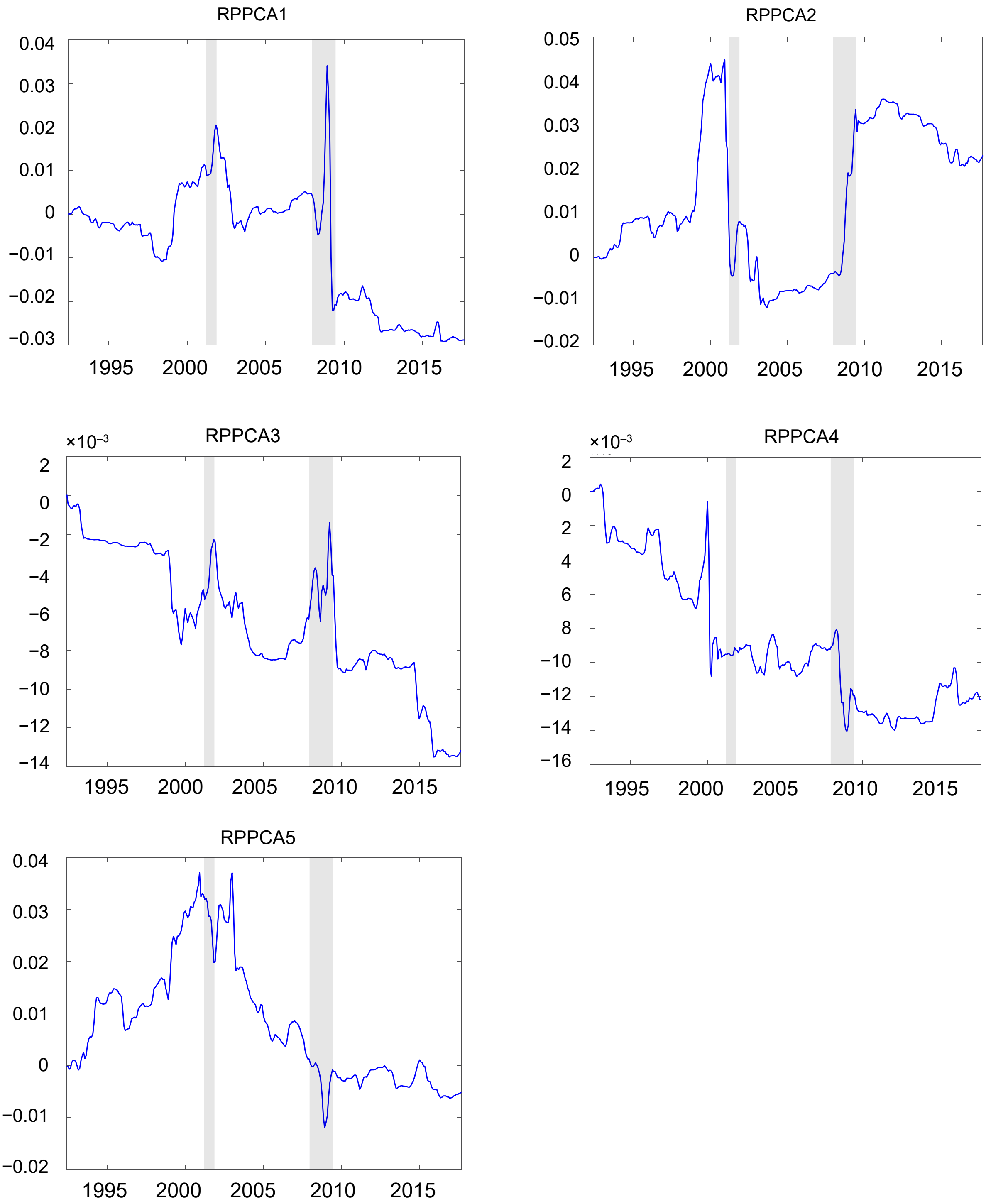
Figure 1: Out-of-sample analysis results for WTI



Notes: we illustrate the difference in the cumulative square prediction error between the historical average benchmark forecast model and each RP-PCA factor forecast. The dependent variable is the WTI spot price, and the forecasting horizon is one month ahead. The vertical bars depict NBER-dated recessions.

Source: author's calculations

Figure 2: Out-of-sample analysis results for Brent



Notes: we illustrate the difference in the cumulative square prediction error between the historical average benchmark forecast model and each RP-PCA factor forecast. The dependent variable is the Brent spot price, and the forecasting horizon is one quarter ahead. The vertical bars depict NBER-dated recessions.

