LCERPA

Laurier Centre for Economic Research & Policy Analysis

LCERPA Working paper No. 2025-7

June 2025

A Reappraisal of Real-time Forecasts of the Real Price of Oil

Eric Benyo
Wilfrid Laurier University

Reinhard Ellwanger
Bank of Canada

Stephen Snudden
Wilfrid Laurier University

A Reappraisal of Real-time Forecasts of the Real Price of Oil*

Eric Benyo† Reinhard Ellwanger‡ Stephen Snudden§

June 19, 2025

Abstract

We replicate Baumeister and Kilian (2012) to reappraise real-time forecasts of the real price of crude oil against the end-of-month no-change forecast, the equivalent naive benchmark used for asset prices. We find no consistently significant improvements in the predictive accuracy of model-based forecasts over this naive benchmark at short horizons. Only futures-based forecasts consistently outperform the end-of-month no-change forecast, and only at longer horizons. These results challenge the consensus on the predictability of the real price of crude oil and the merits of alternative forecast approaches. Our findings motivate broader reassessment and replication of forecasting models of temporally aggregated series.

JEL classification: C1, C53, Q47
Keywords: Forecasting and Prediction Methods, Oil Price Forecast

*The views expressed in this paper are those of the authors; no responsibility for them should be attributed to the Bank of Canada.

†Department of Economics, Wilfrid Laurier University

‡Bank of Canada, International Economic Analysis Department

§Corresponding author; Department of Economics, Wilfrid Laurier University, 64 University Ave W, Waterloo, Ontario, Canada, N2L 3C7; ORCID iD 0000-0002-4033-3235; Email: ssnudden@wlu.ca

1 Introduction

Baumeister and Kilian (2012) provided the first evidence that model-based forecasts could accurately predict the real price of crude oil in real time. Their study demonstrated that various forecasting models, including those based on economic fundamentals, outperformed the no-change forecast at short horizons, with mean-squared prediction error (MSPE) improvements of up to 25 percent. This result stood "in striking contrast to related results in the literature for asset prices,"[1] and spurred a wide range of applications that extended these findings to other models and settings (see, e.g., the survey by Filippidis et al., 2024).

In Baumeister and Kilian (2012), the real price of crude oil is constructed as a monthly average of daily prices. This approach has a drawback: such average prices can exhibit spurious predictability when compared to monthly average no-change forecasts (Weiss, 1984; Marcellino, 1999), even if the underlying daily data follows a random walk (Working, 1960). In this setting, only the end-of-month no-change forecast corresponds to the traditional random walk forecast widely used in economics and finance, and it is also substantially more accurate than the monthly average no-change forecast (Ellwanger and Snudden, 2023). Our study is the first to empirically examine the broader implications of using a corrected no-change benchmark for forecasts of averaged macroeconomic series.

We systematically replicate all real-time forecasts in Baumeister and Kilian (2012) to reevaluate them against the end-of-month no-change forecast. We use an independently updated real-time dataset that it is novel for including all revisions in the historical vintages. This approach allows us to reassess Baumeister and Kilian (2012)'s findings in the original sample and a more recent period. The goal is to revisit the conclusions regarding the predictability of the real price of crude oil and the merits of the alternative forecast approaches under a corrected methodology that appropriately represents the classical random walk null hypothesis.

Our main finding is that the conclusions of Baumeister and Kilian (2012) no longer hold when their forecasts are compared to the end-of-month no-change forecasts. In the original sample, none of their model-based forecasts demonstrates real-time improvements over the traditional random-walk forecast that are significant and consistent across the two forecast criteria considered (mean squared prediction error and directional accuracy). Moreover, we find that employing the appropriate no-change forecast can change the relative performance of different forecast approaches. Only

[1] See Baumeister and Kilian (2012), page 326.

the futures-based forecast shows consistent and significant improvements in both mean-squared precision and directional accuracy, and only at the one-year horizon. Similar results are obtained for the extended sample period.

While Baumeister and Kilian (2012) has been refined in the subsequent literature, replicating and correcting this seminal work is essential, as its findings have shaped the foundation of oil price forecasting literature and remain highly influential. Many recent studies build on its conclusions, meaning any revisions to the original analysis have far-reaching implications. Moreover, while prior studies such as Baumeister and Kilian (2015) and Funk (2018) have replicated parts of Baumeister and Kilian (2012)'s real-time forecasting approaches, they all used an inappropriate period average no-change benchmark. By addressing the methodological correction in this foundational work, we also highlight the importance of reassessing the validity of more recent conclusions that rely on similar methods.

A key contribution of our paper is to show that proper replication is essential for fully assessing the validity of existing conclusions in studies that used an incorrect benchmark. Unlike Ellwanger and Snudden (2023), who reported only the relative MSPE performance of no-change forecasts, replications provide a deeper evaluation by enabling a reassessment of directional accuracy and retesting of the statistical significance of forecast improvements. Because success ratios for directional accuracy are non-transitive with respect to the no-change benchmark, the relative ranking of forecasts may change under the corrected methodology. Moreover, statistical tests typically rely on the full series of loss function differentials, which are not available from MSPE ratios alone. These issues can only be addressed by fully replicating all forecasts and using the appropriate benchmark when constructing success ratios and statistical tests.

