The Illusion of Oil Return Predictability: The Choice of Data Matters![⋆]

Thomas Conlon[a], John Cotter[a,b], Emmanuel Eyiah-Donkor[c,*]

[a]*Michael Smurfit Graduate Business School, University College Dublin, Co. Dublin, Ireland*
[b]*Anderson School of Management, University of California, Los Angeles, USA*
[c]*Rennes School of Business, 2 Rue Robert D'Arbrissel, 35065 Rennes, France*

Abstract

Previous studies document statistically significant evidence of crude oil return predictability by several forecasting variables. We suggest that this evidence is misleading, and follows from the common use of within-month averages of daily oil price data in return predictive regressions. Averaging introduces a bias in the estimates of the first-order autocorrelation coefficient and variance of returns. Consequently, estimates of regression coefficients are inefficient and associated t-statistics are overstated, leading to false inference about the true extent of return predictability. On the contrary, using end-of-month data, we do not find convincing evidence for the predictability of oil returns. Our results highlight and provide a cautionary tale on how the choice of data could influence hypothesis testing for return predictability.

JEL classification: C22; C32; C53; Q47

Keywords: Averaged crude oil data; Spurious autocorrelation; Return predictability; Out-of-sample forecasts; Statistical inference

1. Introduction

Empirical support for the predictability of monthly crude oil spot returns based on various financial, macroeconomic fundamental, commodity market, and technical indicator variables has been well documented (see, for example, Chinn and Coibion, 2014; Yin and Yang, 2016; Zhang et al., 2018; Zhang et al., 2019; and the references therein).[1] The

[⋆]The authors acknowledge the support of Science Foundation Ireland under grant numbers 16/SPP/3347 and 17/SP/5447. We also acknowledge comments from Gregory Connor, Paulo Guasoni, Conall O'Sullivan, and conference participants at the 2019 INFINITI Conference on International Finance Glasgow, Scotland, the Forecasting Financial Market 2019 Conference in Venice, Italy, and the 2019 International Conference on Computational and Financial Econometrics in London, UK.

[*]Corresponding author. *Telephone number:* +33626756191.
Email addresses: `conlon.thomas@ucd.ie` (Thomas Conlon), `john.cotter@ucd.ie` (John Cotter), `emmanuel.eyiah-donkor@rennes-sb.com` (Emmanuel Eyiah-Donkor)

[1]While the focus of the current paper is on forecasting crude oil returns, there is also a voluminous literature, including Ye et al., 2006; Alquist et al., 2013; Chen, 2014; Baumeister and Kilian, 2015; and Baumeister et al., 2018, that forecast the price of oil in levels using monthly average price of crude oil spot.

This version: 5th May 2021

sheer number of papers devoted to forecasting the spot price and return of crude oil is not surprising considering the crucial importance of reliable forecasts for policy-making, explaining fluctuations in and projecting economic activity, and for risk management purposes by firms engaged in the production, marketing, and processing of crude oil (Black, 1976).[2]

The model typically used by the literature in examining oil return predictability is an ordinary least squares (OLS) regression where returns are regressed on a constant and the lagged values of one or more forecasting variables. Significant in-sample t-tests or some measure of out-of-sample tests are then interpreted as evidence of return predictability. In particular, the time-series of returns used in the model are calculated using within-month averages of daily prices (in which case we have monthly average returns). Studies on oil return predictability use monthly averaged returns and find predictability, whereas end-of-month returns are used for other assets with little or no predictability reported.[3]

In this paper, we comprehensively re-examine the ability of 40 popular economic and technical indicator predictor variables to forecast crude oil returns, both in- and out-of-sample, for the two data series: monthly average and end-of-month returns. The purpose is to highlight biases in the estimates of some statistical properties of the commonly used monthly average crude oil returns in predictive regressions, the econometric estimation problems, and the implications for hypothesis testing of return predictability. Returns calculated from within-month averages of daily crude oil prices, besides introducing a bias in the estimates of the first-order autocorrelation coefficient and variance of returns, will generate inefficient estimates of regression slope coefficients and results in serially correlated residuals, leading to biased estimates of standard errors. As a result, evidence against the null hypothesis of no return predictability will appear more statistically significant than they really are.

Although the aforementioned problems have been well documented in the literature for a long time (see, for example, Working, 1960; Cowles, 1960; Daniels, 1966; Rosenberg,

[2]For example, crude oil forecasts serve as a key input in gauging inflation expectations, and large fluctuations in crude oil prices have been shown to have a substantial impact on financial markets and the real economy (Hamilton, 1983, 2009; Baumeister and Peersman, 2013; Hou et al., 2016; Kilian and Vigfusson, 2017).

[3]Table A.1 of the online appendix presents a synopsis review of studies on return predictability across various asset classes, including stocks, bonds, currencies, and commodities, the price data series used in computing returns, the journal that published the article and whether or not they found evidence of predictability.

1971; Schwert, 1990; Wilson et al., 2001), it is surprising that the vast majority of the literature examining the predictability of crude oil returns continue to use averaged returns data. What solid theoretical argument supports this choice is not exactly clear. Perhaps, it simply stems from some kind of "herd behaviour" in empirical research, namely, an initial crude oil predictability study used monthly average returns and then various other studies followed. Ye et al. (2006) is the only study we are aware of that provides some rationale for the use of monthly average price as follows: (i) the average price mitigates one-day market perturbations resulting from rumours, and is less noisy;[4] (ii) the average generates better predictability results; and (iii) there is a high correlation between monthly average and end-of-month prices. We disagree with these reasons as they do not immunize returns calculated from monthly average prices from the severe consequences for econometric model estimation and predictability inference.

The reliance on monthly average prices is also problematic from the point of view of policy-making, investment decision making, and risk management. For example, because average prices do not represent actual settlement prices of crude oil, their use has the potential to pose serious problems for testing the informational efficiency of the market for crude oil. Consequently, this could affect the investments of market participants who deploy trading strategies aimed at exploiting market inefficiencies to make excess profits. Ignoring these problems may lead to very serious and damaging errors in the analysis of hedging and speculation in the crude oil market since the average price will not be available to the decision makers throughout the month. So, whether for budgeting, policy-making, risk management, or investment decisions, end-of-period prices (and therefore end-of-period returns) will be the most appropriate proxy for the instantaneous price

[4]Even though this point applies to all financial markets, the vast predictability literature does not use monthly average returns (see Table A.1 of the online appendix). For example, from the point of view of investment strategies in crude oil futures markets, an investor would, say, buy crude oil (taking a long position in the front futures contract at the end of month t) and sell crude oil (close the open position by taking a short position at the end of month $t+1$; if the price at the end of month $t+1$ is larger than the price at the end of month t then she makes a monthly profit of $Z_t = (P_{t+1} - P_t)/P_t$, where P_t and P_{t+1} are the aforementioned end-of-month prices. Accordingly, it seems then that if the investor was to base her trading decision on predictions of future monthly returns, the appropriate object of the predictability regressions shall be $Z_t = (P_{t+1} - P_t)/P_t$, where P_t and P_{t+1} are the end-of-month prices. It is difficult to fathom just why the investor might be interested in predictive regressions where the object to forecast is $W_t = (P_{t+1} - P_t)/P_t$, where P_t and P_{t+1} are not the prices that would define her actual profit (or loss) but instead the average of within-month daily prices. It seems to be that if the profits are defined by a random variable Z_t then the object of predictions should be Z_t and not something else like W_t. This is possibly why the bulk of papers in the empirical finance literature use end-of-month returns. We thank the reviewer for this comment.

quote as it reflects market conditions in real-time. Therefore, the correct returns series to use for studying predictability is end-of-month returns and not monthly average returns.

Further, we attempt to remedy the econometric issues of inefficiency of slope coefficient estimates, biased estimates of standard errors, and the severe consequence of false inference for the return predictability hypothesis. We follow standard econometric procedures by implementing two remedies: (i) we accept the efficiency loss in the OLS estimator and test for the significance of the estimated slope coefficients using t-statistics that are robust to heteroskedasticity and autocorrelation in estimated regression residuals (Newey and West, 1987); (ii) we implement a generalised least squares (GLS) estimator for slope parameters. This is motivated by the fact that in the presence of serial correlation in the regression errors, the OLS estimator is inefficient and GLS becomes the efficient estimator.

Studies that have looked at related issues include Bork et al. (2018) in the context of forecasting commodity index returns and Benmoussa et al. (2020) who examine the accuracy of model-based forecasts of the real price of crude oil using a new benchmark forecast calculated from end-of-period prices.[5] Our paper differs from these studies in that, apart from highlighting the spurious predictability of crude oil returns calculated from monthly average prices using a large set of predictors, we also implement remedial measures aimed at purging these spurious findings to shed more light on the crucial importance of the choice of returns data when examining return predictability.

Our empirical results can be summarized as follows. First, averaged crude oil price data introduces an upward bias in the estimate of the first-order autocorrelation coefficient in monthly average returns. Estimates of variance and covariance of returns with predictors are also biased downward compared to returns computed from end-of-month prices. For example, monthly average (end-of-month) returns have a first-order autocorrelation coefficient of 0.286 (0.149) and a standard deviation of 8.28% (9.16%). These agree with the findings in Working (1960) and Schwert (1990).

[5] Bork et al. (2018) highlight that the predictability findings in Chen et al. (2010) may be spurious because the commodity index returns used by Chen et al. (2010) were computed from monthly average prices which induces autocorrelation in returns. The study of Benmoussa et al. (2020) highlight that the choice of benchmark forecast matters when examining the predictive accuracy of model-based forecast of the real price of crude oil. They show that a new no-change benchmark forecast based on end-of-period prices generates more accurate forecasts than the model-based forecasts, reversing a previous conclusion where the benchmark forecast was the no-change average crude oil price.

Second, most of the individual economic and technical indicator predictor variables display statistically significant predictive ability at conventional significance levels, both in- and out-of-sample, for monthly average crude oil returns compared to forecasts from the random walk with drift benchmark model. Consistent with findings in Baumeister and Kilian (2014b), Baumeister and Kilian (2015), Yin and Yang (2016), Zhang et al. (2018), among others, we also find that combination forecasts of monthly average returns substantially improve upon the individual forecasts by generating more accurate and stable forecasts. These conclusions, however, are completely reversed, however, when end-of-month crude oil returns are used as the dependent variable in our individual and combination predictive models. The misleading inference for the predictability of monthly average crude oil returns can be attributed to the biased estimates of the statistical properties of monthly averaged returns data which, when used in predictive regressions, lead to estimates of the slope coefficient that are inefficient and estimates of associated standard errors that are biased. This result is reminiscent of findings in the existing literature that highlight how some of these biases could potentially lead to discovering highly significant predictive relationships that otherwise would not exist (see Kendall and Hill, 1953; Working, 1960; Cowles, 1960; Box and Newbold, 1971; Granger and Newbold, 1974; Phillips, 1986; Granger et al., 2001; Valkanov, 2003; Ferson et al., 2003; among others).

Our third major finding is that our earlier results about tests of predictability for monthly average returns remain largely unchanged even after testing the significance of slope coefficient estimates using test statistics which are robust to heteroskedasticity and autocorrelation in the estimated regression residuals; and after correcting the bias in slope coefficient estimates and associated standard errors using feasible generalized least squares estimators.

The rest of the paper is organised as follows. In Section 2, we highlight the biases in the estimates of statistical properties of returns from averaged data and their implications for hypothesis tests of return predictability when used in predictive regressions. Section 3 describes the crude oil price data used in computing returns, the predictor variables, and offers preliminary data analysis. In Section 4, we describe the methodology for predicting and evaluating crude oil return forecasts. The empirical analysis of in-sample and out-of-sample tests of crude oil return predictability is detailed in Section 5. Section 6 provides a discussion of remedies for the spurious autocorrelation in monthly average returns and the associated econometric model estimation issues. We offer concluding remarks in Section 7.

2. Background and Problem Statement

Before detailing the data and methodology for predicting crude oil returns, we first illustrate the econometric model estimation and inferential problems underpinning the use of monthly average returns data in predictive regressions. Suppose T monthly observations of asset prices are available, where P_t denotes the month t price of the asset. Define monthly log returns as

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right). \tag{1}$$

Studies that examine the predictability of asset returns differ depending upon the form of the price data used in (1): end-of-month prices or within-month averages of daily prices where $P_t = (1/n)\sum_{i=1}^{n} P_i$ and n is the number of trading days in the month.

The data commonly used in crude oil predictability studies is the West Texas Intermediate (WTI) crude oil prices available from website of the U.S. Energy Information Administration (EIA).[6] A note to the release of energy spot prices, including crude oil, by the EIA has the following explanatory notes:

> *Weekly, monthly, and annual prices are calculated by EIA from daily data by taking an unweighted average of the daily closing spot prices for a given product over the specified time period.*

First, and as already indicated, returns calculated from averaged data face three biases: estimates of the variance and the first-order autocorrelation coefficient are biased downward and upwards, respectively (Working (1934, 1960)), and estimates of covariance of averaged returns with other variables will be downward biased (Schwert, 1990). Working (1960) shows that the variance of the rates of change in a time-series of the average of successive data points within a given time interval is

$$\mathrm{Var}(r_t) = \left(\frac{2m^2+1}{3m^2}\right) \times \mathrm{Var}(\tilde{r}_t),$$

[6] https://www.eia.gov/

where m is the number of points within the interval (for example, m could be the number of trading days or the number of weeks within a given month), Var$(\cdot)$ is the variance operator and $\tilde{r}_t$ is the end-of-month return. The term in the first bracket which is the variance reduction factor, approaches 2/3 as n increases to infinity. Assuming that there are, on average, 21 trading days within a month, this means that the variance (standard deviation) of monthly average returns should be increased by a factor of 1.5 (1.225) to make them comparable to the variance (standard deviation) of end-of-month returns. As such, the variance of average returns is understated or downward biased by approximately 33%. Working (1960) further shows that the use of average returns to calculate the autocorrelation coefficient leads to an upward bias in the estimated first-order autocorrelation, ρ, given by:

$$\rho \equiv \operatorname{corr}(r_t, r_{t-1}) = \frac{m^2 - 1}{2(2m^2 + 1)},$$

where corr$(\cdot)$ is the correlation operator and m determines the upward bias. For example, for $m = 21$, $\rho \approx 0.25$ meaning averaged data would have first-order autocorrelation of an amount approximately .25 greater than that of the end-of-period data. Similar findings are reported in Cowles (1960), Daniels (1966), and Rosenberg (1971). These biases have been confirmed in Schwert (1990) and Wilson et al. (2001). Schwert (1990) studies CRSP monthly returns of NYSE and AMEX stocks, where returns are calculated using the average of the high and low prices within the month, whereas Wilson et al. (2001) use U.S. S&P 500 Composite Index returns from 1957 to 2001 calculated for three different types of monthly average prices: median high and low, weekly and daily. Schwert (1990) further extended the analysis to show that estimates of covariance of averaged returns with other variables will be downward biased compared to estimates based on the end-of-period returns data.

Second, suppose we are interested in knowing whether the month t value of a candidate variable, x_t, is useful for predicting the month $t+1$ value of log crude oil returns, r_{t+1}. A simple model for assessing the predictive content of x_t is the OLS regression:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}, \tag{2}$$

where the constant, α, and the slope coefficient, β, are unknown parameters to be estimated, and ε_{t+1} is an error term. The standard assumptions underlying the OLS estimator of the linear regression model are that the errors ε_{t+1} are independent of x_t ($\mathrm{E}[\varepsilon_{t+1}|x_t] = 0$) and are independent and identically distributed as normal with zero mean and constant variance (homoskedastic), and serially uncorrelated over time ($\mathrm{E}[\varepsilon_{t+s}, \varepsilon_t] = 0,\ s \neq 0$). If $\beta \neq 0$, then today's value of x can be used to predict the value of r for the next month. The null hypothesis of no-predictability, that is x_t has no predictive content for r_{t+1} and therefore $\beta = 0$, can be tested using the t-statistic for the significance of $\hat{\beta}$, the estimator for β.

As a result of the bias in the estimates of the first-order autocorrelation coefficient and variance of monthly average returns, when used in (2) above, the standard OLS assumptions underlying the model, especially the assumption of serially uncorrelated errors, will typically fail. The consequence is that although $\hat{\beta}$ is still a consistent for β, it is no longer the best linear unbiased estimator. The estimator is inefficient and estimates of the associated standard errors are biased, thus conventional test statistics based on them will be invalid (even under large sample sizes) giving rise to highly unreliable inferences when used in hypothesis testing for predictability (see Greene, 2017). Using averaged returns in the predictive regression model poses an even bigger problem for forecasting: return forecasts will be sub-optimal (Rosenberg, 1971; Box and Newbold, 1971; Granger and Newbold, 1974). Given these problems, it is likely that the evidence against the no-predictability hypothesis documented in the majority of the crude oil return predictability studies is misleading.