Source: author's calculations

6. The Economic Value of the Second RP-PCA Factor

6.1 Economic significance

Consistent with many previous works (e.g., Rapach and Zhou, 2013; Neely et al., 2014; Yin and Yang, 2016; Zhang et al., 2019; Dai and Zhu 2020), we evaluate the out-of-sample utility and Sharpe ratio gains as proxies for the economic value of useful economic predictors in asset allocation.

Investors should construct their trading strategy consisting of stocks and risk-free assets via forecasts on the basis of the information contained in useful economic predictors relatively to the HA forecasts. The weight of stocks in the investors' portfolio should be allocated as follows:

$$w_t = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}} \tag{7}$$

where γ denotes the risk aversion coefficient with a value of three[24], $\hat{r}_{t+1}$ is the forecast of excess returns for investors on the basis of the information from useful economic predictors up to time t and where $\hat{\sigma}^2_{t+1}$ is the realized variance forecast of excess returns via a five-year moving window (see Campbell and Thompson, 2008). Furthermore, we restrict the weight range of stocks from 0.5 to 1.5 to impose a short-selling constraint.

Next, we can evaluate the certainty-equivalent (CE) returns:

$$CER_P = E(R_p) - \frac{\gamma}{2} \sigma_P^2, \tag{8}$$

where $E(R_P)$ and σ_P^2 represent the mean and variance of the portfolio's return, respectively. The CE gain is the difference between the CE return of the investor via a predictive model forecast that is based on information from useful economic predictors and the CE return of the investor via the HA benchmark model forecast. We also calculate the annual CE return gain, where we multiply this difference by 1200.

Moreover, we calculate the Sharpe ratio as

$$SR = \frac{E(R_P)}{\sigma_P}, \tag{10}$$

where $E(R_p)$ and σ_P denote the mean and standard deviation of the portfolio's return, respectively.

24 The results appear similar to other values of the risk aversion coefficient.

Table 7: Sharpe ratio gains

	Sharpe ratio gain			
	WTI spot price		Brent spot price	
RPPCA factors	***H*=1**	***H*=3**	***H*=1**	***H*=3**
RPPCA1	0.354	0.361	0.369	0.379
RPPCA2	0.762	0.798	0.723	0.759
RPPCA3	0.012	0.159	−0.073	0.103
RPPCA4	0.054	0.203	−0.069	0.219
RPPCA5	0.085	0.342	0.144	0.348
Welch and Goyal's (2008) benchmark factors				
DP	0.118	0.073	0.073	0.059
DY	0.211	−0.003	0.155	−0.006
EP	0.198	0.179	0.229	0.180
DE	0.214	0.001	0.145	−0.021
RVOL	0.267	0.088	0.201	0.105
BM	0.031	−0.007	0.028	−0.010
NTIS	−0.026	−0.068	−0.049	−0.051
TBL	−0.064	−0.021	−0.096	−0.045
LTY	−0.012	0.040	−0.038	0.039
LTR	0.643	−0.116	0.631	−0.098
TMS	−0.318	−0.258	−0.357	−0.178
DFY	−0.092	−0.087	−0.142	−0.145
DFR	0.731	0.410	0.849	0.439
INFL	0.061	0.074	−0.124	0.071
Other popular benchmark factors				
EPU	0.307	0.116	0.192	0.089
MS	0.298	0.074	0.040	−0.012
IPI	0.208	0.077	0.051	0.224
COP	−0.212	0.014	−0.269	−0.036
OI	−0.251	0.057	−0.178	0.067
KI	0.220	0.176	0.092	0.152

Notes: this table reports the annualized Sharpe ratio gains (SR) for the WTI spot price and Brent spot price in forecasting *H* horizons, where *H* represents 1 month and 3 months (i.e., one quarter) forecasting horizons.