The issue of spurious predictability has historical precedent in the return forecasting literature. Cowles and Jones (1937) initially rejected the efficient market hypothesis using returns constructed from averaged data. In response to Working (1960)'s finding that the predictability in these returns could be spurious, Cowles (1960) replicated their results using non-averaged returns and conceded that the original conclusions were invalid. This replication was crucial for establishing the efficient market hypothesis as the workhorse model of finance, and the adoption of point-sampling in return forecasting. Similarly, our replication exposes the issue of spurious predictability in macroeconomic forecasts and motivates further replications in oil price forecasting and other areas that rely on similar methods.

2 Method

2.1 Real-time Forecasts

Following Baumeister and Kilian (2012), we forecast the monthly average real price of West Texas Intermediate (WTI) crude oil and the monthly Refiners' Acquisition Cost (RAC) of crude oil. Real-time forecasts are computed for 1- to 12-month horizons using information available at the end of each month.[2] All forecasts follow the methods and specifications outlined in Baumeister and Kilian (2012). Consistent with their approach, forecasting models are estimated using an expanding window with data starting from 1973M1. Forecasts are evaluated out of sample, beginning from 1992M1.

Our replications comprise all forecasting models detailed in Baumeister and Kilian (2012). This includes autoregressive moving average (ARMA) models, classical and Bayesian autoregressive (BAR) models, and (Bayesian) vector autoregressive ((B)VAR) models. The VAR models are estimated with 4 variables: the real price of crude oil, the growth rate of global oil production, a proxy for changes in global crude oil inventories, and the Real Economic Activity (REA) index as a measure of global economic activity. Our analysis also includes forecasts based on the growth rate of the Commodity Research Bureau's (CRB) index of industrial raw material prices and futures prices. All lag specifications and estimation procedures, including Bayesian priors, follow those of Baumeister and Kilian (2012).

As in Baumeister and Kilian (2012), we report two forecast evaluation criteria relative to the monthly average no-change forecast: the MSPE ratio and the success ratio for directional accuracy. The success ratio measures how often the forecasts correctly predict the direction of change in the monthly real price of crude oil. A key innovation of this study is to also compute these ratios relative to the end-of-month no-change forecast, the last observed daily price contained in the forecaster's information set. Inference about the forecast accuracy relative to the alternative no-change forecast uses the test of Diebold and Mariano (1995) for the MSPE ratios and Pesaran and Timmermann (2009) for the directional accuracy.[3]

[2] Online appendix A2.1 examines forecasts for the 24-month horizon.

[3] With real-time data and iterative out-of-sample forecasts, the assumptions of the Diebold-Mariano test are not fulfilled in our setup (Diebold, 2015; Kilian, 2015). For this reason, no tests for the MSPE ratios are presented in Baumeister and Kilian (2012). We report the p-values with this caveat in mind to facilitate comparisons with other studies reporting the test. The conclusions of this paper do not rely on the validity of the test.

2.2 Real-time Data

Baumeister and Kilian (2012) introduced a novel real-time dataset of oil-market and other economic variables. The original dataset spans 1991M1 to 2010M12. For this study, we extend the original dataset to 2021M5, covering the initial phase of the COVID-19 period.[4] Unlike previous efforts to extend the vintages (e.g., Garratt et al., 2019), our dataset is complete in the sense that it includes all monthly data vintages and spans all revisions across the entire history of each vintage.

Daily and monthly nominal WTI crude oil prices are available in real time and obtained from the U.S. Energy Information Administration (EIA). The EIA's monthly average nominal price is a simple mean of daily closing prices for the month. The end-of-month price is the last observed daily price, corresponding to the closing price of the last trading day in the month. The RAC nominal prices are the monthly average refiners import acquisition cost of crude oil and sourced in real-time from vintages of the U.S. Energy Information Administration's Monthly Energy Review (MER). This data series typically has a 2-month publication delay and is subject to sizable revisions.

Consistent with Baumeister and Kilian (2012), the monthly average WTI price is backcasted using the growth rate of the monthly average RAC until 1974M1, and uses the imputation of Mork (1989) for 1973. Nowcasts of RAC follow Baumeister and Kilian (2012) and apply the growth rate in monthly average WTI to fill in missing recent observations. End-of-month observations for the RAC follow Ellwanger and Snudden (2023) and are imputed by applying the ratio of the end-of-month WTI oil price relative to the monthly average WTI oil price to the real-time monthly RAC.

The real-time vintages of CPI are obtained from the real-time database maintained by the Philadelphia Federal Reserve. Missing real-time observations for the consumer price index are nowcasted employing the average historical growth rate. Both monthly average and end-of-month nominal prices are converted to real prices using real-time vintages of the monthly seasonally adjusted U.S. consumer price index (CPI).[5]

Finally, we employ a real-time version of Kilian (2009)'s real economic activity index, calculated with the corrected formula (Kilian, 2019). The underlying nominal shipping rates are based on the rates collected by Kilian (2009) for 1968M1 to 1986M1 and on the Baltic Dry Index obtained from Bloomberg thereafter.

[4] Additional details are provided in an online data appendix A1.