3. Data

3.1. Crude Oil Returns

Daily closing prices and monthly averages of the daily closing WTI crude oil spot prices are obtained from the website of the EIA.[7] From the daily prices, we build end-of-month price series. The price series, which are originally in nominal terms[8], are then deflated by the seasonally adjusted U.S consumer price index obtained from the St Louis Federal Reserve Economic Data (FRED). Log returns are calculated using the real crude

[7]The EIA defines the spot price as the price for a one-time open market transaction for immediate delivery of a specific quantity of crude oil at a specific location where the commodity is purchased "on the spot" at current market rates.

[8]Results based on nominal returns are very similar to those based on real returns and are available upon request.

oil prices. For the remainder of this paper, unless otherwise stated, returns refer to log returns. Our predictability analysis focuses on monthly real crude oil spot returns from January 1987 to December 2016, providing a total of 360 observations. This sample overlaps with the period used by many of the crude oil return predictability studies we cite in this paper.

Panel A of Table 1 presents descriptive statistics for returns. The monthly average and end-of-month return series are quite different under a number of summary headings. The mean and standard deviation of monthly average returns are lower than those of end-of-month returns. For example, the standard deviation of 8.28% for monthly average returns is about 11% lower when compared to a standard deviation of 9.16% for end-of-month returns. Monthly average returns are more left skewed and fat-tailed than end-of-month returns. Monthly average returns also have a first-order autocorrelation coefficient of 0.284, almost double the autocorrelation of 0.129 for end-of-month returns. Figure 1 plots the sample autocorrelation function (acf) up to 36 lags with 95% confidence bands for the two return series. The figure show that the first-order autocorrelation coefficient is significant at the 5% level for both monthly average and end-of-month returns. This is supported by the Lagrange multiplier test for serial correlation which indicates a rejection of the null hypothesis that the first (first 12) autocorrelation coefficient(s) is (are jointly) equal to zero for both returns series. The significantly high levels of the autocorrelation coefficient, especially for the monthly average returns, and as earlier noted may result in estimates of predictive regression slope coefficients that are inefficient and associated standard errors that are biased, leading to unduly high t-statistics for testing the significance of slope coefficients.

The augmented Dickey-Fuller test for a unit root reported in the last column of Panel A of Table 1 indicates that both monthly average and end-of-month returns are stationary. Figure 2 plots the autocorrelation function of squared returns. The figure shows evidence of heteroskedasticity and, therefore, test statistics that account for this feature of the data, as well as autocorrelation, should be used when testing for predictability. The descriptive statistics and qualitative features of monthly average returns confirm the predictions of Working (1960) and Schwert (1990) that averaging returns leads to a downward and upward bias in the estimates of the variance and the first-order autocorrelation coefficient of returns, respectively.

The estimates of covariance of returns with predictors are reported in Panel C of Table 1. Covariance estimates of monthly average returns are biased downward compared to those for end-of-month returns. This is expected to influence the estimates of beta in predictive regressions since the covariance formula, which include estimates of the standard deviation of returns and predictors, respectively, is a key component in the calculation of beta.

[Insert Table 1 about here]

[Insert Figure 1 about here]

[Insert Figure 2 about here]

3.2. Predictor Variables

We consider a set of 40 predictor variables: 28 economic and 12 technical indicator variables that have been used previously in studies on crude oil return predictability. A list of the predictors along with a brief description is provided below.

The first set of 28 economic predictors (see, for example, Baumeister and Kilian, 2014a; Baumeister et al., 2018; among others) are:

- Futures return: the log return on WTI crude oil futures traded on the New York Merchandile Exchange (NYMEX). Returns are calculated using the end-of-month settlement prices of the generic first month maturity future contract, which is constructed by rolling over to the next nearest to maturity contract at the last trading day of the month prior to the delivery month;
- Basis: the log difference between the end-of-month settlement prices of the first two nearest-to-maturity WTI crude oil futures contracts on the NYMEX;[9]
- Hedging pressure (HP): an equally weighted average of hedging pressure for each of the commodities that is a constituent of the S&P Goldman Sach's commodity index. Hedging pressure for each commodity is defined as the ratio of the difference between the dollar value of short and long hedge positions held by commercial traders to the total of the number of hedge positions;

[9]Theoretically, the basis of a commodity is defined as the difference between its contemporaneous spot price and futures price with some maturity. Empirically, because spot and futures contracts are traded on separated markets and the nearest futures price is very close to the spot price due to the no-arbitrage condition, the literature usually uses the nearest futures price to proxy the spot price to compute the basis.

- Price pressure (PP): the rate of change in hedging pressure;
- Open interest (OI): log growth rate of aggregate market open interest. To construct this variable, we aggregate dollar open interest within each of the commodities that is a constituent of the S&P Goldman Sach's commodity index, and then compute the monthly log growth rate. Finally, we compute the aggregate open interest growth as an equally weighted average of the growth rate of open interest across all commodities.
- Spot crack spread (SCS): the log growth in crack spot spread, where crack spread is defined as the sum of two-thirds of gasoline spot price and one-third of heating oil spot price minus crude oil spot price;
- Gasoline spot spread (GSS): log growth in gasoline spot spread, where gasoline spot spread is defined as the difference between gasoline and crude oil spot prices;
- Heating oil spot spread (HSS): the log growth heating oil spot spread, where heating oil spot spread is defined as the difference between heating and crude oil spot prices;
- Global oil inventory (GOI): Log growth of global crude oil inventory. The inventory data used in calculating this variable is constructed by multiplying U.S. crude oil inventories by the ratio of OECD petroleum inventories to U.S. petroleum inventories. Petroleum inventories are defined to include both stocks of crude oil and stocks of refined products;
- Global oil production (GOP): Log growth in global crude oil production. Data on global crude oil production is downloaded from the database of the EIA;
- Commodity currencies: The exchange rate of the currencies of Australia (AUS), Canada (CAN), New Zealand (NZ), and South Africa (SA) against the U.S. dollar;
- Return on S&P 500 index (S&P 500): the monthly log return on the S&P 500 index;
- Treasury bill rate (TBL): the yield on the U.S. 3-month Treasury bill (secondary market);
- Changes in Treasury bill rate (CTBL): changes in the treasury bill rate;
- Yield spread (YS): the yield on Aaa-rated bonds minus the yield on the 3-month treasury bill rate;
- Default premium (DFY): the yield on Baa-rated bond minus yield on long-term government bond;

- Term spreads (TMS1Y; TMS2Y; TMS5Y): the difference between the yield on 2- and 1-year government bonds; the difference between the yield on 5- and 2-year government bonds; and the difference between the yield on 10- and 5-year government bonds;
- VIX: Chicago Board Options Exchange volatility (CBOE) index. The VIX data is only available from January 1990. Prior to this date, we use data on the CBOE S&P 100 volatility index;
- Real economic activity (REA): the global real activity index is constructed from data on global dry cargo ocean shipping freight rates as described in Kilian (2009);
- Baltic dry index (BDI);
- Inflation (INFL): the log growth in the U.S. consumer price index;
- Capacity utilization (CAPUTIL): log degree of capacity utilization in U.S. manufacturing;
- Industrial production (INDPRO): log growth in monthly U.S. industrial production.[10]

In predictive regressions, the macroeconomic variables INFL, CAPUTIL, and INDPRO are lagged by an additional month to account for publication delays.

The second set of predictors we consider are 12 technical indicators based on three trading rules, namely, moving-average, momentum, and on-balance volume moving average (see, for example, Miffre and Rallis, 2007; Fuertes et al., 2010; Szakmary et al., 2010; Yin and Yang, 2016; among others). We use the end-of-month settlement prices and volume data on the generic first month to maturity WTI crude oil futures on the NYMEX, also from Bloomberg, to generate these technical indicators.

The moving average (MA) rule attempts to detect trends in the market prices. It generates a buy (sell) signal ($s_{i,t} = 1$ ($s_{i,t} = 0$)) at the end of month t if the short-term

[10]The sources of data for constructing the economic variables are as follows: Futures return, Basis, commodity currencies, and BDI are from Bloomberg; HP, PP, and OI are from the Commodity Futures Trading Commission (CFTC); SCS, GSS, HSS, GOI, and GOP are from the EIA; S&P 500, TBL, CTBL, YS, DFY, and INFL are available on Amit Goyal's website at http://www.hec.unil.ch/agoyal/; TMSI1Y, TMS2Y, TMS5Y, VIX, CAPUTIL, and INDPRO are from the St Louis Federal Reserve Economic Data at https://fred.stlouisfed.org/; REA is available at the Federal Reserve Bank of Dallas at https://www.dallasfed.org/research/igrea.

moving average of prices is higher (lower) than the long-term moving average of prices:

$$s_{i,t} = \begin{cases} 1, & \text{if } \mathrm{MA}_{k,t} \geq \mathrm{MA}_{l,t}, \\ 0, & \text{if } \mathrm{MA}_{k,t} < \mathrm{MA}_{l,t}, \end{cases} \tag{3}$$

where $\mathrm{MA}_{j,t} = (1/j)\sum_{j=0}^{j-1} P_{t-1}$, $j = k, l$. P_t is the level of crude oil prices, and $k\,(l)$ is the length of the short (long) look-back periods for comparing moving averages, $\mathrm{MA}(k < l)$. The MA indicator with length k and l is denoted by $\mathrm{MA}(k, l)$. Because the MA rule detects movement in prices, we should therefore expect the short MA to be more sensitive to recent movements in crude oil prices compared to the long MA. In our empirical analysis, we consider MA rules with $k = 1, 2$ and $l = 9, 12$.

The momentum (MOM) rule generates a buy or sell signal at the time t ($s_{i,t} = 1$ or $s_{i,t} = 0$) depending on whether the current crude oil price is higher than its price m periods ago. That is, a momentum rule generates the following signal:

$$s_{i,t} = \begin{cases} 1, & \text{if } P_t \geq P_{t-m}, \\ 0, & \text{if } P_t < P_{t-m}. \end{cases} \tag{4}$$

Intuitively, if the current crude oil price is higher than its price level m periods ago, this indicates "positive" momentum and relatively high expected excess returns, and will therefore generate a buy signal. We denote the momentum indicator that compares P_t to P_{t-m} by $\mathrm{MOM}(m)$, and we compute monthly signals for $m = 3, 6, 9, 12$.

The on-balance volume moving average (VOL) rule employs volume data together with past prices to identify market trends. We first define on-balance volume (OBV) as

$$\mathrm{OBV}_t = \sum_{k=1}^{t} \mathrm{VOL}_k D_k, \tag{5}$$

where VOL_k is a measure of trading volume during period k and D_k is a binary variable that takes a value 1 if $P_k - P_{k-1} \geq 0$ and -1 otherwise. We then form a trading signal, ($s_{i,t} = 1$ or $s_{i,t} = 0$, respectively) at month t from OBV_t by comparing two moving averages as

$$s_{i,t} = \begin{cases} 1, & \text{if } \mathrm{MA}_{k,t}^{\mathrm{OBV}} \geq \mathrm{MA}_{l,t}^{\mathrm{OBV}}, \\ 0, & \text{if } \mathrm{MA}_{k,t}^{\mathrm{OBV}} < \mathrm{MA}_{l,t}^{\mathrm{OBV}}, \end{cases} \tag{6}$$

where $\mathrm{MA}_{j,t}^{\mathrm{OBV}} = (1/j)\sum_{i=0}^{j-1}\mathrm{OBV}_{t-i}$, $j = k, l$. The intuition behind this rule is that recent high volume together with recent price increases in crude oil, for example, indicate a strong positive market trend and therefore generates a buy signal. We analyse VOL rules for months $k = 1, 2$ and $l = 9, 12$.

Panel A of Table 2 reports summary statistics for the economic variables. Of these variables, HP, S&P 500, TBL, YS, DFY, TMS1Y, TMS2Y, TMS5Y, VIX, and REA exhibit a high degree of persistence. However, the autocorrelations decay at a rate that is consistent with the assumption that each of the time-series is stationary. This assumption is largely confirmed by an augmented Dickey-Fuller test for a unit root which rejects the null hypothesis of a unit root in favour of the alternative that each time series of predictors is stationary.

Panel B of Table 2 reports the summary statistics for the technical indicators. Autocorrelation coefficient estimates indicate that each of the series are weakly persistent. Similar to the conclusion for the economic variables, the decay rate of the autocorrelations suggest that the series are stationary which is confirmed by the rejection of the augmented Dickey-Fuller test for unit root.

[Insert Table 2 about here]

4. Methodology

4.1. Return Prediction Models

Following the oil return predictability literature, we estimate an OLS predictive regression model as

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1}, \qquad i = 1, \cdots, N, \tag{7}$$

where r_{t+1} is the monthly log return on crude oil, $x_{i,t}$ is a predictor listed in Table 2, and ε_{t+1} is an error term.

Recent studies such as Baumeister and Kilian (2015), Zhang et al. (2018), among others, find that forecast combination methods improve upon individual forecasts of crude oil returns by generating more accurate and stable forecasts when compared to the random walk in out-of-sample predictability tests. The reasons often cited for the use of combination forecasts is that they provide a means to diversify estimation risk of the parameters of the individual predictive models and uncertainty of these models resulting from structural changes in the data (see, for example, Hendry and Clements (2004)). Because these two circumstances are difficult to model fully, the advantageous route is to use combination forecasts.

Our combination forecasts, $\hat{r}_{t+1}^{\text{Comb}}$, take the following form:

$$\hat{r}_{t+1}^{\text{Comb}} = \sum_{i=1}^{N} w_{i,t}\hat{r}_{i,t+1}, \tag{8}$$

where $\hat{r}_{i,t+1} = \hat{\alpha}_i + \hat{\beta}_i x_{i,t}$ denotes the forecast of r_{t+1} generated at time t using the i^{th} predictor, $w_{i,t}$ is the weight assigned to the i^{th} forecast with $\sum_{i=1}^{N} w_{i,t} = 1$ and N is the number of predictor variables.

The combination forecasts we consider differ in the way weights are assigned to the individual forecasts and include (i) the mean combination forecast which assigns equal weights, $w_{i,t} = 1/N, i = 1, ..., N$, to each of the individual forecasts; (ii) the trimmed mean forecast sets the $w_{i,t} = 0$ for the smallest and largest forecasts and $w_{i,t} = 1/(N-2)$ for the remaining individual forecasts; (iii) the median combination forecast is the sample median of the N individual forecasts; (iv) the weighted-mean forecast proposed by Bates and Granger (1969) specifies the combination weights to be proportional to the inverse of the estimated residual variance for the individual predictive regressions, $w_{1,t} = \frac{1/(\hat{\sigma}_{1,t}^2)}{\sum_{i=1}^{N} 1/(\hat{\sigma}_{i,t}^2)}$; and (v) the discounted mean squared forecast error (DMSFE) combination forecast following Stock and Watson (2004). Here, the combination weights are specified as functions of the historical performance of the individual predictive model forecasts over a holdout out-of-sample period,

$$w_{i,t}^{\text{DMSFE}} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}}, \quad \phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} \left(r_{s+1} - \hat{r}_{i,s+1}\right), \tag{9}$$

where $m+1$ indicates the start of the out-of-sample holdout period, and θ is a discount factor.[11] When $\theta < 1$, greater importance is attached to the individual predictive model forecast with lower mean square forecast error (MSFE). That is, the individual predictive model that generates the smallest MSFE is assigned a greater weight because it signals

[11] In practice, the DMSFE forecast requires a holdout out-of-sample period to estimate the combining weights because there are no past individual forecasts to be used to form the weight at the start of the forecast evaluation period. We therefore proceed by assigning equal weights to the first forecast over the out-of-sample period.

better forecasting performance. The special case where there is no discounting ($\theta = 1$) and forecasts are uncorrelated leads to the optimal combination weights in Bates and Granger (1969). We use a θ value of 0.9. As our final combination method, we generate out-of-sample forecasts of crude oil returns by estimating a diffusion index model following Stock and Watson (2002):

$$\hat{r}_{t+1}^{\text{PC}} = \hat{\alpha} + \sum_{k=1}^{K} \hat{\beta}_{k,t} F_{k,t}, \tag{10}$$

where $F_{k,t}$ is the kth principal component estimated from the N predictors. Diffusion indexes provide a convenient way of extracting common factor from a large number of potential predictors. To keep the model parsimonious, the number of principal components is set to a maximum of 3 and are selected using the adjusted R^2 model selection criterion.[12]

4.2. In-sample Predictability

We evaluate the in-sample predictability of each our predictors for returns by testing the significance of the slope coefficient, β_i, in (7) estimated over the full sample. Under the null hypothesis of no predictability, $\beta_i = 0$, expected crude oil returns equals a constant, α. We test $H_0 : \beta_i = 0$ against the $H_A : \beta_i \neq 0$ using a heteroskedasticity-consistent t-statistic corresponding to $\hat{\beta}_i$, the OLS estimate of β_i in (7). If the test rejects the null, then β is significantly different from zero and therefore the predictor contains useful information for explaining crude oil returns over the full sample.