Source: author's calculations

Table 8: Utility gains

	Utility Gains			
	WTI spot price		Brent spot price	
RPPCA factors	***H*=1**	***H*=3**	***H*=1**	***H*=3**
RPPCA1	1.087	2.025	1.502	2.351
RPPCA2	6.483	8.079	6.356	7.680
RPPCA3	−1.106	0.139	−1.656	−0.408
RPPCA4	−0.555	1.094	−1.356	1.303
RPPCA5	−2.164	−0.096	−1.393	−0.317
Welch and Goyal's (2008) benchmark factors				
DP	−1.063	−0.979	−1.508	−1.998
DY	−1.042	−2.035	−1.790	−3.070
EP	0.204	0.416	0.597	0.359
DE	0.414	−0.416	0.093	−0.969
RVOL	0.581	−0.006	0.136	−0.270
BM	−0.654	−0.853	−0.528	−1.597
NTIS	−1.850	−7.333	−1.530	−7.281
TBL	−2.537	−5.692	−2.618	−4.815
LTY	−2.428	−6.272	−2.549	−5.555
LTR	4.058	−2.756	4.266	−2.685
TMS	−2.311	−2.702	−2.504	−3.752
DFY	−3.304	−10.976	−3.763	−9.576
DFR	6.077	2.550	8.516	3.069
INFL	−1.852	−2.462	−2.505	−2.791
Other popular benchmark factors				
EPU	−0.590	−1.315	−2.665	−2.248
MS	1.297	0.128	−0.441	−0.525
IPI	0.959	0.714	0.340	1.721
COP	−3.043	0.110	−3.168	−1.382
OI	−4.516	−1.338	−4.307	−2.118
KI	−2.711	−3.261	−3.418	−6.106

Notes: this table reports the annualized utility gains for the WTI spot price and Brent spot price in forecasting *H* horizons, where *H* represents 1-month and 3-month (i.e., one-quarter) forecasting horizons.

Source: author's calculations

Similarly, we also compute the SR gain as the difference between the SR using a predictive model forecast based on information from useful economic predictors and that of the HA forecasts.

For the analysis of SR and CE gains and in Tables 7 and 8, in addition to the corporate bond yield, the second RP-PCA factor achieves the highest values of SR gains and CE gains. The values of SR gains and CE gains for the second RP-PCA factor are quite high and positive; however, other competing benchmark factors are broadly low and even negative. In other words, the second RP-PCA factor largely outperforms most of the other popular benchmark factors (including Welch and Goyal's benchmark economic factors) with respect to the SR and CE gains. Hence, using the second factor of RP-PCA can create much greater economic gains than the utilization of most of the other benchmark economic factors.

6.2 Market sentiment-related interpretation

Following Yin and Yang (2016) and Dai and Kang (2021), we also provide another possible explanation for the predictive power of the second RP-PCA factor from a market sentiment perspective. In particular, Deeney et al. (2015) noted that market sentiment influences prices in actual oil markets. Furthermore, if the crude oil price predictors also forecast sentiment quite well, the forecasting power of the useful economic predictor may stem from its ability to predict market sentiment.

Following Deeney et al. (2015), we construct the market sentiment for the crude oil market:

$$\begin{aligned} &\textit{WTI sentiment} = 0.36\,\textit{Trading Volume}_{t-1} - 0.44\,\textit{WTI Volatility}_t \\ &-0.53\,\textit{Put Call Ratio}_{t-1} + 0.59\,\textit{SpecWTI}_t - 0.22\,\textit{VIX}_{t-1} \end{aligned} \tag{11}$$

where the trading volume refers to WTI crude oil futures' trading volumes; volatility denotes the 30-day volatility of historical futures returns; the put-call ratio for oil futures options is the third proxy; Spec means speculation, indicating the number of noncommercial futures positions from the CFTC; and VIX is considered as a measure of investors' fear or anxiety.

Table 9: WTI sentiment regression results

Panel A:	**Slope coefficient**	***t*-statistic**	***R*-square**
WTI sentiment index	0.028	8.23	0.168
Panel B:			
RPPCA2	0.328	5.28	0.077

Notes: this table reports the in-sample regression of the WT sentiment index. where r_t, s_t and x_{t-1}, and represent WTI oil return, the WTI sentiment index and the second RP-PCA factor in month $t-1$, respectively.
Source: author's calculations

In Table 9, we regress the crude oil price on the WTI sentiment index, obtaining reasonably high t-statistic (t-statistic = 8.23) for the slope and corresponding high R-squared value (R-squared = 0.168). This result suggests that oil market sentiment can effectively capture fluctuations in the WTI market. Moreover, when we regress the WTI sentiment index on the second RP-PCA factor, we also obtain a significant t-statistic (t-statistic = 5.28) for the slope and corresponding high R-squared (R-squared = 0.077), implying that the second RP-PCA index could robustly capture WTI sentiment. Overall, the predictive power of the second RP-PCA factor for crude oil prices may stem from its ability to predict WTI sentiment.