[5] For robustness, we tested the sensitivity of the forecasts to deflating the end-of-month nominal oil prices with an interpolated daily CPI (see online appendix A2.2). We found no notable difference in forecast performance.

3 Replication Results

3.1 Original Sample

Table 1 and 2 report the real-time forecast replications for the real prices of WTI and RAC crude oil, covering the original sample period up to 2010M6. The first set of columns reports the forecast criteria and corresponding p-values relative to the monthly average no-change forecast, consistent with Baumeister and Kilian (2012). The second set of columns reports the forecast criteria and corresponding p-values relative to the end-of-month no-change forecast. A MSPE ratio below one and a success ratio above 0.5 indicate accuracy improvements relative to the respective no-change forecasts.

The first set of columns in Table 1 shows that we can replicate Baumeister and Kilian (2012)'s results for the WTI qualitatively. At the one-month horizon, most models outperform the monthly average-price no-change forecasts in terms of the MSPE ratio, with values as low as 0.78 for the CRB-based forecast. The gains in forecast accuracy are more consistent at the 3-month-ahead prediction, where almost all models outperform the monthly average no-change forecast across both criteria. Notably, models based on economic fundamentals such as the CRB index and the BVAR perform best in terms of the MSPE and success ratio.

In addition to the model-based forecasts, the first row compares the performance of the end-of-month no-change forecast to the monthly average no-change forecast. The results show that the end-of-month no-change forecast outperforms the conventional monthly average no-change forecast for all horizons, with large accuracy improvements at short horizons. This confirms that the end-of-month no-change forecast is a more stringent benchmark than the conventional monthly average no-change forecast for the real price of crude oil (Ellwanger and Snudden, 2023).