4.3. Out-of-sample Predictability

To generate out-of-sample forecasts of returns, we proceed as follows. Suppose T observations are available for returns and predictors. We split the sample into two parts, use the first R observations (January 1987 to December 1996) as the initial estimation sample and the remaining $P = T - R$ observations (January 1997 to December 2016) as the out-of-sample period. Specifically, we first estimate our models using January 1987 to December 1996, and use the estimated coefficients to forecast crude oil returns for January 1997:

$$\hat{r}_{t+1} = \hat{\alpha}_i + \hat{\beta}_i x_{i,t}, \qquad i = 1, \cdots, N, \tag{11}$$

[12] We obtain similar results when we use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the OLS estimates of α_i and β_i in (7), respectively, from regressing $\{r_{t+1}\}_{t=1}^{R-1}$ on a constant and $\{x_{i,t}\}_{t=1}^{R-1}$.

We next include January 1997 in the estimation sample and use the corresponding coefficient estimates to forecast returns for February 1997. We proceed in this recursive estimation fashion,[13] re-estimating the model parameters using all previous observations, until the end of the sample in December 2016, giving rise to a time-series of P one-step-ahead out-of-sample forecasts of returns $\{\hat{r}_{t+1}\}_{t=R}^{T-1}$.

Following the convention in the return predictability literature, we evaluate the out-of-sample predictive accuracy of the forecasts from individual and combination models relative to a benchmark model. We use the Campbell and Thompson (2008) out-of-sample R^2 statistic, R^2_{OS}, which measures the proportional reduction in mean square forecast error (MSFE) for an alternative forecast relative to the MSFE of the benchmark model. That is

$$R^2_{OS} = 1 - \frac{\text{MSFE}_{model}(\hat{r}_t)}{\text{MSFE}_{bench}(\bar{r}_t)} = 1 - \frac{\sum_{t=R+1}^{T} (r_t - \hat{r}_t)^2}{\sum_{t=R+1}^{T} (r_t - \bar{r}_t)^2}, \tag{12}$$

where $\text{MSFE}_{model} = \frac{1}{T-R} \sum_{t=R+1}^{T} (r_t - \hat{r}_t)^2$, r_t is the realized return at time t and $\hat{r}_t(\bar{r}_t)$ is the crude oil predictive (benchmark) model forecast at time t. A positive R^2_{OS} value implies that the individual or combination model generates a more accurate forecast than the benchmark model. We evaluate the statistical significance of the R^2_{OS} statistic using the p-value of the MSFE-adjusted statistic of Clark and West (2007). This statistic tests the null hypothesis that the MSFE of the benchmark forecast is less than or equal to the MSFE of the individual or combination forecast against the one-sided (upper-tailed) alternative hypothesis that the benchmark MSFE is greater than the MSFE of the alternative forecast.

As a choice of benchmark, we use the random walk with drift (RWWD) forecast which means crude oil returns are independent of the predictors. Accordingly, at the end of month R, the forecasted return for month $R + 1$ is simply the average of the prior returns over the estimation window. That is, $\bar{r}_{R+1} = \hat{\alpha} = (1/R) \sum_{t=1}^{R} r_t$. This benchmark forecast is a popular choice and has consistently been used across studies on

[13] Results based on a rolling window estimation approach (which are available upon request) are very similar to those from the recursive approach.

return predictability (see, for example, Lin et al., 2017; Ahmed and Tsvetanov, 2016; Alquist and Kilian, 2010; Campbell and Thompson, 2008; and the references therein).

5. Return Predictability Analysis

5.1. In-sample Tests

Table 3 reports the in-sample predictive regression model (Equation 7) estimation results for monthly average and end-of-month returns based on each of the 40 predictors over the full sample period, January 1987 to December 2016. The table reports estimates of the slope coefficient ($\hat{\beta}$) and the associated heteroskedasticity-consistent standard errors (se($\hat{\beta}$)), the statistic of the two-tailed alternative test (t-stat) for the significance of $\hat{\beta}$, the coefficient of determination (R^2), and the Durbin-Watson statistic (DW) for testing the null hypothesis of no serial correlation of order one in the estimated regression residuals. Also reported in the table are averages of the absolute values of these statistics across the economic and technical indicators variables, respectively.

Panel A of Table 3 reports results based on the economic variables. From the table, almost all the estimates of the slope coefficients (and associated standard errors) for monthly average returns are greater (less) than those of end-of-month returns. The t-test for the null hypothesis of no-predictability of monthly average returns reveals a rejection of the null at conventional levels for 13 economic predictors, namely lagged monthly average return, futures return, PP, SCS, GSS, HSS, AUS, CAN, NZ, SA, CTBL, and BDI. Of the predictors, only 8, namely the lagged end-of-month return, the futures return, Basis, SCS, GSS, HSS, GOP, and CTBL are significant in-sample predictors for end-of-month returns albeit with much lower t-statistics. The predictability findings are also confirmed by the R^2 statistic where higher values are recorded for monthly average returns than for end-of-month returns. For example, the futures return displays a significant t-stat (R^2) of 11.17 (34.99%) for monthly average returns compared to 2.25 (2.27%) for end-of-month returns. These findings are further supported by the average t-statistics across the predictors which are significant for monthly average returns but not end-of-month returns. For both return series, the DW statistic fails to reject the null of serial correlation of order one in the estimated regression residuals for almost all predictors, although the rejection is much stronger for monthly average returns.

Panel B of Table 3 reports the estimation results based on the technical indicator variables. The null of no-predictability for monthly average crude oil returns is rejected based on the t-test at conventional levels for 10 out of the 12 technical indicator predictors, namely, MA$(1,9)$, MA$(1,12)$, MA$(2,9)$, MOM(3), MOM(6), MOM(9), MOM(12), VOL$(1,9)$, VOL$(1,12)$, VOL$(2,12)$. The t-test results for the end-of-month returns indicates a failure to reject the null of no-predictability for all the 12 technical indicators at conventional levels. This is supported by the comparatively low R^2 statistics for end-of-month returns. Essentially, what the test results tell us is that, these variables have statistically significant predictive power for monthly average crude oil returns, whereas none of the same variables contain any useful information for predicting end-of-month crude oil returns beyond a constant.

[Insert Table 3 about here]

The different inference for return predictability depending on the returns data used, especially the misleading inference for the predictability of monthly average returns, can be attributed to the bias in the estimates of the first-order autocorrelation coefficient and variance of monthly average returns reported in Table 1 leading to slope coefficient estimates that are inefficient along with bias in the estimates of the associated standard errors.

5.2. Out-of-sample Tests

Tables 4 and 5 present the out-of-sample predictability results for monthly average and end-of-month returns based on each of the economic and technical indicator predictor variables individually and their combinations, respectively. The tables report the MSFE, R^2_{OS}, and the MSFE-adjusted statistic for the significance of the R^2_{OS}. The statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. The tables also report averages of these statistics across the economic and technical indicators variables, respectively. The forecast evaluation period is January 1997 to December 2016.

Panel A of Table 4 report results for the return forecasts based on the individual economic variables. As can be seen from the table, the 10 economic variables that were found to be significant in-sample predictors for monthly average returns are also significant in the out-of-sample tests, and vice versa, at the same significance level based on a one-sided alternative test. The results for the combination forecasts in Panel B of Table 4 indicate that all the combination forecasts add substantial improvements in out-of-sample predictive performance over the RWWD forecast. All combination forecasts of monthly average returns have R^2_{OS} values that are statistically significant at the 1% level. On the contrary, only two of the variables found to be significant in-sample predictors for end-of-month returns, namely Basis and CTBL, are significant in the out-of-sample tests of predictability at the 10% and 5% levels, respectively. All other individual and combination forecasts are statistically insignificant and fail to add any improvement to the forecast from the RWWD model. This perhaps is not surprising considering that it is well documented in the return predictability literature that most economic variables that pass in-sample tests of predictability fail in out-of-sample tests (see, Welch and Goyal, 2008). Not even the combination forecasts, which are expected to guard against model uncertainty and parameter instability of individual predictive model forecasts, display statistically significant predictability for end-of-month returns. The reported findings are further supported by the average R^2_{OS} across the predictors which are statistically significant for monthly average returns but not end-of-month returns.

[Insert Table 4 about here]

Table 5 reports results for individual and combination forecasts based on the 12 technical indicator variables. All the 8 out of the 10 variables that display significant in-sample predictability for monthly average returns are also significant in out-of-sample tests of predictability. The R^2_{OS} values for all combination forecasts are also significant at the 5% level offering substantial improvement over the performance of most of the individual forecasts. Consistent with the in-sample predictability tests, R^2_{OS} values for all the individual and combination forecasts of technical indicators for end-of-month returns are statistically insignificant. The lack of predictability for end-of-month returns based on the technical indicators provides a strong warning about the dangers of using monthly averaged returns.

[Insert Table 5 about here]

Overall, our findings concerning in-sample and out-of-sample tests of monthly average return predictability confirm the predictions and findings of Working (1960), Schwert

(1990), Wilson et al. (2001), and the voluminous literature that highlight how the biases in the estimates of the first-order autocorrelation coefficient and variance of monthly average returns and the associated econometric estimation problems could influence hypothesis tests of return predictability.

6. Remedies for the Spurious Autocorrelation in Monthly Average Returns

In this section, we consider two remedial measures to deal with the biased estimate of the first-order autocorrelation of monthly average returns and the associated econometric inferential issues for testing return predictability highlighted thus far. In what follows, we detail and present predictability results using test statistics robust to heteroskedasticity and autocorrelation in the estimated predictive regression residuals, and an alternative model specification and estimation approach that directly deals with the presence of serial correlation in the regression errors.[14]

6.1. Tests for Predictability using HAC t-statistics

The Durbin-Watson statistics reported for the in-sample predictability results in Table 3 for testing the null hypothesis of no serial correlation of order one in the estimated regression residuals failed to reject the null in favour of the alternative hypothesis. As suggested by Greene (2017), if the researcher is uncomfortable with explicitly modelling the serial correlation because of specification issues, she can test the significance of β using t-statistics computed using heteroskedasticity-and-autocorrelation-consistent (HAC) standard errors à la Newey and West (1987) with 3 or 4 lags.

Table 6 reports in-sample predictability results for monthly average returns using heteroskedasticity and autocorrelation-consistent t-statistics. As a basis for comparison, the table also repeat the results for end-of-month returns that are generated using the OLS estimators of the slope coefficients and heteroskedasticity-consistent standard errors reported in Table 3. As can be seen from the table, the earlier reported findings of monthly average return predictability based on the economic and technical indicator variables are

[14] We have also included in the online Appendix as an additional remedy predictability results based on filtered returns. That is, returns generated by a filtering procedure in Schwert (1990) that adjust the biased estimates of variance and first-order autocorrelation coefficient of monthly average returns to bring them to levels closer to those of end-of-month returns. Although the procedure work well in dealing with the biases in returns, they do not change very much our earlier findings of predictability of monthly average returns. This is not surprising since as noted by Schwert, the procedure does not deal suitably with cross-correlations, which is important in our regression setting.

robust to correcting test statistics for residual autocorrelation.[15] Clearly, correcting test statistics for autocorrelation using HAC standard errors does not weaken the evidence of predictability for monthly average returns to levels similar to those for end-of-month returns, indicating that it is not sufficient to alleviate the inferential issues associated with the use of monthly average returns.

[Insert Table 6 about here]

6.2. Feasible Generalized Least Squares Estimation

It is well known that serial correlation in the estimated regression residuals has two consequences for the OLS estimators for β. That is, (a) OLS is no longer the best linear unbiased estimator and thus inefficient and (b) the usual OLS standard errors are biased. Against this backdrop, the generalized least squares (GLS) estimator of β is the most efficient. The difficulty with implementing GLS estimation is not knowing the true order of autocorrelation. This is not a concern in the present context because as can be seen from Figure 1, the autocorrelation in monthly average returns is of the first order.

To test for in-sample predictability, we implement the feasible GLS estimation based on the Prais and Winsten (1954) estimator that includes the first observation of the return series. Since this procedure is well known, we leave out the details and refer the reader to Chapter 20 of Greene (2017). The model is given by:

$$\begin{aligned} r_{t+1} &= \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1}, \\ \varepsilon_{t+1} &= \rho_i \varepsilon_t + u_t, \end{aligned} \tag{13}$$

where $\rho_i < 1$ is the first-order autocorrelation coefficient.

The in-sample predictability results for monthly average returns based on the economic and technical indicator predictor variables are reported in Table 7. The significance of the slope coefficients are tested using heteroskedasticity-consistent t-statistics.

[15] We note that to deal with the efficiency loss in the OLS estimator, one would need a larger sample size, say 500 or more monthly observations. It is possible that if the sample were large enough, predictability findings from monthly average and end-of-month returns would be fairly similar. We address this issue and increase our sample size by using weekly data. We estimated the same predictive regressions using weekly data, enabling us to have a much larger sample. The predictors were limited to only market-based variables (that is, 19 economic variables and all our 12 technical indicators) for which real time data is at the weekly frequency. The in-sample and out-of-sample predictability results (which are available upon request) from this exercise do not alter our earlier conclusions based on monthly data.

We also include, for comparison, results for end-of-month returns generated using the OLS estimators for the slope coefficients and heteroskedasticity-consistent standard errors reported in Table 3. As can be seen from the table, the evidence of predictability slightly weakens but do not resemble those for end-of-month returns. The p-values associated with DW statistic, however, indicate a failure to reject the null hypothesis of no serial correlation of order one in the estimated regression residuals in favour of the alternative for all predictors, giving support that the GLS estimation procedure remedies the residual autocorrelation.

The out-of-sample predictability for monthly average returns based on predictor variables are reported in Tables 8 and 9, respectively. The tables also include, for comparison, out-of-sample results for end-of-month returns that are generated using the OLS estimators for the slope coefficients earlier reported. As can be seen from the tables, the predictability findings for monthly average returns are largely consistent with the earlier findings for monthly average returns based on their OLS counterparts reported in Tables 4 and 5 albeit slightly weaker and nowhere close to those of end-of-month returns. It is also interesting to note that despite the negative R^2_{oos} statistics for almost individual and combination forecasts of monthly average returns, the MSFE-adjusted statistics indicate that their MSFEs are significantly less than that of the benchmark random walk with drift forecast. This might seem strange at first, but as noted by Clark and West (2007) this is possible especially when comparing nested model forecasts.

Overall, while the GLS estimation procedure only slightly weaken the evidence of predictability for monthly average returns, it does not bring it to levels similar to those for end-of-month returns, indicating that the procedure does not completely remedy the econometric inferential issues associated with the use monthly average returns in predictive regressions earlier documented.

[Insert Table 7 about here]

[Insert Table 8 about here]

[Insert Table 9 about here]

7. Conclusions

In this paper, we re-examine the evidence for monthly crude oil return predictability using 40 individual macroeconomic fundamental, commodity market, and technical indicator variables. Two sets of monthly crude oil spot returns data are considered, namely monthly average returns (calculated from within-month averages of daily closing prices) and end-of-month returns (calculated from end-of-month closing prices). The former is the data set of choice used in almost all studies on crude oil return predictability, while the latter is ubiquitous for most predictability studies on stocks, bonds, currencies and other commodities.

Using data on WTI crude oil returns and our set of predictors from January 1987 to December 2016, we find that most of the individual economic and technical indicator variables and their combinations display statistically significant predictive power for monthly average returns in both in- and out-of-sample tests of predictability. These findings are consistent with the results in the extant literature on crude oil return predictability. On the contrary, these predictability findings are completely reversed when we use end-of-month returns as the dependent variables in our predictive models. Specifically, we find no convincing evidence of predictive ability of the forecasting variables for end-of-month returns in both in-sample and out-of sample tests of predictability.

We argue that the evidence for monthly average crude oil return predictability documented in previous studies appears more significant than it really is, and is an artefact of biases in the estimates of statistical properties of monthly average returns induced by using averaged crude oil price data to calculate returns. Specifically, averaged returns data introduces a spurious upward bias in the estimate of the first-order autocorrelation coefficient in returns, and generates a downward bias in estimates of variance and covariance of returns with predictors. As a result, when used in predictive regressions, estimated slope coefficients are inefficient and associated standard errors are biased leading to false inference about the true extent of predictability. These findings accord with the results in Working (1960), Cowles (1960), Daniels (1966), Rosenberg (1971), Schwert (1990), Wilson et al. (2001), and the literature on the spurious regression problem (see Granger and Newbold, 1974; Granger et al., 2001; Ferson et al., 2003) that highlights how these statistical biases could lead to false inference when testing for return predictability.

The biases in the estimates of statistical properties of returns and the misleading econometric inference for return predictability induced by averaging the price data are so severe that remedial measures, such as calculating test statistics using heteroskedasticity and autocorrelation consistent standard errors and implementing feasible generalized least squares estimation to generate more efficient slope coefficient estimates, fail to comprehensively reverse the misleading inference for crude oil return predictability.