7. Conclusion

It is notoriously difficult to use financial ratios to predict crude oil prices. To address this issue, our study alternatively utilizes the second RP-PCA factor as a sound new predictor that forecasts WTI and Brent crude oil prices robustly. Throughout a range of rigorous tests, we discover that the second RP-PCA factor substantially outperforms other RP-PCA factors and many other popular conventional economic/financial predictors; and the second RP-PCA factor can also generate much greater economic profit. Our viewpoint is consistent with Lettau and Pelger's (2020) detailed findings and Cochrane's (2005) theoretical motivation for any candidate pricing factors in Ross APT model. Finally, using the intermediate role of oil market sentiment, we reveal that the significant ability of the second RP-PCA to forecast WTI and Brent prices may stem from its ability to forecast the oil market sentiment index.

Following Shi's (2023) study on the use of the aligned RP-PCA index in predicting aggregate stock returns, our study also employs ex post whole sample data in all of our analyses. Similarly, Lettau and Ludvigson's (2001) consumption-to-wealth ratio also typically suffers from "look-ahead bias" in out-of-sample forecasting. According to previous studies (Lettau and Ludvigson, 2001; Lettau and Ludvigson, 2005[25]; Shi, 2023), forecasting power cannot be spurious or generally distorted merely because one uses an ex post whole sample to estimate each parameter. Rather, many asset pricing studies accept ex post sample evidence as robust evidence (e.g., Campbell and Vuolteenaho, 2004; Petkova, 2006; Maio, 2013). Hence, our study at least indicates the distinct forecasting power of the second RP-PCA factor using ex post evidence. Finally, future research may build a reasonable link between RP-PCA factors and other dominant proxies of economic states, such as the future price of bonds, currencies, and commodities.

References

Alquist, R., Kilian, L., Vigfusson, R. J. (2013). Forecasting the Price of Oil. In *Handbook of Economic Forecasting* (2, 427–507). Elsevier.

Aastveit, K. A., Bjørnland, H. C., Thorsrud, L. A. (2015). What Drives Oil Prices? Emerging versus Developed Economies. *Journal of Applied Econometrics*, 30(7), 1013–1028. https://doi.org/10.1002/jae.2406

Asness, C. S., Frazzini, A., Pedersen, L. H. (2019). Quality Minus Junk. *Review of Accounting Studies*, 24(1), 34–112. https://doi.org/10.1007/s11142-018-9470-2

Baumeister, C., Kilian, L. (2012). Real-time Forecasts of the Real Price of Oil. *Journal of Business & Economic Statistics*, 30(2), 326–336. https://doi.org/10.1080/07350015.2011.648859

Baumeister, C., Kilian, L. (2014). What Central Bankers Need to Know about Forecasting Oil Prices. *International Economic Review*, 55(3), 869–889. https://doi.org/10.1111/iere.12074

Boons, M. (2016). State Variables, Macroeconomic Activity, and the Cross-section of Individual Stocks. *Journal of Financial Economics*, 119(3), 489–511. https://doi.org/10.1016/j.jfineco.2015.05.010

Brahmasrene, T., Huang, J. C., Sissoko, Y. (2014). Crude Oil prices and Exchange Rates: Causality, Variance Decomposition and Impulse Response. *Energy Economics*, 44, 407–412. https://doi.org/10.1016/j.eneco.2014.05.011

Campbell, J. Y., Vuolteenaho, T. (2004). Bad Beta, Good Beta. *American Economic Review*, 94(5), 1249–1275. https://doi.org/10.1257/0002828043052240

Campbell, J. Y., Thompson, S. B. (2008). Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies,* 21(4), 1509–1531.

Chen, S. S. (2014). Forecasting Crude Oil Price Movements with Oil-sensitive Stocks. *Economic Inquiry*, 52(2), 830–844. https://doi.org/10.1111/ecin.12053

25 To justify the valid predictive role of "cay", Lettau and Ludvigson (2005) provide crucial checks for defending the robustness of their out-of-sample estimations from the criticism of bias.

Chen, S. S., Chen, H. C. (2007). Oil Prices and Real Exchange rates. *Energy Economics*, 29(3), 390-404. https://doi.org/10.1016/j.eneco.2006.08.003

Chib, S., Zhao, L., Zhou, G. (2024). Winners from Winners: A Tale of Risk Factors. *Management Science*, 70(1), 396–414. https://doi.org/10.1287/mnsc.2023.4668

Clark, T. E., West, K. D. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, 138(1), 291–311. https://doi.org/10.1016/j.jeconom.2006.05.023

Cochrane, J. H. (2005). *Asset Pricing*. Princeton: Princeton University Press.

Cochrane, J. H. (2007). *Financial Markets and the Real Economy*. In R. Mehra (Ed.), Handbook of the Equity Premium, Amsterdam. The Netherlands: Elsevier.