Table 1. Forecasts of WTI Crude Oil, 1992M1–2010M6

	Versus Monthly Average No-Change					Versus End-of-Month No-Change				
Horizon	1	3	6	9	12	1	3	6	9	12
Model					MSPE Ratio					
EoM NC	**0.64 (0.005)**	**0.88 (0.043)**	**0.94 (0.032)**	**0.95 (0.041)**	**0.96 (0.060)**					
ARMA(1,1)	**0.91 (0.248)**	**0.92 (0.258)**	**0.94 (0.320)**	**0.94 (0.329)**	**0.92 (0.304)**	1.42 (1.000)	1.05 (0.740)	1.01 (0.519)	**0.99 (0.469)**	**0.96 (0.418)**
AR(12)	**0.92 (0.266)**	**0.95 (0.321)**	1.01 (0.528)	1.02 (0.602)	1.00 (0.498)	1.43 (1.000)	1.08 (0.829)	1.08 (0.768)	1.08 (0.758)	1.04 (0.633)
BAR(12)	**0.91 (0.238)**	**0.94 (0.293)**	**0.99 (0.475)**	1.01 (0.530)	**0.98 (0.442)**	1.42 (1.000)	1.07 (0.802)	1.06 (0.727)	1.06 (0.703)	1.03 (0.584)
VAR(12)	**0.85 (0.225)**	**0.85 (0.247)**	1.02 (0.604)	1.09 (0.813)	1.08 (0.821)	1.33 (0.974)	**0.97 (0.440)**	1.09 (0.841)	1.14 (0.915)	1.13 (0.911)
BVAR(12)	**0.87 (0.198)**	**0.91 (0.209)**	1.06 (0.762)	1.13 (0.789)	1.15 (0.854)	1.36 (0.994)	1.04 (0.669)	1.13 (0.914)	1.19 (0.876)	1.20 (0.921)
AR(24)	**0.95 (0.345)**	**0.97 (0.372)**	**0.98 (0.391)**	**0.97 (0.344)**	**0.94 (0.293)**	1.49 (1.000)	1.10 (0.911)	1.04 (0.680)	1.02 (0.563)	**0.98 (0.444)**
BAR(24)	**0.90 (0.220)**	**0.93 (0.258)**	**0.95 (0.315)**	**0.95 (0.303)**	**0.93 (0.268)**	1.41 (1.000)	1.06 (0.760)	1.02 (0.566)	1.00 (0.499)	**0.97 (0.410)**
VAR(24)	1.13 (0.764)	1.29 (0.830)	1.33 (0.951)	1.35 (0.891)	1.57 (0.926)	1.77 (1.000)	1.47 (0.931)	1.43 (0.975)	1.42 (0.924)	1.64 (0.943)
BVAR(24)	**0.87 (0.191)**	**0.94 (0.297)**	1.09 (0.722)	1.22 (0.774)	1.35 (0.846)	1.37 (0.995)	1.08 (0.803)	1.17 (0.834)	1.29 (0.828)	1.41 (0.881)
Futures	1.01 (0.664)	**0.99 (0.439)**	1.00 (0.490)	**0.98 (0.369)**	**0.92 (0.181)**	1.57 (0.997)	1.14 (0.965)	1.07 (0.908)	1.03 (0.634)	**0.96 (0.337)**
CRB Index	**0.78 (0.116)**	**0.80 (0.088)**	1.17 (0.906)	1.20 (0.882)	1.20 (0.832)	1.22 (0.957)	**0.92 (0.204)**	1.25 (0.955)	1.27 (0.934)	1.25 (0.890)
					Success Ratio					
EoM NC	**0.71 (0.000)**	**0.62 (0.001)**	**0.58 (0.028)**	**0.57 (0.065)**	**0.58 (0.031)**					
ARMA(1,1)	**0.52 (0.321)**	**0.53 (0.151)**	0.47 (0.789)	0.47 (0.729)	**0.51 (0.480)**	0.46 (0.864)	**0.51 (0.336)**	0.45 (0.858)	0.47 (0.752)	0.48 (0.707)
AR(12)	0.50 (0.540)	**0.53 (0.271)**	0.46 (0.833)	0.43 (0.950)	0.47 (0.809)	0.47 (0.854)	**0.52 (0.316)**	0.48 (0.749)	0.43 (0.948)	0.49 (0.723)
BAR(12)	0.48 (0.716)	**0.53 (0.215)**	0.45 (0.871)	0.44 (0.941)	0.49 (0.702)	0.45 (0.938)	**0.51 (0.353)**	0.48 (0.771)	0.43 (0.956)	0.47 (0.804)
VAR(12)	0.50 (0.486)	**0.58 (0.049)**	**0.53 (0.357)**	**0.56 (0.199)**	**0.59 (0.050)**	0.48 (0.752)	**0.53 (0.192)**	**0.52 (0.389)**	**0.53 (0.383)**	**0.57 (0.090)**
BVAR(12)	0.49 (0.686)	**0.58 (0.079)**	**0.54 (0.438)**	**0.54 (0.506)**	**0.56 (0.241)**	0.46 (0.896)	**0.54 (0.236)**	**0.52 (0.564)**	0.50 (0.728)	**0.51 (0.683)**
AR(24)	0.49 (0.606)	**0.51 (0.412)**	0.50 (0.450)	0.45 (0.869)	0.48 (0.713)	0.47 (0.839)	**0.52 (0.306)**	0.50 (0.507)	0.47 (0.691)	0.42 (0.940)
BAR(24)	0.49 (0.643)	**0.53 (0.271)**	0.50 (0.498)	0.45 (0.873)	0.47 (0.744)	0.48 (0.683)	**0.52 (0.321)**	0.50 (0.491)	0.45 (0.847)	0.44 (0.895)
VAR(24)	**0.52 (0.346)**	**0.57 (0.048)**	0.48 (0.752)	0.44 (0.942)	0.46 (0.848)	0.48 (0.765)	**0.55 (0.082)**	0.50 (0.629)	0.44 (0.930)	0.47 (0.826)
BVAR(24)	**0.52 (0.442)**	**0.58 (0.083)**	**0.58 (0.130)**	**0.51 (0.614)**	**0.54 (0.346)**	0.49 (0.557)	**0.56 (0.130)**	**0.55 (0.319)**	**0.52 (0.508)**	**0.53 (0.443)**
Futures	0.47 (0.492)	0.49 (0.436)	0.50 (0.304)	**0.54 (0.088)**	***0.55 (0.069)***	0.47 (0.762)	0.48 (0.503)	0.43 (0.814)	**0.54 (0.064)**	***0.56 (0.005)***
CRB Index	**0.54 (0.175)**	**0.60 (0.012)**	**0.62 (0.022)**	**0.54 (0.317)**	**0.55 (0.252)**	**0.51 (0.364)**	**0.53 (0.253)**	**0.56 (0.148)**	**0.53 (0.380)**	**0.52 (0.456)**

Note: Replication of Baumeister and Kilian (2012)'s real-time, out-of-sample-forecasts of the monthly average real price of WTI crude oil, 1992M1–2010M6. Brackets report the p-values for the serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and of Diebold and Mariano (1995) for equal MSFEs relative to the corresponding no-change forecast. Boldface values indicate gains relative to the respective no-change forecast. Statistically significant improvements over the end-of-month no-change forecast at the five percent significance level are italicized.