Our paper highlights the econometric issues associated with the use of monthly average returns in predictive regressions and how they invalidate test statistics for testing the hypotheses of return predictability if ignored by econometricians.

References

Ahmed, S., Tsvetanov, D., 2016. The predictive performance of commodity futures risk factors. Journal of Banking & Finance 71, 20–36.

Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? Journal of Applied Econometrics 25, 539–573.

Alquist, R., Kilian, L., Vigfusson, R.J., 2013. Forecasting the price of oil, in: Handbook of Economic Forecasting. Elsevier. volume 2, pp. 427–507.

Bates, J.M., Granger, C.W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Baumeister, C., Kilian, L., 2014a. A general approach to recovering market expectations from futures prices with an application to crude oil. CFS Working Paper, No 466 .

Baumeister, C., Kilian, L., 2014b. What central bankers need to know about forecasting oil prices. International Economic Review 55, 869–889.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: a forecast combination approach. Journal of Business & Economic Statistics 33, 338–351.

Baumeister, C., Kilian, L., Zhou, X., 2018. Are product spreads useful for forecasting oil prices? an empirical evaluation of the verleger hypothesis. Macroeconomic Dynamics 22, 562–580.

Baumeister, C., Peersman, G., 2013. Time-varying effects of oil supply shocks on the us economy. American Economic Journal: Macroeconomics 5, 1–28.

Benmoussa, A.A., Ellwanger, R., Snudden, S., 2020. The new benchmark for forecasts of the real price of crude oil. Bank of Canada Staff Working Paper. 2020. Available at https://www.bankofcanada.ca/wp-content/uploads/2020/09/swp2020-39.pdf .

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3, 167–179.

Bork, L., Rovira Kaltwasser, P., Sercu, P., 2018. Can exchange rates forecast commodity prices? a reconsideration. February 03, 2018, Available at SSRN: https://ssrn.com/abstract=2473624 .

Box, G.E., Newbold, P., 1971. Some comments on a paper of coen, gomme and kendall. Journal of the Royal Statistical Society: Series A (General) 134, 229–240.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chen, S.S., 2014. Forecasting crude oil price movements with oil-sensitive stocks. Economic Inquiry 52, 830–844.

Chen, Y.C., Rogoff, K.S., Rossi, B., 2010. Can exchange rates forecast commodity prices? The Quarterly Journal of Economics 125, 1145–1194.

Chinn, M.D., Coibion, O., 2014. The predictive content of commodity futures. Journal of Futures Markets 34, 607–636.

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cowles, A., 1960. A revision of previous conclusions regarding stock price behavior. Econometrica: Journal of the Econometric Society , 909–915.

Daniels, H., 1966. Autocorrelation between first differences of mid-ranges. Econometrica: Journal of the Econometric Society , 215–219.

Ferson, W.E., Sarkissian, S., Simin, T.T., 2003. Spurious regressions in financial economics? The Journal of Finance 58, 1393–1413.

Fuertes, A.M., Miffre, J., Rallis, G., 2010. Tactical allocation in commodity futures markets: Combining momentum and term structure signals. Journal of Banking & Finance 34, 2530–2548.

Granger, C.W., Hyung, N., Jeon, Y., 2001. Spurious regressions with stationary series. Applied Economics 33, 899–904.

Granger, C.W., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 111–120.

Greene, W.H., 2017. Econometric Analysis, Eighth Edition. Pearson.

Hamilton, J.D., 1983. Oil and the macroeconomy since world war ii. Journal of Political Economy 91, 228–248.

Hamilton, J.D., 2009. Causes and consequences of the oil shock of 2007–08. Brookings Papers on Economic Activity .

Hendry, D.F., Clements, M.P., 2004. Pooling of forecasts. The Econometrics Journal 7, 1–31.

Hou, K., Mountain, D.C., Wu, T., 2016. Oil price shocks and their transmission mechanism in an oil-exporting economy: A var analysis informed by a dsge model. Journal of International Money and Finance 68, 21–49.

Kendall, M.G., Hill, A.B., 1953. The analysis of economic time-series-part i: Prices. Journal of the Royal Statistical Society. Series A (General) 116, 11–34.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. The American Economic Review 99, 1053–1069.

Kilian, L., Vigfusson, R.J., 2017. The role of oil price shocks in causing us recessions. Journal of Money, Credit and Banking 49, 1747–1776.

Lin, H., Wu, C., Zhou, G., 2017. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science .

Miffre, J., Rallis, G., 2007. Momentum strategies in commodity futures markets. Journal of Banking & Finance 31, 1863–1886.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Phillips, P.C., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311–340.

Prais, S., Winsten, C., 1954. Trend estimation and serial correlation. Cowles Commission discussion paper no. 383, Chicago .

Rosenberg, B., 1971. Statistical analysis of price series obscured by averaging measures. Journal of Financial and Quantitative Analysis , 1083–1094.

Schwert, G.W., 1990. Indexes of u.s. stock prices from 1802 to 1987. Journal of Business , 399–426.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Szakmary, A.C., Shen, Q., Sharma, S.C., 2010. Trend-following trading strategies in commodity futures: A re-examination. Journal of Banking & Finance 34, 409–426.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68, 201–232.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Wilson, J.W., Jones, C.P., Lundstrum, L.L., 2001. Stochastic properties of time-averaged financial data: Explanation and empirical demonstration using monthly stock prices. Financial Review 36, 175–190.

Working, H., 1934. A random-difference series for use in the analysis of time series. journal of the American Statistical Association 29, 11–24.

Working, H., 1960. Note on the correlation of first differences of averages in a random chain. Econometrica 28, 916–918.

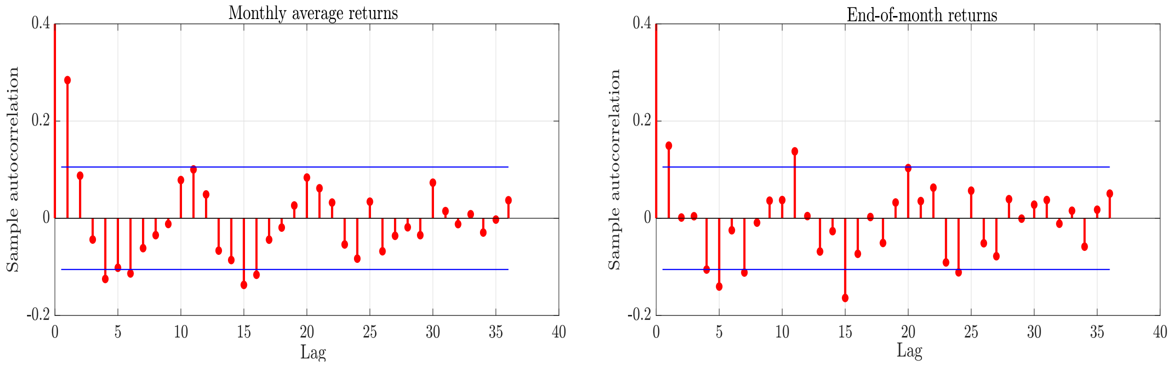
Ye, M., Zyren, J., Shore, J., 2006. Forecasting short-run crude oil price using high-and low-inventory variables. Energy Policy 34, 2736–2743.

Yin, L., Yang, Q., 2016. Predicting the oil prices: Do technical indicators help? Energy Economics 56, 338–350.

Zhang, Y., Ma, F., Shi, B., Huang, D., 2018. Forecasting the prices of crude oil: An iterated combination approach. Energy Economics 70, 472–483.

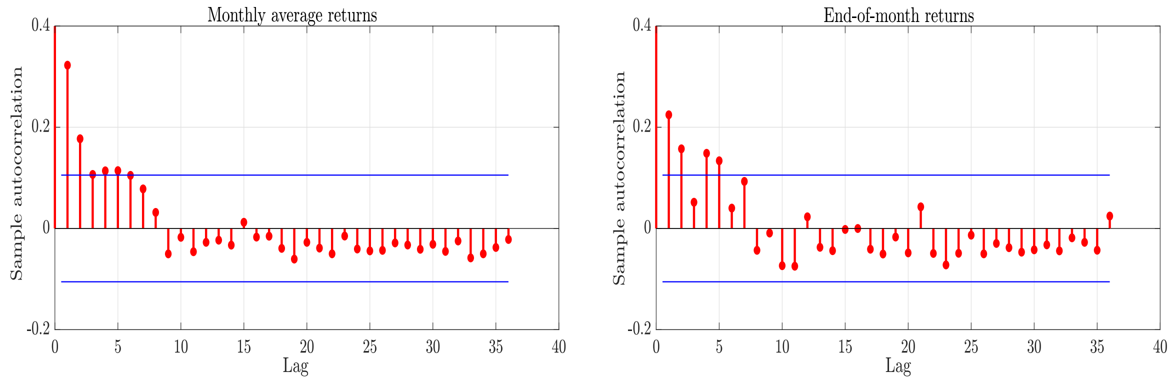
Zhang, Y., Ma, F., Wang, Y., 2019. Forecasting crude oil prices with a large set of predictors: Can lasso select powerful predictors? Journal of Empirical Finance 54, 97–117.

Figure 1: Sample Autocorrelation Functions of Crude Oil Returns



Notes. This figure plots the the sample autocorrelation functions of monthly average and end-of-month crude oil returns. The sample period is 1987:01-2016:12.

Figure 2: Sample Autocorrelation Function of Squared Crude Oil Returns



Notes. This figure plots the the sample autocorrelation functions of squared monthly average and end-of-month crude oil returns. The sample period is 1987:01-2016:12.

Table 1: Summary Statistics for Returns

Panel A: Summary statistics and autocorrelations

Returns	Mean	Std. Dev.	Skew	Kurt	Min.	Max.	Autocorrelations $\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	LM(1)	LM(12)	p(ADF)
Monthly average	0.107	8.282	−0.237	4.993	−32.366	38.348	0.284	0.088	−0.044	0.00	0.01	0.00
End-of-month	0.087	9.161	−0.101	4.505	−38.253	35.788	0.149	0.001	0.004	0.03	0.01	0.00

Panel B: Return correlations

	y_1	y_2
Monthly average (y_1)	1.00	0.72
End-of-month (y_2)		1.00

Panel C: Estimates of covariance of returns with predictors

Panel C1			Panel C2		
Economic variable (x)	Cov(y_1, x)	Cov(y_2, x)	Technical indicator (z)	Cov(y_1, z)	Cov(y_2, z)
Futures return	55.23	84.34	MA(1, 9)	183.15	176.07
Basis	3.71	3.50	MA(1, 12)	166.44	152.67
HP	13.57	14.65	MA(2, 9)	142.28	104.81
PP	10.50	18.23	MA(2, 12)	117.19	79.51
OI	6.63	7.58	MOM(3)	213.71	198.50
SCS	55.52	84.82	MOM(6)	150.37	139.27
GSS	55.41	84.77	MOM(9)	92.43	103.99
HSS	55.74	84.91	MOM(12)	122.14	118.32
GOI	−0.71	−0.81	VOL(1, 9)	167.75	182.02
GOP	−0.76	−0.18	VOL(1, 12)	159.95	158.00
AUS	6.20	10.06	VOL(2, 9)	130.87	85.34
CAN	4.57	7.28	VOL(2, 12)	114.93	91.84
NZ	6.19	7.81	Average	146.77	132.53
SA	6.59	7.27			
S&P 500	1.51	2.89			
TBL	0.82	0.68			
CTBL	0.17	0.23			
YS	−0.92	−0.91			
DFY	−0.34	−0.27			
TMS1Y	−0.07	−0.09			
TMS2Y	−0.06	−0.04			
TMS5Y	−0.14	−0.16			
VIX	−7.35	−9.05			
REA	22.36	18.63			
BDI	30.52	34.94			
INFL	0.17	0.04			
CAPUTIL	1.00	1.18			
INDPRO	0.79	0.80			
Average	12.41	17.36			

Notes. This table reports the summary statistics for two measures of monthly crude oil returns (in percent): monthly average returns and end-of-month returns. We report the mean, standard deviation, skewness, kurtosis, minimum, maximum, and sample autocorrelation coefficient coefficient up to 3 lags. LM(1) (LM(12)) is the p-values (based on heteroskedasticity robust standard errors) associated with the Lagrange multiplier test of the null hypothesis that the first (first 12) autocorrelation coefficient is (are jointly) equal to zero. p(ADF) is the p-value associated with the augmented Dickey-Fuller test of unit root. For all p-values, 0.00 indicates values less than 0.001. In the Panels, y_1 and y_2 denote monthly average returns and end-of-month returns, respectively. The table also reports the covariance between returns and the economic and technical indicator predictors variables. The variable Average is the average of the absolute covariances across all predictors. The sample period is 1987:01-2016:12.

Table 2: Summary Statistics for Predictors

							Autocorrelations					
Variable	Mean	Std. Dev.	Skew	Kurt	Min.	Max.	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	LM(1)	LM(12)	p(ADF)
Panel A: Economic predictor variables												
Futures return	0.305	9.231	−0.170	4.658	−39.484	36.893	0.164	0.000	0.009	0.02	0.01	0.00
Basis	−0.023	1.973	−0.296	6.049	−10.144	6.345	0.764	0.600	0.501	0.00	0.00	0.00
HP	4.067	9.516	0.563	3.167	−23.543	32.963	0.867	0.794	0.737	0.00	0.00	0.00
PP	0.073	4.729	−0.010	4.827	−16.746	18.258	−0.213	−0.096	−0.107	0.00	0.00	0.00
OI	0.784	6.741	0.279	5.718	−26.133	28.345	−0.133	−0.089	−0.050	0.04	0.14	0.00
SCS	0.304	9.265	−0.126	4.592	−39.071	37.158	0.162	0.009	0.007	0.02	0.01	0.00
GSS	0.303	9.262	−0.121	4.598	−38.930	37.450	0.161	0.010	0.007	0.02	0.01	0.00
HSS	0.304	9.275	−0.134	4.580	−39.354	36.575	0.162	0.008	0.006	0.02	0.01	0.00
GOI	0.103	1.119	0.267	3.275	−2.900	4.757	−0.051	−0.030	0.008	0.37	0.00	0.00
GOP	0.103	0.989	−0.560	9.768	−6.042	4.527	−0.070	−0.054	−0.059	0.43	0.16	0.00
AUS	0.022	3.339	−0.573	5.307	−17.327	9.310	0.028	−0.034	0.129	0.68	0.47	0.00
CAN	0.008	2.217	−0.606	8.084	−13.460	9.218	−0.064	0.050	−0.051	0.38	0.09	0.00
NZ	0.078	3.432	−0.446	5.095	−14.166	12.581	−0.033	−0.029	0.214	0.61	0.18	0.00
SA	−0.510	4.021	−0.448	4.491	−15.764	11.500	0.018	0.016	0.073	0.79	0.48	0.00
S&P 500	0.715	4.325	−0.809	5.640	−21.763	13.177	0.073	−0.055	0.016	0.30	0.95	0.00
TBL	3.225	2.518	0.124	1.798	0.010	8.820	0.995	0.986	0.976	0.00	0.00	0.67
CTBL	−0.014	0.187	−1.069	5.997	−0.860	0.460	0.476	0.264	0.340	0.00	0.00	0.00
YS	4.236	1.508	0.116	2.539	1.250	9.020	0.985	0.957	0.929	0.00	0.00	0.11
DFY	0.975	0.385	3.081	16.730	0.550	3.380	0.961	0.891	0.813	0.00	0.00	0.00
TMS1Y	0.343	0.269	0.062	2.846	−0.350	1.060	0.943	0.864	0.784	0.00	0.00	0.00
TMS2Y	0.621	0.500	0.165	1.785	−0.180	1.550	0.981	0.953	0.925	0.00	0.00	0.15
TMS5Y	0.549	0.422	0.236	1.930	−0.270	1.460	0.984	0.961	0.938	0.00	0.00	0.24
VIX	20.120	7.755	1.860	8.365	10.420	61.410	0.815	0.673	0.575	0.00	0.00	0.00
REA	0.799	27.248	−0.111	4.792	−133.649	66.362	0.947	0.866	0.796	0.00	0.00	0.00
BDI	0.088	18.663	−1.417	13.006	−132.979	67.107	0.137	−0.051	−0.009	0.15	0.11	0.00
INFL	0.218	0.274	−1.176	10.929	−1.699	1.216	0.410	0.056	−0.021	0.00	0.00	0.00
CAPUTIL	−0.018	0.743	−0.723	4.961	−3.184	2.003	0.229	0.272	0.211	0.00	0.02	0.00
INDPRO	−0.053	2.194	−8.768	95.279	−26.491	3.680	0.017	0.007	0.017	0.43	0.83	0.00
Panel B: Technical indicator predictor variables												
MA(1, 9)	55.556	49.760	−0.224	1.050	0.000	100.000	0.638	0.456	0.330	0.00	0.00	0.00
MA(1, 12)	56.389	49.659	−0.258	1.066	0.000	100.000	0.727	0.600	0.451	0.00	0.00	0.00
MA(2, 9)	56.667	49.623	−0.269	1.072	0.000	100.000	0.749	0.555	0.360	0.00	0.00	0.00
MA(2, 12)	58.611	49.321	−0.350	1.122	0.000	100.000	0.780	0.607	0.467	0.00	0.00	0.00
MOM(3)	56.111	49.694	−0.246	1.061	0.000	100.000	0.535	0.206	−0.056	0.00	0.00	0.00
MOM(6)	56.944	49.584	−0.280	1.079	0.000	100.000	0.658	0.463	0.269	0.00	0.00	0.00
MOM(9)	58.611	49.321	−0.350	1.122	0.000	100.000	0.655	0.561	0.479	0.00	0.00	0.00
MOM(12)	57.778	49.460	−0.315	1.099	0.000	100.000	0.695	0.511	0.452	0.00	0.00	0.00
VOL(1, 9)	59.167	49.221	−0.373	1.139	0.000	100.000	0.423	0.410	0.258	0.00	0.00	0.00
VOL(1, 12)	60.833	48.880	−0.444	1.197	0.000	100.000	0.544	0.542	0.388	0.00	0.00	0.00
VOL(2, 9)	58.889	49.272	−0.361	1.131	0.000	100.000	0.677	0.434	0.329	0.00	0.00	0.00
VOL(2, 12)	58.333	49.369	−0.338	1.114	0.000	100.000	0.792	0.642	0.560	0.00	0.00	0.00

Notes. This table reports the summary statistics for monthly predictor variables (in percent). We report the mean, standard deviation, skewness, kurtosis, minimum, maximum, and autocorrelations coefficient up to 3 lags. LM(1) (LM(12)) is the p-values (based on heteroskedasticity robust standard errors) associated with the Lagrange multiplier test of the null hypothesis that the first (first 12) autocorrelation coefficient(s) is (are jointly) equal to zero. p(ADF) is the p-value associated with the augmented Dickey-Fuller test of unit root. For all p-values, 0.00 indicates values less than 0.001. The sample period is 1987:01-2016:12.