Cooper, I., Maio, P. (2019). Asset Growth, Profitability, and Investment Opportunities. *Management Science*, 65(9), 3988–4010. https://doi.org/10.1287/mnsc.2018.3036

Dai, Z., Kang, J. (2021). Bond Yield and Crude Oil Prices Predictability. *Energy Economics*, 97, 105205. https://doi.org/10.1016/j.eneco.2021.105205

Dai, Z., Zhu, H. (2020). Stock Return Predictability from a Mixed Model Perspective. *Pacific-Basin Finance Journal*, 60, 101267. https://doi.org/10.1016/j.pacfin.2020.101267

Daniel, K., Hirshleifer, D., Sun, L. (2020). Short-and Long-horizon Behavioral Factors. *The Review of Financial Studies*, 33(4), 1673–1736.

Deeney, P., Cummins, M., Dowling, M., Bermingham, A. (2015). Sentiment in Oil Markets. *International Review of Financial Analysis*, 39, 179–185. https://doi.org/10.1016/j.irfa.2015.01.005

Dickey, D. A., Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Dong, X., Li, Y., Rapach, D. E., Zhou, G. (2022). Anomalies and the Expected Market Return. *The Journal of Finance*, 77(1), 639–681. https://doi.org/10.1111/jofi.13099

Elliott, G., Müller, U. K. (2006). Efficient Tests for General Persistent Time Variation in Regression Coefficients. *The Review of Economic Studies*, 73(4), 907–940.

Fama, E. F., French, K. R. (2015). A Five-factor Asset Pricing Model. *Journal of Financial Economics*, 116(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Hamilton, J. D., Herrera, A. M. (2004). Comment: Oil Shocks and Aggregate Macroeconomic Behavior: The Role of Monetary Policy. *Journal of Money, Credit and Banking*, 265–286.

He, M., Zhang, Y., Wen, D., Wang, Y. (2021). Forecasting Crude Oil Prices: A Scaled PCA Approach. *Energy Economics*, 97, 105189. https://doi.org/10.1016/j.eneco.2021.105189

Herrera, A. M., Pesavento, E. (2009). Oil Price Shocks, Systematic Monetary Policy, and the "Great Moderation". *Macroeconomic Dynamics*, 13(1), 107–137. https://doi.org/10.1017/S1365100508070454

Hou, K., Xue, C., Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, 28(3), 650–705.

Hou, K., Mo, H., Xue, C., Zhang, L. (2021). An Augmented Q-factor Model with Expected Growth. *Review of Finance*, 25(1), 1–41. https://doi.org/10.1093/rof/rfaa004

Huang, D., Jiang, F., Tu, J., Zhou, G. (2015). Investor Sentiment Aligned: A Powerful Predictor of Stock Returns. *The Review of Financial Studies*, 28(3), 791–837.

Kilian, L. (2009). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *American Economic Review*, 99(3), 1053–1069. https://doi.org/10.1257/aer.99.3.1053

Leitch, G., Tanner, J. E. (1991). Economic Forecast Evaluation: Profits versus the Conventional Error Measures. *The American Economic Review*, 580–590.

Lettau, M., Ludvigson, S. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *The Journal of Finance*, 56(3), 815–849. https://doi.org/10.1111/0022-1082.00347

Lettau, M., Ludvigson, S. C. (2005). Tay's as Good as Cay: Reply. *Finance Research Letters*, 2(1), 15–22. https://doi.org/10.1016/j.frl.2004.10.002

Lettau, M., Pelger, M. (2020). Factors That Fit the Time Series and Cross-section of Stock Returns. *The Review of Financial Studies*, 33(5), 2274–2325.

Lv, W., Qi, J. (2022). Stock Market Return Predictability: A Combination Forecast Perspective. *International Review of Financial Analysis*, 84, 102376. https://doi.org/10.1016/j.irfa.2022.102376

Ma, F., Liu, J., Wahab, M. I. M., Zhang, Y. (2018). Forecasting the Aggregate Oil Price Volatility in a Data-rich Environment. *Economic Modelling*, 72, 320–332. https://doi.org/10.1016/j.econmod.2018.02.009

Maio, P. (2013). Intertemporal CAPM with Conditioning Variables. *Management Science*, 59(1), 122–141. https://doi.org/10.1287/mnsc.1120.1557

Maio, P., Santa-Clara, P. (2012). Multifactor Models and Their Consistency with the ICAPM. *Journal of Financial Economics*, 106(3), 586–613. https://doi.org/10.1016/j.jfineco.2012.07.001

Neely, C. J., Rapach, D. E., Tu, J., Zhou, G. (2014). Forecasting the Equity Risk Premium: the Role of Technical Indicators. *Management Science*, 60(7), 1772–1791. https://doi.org/10.1287/mnsc.2013.1838

Newey, W. K., West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Social Science Electronic Publishing*.