Table 2. Forecasts of RAC Crude Oil, 1992M1–2010M6

	Versus Monthly Average No-Change					Versus End-of-Month No-Change				
Horizon	1	3	6	9	12	1	3	6	9	12
Model					MSPE Ratio					
EoM NC	**0.69 (0.011)**	**0.88 (0.037)**	**0.93 (0.025)**	**0.95 (0.032)**	**0.95 (0.038)**					
ARMA(1,1)	**0.89 (0.241)**	**0.91 (0.227)**	**0.95 (0.319)**	**0.95 (0.357)**	**0.94 (0.327)**	1.30 (0.999)	1.04 (0.703)	1.01 (0.546)	1.01 (0.523)	**0.99 (0.472)**
AR(12)	**0.85 (0.197)**	**0.91 (0.264)**	1.00 (0.498)	1.04 (0.659)	1.02 (0.590)	1.23 (0.986)	1.04 (0.649)	1.07 (0.730)	1.10 (0.798)	1.07 (0.722)
BAR(12)	**0.85 (0.186)**	**0.90 (0.249)**	**0.99 (0.471)**	1.03 (0.616)	1.01 (0.548)	1.23 (0.987)	1.03 (0.623)	1.06 (0.709)	1.09 (0.770)	1.06 (0.689)
VAR(12)	**0.74 (0.142)**	**0.82 (0.238)**	1.04 (0.627)	1.10 (0.893)	1.08 (0.815)	1.08 (0.660)	**0.93 (0.385)**	1.12 (0.795)	1.17 (0.946)	1.13 (0.888)
BVAR(12)	**0.78 (0.146)**	**0.87 (0.200)**	1.07 (0.822)	1.14 (0.881)	1.15 (0.935)	1.14 (0.831)	**0.99 (0.467)**	1.14 (0.961)	1.21 (0.950)	1.21 (0.978)
AR(24)	**0.90 (0.282)**	**0.98 (0.448)**	1.01 (0.558)	1.00 (0.480)	**0.97 (0.373)**	1.31 (0.994)	1.12 (0.893)	1.08 (0.822)	1.05 (0.707)	1.02 (0.552)
BAR(24)	**0.87 (0.223)**	**0.93 (0.311)**	**0.97 (0.395)**	**0.97 (0.362)**	**0.95 (0.310)**	1.26 (0.989)	1.06 (0.724)	1.04 (0.650)	1.02 (0.578)	**0.99 (0.478)**
VAR(24)	**0.88 (0.299)**	1.20 (0.758)	1.39 (0.957)	1.45 (0.894)	1.69 (0.928)	1.27 (0.919)	1.37 (0.895)	1.49 (0.976)	1.53 (0.922)	1.78 (0.943)
BVAR(24)	**0.80 (0.162)**	**0.94 (0.326)**	1.13 (0.790)	1.27 (0.796)	1.40 (0.862)	1.16 (0.858)	1.07 (0.732)	1.20 (0.878)	1.34 (0.845)	1.47 (0.894)
Futures	1.00 (0.609)	**0.98 (0.244)**	**0.98 (0.366)**	**0.96 (0.293)**	**0.90 (0.111)**	1.46 (0.994)	1.11 (0.954)	1.05 (0.853)	1.01 (0.567)	**0.94 (0.256)**
CRB Index	**0.79 (0.142)**	**0.81 (0.100)**	1.17 (0.917)	1.19 (0.874)	1.18 (0.819)	1.15 (0.893)	**0.93 (0.242)**	1.25 (0.960)	1.26 (0.930)	1.24 (0.886)
					Success Ratio					
EoM NC	**0.70 (0.000)**	**0.63 (0.000)**	**0.60 (0.008)**	**0.55 (0.142)**	**0.58 (0.049)**					
ARMA(1,1)	**0.56 (0.022)**	**0.55 (0.042)**	0.50 (0.526)	0.49 (0.645)	**0.52 (0.394)**	**0.54 (0.112)**	**0.54 (0.081)**	0.49 (0.614)	0.46 (0.798)	**0.55 (0.215)**
AR(12)	**0.57 (0.018)**	**0.54 (0.152)**	0.49 (0.659)	0.43 (0.945)	0.49 (0.699)	**0.52 (0.265)**	**0.55 (0.084)**	0.49 (0.661)	0.41 (0.988)	0.49 (0.650)
BAR(12)	**0.56 (0.034)**	**0.53 (0.175)**	**0.51 (0.415)**	0.43 (0.939)	0.49 (0.684)	**0.52 (0.242)**	**0.53 (0.161)**	0.47 (0.819)	0.42 (0.962)	0.50 (0.621)
VAR(12)	**0.57 (0.011)**	**0.58 (0.019)**	**0.53 (0.263)**	**0.55 (0.181)**	**0.60 (0.010)**	**0.55 (0.065)**	**0.56 (0.054)**	0.50 (0.434)	0.49 (0.620)	**0.55 (0.096)**
BVAR(12)	**0.57 (0.031)**	**0.57 (0.047)**	**0.55 (0.217)**	0.49 (0.752)	**0.53 (0.411)**	**0.53 (0.186)**	**0.56 (0.052)**	0.48 (0.839)	0.47 (0.884)	0.50 (0.727)
AR(24)	**0.53 (0.259)**	**0.54 (0.122)**	**0.53 (0.260)**	0.43 (0.929)	0.47 (0.766)	**0.53 (0.216)**	**0.54 (0.113)**	**0.52 (0.311)**	0.44 (0.906)	0.45 (0.882)
BAR(24)	**0.57 (0.020)**	***0.55 (0.088)***	**0.55 (0.181)**	0.48 (0.725)	**0.50 (0.540)**	**0.53 (0.176)**	***0.57 (0.040)***	**0.52 (0.319)**	0.45 (0.867)	0.45 (0.884)
VAR(24)	**0.53 (0.178)**	**0.56 (0.102)**	**0.52 (0.415)**	0.45 (0.919)	0.49 (0.783)	0.50 (0.477)	0.52 (0.412)	0.49 (0.673)	0.42 (0.966)	0.46 (0.906)
BVAR(24)	**0.58 (0.014)**	***0.64 (0.001)***	**0.59 (0.044)**	**0.53 (0.476)**	**0.57 (0.229)**	**0.53 (0.183)**	***0.60 (0.011)***	**0.57 (0.155)**	0.50 (0.604)	**0.55 (0.284)**
Futures	0.45 (0.373)	0.50 (0.371)	**0.51 (0.247)**	**0.55 (0.055)**	***0.59 (0.003)***	**0.51 (0.321)**	0.47 (0.689)	0.44 (0.736)	**0.53 (0.073)**	***0.57 (0.002)***
CRB Index	**0.54 (0.189)**	**0.63 (0.001)**	**0.63 (0.015)**	**0.51 (0.522)**	**0.52 (0.462)**	**0.56 (0.055)**	**0.55 (0.149)**	**0.57 (0.128)**	**0.52 (0.464)**	0.50 (0.672)