Table 3: In-sample Predictability Results

	Monthly average returns					End-of-month returns				
Predictor	$\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW	$\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW
Panel A: Economic variables										
Lagged return	0.286	0.065	4.40***	8.19	1.98	0.150	0.067	2.25**	2.24	1.98
Futures return	0.529	0.047	11.17***	34.99	2.32	0.150	0.066	2.25**	2.27	1.98
Basis	0.087	0.259	0.33	0.04	1.41***	−0.558	0.248	−2.25**	1.44	1.64***
HP	0.036	0.045	0.79	0.17	1.42***	−0.040	0.049	−0.81	0.17	1.68***
PP	0.392	0.086	4.55***	5.00	1.53***	0.151	0.101	1.49	0.60	1.75***
OI	0.049	0.066	0.73	0.16	1.42***	−0.003	0.075	−0.04	0.00	1.69***
SCS	0.520	0.047	10.96***	34.07	2.31	0.145	0.066	2.19**	2.15	1.98
GSS	0.521	0.048	10.94***	34.10	2.31	0.145	0.066	2.18**	2.14	1.97
HSS	0.519	0.047	11.00***	33.98	2.31	0.146	0.066	2.21**	2.18	1.98
GOI	−0.119	0.358	−0.33	0.03	1.41***	0.521	0.453	1.15	0.41	1.68***
GOP	−0.543	0.473	−1.15	0.42	1.44***	−1.147	0.670	−1.71*	1.53	1.72***
AUS	0.544	0.149	3.66***	4.84	1.55***	0.164	0.180	0.91	0.36	1.75***
CAN	0.855	0.223	3.84***	5.29	1.56***	0.167	0.241	0.69	0.16	1.73***
NZ	0.296	0.149	1.99**	1.51	1.49***	0.031	0.159	0.19	0.01	1.70***
SA	0.296	0.125	2.37**	2.08	1.47***	0.103	0.141	0.73	0.20	1.71***
S&P 500	0.097	0.135	0.72	0.26	1.42***	0.072	0.150	0.48	0.12	1.70***
TBL	0.088	0.190	0.46	0.07	1.41***	0.068	0.206	0.33	0.04	1.69***
CTBL	7.324	2.975	2.46**	2.76	1.45***	8.116	3.268	2.48**	2.75	1.71***
YS	−0.373	0.369	−1.01	0.46	1.42***	−0.303	0.353	−0.86	0.25	1.69***
DFY	−0.681	2.028	−0.34	0.10	1.42***	−0.123	1.782	−0.07	0.00	1.69***
TMS1Y	−0.328	1.759	−0.19	0.01	1.41***	−0.053	1.915	−0.03	0.00	1.69***
TMS2Y	0.039	0.983	0.04	0.00	1.41***	0.103	1.088	0.09	0.00	1.69***
TMS5Y	−0.988	1.002	−0.99	0.26	1.41***	−0.882	1.079	−0.82	0.17	1.69***
VIX	−0.118	0.085	−1.38	1.22	1.43***	−0.093	0.083	−1.11	0.61	1.70***
REA	0.012	0.019	0.60	0.15	1.42***	0.008	0.020	0.41	0.06	1.69***
BDI	0.081	0.033	2.47**	3.36	1.49***	0.045	0.035	1.30	0.84	1.74***
INFL	0.530	2.105	0.25	0.03	1.41***	0.052	1.963	0.03	0.00	1.69***
CAPUTL	0.828	0.783	1.06	0.56	1.43***	0.127	0.810	0.16	0.01	1.69***
INDPRO	−0.019	0.167	−0.11	0.00	1.41***	0.038	0.167	0.23	0.01	1.69***
Average	0.590	0.511	2.77***	6.00	1.58***	0.473	0.537	1.02	0.71	1.75***
Panel B: Technical indicator variables										
MA(1, 9)	0.0259	0.0088	2.95***	2.44	1.49***	−0.0025	0.0098	−0.26	0.02	1.68***
MA(1, 12)	0.0279	0.0089	3.15***	2.82	1.50***	0.0074	0.0098	0.75	0.16	1.71***
MA(2, 9)	0.0165	0.0090	1.83*	0.98	1.45***	−0.0011	0.0098	−0.11	0.00	1.69***
MA(2, 12)	0.0104	0.0092	1.14	0.39	1.43***	−0.0045	0.0099	−0.46	0.06	1.68***
MOM(3)	0.0372	0.0087	4.30***	5.00	1.58***	0.0106	0.0096	1.09	0.33	1.73***
MOM(6)	0.0180	0.0090	1.99**	1.17	1.45***	−0.0050	0.0099	−0.50	0.07	1.68***
MOM(9)	0.0163	0.0089	1.82*	0.94	1.42***	−0.0060	0.0098	−0.62	0.11	1.68***
MOM(12)	0.0221	0.0089	2.48**	1.75	1.46***	0.0061	0.0098	0.62	0.11	1.70***
VOL(1, 9)	0.0312	0.0092	3.40***	3.45	1.51***	−0.0012	0.0101	−0.12	0.00	1.69***
VOL(1, 12)	0.0273	0.0095	2.88***	2.61	1.50***	0.0039	0.0103	0.38	0.04	1.70***
VOL(2, 9)	0.0109	0.0092	1.18	0.43	1.44***	−0.0015	0.0101	−0.14	0.01	1.69***
VOL(2, 12)	0.0192	0.0093	2.08**	1.32	1.45***	0.0048	0.0102	0.47	0.07	1.70***
Average	0.0219	0.0090	2.43**	1.94	1.47***	0.0045	0.0099	0.46	0.08	1.69***

Notes. This table reports the in-sample OLS estimation results for the predictive regression models of crude oil returns in Equation (7). The return series are monthly average returns and end-of-month returns. The table reports the slope coefficient, $\hat{\beta}$, and the associated heteroskedasticity-consistent standard errors, se($\hat{\beta}$), the statistic for the two-tailed alternative test, t-stat, for the significance of $\hat{\beta}$. R^2 is the coefficient of determination, and DW is the Durbin-Watson statistic for testing the null hypothesis of no serial correlation of order one in the estimated regression residuals. The variable Average is the average of the absolute values of beta estimates, standard errors, t-stats, R^2, and DW statistics across the predictors. Results are reported for the full sample period 1987:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Out-of-Sample Predictability Results: Economic Variables

	Monthly average returns			End-of-month returns		
Predictor	MSFE	R^2_{OS} (%)	MSFE-adjusted	MSFE	R^2_{OS} (%)	MSFE-adjusted
RWWD	75.51			91.51		
Panel A: Individual predictive model forecasts						
Futures return	52.92	29.92	5.73***	91.27	0.26	1.07
Basis	76.01	−0.67	−0.27	90.57	1.02	1.57*
HP	76.79	−1.70	1.22	91.83	−0.35	−0.35
PP	73.47	2.69	2.84***	91.62	−0.13	0.45
OI	75.61	−0.14	−0.28	91.89	−0.41	−1.65
SCS	53.82	28.72	5.67***	91.50	0.01	0.94
GSS	53.80	28.75	5.67***	91.50	0.00	0.94
HSS	53.89	28.63	5.67***	91.50	0.01	0.95
GOI	75.78	−0.36	−0.89	92.14	−0.69	−0.51
GOP	75.67	−0.21	0.14	92.56	−1.15	0.17
AUS	72.03	4.61	2.60***	93.06	−1.69	−0.65
CAN	71.23	5.67	3.20***	92.49	−1.07	−0.87
NZ	75.30	0.27	1.11	93.87	−2.58	−1.56
SA	74.25	1.67	2.29**	92.32	−0.89	−0.53
S&P 500	76.96	−1.92	−0.44	92.30	−0.86	−0.22
TBL	76.39	−1.17	−1.34	92.44	−1.02	−1.51
CTBL	74.21	1.72	1.52*	89.69	1.98	1.88**
YS	76.81	−1.72	−0.48	92.93	−1.55	−0.82
DFY	78.40	−3.83	−0.07	93.85	−2.57	−0.31
TMS1Y	76.13	−0.82	−0.75	92.17	−0.73	−1.02
TMS2Y	75.94	−0.58	−1.29	92.06	−0.61	−1.36
TMS5Y	76.74	−1.63	−0.32	92.85	−1.47	−0.60
VIX	75.38	0.17	0.57	92.02	−0.57	0.37
REA	76.60	−1.45	−0.42	92.90	−1.52	−0.89
BDI	73.78	2.29	1.71**	92.91	−1.53	0.10
INFL	76.61	−1.46	−0.29	92.59	−1.19	−1.10
CAPUTIL	76.21	−0.92	0.53	92.42	−0.99	−1.03
INDPRO	76.07	−0.74	−0.81	92.07	−0.61	−1.46
Average	72.39	4.14	1.17	92.19	−0.75	−0.28
Panel B: Combination forecasts						
Mean	68.72	8.99	4.79***	91.47	0.04	0.27
Median	74.55	1.27	2.55***	91.56	−0.06	−0.18
Trimmed mean	69.25	8.29	4.78***	91.44	0.08	0.35
Weighted mean	66.71	11.66	5.06***	91.47	0.04	0.28
DMSFE ($\theta = 0.9$)	66.78	11.56	4.51***	91.49	0.01	0.23
PC (IC $= R^2$)	57.83	23.41	5.01***	92.46	−1.04	0.71
Average	67.31	10.86	4.45***	91.65	−0.16	0.28

Notes. This table reports out-of-sample results for the individual and combination forecasts of crude oil returns based on 28 economic variables. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. The variable Average is the average of the MSFE, R^2_{OS}, and MSFE-adjusted statistics across the predictors. Results are reported for monthly average returns and end-of-month returns. The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Out-of-Sample Predictability Results: Technical Indicators

	Monthly average returns			End-of-month returns		
Predictor	MSFE	R^2_{OS} (%)	MSFE-adjusted	MSFE	R^2_{OS} (%)	MSFE-adjusted
RWWD	75.51			91.51		
Panel A: Individual predictive model forecasts						
MA$(1,9)$	73.93	2.08	2.11**	91.93	−0.47	−0.98
MA$(1,12)$	73.83	2.22	2.15**	91.78	−0.30	−0.24
MA$(2,9)$	74.87	0.84	1.28	92.33	−0.90	−0.73
MA$(2,12)$	75.63	−0.16	0.02	92.48	−1.07	−1.15
MOM(3)	72.73	3.68	2.88***	92.16	−0.72	−0.50
MOM(6)	74.75	1.01	1.34*	92.37	−0.94	−0.54
MOM(9)	74.91	0.79	1.23	92.19	−0.74	−0.84
MOM(12)	74.68	1.09	1.54*	91.83	−0.35	−0.71
VOL$(1,9)$	73.29	2.94	2.51***	92.03	−0.58	−0.80
VOL$(1,12)$	73.97	2.04	1.99**	91.95	−0.48	−1.92
VOL$(2,9)$	75.72	−0.28	−0.16	92.72	−1.33	−1.33
VOL$(2,12)$	74.53	1.29	1.63*	92.18	−0.74	−1.58
Average	74.40	1.46	1.54*	92.16	−0.72	−0.94
Panel B: Combination forecasts						
Mean	73.86	2.18	2.19**	92.00	−0.54	−1.85
Median	73.82	2.23	2.21**	92.13	−0.69	−1.63
Trimmed mean	73.93	2.08	2.11**	92.02	−0.56	−1.71
Weighted mean	73.85	2.20	2.20**	92.01	−0.55	−1.85
DMSFE ($\theta = 0.9$)	73.93	2.09	2.10**	92.13	−0.68	−2.47
PC ($\textsc{ic} = R^2$)	74.49	1.35	1.88**	93.06	−1.70	−1.45
Average	73.98	2.02	2.11**	92.23	−0.78	−1.83

Notes. This table reports out-of-sample results for the individual and combination forecasts of crude oil returns based on 18 technical indicator variables. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. The variable Average is the average of the MSFE, R^2_{OS}, and MSFE-adjusted statistics across the predictors. Results are reported for monthly average returns and end-of-month returns. The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: In-sample Predictability Results with HAC Standard Errors

Predictor	Monthly average returns $\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	End-of-month returns $\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)
Panel A: Economic variables								
Lagged return	0.286	0.069	4.14***	8.19	0.150	0.067	2.25**	2.24
Futures return	0.529	0.048	11.01***	34.99	0.150	0.066	2.25**	2.27
Basis	0.087	0.216	0.40	0.04	−0.558	0.248	−2.25**	1.44
HP	0.036	0.057	0.63	0.17	−0.040	0.049	−0.81	0.17
PP	0.392	0.086	4.58***	5.00	0.151	0.101	1.49	0.60
OI	0.049	0.065	0.75	0.16	−0.003	0.075	−0.04	0.00
SCS	0.520	0.047	10.97***	34.07	0.145	0.066	2.19**	2.15
GSS	0.521	0.047	10.96***	34.10	0.145	0.066	2.18**	2.14
HSS	0.519	0.047	10.97***	33.98	0.146	0.066	2.21**	2.18
GOI	−0.119	0.377	−0.32	0.03	0.521	0.453	1.15	0.41
GOP	−0.543	0.432	−1.26	0.42	−1.147	0.670	−1.71*	1.53
AUS	0.544	0.187	2.91***	4.84	0.164	0.180	0.91	0.36
CAN	0.855	0.232	3.69***	5.29	0.167	0.241	0.69	0.16
NZ	0.296	0.193	1.54	1.51	0.031	0.159	0.19	0.01
SA	0.296	0.147	2.01**	2.08	0.103	0.141	0.73	0.20
S&P 500	0.097	0.174	0.55	0.26	0.072	0.150	0.48	0.12
TBL	0.088	0.229	0.38	0.07	0.068	0.206	0.33	0.04
CTBL	7.324	3.976	1.84*	2.76	8.116	3.268	2.48**	2.75
YS	−0.373	0.484	−0.77	0.46	−0.303	0.353	−0.86	0.25
DFY	−0.681	2.804	−0.24	0.10	−0.123	1.782	−0.07	0.00
TMS1Y	−0.328	2.457	−0.13	0.01	−0.053	1.915	−0.03	0.00
TMS2Y	0.039	1.228	0.03	0.00	0.103	1.088	0.09	0.00
TMS5Y	−0.988	1.156	−0.85	0.26	−0.882	1.079	−0.82	0.17
VIX	−0.118	0.115	−1.03	1.22	−0.093	0.083	−1.11	0.61
REA	0.012	0.021	0.57	0.15	0.008	0.020	0.41	0.06
BDI	0.081	0.044	1.84*	3.36	0.045	0.035	1.30	0.84
INFL	0.530	2.327	0.23	0.03	0.052	1.963	0.03	0.00
CAPUTIL	0.828	0.898	0.92	0.56	0.127	0.810	0.16	0.01
INDPRO	−0.019	0.175	−0.11	0.00	0.038	0.167	0.23	0.01
Average	0.590	0.632	2.61***	6.00	0.473	0.537	1.02	0.71
Panel B: Technical indicator variables								
MA(1,9)	0.0259	0.0084	3.08***	2.44	−0.0025	0.0098	−0.26	0.02
MA(1,12)	0.0279	0.0088	3.16***	2.82	0.0074	0.0098	0.75	0.16
MA(2,9)	0.0165	0.0086	1.93*	0.98	−0.0011	0.0098	−0.11	0.00
MA(2,12)	0.0104	0.0090	1.17	0.39	−0.0045	0.0099	−0.46	0.06
MOM(3)	0.0372	0.0090	4.11***	5.00	0.0106	0.0096	1.09	0.33
MOM(6)	0.0180	0.0093	1.94*	1.17	−0.0050	0.0099	−0.50	0.07
MOM(9)	0.0163	0.0089	1.82*	0.94	−0.0060	0.0098	−0.62	0.11
MOM(12)	0.0221	0.0087	2.53**	1.75	0.0061	0.0098	0.62	0.11
VOL(1,9)	0.0312	0.0093	3.37***	3.45	−0.0012	0.0101	−0.12	0.00
VOL(1,12)	0.0273	0.0094	2.92***	2.61	0.0039	0.0103	0.38	0.04
VOL(2,9)	0.0109	0.0097	1.12	0.43	−0.0015	0.0101	−0.14	0.01
VOL(2,12)	0.0192	0.0097	1.98**	1.32	0.0048	0.0102	0.47	0.07
Average	0.0219	0.0091	2.43**	1.94	0.0045	0.0099	0.46	0.08