Pesaran, M. H., Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics*, 10(4), 461–465. https://doi.org/10.1080/07350015.1992.10509922

Petkova, R. (2006). Do the Fama–French Factors Proxy for Innovations in Predictive Variables? *The Journal of Finance*, 61(2), 581–612. https://doi.org/10.1111/j.1540-6261.2006.00849.x

Rapach, D. E., Strauss, J. K., Zhou, G. (2010). Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, 23(2), 821–862.

Rapach, D. E., Wohar, M. E., Rangvid, J. (2005). Macro Variables and International Stock Return Predictability. *International Journal of Forecasting*, 21(1), 137–166. https://doi.org/10.1016/j.ijforecast.2004.05.004

Rapach, D., Zhou, G. (2013). *Forecasting Stock Returns*. In Handbook of Economic Forecasting (Vol. 2, 328-383). Elsevier.

Rapach, D. E., Ringgenberg, M. C., Zhou, G. (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics*, 121(1), 46–65. https://doi.org/10.1016/j.jfineco.2016.03.004

Ross, S. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13(3), 341–360. https://doi.org/10.1016/0022-0531(76)90046-6

Shi, Q., Li, B. (2022). Further Evidence on Financial Information and Economic Activity Forecasts in the United States. *The North American Journal of Economics and Finance*, 60, 101647. https://doi.org/10.1016/j.najef.2022.101647

Shi, Q. (2023). The RP-PCA Factors and Stock Return Predictability: An Aligned Approach. *The North American Journal of Economics and Finance*, 64, 101862. https://doi.org/10.1016/j.najef.2022.101862

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics*, 54(3), 375–421. https://doi.org/10.1016/S0304-405X(99)00041-0

Stambaugh, R. F., Yuan, Y. (2017). Mispricing Factors. *The Review of Financial Studies*, 30(4), 1270–1315.

Welch, I., Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21(4), 1455–1508.

Yin, L., Yang, Q. (2016). Predicting the Oil Prices: Do Technical Indicators Help? *Energy Economics*, 56, 338–350. https://doi.org/10.1016/j.eneco.2016.03.017

Zhang, Y., Ma, F., Shi, B., Huang, D. (2018). Forecasting the Prices of Crude Oil: An Iterated Combination Approach. *Energy Economics*, 70, 472–483. https://doi.org/10.1016/j.eneco.2018.01.027

Zhang, Y., Ma, F., Wang, Y. (2019). Forecasting Crude Oil Prices with a Large Set of Predictors: Can LASSO Select Powerful Predictors? *Journal of Empirical Finance*, 54, 97–117. https://doi.org/10.1016/j.jempfin.2019.08.007

Appendix

Table A1: Elliott and Müller (2006) test results

Regressor	**WTI spot price**			
H=	**1**	**3**	**6**	**12**
RPPCA1	−1.188	−1.987	−2.636	−3.321
RPPCA2	−5.925	−5.639	−6.114	−4.135
RPPCA3	−4.436	−3.431	−4.210	−6.018
RPPCA4	−2.833	−2.686	−3.562	−3.343
RPPCA5	−2.217	−3.476	−4.178	−4.913
Regressor	**Brent spot price**			
H=	1	3	6	12
RPPCA1	−1.141	−2.399	−2.654	−3.283
RPPCA2	−4.926	−5.183	−5.892	−3.856
RPPCA3	−3.123	−2.509	−4.067	−5.919
RPPCA4	−2.494	−3.047	−4.032	−3.331
RPPCA5	−2.609	−4.085	−4.445	−5.455

Notes: We report the qLL statistic of in-sample predictive regression in forecasting the WTI spot price and Brent spot price during the period of 1987:06–2017:12. Elliott and Müller 's (2006) qLL statistic aims to test null that the structure of the model is not stable: the 1%, 5%, and 10% critical values are −11.05, −8.36 and −7.14, respectively (where we reject this null hypothesis for small values).