Note: Replication of Baumeister and Kilian (2012)'s real-time, out-of-sample-forecasts of the monthly average real price of RAC crude oil, 1992M1–2010M6. Brackets report the p-values for the serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and of Diebold and Mariano (1995) for equal MSFEs relative to the corresponding no-change forecast. Boldface values indicate gains relative to the respective no-change forecast. Statistically significant improvements over the end-of-month no-change forecast at the five percent significance level are italicized.

The second set of columns of Table 1 shows little evidence that the forecasts outperform the end-of-month no-change forecast at short forecast horizons. At the 1-month horizon, all models perform considerably worse than the end-of-month no-change forecast in terms of the MSPE, and, with a single exception, also in terms of the success ratio. At the 3-month horizon, some evidence of directional accuracy remains, but the economic significance is greatly reduced, and the statistical significance vanishes. Furthermore, the non-transitivity of directional accuracy is evident, as the ranking of success ratios changes when forecasts are evaluated against the end-of-month no-change benchmark. Notably, only the futures-based forecast significantly outperforms the end-of-month no-change forecast in directional accuracy, and only at the 12-month horizon.

The RAC forecasts in Table 2 present a similar pattern. The comparison against the monthly average-price no-change forecast reveals large forecast improvements of up to 26% in terms of MSPE at the 1-month horizon and 19% at the 3-month horizon. However, all forecasts perform worse than the end-of-month no-change forecast in terms of MSPE at the 1-month horizon. Evidence for predictability at the 3-month horizon is reduced or vanishes completely, with only the CRB index and VAR(12) showing consistent gains in both the MSPE and success ratio.

3.2 Updated Sample

We next examine the robustness of these results by extending the original sample period of Baumeister and Kilian (2012) to 2021M1. To our knowledge, this is the first study to include the COVID-19 crisis in out-of-sample forecasts of the real price of crude oil.[6]

Consistent with other replications of the VAR models, both real-time (Baumeister et al., 2014; Baumeister and Kilian, 2015; Funk, 2018) and non-real-time (Snudden, 2018; Baumeister et al., 2022), the forecast performance of the VAR models at short horizons deteriorates relative to the original sample. This deterioration is less pronounced in terms of directional accuracy for RAC, where the BVAR and CRB index forecasts provide some of the best forecasts at the 3- and 6-month horizons.

[6]The latest study to examine both directional accuracy and mean-squared precision of real-time forecasts is Funk (2018), with a sample ending in 2017M11. While Baumeister et al. (2022) extend the sample to 2018M8, their analysis neither examines directional accuracy nor includes real-time methods.