Notes. This table reports the in-sample estimation results for the predictive regression models of crude oil returns in Equation (7). The return series is either monthly average or end-of-month returns. The table reports the slope coefficient, $\hat{\beta}$, and the associated Newey and West (1987) heteroskedasticity-and-autocorrelation-consistent standard errors, se($\hat{\beta}$), computed with 4 lags, the statistic for the two-tailed alternative test, t-stat, for the significance of $\hat{\beta}$. For comparison, we repeat the results for end-of-month returns that are generated using the OLS estimators for the slope coefficients and heteroskedasticity-consistent standard errors reported in Table 3. The variable Average is the average of the absolute values of beta estimates, standard errors, t-stats, and R^2 statistics across the predictors. Results are computed for the full sample period 1987:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: In-sample Predictability Results based on Feasible Generalized Least Squares

	Monthly average returns					End-of-month returns				
Predictor	$\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW	$\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW
Panel A: Economic variables										
Lagged return	0.297	0.065	4.57***	8.81	1.97	0.150	0.067	2.25**	2.24	1.98
Futures return	0.649	0.047	13.88***	49.77	2.04	0.150	0.066	2.25**	2.27	1.98
Basis	0.057	0.311	0.18	0.01	1.99	−0.558	0.248	−2.25**	1.44	1.64***
HP	0.023	0.053	0.43	0.04	1.99	−0.040	0.049	−0.81	0.17	1.68***
PP	0.252	0.078	3.25***	2.64	2.02	0.151	0.101	1.49	0.60	1.75***
OI	0.007	0.056	0.13	0.00	1.99	−0.003	0.075	−0.04	0.00	1.69***
SCS	0.640	0.047	13.54***	48.69	2.04	0.145	0.066	2.19**	2.15	1.98
GSS	0.640	0.047	13.49***	48.69	2.04	0.145	0.066	2.18**	2.14	1.97
HSS	0.639	0.047	13.63***	48.62	2.04	0.146	0.066	2.21**	2.18	1.98
GOI	−0.038	0.332	−0.11	0.00	1.99	0.521	0.453	1.15	0.41	1.68***
GOP	−0.109	0.353	−0.31	0.02	1.99	−1.147	0.670	−1.71*	1.53	1.72***
AUS	0.329	0.129	2.55**	2.03	1.98	0.164	0.180	0.91	0.36	1.75***
CAN	0.531	0.197	2.70***	2.43	1.99	0.167	0.241	0.69	0.16	1.73***
NZ	0.093	0.123	0.75	0.18	1.99	0.031	0.159	0.19	0.01	1.70***
SA	0.166	0.108	1.54	0.76	1.98	0.103	0.141	0.73	0.20	1.71***
S&P 500	0.037	0.110	0.34	0.04	1.99	0.072	0.150	0.48	0.12	1.70***
TBL	0.083	0.250	0.33	0.04	1.99	0.068	0.206	0.33	0.04	1.69***
CTBL	6.361	3.033	2.10**	1.85	1.99	8.116	3.268	2.48**	2.75	1.71***
YS	−0.457	0.479	−0.95	0.39	2.00	−0.303	0.353	−0.86	0.25	1.69***
DFY	−0.584	2.504	−0.23	0.04	1.99	−0.123	1.782	−0.07	0.00	1.69***
TMS1Y	−0.460	2.184	−0.21	0.01	1.99	−0.053	1.915	−0.03	0.00	1.69***
TMS2Y	0.032	1.288	0.03	0.00	1.99	0.103	1.088	0.09	0.00	1.69***
TMS5Y	−1.240	1.371	−0.90	0.23	1.99	−0.882	1.079	−0.82	0.17	1.69***
VIX	−0.123	0.092	−1.33	0.90	2.00	−0.093	0.083	−1.11	0.61	1.70***
REA	0.008	0.027	0.28	0.04	1.99	0.008	0.020	0.41	0.06	1.69***
BDI	0.053	0.029	1.85*	1.55	1.98	0.045	0.035	1.30	0.84	1.74***
INFL	0.370	2.223	0.17	0.01	1.99	0.052	1.963	0.03	0.00	1.69***
CAPUTIL	0.341	0.690	0.49	0.10	1.99	0.127	0.810	0.16	0.01	1.69***
INDPRO	−0.074	0.111	−0.67	0.05	1.99	0.038	0.167	0.23	0.01	1.69***
Average	0.507	0.565	2.79***	7.52	2.00	0.473	0.537	1.02	0.71	1.75***
Panel B: Technical indicator variables										
MA(1,9)	0.0139	0.0102	1.36	0.56	1.98	−0.0025	0.0098	−0.26	0.02	1.68***
MA(1,12)	0.0173	0.0105	1.65*	0.82	1.98	0.0074	0.0098	0.75	0.16	1.71***
MA(2,9)	0.0060	0.0108	0.56	0.09	1.98	−0.0011	0.0098	−0.11	0.00	1.69***
MA(2,12)	−0.0006	0.0104	−0.06	0.00	1.99	−0.0045	0.0099	−0.46	0.06	1.68***
MOM(3)	0.0227	0.0095	2.39**	1.63	1.97	0.0106	0.0096	1.09	0.33	1.73***
MOM(6)	0.0090	0.0099	0.91	0.23	1.98	−0.0050	0.0099	−0.50	0.07	1.68***
MON(9)	0.0152	0.0100	1.51	0.63	1.98	−0.0060	0.0098	−0.62	0.11	1.68***
MOM(12)	0.0152	0.0102	1.50	0.63	1.99	0.0061	0.0098	0.62	0.11	1.70***
VOL(1,9)	0.0199	0.0090	2.22**	1.30	1.98	−0.0012	0.0101	−0.12	0.00	1.69***
VOL(1,12)	0.0149	0.0098	1.52	0.67	1.98	0.0039	0.0103	0.38	0.04	1.70***
VOL(2,9)	−0.0037	0.0103	−0.36	0.04	1.99	−0.0015	0.0101	−0.14	0.01	1.69***
VOL(2,12)	0.0103	0.0109	0.95	0.27	1.98	0.0048	0.0102	0.47	0.07	1.70***
Average	0.0124	0.0101	1.25	0.57	1.98	0.0045	0.0099	0.46	0.08	1.69***

Notes. This table reports the in-sample feasible generalized least squares estimation results for the predictive regression model in Equation (13) of monthly average and end-of-month crude oil returns. The table reports the slope coefficient, $\hat{\beta}$, and the associated heteroskedasticity-consistent standard errors, se($\hat{\beta}$), the statistic for the two-tailed alternative test, t-stat, for the significance of $\hat{\beta}$. R^2 is the coefficient of determination, and DW is the Durbin-Watson statistic for testing the null hypothesis of no serial correlation of order one in the estimated regression residuals. For comparison, we repeat the results for end-of-month returns that are generated using the OLS estimators for the slope coefficients and heteroskedasticity-consistent standard errors reported in Table 3. The variable Average is the average of the absolute values of beta estimates, standard errors, t-stats, R^2, and DW statistics across the predictors. Results are reported for the full sample period 1987:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Out-of-Sample Predictability Results using FGLS: Economic Variables

Predictor	Monthly average returns MSFE	Monthly average returns R^2_{OS} (%)	Monthly average returns MSFE-adjusted	End-of-month returns MSFE	End-of-month returns R^2_{OS} (%)	End-of-month returns MSFE-adjusted
RWWD	75.51			91.51		
Panel A: Individual predictive model forecasts						
Futures return	57.61	23.70	5.77***	91.27	0.26	1.07
Basis	76.71	−1.59	1.51*	90.57	1.02	1.57*
HP	76.13	−0.83	2.04**	91.83	−0.35	−0.35
PP	71.37	5.48	2.86***	91.62	−0.13	0.45
OI	76.31	−1.06	1.52*	91.89	−0.41	−1.65
SCS	58.53	22.49	5.69***	91.50	0.01	0.94
GSS	58.46	22.58	5.69***	91.50	0.00	0.94
HSS	58.65	22.32	5.69***	91.50	0.01	0.95
GOI	76.64	−1.50	1.47*	92.14	−0.69	−0.51
GOP	76.27	−1.01	1.52*	92.56	−1.15	0.17
AUS	74.54	1.28	2.01**	93.06	−1.69	−0.65
CAN	73.94	2.07	2.21**	92.49	−1.07	−0.87
NZ	76.90	−1.84	1.41*	93.87	−2.58	−1.56
SA	76.10	−0.78	1.64*	92.32	−0.89	−0.53
S&P 500	76.82	−1.73	1.56*	92.30	−0.86	−0.22
TBL	76.94	−1.90	1.46*	92.44	−1.02	−1.51
CTBL	75.02	0.65	1.78**	89.69	1.98	1.88**
YS	77.26	−2.32	1.34*	92.93	−1.55	−0.82
DFY	78.92	−4.52	1.21	93.85	−2.57	−0.31
TMS1Y	76.76	−1.65	1.53*	92.17	−0.73	−1.02
TMS2Y	76.55	−1.38	1.49*	92.06	−0.61	−1.36
TMS5Y	76.85	−1.78	1.41*	92.85	−1.47	−0.60
VIX	76.08	−0.75	1.46*	92.02	−0.57	0.37
REA	77.44	−2.56	1.34*	92.90	−1.52	−0.89
BDI	76.21	−0.94	1.72**	92.91	−1.53	0.10
INFL	76.08	−0.76	1.71**	92.59	−1.19	−1.10
CAPUTIL	76.19	−0.90	1.65**	92.42	−0.99	−1.03
INDPRO	76.39	−1.17	1.52*	92.07	−0.61	−1.46
Average	73.63	2.49	2.22**	92.19	−0.75	−0.28
Panel B: Combination forecasts						
Mean	67.95	10.00	3.29***	91.47	0.04	0.27
Median	75.75	−0.32	1.65**	91.56	−0.06	−0.18
Trimmed mean	68.99	8.63	3.09***	91.44	0.08	0.35
Weighted mean	65.55	13.19	3.79***	91.47	0.04	0.28
DMSFE ($\theta = 0.9$)	66.84	11.48	3.45***	91.49	0.01	0.23
PC ($\textsc{ic} = R^2$)	60.70	19.61	5.46***	92.46	−1.04	0.71
Average	67.63	10.43	3.45***	91.65	−0.16	0.28

Notes. This table reports out-of-sample results for the individual and combination forecasts of monthly average crude oil returns based on 28 economic variables using feasible generalised least squares estimators of the model parameters. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. For comparison, we repeat the results for end-of-month returns that are generated using the OLS estimators for the slope coefficients reported in Table 4. The variable Average is the average of the MSFE, R^2_{OS}, and MSFE-adjusted statistics across the predictors. The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Out-of-Sample Predictability Results using FGLS: Technical Indicators

	Monthly average returns			End-of-month returns		
Predictor	MSFE	R^2_{OS} (%)	MSFE-adjusted	MSFE	R^2_{OS} (%)	MSFE-adjusted
RWWD	75.51			91.51		
Panel A: Individual predictive model forecasts						
MA(1, 9)	75.72	−0.28	1.85**	91.93	−0.47	−0.98
MA(1, 12)	75.66	−0.20	1.98**	91.78	−0.30	−0.24
MA(2, 9)	76.43	−1.22	1.46*	92.33	−0.90	−0.73
MA(2, 12)	77.12	−2.13	1.27	92.48	−1.07	−1.15
MOM(3)	75.00	0.68	2.22**	92.16	−0.72	−0.50
MOM(6)	76.64	−1.50	1.45*	92.37	−0.94	−0.54
MOM(9)	76.01	−0.66	1.67**	92.19	−0.74	−0.84
MOM(12)	75.62	−0.15	1.81**	91.83	−0.35	−0.71
VOL(1, 9)	74.95	0.74	1.97**	92.03	−0.58	−0.80
VOL(1, 12)	75.34	0.23	1.81**	91.95	−0.48	−1.92
VOL(2, 9)	77.46	−2.58	1.22	92.72	−1.33	−1.33
VOL(2, 12)	76.42	−1.21	1.48*	92.18	−0.74	−1.58
Average	76.03	−0.69	1.68**	92.16	−0.72	−0.94
Panel B: Combination forecasts						
Mean	75.75	−0.33	1.70**	92.00	−0.54	−1.85
Median	75.69	−0.24	1.73**	92.13	−0.69	−1.63
Trimmed mean	75.77	−0.35	1.70**	92.02	−0.56	−1.71
Weighted mean	75.75	−0.31	1.71**	92.01	−0.55	−1.85
DMSFE ($\theta = 0.9$)	75.75	−0.31	1.71**	92.13	−0.68	−2.47
PC (IC $= R^2$)	76.19	−0.90	1.98**	93.06	−1.70	−1.45
Average	75.84	−0.44	1.75**	92.23	−0.78	−1.83

Notes. This table reports out-of-sample results for the individual and combination forecasts of monthly average crude oil returns based on 18 technical indicator variables using feasible generalised least squares estimators of the model parameters. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. For comparison, we repeat the results for end-of-month returns that are generated using the OLS estimators for the slope coefficients reported in Table 5. The variable Average is the average of the MSFE, R^2_{OS}, and MSFE-adjusted statistics across the predictors. The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Internet Appendix to
"The Illusion of Oil Return Predictability: The Choice of Data Matters!"

This version: May 5, 2021

This internet appendix presents Table A.1 referenced in the paper, details a procedure to adjust the bias in estimates of the first-order autocorrelation coefficient and variance of monthly average returns following Schwert (1990) to generate a new return series, and provide a discussion of the predictability findings from this new return series.

A. Additional Analysis and Supplementary Tables

A.1. A Review of Price Data Series used in Return Predictability Studies

Table A.1 presents a literature review of the price series used in computing returns in studies of return predictability across commodity, stock, bond, and currency markets discussed in footnote 3 of the paper.

[Insert Table A.1 about here]

A.2. Adjusting the Bias in the Estimates of First-order Autocorrelation Coefficient and Variance of Monthly Average Returns

We attempt to adjust the biased estimates of first-order autocorrelation and variance of monthly average returns by implementing a filtering procedure following Schwert (1990). The motivation for this remedial measure is that if the filtering procedure works well in generating a new return series with properties similar to those of end-of-month returns, then we should expected the predictability findings based on it to also be similar to those of end-of-month returns. Schwert proposes to adjust returns for these biases by estimating a first-order moving average (MA(1)) process,

$$r_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \tag{1}$$

where r_t are the log monthly average returns and θ is the moving average parameter. An analysis of the partial autocorrelation function (pacf) of monthly average returns in Figure A.1, and considering that its significant autocorrelation cuts off at lag 1, indicates that a moving average process of order 1 (MA(1)) would fit the data well. An estimate of returns, which has the same mean, $\hat{\mu}$, as monthly average returns and no or almost zero first-order autocorrelation, is then given by $\hat{r}_t = \hat{\mu} + \hat{\varepsilon}_t$. As noted by Schwert (1990), unfortunately the variance of $\hat{r}_t$ given by $\text{Var}(\hat{r}_t) = \text{Var}(\hat{\varepsilon}_t) = \text{Var}(r_t)/(1+\theta^2)$ will be less than the variance of monthly average returns. To adjust for this bias, we multiply the estimated residuals, $\hat{\varepsilon}_t$, by a factor $[1.11(1+\theta^2)^{1/2})]$ which should result in a standard deviation estimate that is about 11% larger than the standard deviation of monthly average returns. Here, 11% is the approximate percentage amount by which the standard deviation of monthly average returns should be increased to equal that of end-of-month returns over the full sample. Our new estimate of the time-series of returns, which we denote filtered returns, are then given by

$$\hat{r}_t = \hat{\mu} + \hat{\varepsilon}_t[1.11(1+\hat{\theta}^2)^{1/2}], \quad t = 1, \cdots, T, \tag{2}$$

where $\hat{\mu}$, $\hat{\theta}$ and $\hat{\varepsilon}_t$ are estimates of the parameters from (1).