Table A2: Estimations of out-of-sample predictability

	WTI spot price				Brent spot price			
	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
PEAD	−1.276	0.926	−2.314	0.819	−1.319	0.957	−2.672	0.908
FIN	−0.326	0.306	−2.859	0.083	−0.693	0.519	−3.758	0.104
BAB	−0.654	0.297	−1.643	0.235	−0.958	0.411	−1.917	0.257
QMJ	−0.057	0.247	−3.872	0.664	−0.053	0.209	−3.944	0.555
RMW	−0.736	0.366	−2.458	0.445	−1.495	0.543	−3.630	0.502
CMA	0.128	0.108	−1.782	0.047	−0.093	0.130	−2.976	0.069
I/A	0.082	0.163	−2.197	0.090	−0.211	0.214	−3.456	0.116
ROE	−0.372	0.285	−1.776	0.582	−1.106	0.499	−2.288	0.653
EGF	−0.133	0.297	−1.458	0.674	−0.300	0.259	−1.791	0.640
MGMT	**1.077**	**0.026**	−2.017	0.019	**0.674**	**0.036**	−3.698	0.024
PERF	−0.535	0.607	−3.163	0.961	−0.624	0.612	−3.117	0.968

Notes: We present the out-of-sample predictive forecast for the WTI spot price and Brent spot price, using a rich list of recent popular financial ratios. We report the R^2 statistic, which denotes the out-of-sample coefficient of determination (in %) in each column of *H*=1 and 3, where *H* represents the forecasting horizons. We also report the statistics of Clark and West's (2007) *MSFE adjusted p value*. The recent popular financial predictors are defined as follows: *PEAD* = the Daniel et al. (2020) short-horizon behavior factor; *FIN* = the Daniel et al. (2020) long-horizon behavior factors; *qmj* = the Asness et al. (2019) quality ('quality minus junk') factor; *bab* = the Frazzini and Pedersen (2014) betting against beta factor; *CMA* = the Fama and French (2015) investment factor; *RMW* = the Fama and French (2015) profitability factors; *I/A* = Hou et al. (2015) investment factor; *ROE* = Hou et al. (2015) profitability (ROE) factors; *EGF* = Hou et al. (2021) expected growth factor; *MGMT* = Stambaugh and Yuan's (2017) mispricing factor controlled for by management; *PERF* = Stambaugh and Yuan's (2017) mispricing factor related to performance. Forecasts begin 5 years after the sample starts. Bold displays significance at the 10% level. The benchmark is historical mean forecasts.

Table A3: The estimations of out-of-sample predictability

	Initial estimation period: 5 years				Initial estimation period: 15 years			
	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	−0.132	0.192	−4.706	0.407	−0.389	0.250	−7.276	0.619
RPPCA2	**1.593**	**0.009**	**0.927**	**0.000**	**2.374**	**0.038**	**1.292**	**0.048**
RPPCA3	−1.092	0.825	−3.026	0.888	−0.997	0.850	−2.390	0.855
RPPCA4	−1.156	0.721	−2.939	0.767	−0.492	0.749	−1.698	0.777
RPPCA5	−1.443	0.523	−2.183	0.116	−1.795	0.860	−6.410	0.852
Welch and Goyal's (2008) benchmark factors								
DP	−1.308	0.687	−5.471	0.922	−0.041	0.402	−1.268	0.654
DY	−2.350	0.780	−5.306	0.936	−0.320	0.514	−1.508	0.737
EP	−0.894	0.584	−2.925	0.722	−0.343	0.471	−1.065	0.507
DE	−2.577	0.600	−9.831	0.831	−2.870	0.552	−11.334	0.797
RVOL	−3.061	0.829	−6.160	0.949	−1.461	0.697	−2.705	0.852
BM	−1.350	0.897	−5.113	0.987	0.043	0.390	−0.924	0.710
NTIS	−1.286	0.783	−6.438	0.933	−1.231	0.739	−7.088	0.925
TBL	−1.914	0.762	−5.323	0.847	−1.433	0.893	−4.365	0.870
LTY	−1.709	0.738	−5.696	0.884	−1.335	0.785	−4.039	0.722
LTR	**1.547**	**0.008**	−1.230	0.447	**1.171**	**0.046**	−1.366	0.478
TMS	−1.463	0.886	−3.817	0.910	−0.683	0.926	−2.055	0.959
DFY	−6.056	0.605	−18.987	0.987	−4.107	0.456	−16.111	0.941
DFR	**2.669**	**0.012**	−2.141	0.099	**5.289**	**0.007**	−2.178	0.099
INFL	−1.122	0.677	−4.670	0.932	−0.882	0.547	−3.557	0.859
Other popular benchmark factors								
EPU	−0.862	0.160	−5.618	0.522	−3.528	0.306	−7.903	0.634
MS	0.171	0.122	−2.440	0.886	**1.361**	**0.034**	−1.764	0.888
IPI	−0.520	0.536	−2.173	0.928	−0.531	0.623	−1.808	0.918
COP	−1.879	0.954	−1.728	0.886	−1.801	0.923	−1.345	0.834
OI	−4.020	0.909	−1.319	0.631	−5.149	0.893	−0.199	0.388
KI	0.748	0.130	−6.525	0.620	2.068	0.106	−5.799	0.512