Table 3. Forecasts of WTI Crude Oil, 1992M1–2021M1

	Versus Monthly Average No-Change					Versus End-of-Month No-Change				
Horizon	1	3	6	9	12	1	3	6	9	12
Model	MSPE Ratio									
EoM NC	**0.59 (0.000)**	**0.89 (0.012)**	**0.95 (0.029)**	**0.96 (0.037)**	**0.96 (0.028)**					
ARMA(1,1)	**0.91 (0.130)**	**0.94 (0.227)**	**0.95 (0.287)**	**0.94 (0.295)**	**0.93 (0.265)**	1.54 (1.000)	1.06 (0.873)	1.00 (0.495)	**0.98 (0.44)**	**0.97 (0.396)**
AR(12)	**0.92 (0.179)**	**0.96 (0.317)**	**0.99 (0.434)**	1.00 (0.496)	**0.98 (0.411)**	1.57 (1.000)	1.09 (0.936)	1.04 (0.705)	1.04 (0.688)	1.02 (0.587)
VAR(12)	**0.91 (0.139)**	**0.96 (0.283)**	**0.98 (0.389)**	**0.99 (0.441)**	**0.97 (0.371)**	1.54 (1.000)	1.08 (0.922)	1.03 (0.664)	1.03 (0.640)	1.01 (0.546)
BAR(12)	1.03 (0.582)	1.05 (0.616)	1.10 (0.849)	1.14 (0.901)	1.16 (0.904)	1.74 (1.000)	1.18 (0.896)	1.16 (0.950)	1.19 (0.954)	1.21 (0.950)
BVAR(12)	**0.96 (0.336)**	1.02 (0.598)	1.09 (0.832)	1.13 (0.836)	1.15 (0.884)	1.62 (1.000)	1.16 (0.967)	1.14 (0.941)	1.18 (0.909)	1.20 (0.942)
AR(24)	**0.98 (0.388)**	1.00 (0.481)	1.00 (0.488)	1.00 (0.482)	**0.98 (0.410)**	1.66 (1.000)	1.13 (0.989)	1.05 (0.772)	1.04 (0.685)	1.02 (0.583)
VAR(24)	**0.92 (0.155)**	**0.96 (0.323)**	**0.98 (0.377)**	**0.98 (0.407)**	**0.97 (0.356)**	1.56 (1.000)	1.09 (0.945)	1.03 (0.648)	1.02 (0.606)	1.01 (0.526)
BAR(24)	1.34 (0.986)	1.41 (0.968)	1.43 (0.991)	1.45 (0.978)	1.57 (0.978)	2.28 (1.000)	1.59 (0.995)	1.50 (0.996)	1.51 (0.988)	1.64 (0.986)
BVAR(24)	**0.98 (0.435)**	1.08 (0.781)	1.15 (0.880)	1.25 (0.873)	1.34 (0.919)	1.66 (1.000)	1.22 (0.988)	1.21 (0.941)	1.30 (0.913)	1.39 (0.946)
Futures	1.00 (0.425)	**0.96 (0.078)**	**0.95 (0.130)**	**0.92 (0.095)**	**0.85 (0.025)**	1.69 (1.000)	1.09 (0.959)	1.00 (0.538)	**0.96 (0.259)**	**0.89 (0.078)**
CRB Index	**0.86 (0.096)**	**0.90 (0.144)**	1.17 (0.955)	1.22 (0.952)	1.19 (0.903)	1.45 (1.000)	1.01 (0.558)	1.23 (0.983)	1.28 (0.977)	1.24 (0.950)
	Success Ratio									
EoM NC	**0.71 (0.000)**	**0.61 (0.000)**	**0.55 (0.057)**	**0.56 (0.029)**	**0.57 (0.007)**					
ARMA(1,1)	**0.54 (0.066)**	**0.53 (0.088)**	0.50 (0.403)	**0.51 (0.341)**	**0.54 (0.155)**	0.47 (0.895)	0.50 (0.353)	0.49 (0.495)	**0.51 (0.380)**	0.50 (0.486)
AR(12)	**0.51 (0.320)**	**0.51 (0.362)**	0.50 (0.492)	0.49 (0.566)	**0.51 (0.387)**	0.46 (0.955)	0.50 (0.512)	0.49 (0.588)	0.47 (0.740)	**0.52 (0.348)**
VAR(12)	0.50 (0.470)	**0.51 (0.306)**	0.50 (0.504)	0.50 (0.502)	**0.53 (0.278)**	0.46 (0.958)	0.50 (0.389)	0.49 (0.581)	0.48 (0.712)	0.50 (0.476)
BAR(12)	0.49 (0.596)	**0.53 (0.140)**	**0.52 (0.234)**	**0.55 (0.084)**	**0.55 (0.071)**	0.45 (0.950)	**0.51 (0.315)**	0.50 (0.423)	**0.53 (0.221)**	**0.54 (0.101)**
BVAR(12)	0.48 (0.682)	**0.52 (0.261)**	**0.53 (0.250)**	**0.54 (0.179)**	**0.54 (0.205)**	0.45 (0.960)	**0.51 (0.342)**	0.50 (0.459)	**0.51 (0.427)**	**0.51 (0.432)**
AR(24)	0.50 (0.489)	**0.51 (0.327)**	**0.51 (0.325)**	0.48 (0.621)	0.49 (0.533)	0.46 (0.959)	**0.51 (0.333)**	0.50 (0.504)	0.49 (0.595)	0.45 (0.865)
VAR(24)	0.50 (0.481)	**0.51 (0.327)**	0.50 (0.428)	0.48 (0.648)	0.48 (0.635)	0.47 (0.847)	0.50 (0.407)	0.49 (0.535)	0.46 (0.798)	0.45 (0.856)
BAR(24)	0.49 (0.558)	**0.53 (0.140)**	0.50 (0.489)	0.47 (0.745)	0.45 (0.827)	0.47 (0.837)	**0.53 (0.178)**	0.48 (0.619)	0.47 (0.726)	0.46 (0.814)
BVAR(24)	0.50 (0.499)	**0.53 (0.200)**	**0.55 (0.114)**	**0.51 (0.443)**	**0.51 (0.418)**	0.48 (0.752)	**0.52 (0.236)**	**0.52 (0.319)**	**0.51 (0.438)**	0.49 (0.537)
Futures	0.47 (0.458)	0.50 (0.436)	**0.56 (0.074)**	***0.61 (0.002)***	***0.64 (0.000)***	0.47 (0.835)	0.49 (0.487)	0.49 (0.422)	***0.57 (0.014)***	***0.61 (0.000)***
CRB Index	**0.54 (0.083)**	***0.57 (0.030)***	***0.60 (0.010)***	**0.56 (0.129)**	**0.57 (0.067)**	0.49 (0.601)	**0.53 (0.222)**	**0.55 (0.129)**	**0.54 (0.187)**	**0.55 (0.121)**

Note: Replication of Baumeister and Kilian (2012)'s real-time, out-of-sample forecasts of the monthly average real price of RAC crude oil, 1992M1–2021M1. Brackets report the p-values for the serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and of Diebold and Mariano (1995) for equal MSFEs relative to the corresponding no-change forecast. Boldface values indicate gains relative to the respective no-change forecast. Statistically significant improvements over the end-of-month no-change forecast at the five percent significance level are italicized.