The descriptive statistics for filtered returns are reported in Panel A of Table A.2. Filtered returns have almost the same mean as monthly average returns but a standard deviation that is about 11% higher than that of monthly average returns but comparable to that of end-of-month returns. The first-order autocorrelation is almost zero and confirmed by the autocorrelation function (acf) plot in Figure A.2. The Lagrange multiplier test of serial correlation fails to reject the null hypothesis that the first (first 12) autocorrelation coefficient(s) is (are jointly) equal to zero. The augmented Dickey-Fuller test for the null hypothesis of a unit root in filtered returns is also rejected in favour of the alternative that returns are stationary.

The estimates of covariances of filtered returns with predictors reported in Panel C of Table A.2 are very similar in magnitude to those for monthly average returns. Clearly, whereas the filtering process adjusts the downward and upward bias in the estimates of the first-order autocorrelation and variance, respectively, of monthly average returns, the bias in covariance estimates with predictors persist, more so for the economic variables than the technical indicators. This could be explained by the fact that, and as noted by Schwert (1990), the filtering procedure does not deal suitably with cross-correlation.

[Insert Table A.2 about here]

[Insert Figure A.1 about here]

[Insert Figure A.2 about here]

The in-sample predictability results for filtered returns based on the economic variables are reported in Panel A of Table A.3. For comparison, we also reported the results for end-of-month returns. The magnitude of estimated slope coefficients and associated heteroskedasticity-consistent standard errors are not even close to those for end-of-month returns but instead are largely similar to those for monthly average returns. As indicated by the t-test, even after adjusting monthly average returns for the bias in the estimates of the first-order autocorrelation coefficient and variance, inferences about the extent of predictability for filtered returns based on the predictor variables are very similar to the conclusions for monthly average returns. Ten economic variables that were found to be significant in-sample predictors for monthly average returns, namely Futures returns, PP, SCS, GSS, HSS, AUS, CAN, SA, CTBL, and BDI, are also significant in-sample predictors for filtered returns. The Durbin-Watson statistics rejects the null hypothesis of serial correlation in the estimated regression residuals.

The in-sample predictability results for filtered returns based on the 12 technical indicator variables are reported in Panel B of Table A.3. The results show that only 3 (out of the 10 variables that displayed significant in-sample predictability for monthly average returns) of the 12 variables display statistically significant predictive ability for returns. The variables are MA$(1, 12)$, MOM(3), and VOL(1 9). Again, these variables display much weaker evidence of predictability for filtered returns compared to monthly average returns but nowhere close to the predictability findings for end-of-month returns. These in-sample predictability tests indicate that adjusting returns for the bias in the estimates of first-order autocorrelation coefficient and variance of monthly average returns substantially weakens, and in some cases, eliminates the statistical evidence of predictability. This finding may very well indicate that, since the technical indicator variables are calculated using end-of-period data and not time-averaged data, the bias in estimates of covariance between filtered returns and the predictors are potentially less severe if not eliminated.

[Insert Table A.3 about here]

Table A.4 presents results for the out-of-sample performance of filtered return forecasts based on the economic variables. For comparison, we also reported the results for end-of-month returns. The results are similar to findings reported for both the individual and combination forecasts of monthly average returns. Purging the monthly average returns for the aforementioned statistical biases, however, reduces the magnitude of the R^2_{OS} values by more than 50% percent with consistently lower t-statistics compared to those reported for monthly average returns. For example, the mean combination forecast of monthly average records an R^2_{OS} (t-stat) value of about 9% (5%) compared to about 4% (3%) for filtered returns, although both values are statistically significant at the 1% level.

The results for performance of return forecasts based on the technical indicators reported in Table A.5 tells a different story. We can see that, of the 3 variables that were found to be significant for the in-sample tests of filtered return predictability, none are statistically significant in out-of-sample tests. Furthermore, all the other in-sample insignificant variables are also insignificant in the out-of-sample tests. The performance of the combination forecasts also display a similar pattern. Of the 6 combination forecasts of monthly average returns with R^2_{OS} values significant at the 1% level, none are significant for filtered returns and even this occurs at 10% level.

The conclusions that can be gleaned from the analysis of the in- and out-of-sample predictability results for filtered returns is that, adjusting monthly average returns for the bias in the estimates of first-order autocorrelation coefficient and variance leads to a weakening, and in some cases, a total elimination of the forecasting power of the predictors. Whereas the filtering procedure reduces the magnitude of the predictive power of the economic variables by more than 50% percent, predictability is largely absent for almost all the technical indicator variables and their combinations, re-echoing the findings in Schwert (1990) and Wilson, Jones and Lundstrum (2001). A potential reason for these findings is that although the filtering procedure implemented eliminates the bias in the estimates of statistical properties of monthly average returns, the resulting filtered returns when used in predictive regressions does not completely alleviate the severe econometric problems of inefficiency of estimated slope coefficients and bias in the estimates of associated standard errors.

References

Acharya, V.V., Lochstoer, L.A., Ramadorai, T., 2013. Limits to arbitrage and hedging: Evidence from commodity markets. Journal of Financial Economics 109, 441–465.

Ahmed, S., Liu, X., Valente, G., 2016. Can currency-based risk factors help forecast exchange rates? International Journal of Forecasting 32, 75–97.

Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? Journal of Applied Econometrics 25, 539–573.

Alquist, R., Kilian, L., Vigfusson, R.J., 2013. Forecasting the price of oil, in: Handbook of Economic Forecasting. Elsevier. volume 2, pp. 427–507.

Anatolyev, S., Gospodinov, N., Jamali, I., Liu, X., 2017. Foreign exchange predictability and the carry trade: A decomposition approach. Journal of Empirical Finance 42, 199–211.

Baumeister, C., Kilian, L., 2014. What central bankers need to know about forecasting oil prices. International Economic Review 55, 869–889.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: a forecast combination approach. Journal of Business & Economic Statistics 33, 338–351.

Baumeister, C., Kilian, L., Zhou, X., 2018. Are product spreads useful for forecasting oil prices? an empirical evaluation of the verleger hypothesis. Macroeconomic Dynamics 22, 562–580.

Bessembinder, H., Chan, K., 1992. Time-varying risk premia and forecastable returns in futures markets. Journal of Financial Economics 32, 169–193.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chen, Y.C., Rogoff, K.S., Rossi, B., 2010. Can exchange rates forecast commodity prices? The Quarterly Journal of Economics 125, 1145–1194.

Chinn, M.D., Coibion, O., 2014. The predictive content of commodity futures. Journal of Futures Markets 34, 607–636.

Choi, Y., Jacewitz, S., Park, J.Y., 2016. A reexamination of stock return predictability. Journal of Econometrics 192, 168–189.

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cochrane, J.H., Piazzesi, M., 2005. Bond risk premia. American Economic Review , 138–160.

Della Corte, P., Sarno, L., Tsiakas, I., 2008. An economic evaluation of empirical exchange rate models. The Review of Financial Studies 22, 3491–3530.

Ferreira, M.A., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics 100, 514–537.

Gargano, A., Timmermann, A., 2014. Forecasting commodity price indexes using macroeconomic and financial predictors. International Journal of Forecasting 30, 825–843.

Gorton, G.B., Hayashi, F., Rouwenhorst, K.G., 2013. The fundamentals of commodity futures returns. Review of Finance , 35–105.

Greenwood, R., Hanson, S.G., 2013. Issuer quality and corporate bond returns. The Review of Financial Studies 26, 1483–1525.

Hamilton, J.D., 2009. Understanding crude oil prices. The Energy Journal 30.

Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? Journal of Financial Economics 105, 473–490.

Lanne, M., 2002. Testing the predictability of stock returns. Review of Economics and Statistics 84, 407–415.

Levich, R.M., Potì, V., 2015. Predictability and 'good deals' in currency markets. International Journal of Forecasting 31, 454–472.

Li, J., Tsiakas, I., Wang, W., 2015. Predicting exchange rates out of sample: Can economic fundamentals beat the random walk? Journal of Financial Econometrics 13, 293–341.

Lin, H., Wang, J., Wu, C., 2014. Predictions of corporate bond excess returns. Journal of Financial Markets 21, 123–152.

Lin, H., Wu, C., Zhou, G., 2017. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science .

Ludvigson, S.C., Ng, S., 2009. Macro factors in bond risk premia. The Review of Financial Studies 22, 5027–5067.

Molodtsova, T., Papell, D.H., 2009. Out-of-sample exchange rate predictability with taylor rule fundamentals. Journal of international economics 77, 167–180.

Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Management Science 60, 1772–1791.

Rapach, D.E., Ringgenberg, M.C., Zhou, G., 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121, 46–65.

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, 821–862.

Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: what is the role of the united states? The Journal of Finance 68, 1633–1662.

Rossi, B., 2013. Exchange rate predictability. Journal of Economic Literature 51, 1063–1119.

Sarno, L., Schneider, P., Wagner, C., 2016. The economic value of predicting bond risk premia. Journal of Empirical Finance 37, 247–267.

Schwert, G.W., 1990. Indexes of us stock prices from 1802 to 1987. Journal of Business , 399–426.

Wang, Y., Liu, L., Wu, C., 2017. Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. Energy Economics 66, 337–348.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.

Wilson, J.W., Jones, C.P., Lundstrum, L.L., 2001. Stochastic properties of time-averaged financial data: Explanation and empirical demonstration using monthly stock prices. Financial Review 36, 175–190.

Ye, M., Zyren, J., Shore, J., 2005. A monthly crude oil spot price forecasting model using relative inventories. International Journal of Forecasting 21, 491–501.

Ye, M., Zyren, J., Shore, J., 2006. Forecasting short-run crude oil price using high-and low-inventory variables. Energy Policy 34, 2736–2743.

Yi, Y., Ma, F., Zhang, Y., Huang, D., 2018. Forecasting the prices of crude oil using the predictor, economic and combined constraints. Economic Modelling 75, 237–245.

Yin, L., Yang, Q., 2016. Predicting the oil prices: Do technical indicators help? Energy Economics 56, 338–350.

Zhang, Y., Ma, F., Shi, B., Huang, D., 2018. Forecasting the prices of crude oil: An iterated combination approach. Energy Economics 70, 472–483.

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.

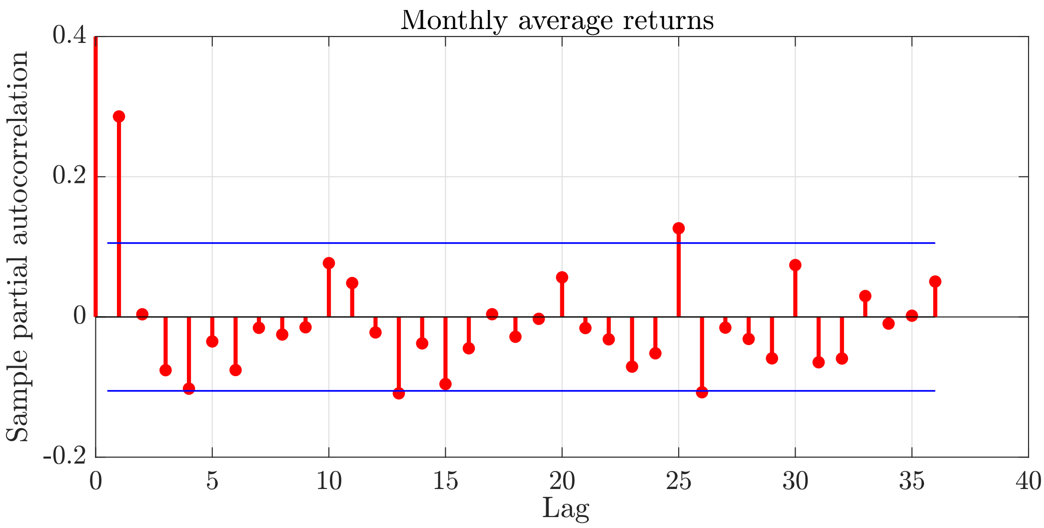
Zhong, X., Wang, J., 2018. Prospect theory and corporate bond returns: An empirical study. Journal of Empirical Finance 47, 25–48.

Table A.1: Literature Review: Price Series used in Studying Predictability

Article	Journal published	Price series	Evidence of Predictability
Panel A: Crude oil spot price/return predictability studies			
Zhang, Ma and Wang (2019)	Empirical Finance	Monthly averages	Yes
Yi, Ma, Zhang and Huang (2018)	Economic Modelling	Monthly averages	Yes
Baumeister, Kilian and Zhou (2018)	Macroeconomic Dynamics	Monthly averages	Yes
Zhang, Ma, Shi and Huang (2018)	Energy Economics	Monthly averages	Yes
Wang, Liu and Wu (2017)	Energy Economics	Monthly averages	Yes
Yin and Yang (2016)	Energy Economics	Monthly averages	Yes
Baumeister and Kilian (2015)	Journal of Business and Economic Statistics	Monthly averages	Yes
Chinn and Coibion (2014)	Futures Market	End-of-month	Yes
Baumeister and Kilian (2014)	International Economic Review	Monthly averages	Yes
Alquist, Kilian and Vigfusson (2013)	Handbook of Economic Forecasting	End-of-month	No
Alquist and Kilian (2010)	Journal of Applied Econometrics	End-of-month	No
Hamilton (2009)	The Energy Journal	Monthly averages	No
Ye, Zyren and Shore (2006)	Energy Policy	Monthly averages	Yes
Ye, Zyren and Shore (2005)	International Journal of Forecasting	Monthly averages	Yes
Panel B: Other commodity return predictability studies			
Gargano and Timmermann (2014)	International Journal of Forecasting	End-of-month	Yes
Chinn and Coibion (2014)	Journal of Futures Markets	End-of-month	Yes
Acharya, Lochstoer and Ramadorai (2013)	Journal of Financial Economics	End-of-month	Yes
Gorton, Hayashi and Rouwenhorst (2013)	Review of Finance	End-of-month	Yes
Hong and Yogo (2012)	Journal of Financial Economics	End-of-month	Yes
Chen, Rogoff and Rossi (2010)	The Quarterly Journal of Economics	End-of-month	Yes
Bessembinder and Chan (1992)	Journal of Financial Economics	End-of-month	Yes
Panel C: Equity risk premium predictability studies			
Choi, Jacewitz and Park (2016)	Journal of Econometrics	End-of-month	No
Rapach, Ringgenberg and Zhou (2016)	Journal of Financial Economics	End-of-month	Yes
Neely, Rapach, Tu and Zhou (2014)	Management Science	End-of-month	Yes
Rapach, Strauss and Zhou (2013)	Journal of Finance	End-of-month	Yes
Ferreira and Santa-Clara (2011)	Journal of Financial Economics	End-of-month	Yes
Rapach, Strauss and Zhou (2010)	Review of Financial Studies	End-of-month	Yes
Campbell and Thompson (2008)	The Review of Financial Studies	End-of-month	Yes
Welch and Goyal (2008)	Journal of Financial Economics	End-of-month	No
Lanne (2002)	The Review of and Economics Statistics	End-of-month	No
Panel D: Bond return predictability studies			
Zhong and Wang (2018)	Journal of Empirical Finance	End-of-month	Yes
Lin, Wu and Zhou (2017)	Management Science	End-of-month	Yes
Sarno, Schneider and Wagner (2016)	Journal of Empirical Finance	End-of-month	Yes
Lin, Wang and Wu (2014)	Journal of Financial Markets	End-of-month	Yes
Greenwood and Hanson (2013)	The Review of Financial Studies	End-of-month	Yes
Ludvigson and Ng (2009)	The Review of Financial Studies	End-of-month	Yes
Cochrane and Piazzesi (2005)	American Economic Review	End-of-month	Yes
Panel E: Currency return predictability studies			
Anatolyev, Gospodinov, Jamali and Liu (2017)	Journal of Empirical Finance	End-of-month	Yes
Ahmed, Liu and Valente (2016)	International Journal of Forecasting	End-of-month	No
Li, Tsiakas and Wang (2015)	Journal of Financial Econometrics	End-of-month	Yes
Levich and Potì (2015)	International Journal of Forecasting	End-of-month	Yes
Rossi (2013)	Journal of Economic Literature	End-of-month	Yes
Molodtsova and Papell (2009)	Journal of International Economics	End-of-month	No
Della Corte, Sarno and Tsiakas (2008)	The Review of Financial Studies	End-of-month	Yes

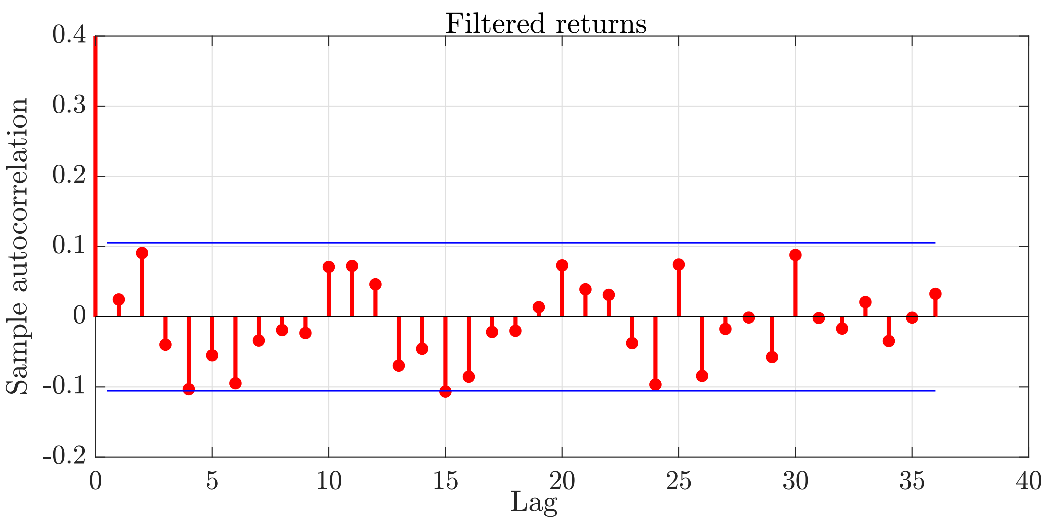
Notes. This table lists select articles on studies of return predictability across commodity, stock, bond and currency markets, the journals that published the articles, the prices series used in computing returns (monthly averages of daily prices or end-of-month prices) in the articles, and whether or not they found evidence of return predictability.