Notes: We present the out-of-sample predictive forecast for the WTI spot price. We report the R^2 statistic, which denotes the out-of-sample coefficient of determination (in %) in each column of *H*=1 and 3, where *H* represents the forecasting horizons. We also report the statistics of Clark and West's (2007) *MSFE adjusted p value.* Forecasts begin 5 years and 15 years after the sample starts, respectively. Bold displays significance at the 10% level. The benchmark is the no-change forecast model.

Table A4: Estimations of out-of-sample predictability

	Initial estimation period: 5 years				**Initial estimation period: 15 years**			
	***H*=1**	***P value***	***H*=3**	***P value***	***H*=1**	***P value***	***H*=3**	***P value***
RPPCA1	−0.785	0.299	−3.182	0.342	−0.954	0.366	−5.796	0.584
RPPCA2	**0.067**	**0.078**	**1.416**	**0.000**	**1.260**	**0.072**	**1.706**	**0.033**
RPPCA3	−0.657	0.839	−1.274	0.782	−0.587	0.880	−0.928	0.753
RPPCA4	−0.882	0.933	−1.303	0.557	−0.190	0.965	−0.610	0.697
RPPCA5	−0.979	0.540	−0.523	0.045	−1.083	0.796	−4.755	0.728
Welch and Goyal's (2008) benchmark factors								
DP	−1.408	0.830	−3.869	0.756	0.071	0.389	−0.065	0.443
DY	−1.774	0.771	−3.352	0.824	−0.096	0.501	−0.285	0.530
EP	−0.632	0.656	−1.171	0.492	−0.317	0.447	0.210	0.201
DE	−1.773	0.768	−6.444	0.756	−2.119	0.718	−7.866	0.747
RVOL	−1.424	0.810	−3.803	0.891	−0.882	0.761	−1.044	0.725
BM	−1.129	0.833	−3.426	0.875	0.164	0.239	0.180	0.312
NTIS	−0.617	0.784	−4.392	0.941	−0.677	0.708	−5.374	0.935
TBL	−0.998	0.945	−3.329	0.898	−1.006	0.985	−2.875	0.900
LTY	−1.264	0.993	−4.105	0.927	−0.962	0.940	−2.729	0.747
LTR	**1.224**	**0.017**	−0.066	0.170	−0.209	0.137	−0.464	0.301
TMS	−0.676	0.873	−1.844	0.721	−0.296	0.913	−0.611	0.868
DFY	−3.813	0.715	−14.874	0.973	−3.132	0.624	−12.414	0.932
DFR	**2.940**	**0.017**	−0.630	0.083	**5.274**	**0.014**	−1.085	0.102
INFL	−1.058	0.880	−3.003	0.746	−1.125	0.783	−2.057	0.801
Other popular benchmark factors								
EPU	−1.215	0.150	−3.554	0.379	−3.098	0.324	−5.591	0.548
MS	−0.485	0.512	−0.918	0.863	0.318	0.227	−0.518	0.702
IPI	−0.185	0.470	−0.482	0.667	−0.300	0.625	−0.472	0.768
COP	−1.144	0.979	−0.172	0.494	−1.215	0.955	−0.110	0.511
OI	−2.913	0.910	0.146	0.292	−3.806	0.905	1.134	0.203
KI	0.031	0.197	−4.459	0.555	0.723	0.162	−4.157	0.513

Notes: The out-of-sample forecasting model is Rapach et al. (2010) ARDL (autoregressive distributed lag) model, where we utilize first-order lag as the setup of Rapach et al (2010) model. We report the R^2 statistic, which denotes the out-of-sample coefficient of determination (in %) in each column of H=1 and 3, where H represents the forecasting horizons. We also report the statistics of Clark and West's (2007) *MSFE adjusted p value.* Forecasts begin 5 years and 15 years after the sample starts, respectively. Bold displays significance at the 10% level. The benchmark is the first-order autoregression (AR) forecast model.