Figure A.1: Sample and Partial Autocorrelation Functions for Crude Oil Returns



Notes. This figure plots the the sample partial autocorrelation function of monthly average crude oil returns. The sample period is 1987:01-2016:12.

Figure A.2: Sample Autocorrelation Function of Filtered Crude Oil Returns



Notes. This figure plots the the sample autocorrelation function of filtered crude oil returns. Filtered returns are estimated following the filtering procedure in Schwert (1990). The sample period is 1987:01-2016:12.

Table A.2: Summary Statistics for Filtered Returns

Panel A: Summary statistics and autocorrelations

							Autocorrelations					
Returns	Mean	Std. Dev.	Skewness	Kurtosis	Min.	Max.	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	LM(1)	LM(12)	p(ADF)
Monthly average	0.107	8.282	−0.237	4.993	−32.366	38.348	0.284	0.088	−0.044	0.00	0.01	0.00
End-of-month	0.087	9.161	−0.101	4.505	−38.253	35.788	0.149	0.001	0.004	0.03	0.01	0.00
Filtered	0.120	9.141	−0.001	4.274	−31.748	40.417	0.025	0.091	−0.040	0.71	0.33	0.00

Panel B: Correlations between returns

	y_1	y_2	y_3
Monthly average (y_1)	1.00	0.72	0.97
End-of-month (y_2)		1.00	0.71
Filtered (y_3)			1.00

Panel C: Estimates of covariance of returns with predictors

Panel C1		Panel C2	
Economic variable (x)	Cov(y_3, x)	Technical indicator (z)	Cov(y_3, z)
Futures return	60.24	MA(1, 9)	165.89
Basis	3.23	MA(1, 12)	150.04
HP	14.91	MA(2, 9)	117.57
PP	14.96	MA(2, 12)	95.97
OI	8.79	MOM(3)	187.81
SCS	60.51	MOM(6)	140.10
GSS	60.35	MOM(9)	83.09
HSS	60.81	MOM(12)	114.34
GOI	−0.51	VOL(1, 9)	154.04
GOP	−0.89	VOL(1, 12)	147.47
AUS	6.46	VOL(2, 9)	104.62
CAN	4.58	VOL(2, 12)	95.41
NZ	6.13		
SA	5.86		
S&P 500	1.45		
TBL	0.68		
CTBL	0.15		
YS	−0.81		
DFY	−0.28		
TMS1Y	−0.06		
TMS2Y	−0.05		
TMS5Y	−0.13		
VIX	−6.67		
REA	17.10		
BDI	28.71		
INFL	−0.04		
CAPUTIL	0.99		
INDPRO	0.55		

Notes. This table reports the summary statistics for three measures of monthly crude oil returns (in percent). We report the mean, standard deviation, skewness, kurtosis, minimum, maximum, and sample autocorrelation coefficient coefficient up to 12 lags. The return series are monthly average returns, end-of-month returns, and filtered returns. LM(1) (LM(12)) is the p-values (calculated using heteroskedasticity-consistent standard errors) associated with the Lagrange multiplier test of the null hypothesis that the first (first 12) autocorrelation coefficient is (are jointly) equal to zero. p(ADF) is the p-value associated with the augmented Dickey-Fuller test of unit root. For all p-values 0.00 indicates that less 0.001. In Panel B, y_1, y_2, and y_3 denote monthly average, end-of-month, and filtered returns, respectively. The sample period is 1987:01-2016:12.

Table A.3: In-sample Predictability Results for Filtered Returns

Predictor	Filtered returns $\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW	End-of-month returns $\hat{\beta}$	se($\hat{\beta}$)	t-stat	R^2 (%)	DW
Panel A: Economic variables										
Lagged return	0.025	0.066	0.38	0.06	1.98	0.150	0.067	2.25**	2.24	1.98
Futures return	0.424	0.055	7.67***	18.49	2.52	0.150	0.066	2.25**	2.27	1.98
Basis	−0.122	0.296	−0.41	0.07	1.92	−0.558	0.248	−2.25**	1.44	1.64***
HP	0.001	0.048	0.02	0.00	1.93	−0.040	0.049	−0.81	0.17	1.68***
PP	0.280	0.098	2.87***	2.09	2.02	0.151	0.101	1.49	0.60	1.75***
OI	0.006	0.071	0.08	0.00	1.93	−0.003	0.075	−0.04	0.00	1.69***
SCS	0.415	0.055	7.48**	17.78	2.51	0.145	0.066	2.19**	2.15	1.98
GSS	0.416	0.056	7.48***	17.84	2.51	0.145	0.066	2.18**	2.14	1.97
HSS	0.413	0.055	7.48***	17.66	2.51	0.146	0.066	2.21**	2.18	1.98
GOI	−0.038	0.397	−0.10	0.00	1.93	0.521	0.453	1.15	0.41	1.68***
GOP	−0.398	0.447	−0.89	0.19	1.94	−1.147	0.670	−1.71*	1.53	1.72***
AUS	0.470	0.157	2.99***	2.97	2.03	0.164	0.180	0.91	0.36	1.75***
CAN	0.739	0.241	3.06***	3.24	2.03	0.167	0.241	0.69	0.16	1.73***
NZ	0.202	0.156	1.29	0.58	1.97	0.031	0.159	0.19	0.01	1.70***
SA	0.248	0.133	1.86*	1.20	1.96	0.103	0.141	0.73	0.20	1.71***
S&P 500	0.092	0.143	0.64	0.19	1.93	0.072	0.150	0.48	0.12	1.70***
TBL	0.067	0.206	0.33	0.03	1.93	0.068	0.206	0.33	0.04	1.69***
CTBL	7.311	3.077	2.38*	2.26	1.95	8.116	3.268	2.48**	2.75	1.71***
YS	−0.335	0.398	−0.84	0.31	1.93	−0.303	0.353	−0.86	0.25	1.69***
DFY	−0.318	2.162	−0.15	0.02	1.93	−0.123	1.782	−0.07	0.00	1.69***
TMS1Y	−0.159	1.910	−0.08	0.00	1.93	−0.053	1.915	−0.03	0.00	1.69***
TMS2Y	0.110	1.060	0.10	0.00	1.92	0.103	1.088	0.09	0.00	1.69***
TMS5Y	−0.937	1.122	−0.84	0.19	1.93	−0.882	1.079	−0.82	0.17	1.69***
VIX	−0.108	0.088	−1.22	0.84	1.94	−0.093	0.083	−1.11	0.61	1.70***
REA	0.007	0.022	0.34	0.05	1.93	0.008	0.020	0.41	0.06	1.69***
BDI	0.071	0.034	2.10**	2.10	1.99	0.045	0.035	1.30	0.84	1.74***
INFL	0.743	2.310	0.32	0.05	1.92	0.052	1.963	0.03	0.00	1.69***
CAPUTL	0.483	0.857	0.56	0.16	1.94	0.127	0.810	0.16	0.01	1.69***
INDPRO	−0.053	0.136	−0.39	0.02	1.92	0.038	0.167	0.23	0.01	1.69***
Panel B: Technical indicator variables										
MA(1, 9)	0.0126	0.0098	1.28	0.47	1.96	−0.0025	0.0098	−0.26	0.02	1.68***
MA(1, 12)	0.0164	0.0098	1.67*	0.80	1.97	0.0074	0.0098	0.75	0.16	1.71***
MA(2, 9)	0.0068	0.0100	0.68	0.14	1.94	−0.0011	0.0098	−0.11	0.00	1.69***
MA(2, 12)	0.0019	0.0101	0.19	0.01	1.93	−0.0045	0.0099	−0.46	0.06	1.68***
MOM(3)	0.0231	0.0097	2.38**	1.59	2.01	0.0106	0.0096	1.09	0.33	1.73***
MOM(6)	0.0060	0.0100	0.61	0.11	1.94	−0.0050	0.0099	−0.50	0.07	1.68***
MOM(9)	0.0100	0.0099	1.01	0.29	1.93	−0.0060	0.0098	−0.62	0.11	1.68***
MOM(12)	0.0134	0.0099	1.36	0.53	1.95	0.0061	0.0098	0.62	0.11	1.70***
VOL(1, 9)	0.0194	0.0102	1.92*	1.10	1.97	−0.0012	0.0101	−0.12	0.00	1.69***
VOL(1, 12)	0.0155	0.0105	1.48	0.69	1.97	0.0039	0.0103	0.38	0.04	1.70***
VOL(2, 9)	0.0016	0.0101	0.16	0.01	1.93	−0.0015	0.0101	−0.14	0.01	1.69***
VOL(2, 12)	0.0121	0.0102	1.19	0.43	1.95	0.0048	0.0102	0.47	0.07	1.70***

Notes. This table reports the in-sample OLS estimation results for filtered returns (monthly average returns adjusted for bias in the estimates of variance and first-order autocorrelation following the filtering procedure in Schwert (1990)). For comparison, the results for end-of-month returns in Table 3 of the paper is also provided. The table reports the slope coefficient, $\hat{\beta}$, and the associated heteroskedasticity-consistent standard errors, se($\hat{\beta}$), the statistic for the two-tailed alternative test, t-stat, for the significance of $\hat{\beta}$. R^2 is the coefficient of determination, and DW is the Durbin-Watson statistic for testing the presence of serial correlation in the estimated predictive regression residuals. Results are reported for the full sample period 1987:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A.4: Out-of-Sample Predictability Results for Filtered Returns: Economic Variables

Predictor	Filtered returns			End-of-month returns		
	MSFE	R^2_{OS} (%)	MSFE-adjusted	MSFE	R^2_{OS} (%)	MSFE-adjusted
RWWD	96.54			91.51		
Panel A: Individual predictive model forecasts						
Futures return	84.63	12.34	4.28***	91.27	0.26	1.07
Basis	97.09	−0.56	−2.87	90.57	1.02	1.57*
HP	98.05	−1.56	−0.34	91.83	−0.35	−0.35
PP	98.06	−1.57	0.33	91.62	−0.13	0.45
OI	96.87	−0.34	−1.42	91.89	−0.41	−1.65
SCS	85.41	11.53	4.16***	91.50	0.01	0.94
GSS	85.37	11.57	4.16***	91.50	0.00	0.94
HSS	85.49	11.44	4.15***	91.50	0.01	0.95
GOI	97.10	−0.58	−2.16	92.14	−0.69	−0.51
GOP	96.75	−0.21	−0.18	92.56	−1.15	0.17
AUS	94.59	2.02	1.95**	93.06	−1.69	−0.65
CAN	93.70	2.95	2.47***	92.49	−1.07	−0.87
NZ	97.28	−0.76	0.03	93.87	−2.58	−1.56
SA	95.98	0.58	1.34*	92.32	−0.89	−0.53
S&P 500	97.98	−1.49	−0.54	92.30	−0.86	−0.22
TBL	97.54	−1.04	−1.65	92.44	−1.02	−1.51
CTBL	95.51	1.07	1.33*	89.69	1.98	1.88**
YS	98.00	−1.51	−0.64	92.93	−1.55	−0.82
DFY	99.87	−3.45	−0.37	93.85	−2.57	−0.31
TMS1Y	97.29	−0.77	−1.04	92.17	−0.73	−1.02
TMS2Y	97.08	−0.56	−1.59	92.06	−0.61	−1.36
TMS5Y	97.88	−1.39	−0.50	92.85	−1.47	−0.60
VIX	96.92	−0.39	0.37	92.02	−0.57	0.37
REA	98.06	−1.57	−1.01	92.90	−1.52	−0.89
BDI	95.95	0.61	1.36*	92.91	−1.53	0.10
INFL	97.75	−1.25	−0.93	92.59	−1.19	−1.10
CAPUTIL	97.97	−1.48	−0.03	92.42	−0.99	−1.03
INDPRO	96.80	−0.27	−1.03	92.07	−0.61	−1.46
Panel B: Combination forecasts						
Mean	93.13	3.53	3.17***	91.47	0.04	0.27
Median	96.30	0.25	0.95	91.56	−0.06	−0.18
Trimmed mean	93.34	3.31	3.20***	91.44	0.08	0.35
Weighted mean	92.69	3.99	3.29***	91.47	0.04	0.28
DMSFE ($\theta = 0.9$)	93.58	3.07	2.88***	91.49	0.01	0.23
PC ($\textsc{ic} = R^2$)	89.79	7.00	3.31***	92.46	−1.04	0.71

Notes. This table reports out-of-sample results for the individual and combination forecasts of crude oil returns based on 28 economic variables. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. Results are reported for filtered returns and end-of-month returns (for comparison). The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table A.5: Out-of-Sample Predictability Results for Filtered Returns: Technical Indicators

	Filtered returns			End-of-month returns		
Predictor	MSFE	R^2_{OS} (%)	MSFE-adjusted	MSFE	R^2_{OS} (%)	MSFE-adjusted
RWWD	96.54			91.51		
Panel A: Individual predictive model forecasts						
MA$(1, 9)$	96.55	−0.01	0.31	91.93	−0.47	−0.98
MA$(1, 12)$	96.41	0.14	0.66	91.78	−0.30	−0.24
MA$(2, 9)$	97.00	−0.47	−0.64	92.33	−0.90	−0.73
MA$(2, 12)$	97.54	−1.03	−1.69	92.48	−1.07	−1.15
MOM(3)	96.11	0.45	1.12	92.16	−0.72	−0.50
MOM(6)	97.17	−0.65	−0.71	92.37	−0.94	−0.54
MOM(9)	96.77	−0.23	−0.25	92.19	−0.74	−0.84
MOM(12)	96.67	−0.13	0.30	91.83	−0.35	−0.71
VOL$(1, 9)$	96.07	0.49	1.00	92.03	−0.58	−0.80
VOL$(1, 12)$	96.46	0.09	0.49	91.95	−0.48	−1.92
VOL$(2, 9)$	97.76	−1.27	−1.56	92.72	−1.33	−1.33
VOL$(2, 12)$	96.66	−0.12	0.04	92.18	−0.74	−1.58
Panel B: Combination forecasts						
Mean	96.51	0.03	0.25	92.00	−0.54	−1.85
Median	96.57	−0.03	0.13	92.13	−0.69	−1.63
Trimmed mean	96.54	0.01	0.21	92.02	−0.56	−1.71
Weighted mean	96.51	0.03	0.25	92.01	−0.55	−1.85
DMSFE ($\theta = 0.9$)	96.64	−0.10	0.00	92.13	−0.68	−2.47
PC ($\textsc{ic} = R^2$)	98.17	−1.68	−0.57	93.06	−1.70	−1.45

Notes. This table reports out-of-sample results for the individual and combination forecasts of crude oil returns based on 12 technical indicator variables. RWWD is the random walk with drift benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the RWWD forecast. Statistical significance for the R^2_{OS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the RWWD forecast MSFE is less than or equal to the MSFE of the competing forecast against the one-sided (upper tailed) alternative hypothesis that the RWWD forecast MSFE is greater than the MSFE of the competing forecast. Results are reported for filtered returns and end-of-month returns (for comparison). The out-of-sample forecast evaluation period is 1997:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.