Table 4. Forecasts of RAC Crude Oil, 1992M1–2021M1

	Versus Monthly Average No-Change					Versus End-of-Month No-Change				
Horizon	1	3	6	9	12	1	3	6	9	12
Model					MSPE Ratio					
EoM NC	**0.70 (0.001)**	**0.88 (0.010)**	**0.95 (0.027)**	**0.96 (0.028)**	**0.96 (0.018)**					
ARMA(1,1)	**0.93 (0.221)**	**0.93 (0.196)**	**0.95 (0.293)**	**0.95 (0.318)**	**0.94 (0.284)**	1.33 (1.000)	1.06 (0.846)	1.00 (0.517)	1.00 (0.489)	**0.98 (0.442)**
AR(12)	**0.94 (0.299)**	**0.96 (0.332)**	**0.99 (0.473)**	1.01 (0.540)	1.00 (0.480)	1.35 (1.000)	1.08 (0.872)	1.05 (0.705)	1.05 (0.718)	1.04 (0.654)
VAR(12)	**0.93 (0.252)**	**0.95 (0.312)**	**0.99 (0.451)**	1.00 (0.512)	0.99 (0.453)	1.33 (1.000)	1.08 (0.858)	1.04 (0.687)	1.05 (0.696)	1.03 (0.630)
BAR(12)	**0.95 (0.383)**	1.03 (0.571)	1.14 (0.858)	1.16 (0.911)	1.15 (0.876)	1.36 (0.989)	1.17 (0.823)	1.00 (0.934)	1.21 (0.951)	1.20 (0.924)
BVAR(12)	**0.89 (0.212)**	1.00 (0.503)	1.10 (0.863)	1.13 (0.881)	1.13 (0.890)	1.28 (0.992)	1.13 (0.880)	1.15 (0.954)	1.18 (0.942)	1.18 (0.949)
AR(24)	**0.99 (0.478)**	1.02 (0.601)	1.04 (0.702)	1.02 (0.617)	1.01 (0.545)	1.42 (1.000)	1.16 (0.983)	1.09 (0.900)	1.06 (0.804)	1.05 (0.716)
VAR(24)	**0.95 (0.320)**	**0.99 (0.439)**	1.01 (0.535)	1.00 (0.513)	1.00 (0.479)	1.36 (1.000)	1.11 (0.936)	1.06 (0.773)	1.05 (0.712)	1.04 (0.654)
BAR(24)	1.14 (0.794)	1.36 (0.957)	1.54 (0.997)	1.58 (0.986)	1.69 (0.984)	1.63 (1.000)	1.54 (0.993)	1.62 (0.999)	1.65 (0.991)	1.76 (0.989)
BVAR(24)	**0.92 (0.262)**	1.07 (0.732)	1.20 (0.933)	1.30 (0.907)	1.39 (0.939)	1.31 (0.995)	1.21 (0.968)	1.27 (0.967)	1.36 (0.937)	1.45 (0.960)
Futures	**0.99 (0.263)**	**0.95 (0.024)**	**0.93 (0.057)**	**0.89 (0.045)**	***0.83 (0.007)***	1.42 (1.000)	1.07 (0.933)	**0.98 (0.326)**	**0.93 (0.161)**	***0.87 (0.034)***
CRB Index	**0.92 (0.248)**	**0.89 (0.139)**	1.13 (0.920)	1.18 (0.917)	1.11 (0.805)	1.31 (0.999)	1.01 (0.546)	1.19 (0.964)	1.23 (0.957)	1.16 (0.889)
					Success Ratio					
EoM NC	**0.71 (0.000)**	**0.61 (0.000)**	**0.58 (0.006)**	**0.55 (0.072)**	**0.57 (0.009)**					
ARMA(1,1)	**0.57 (0.001)**	**0.55 (0.006)**	**0.51 (0.228)**	**0.53 (0.204)**	**0.57 (0.044)**	**0.53 (0.169)**	**0.53 (0.103)**	**0.52 (0.195)**	**0.51 (0.330)**	**0.55 (0.068)**
AR(12)	**0.57 (0.001)**	**0.54 (0.047)**	0.50 (0.410)	0.50 (0.392)	**0.55 (0.072)**	**0.53 (0.144)**	**0.52 (0.157)**	**0.52 (0.241)**	0.48 (0.692)	**0.52 (0.279)**
VAR(12)	**0.57 (0.002)**	**0.54 (0.055)**	**0.52 (0.234)**	0.50 (0.404)	**0.56 (0.056)**	**0.53 (0.143)**	**0.51 (0.247)**	0.50 (0.410)	0.49 (0.602)	**0.53 (0.158)**
BAR(12)	**0.56 (0.005)**	**0.56 (0.020)**	**0.54 (0.107)**	**0.57 (0.034)**	***0.60 (0.001)***	**0.55 (0.057)**	**0.53 (0.127)**	**0.52 (0.186)**	**0.51 (0.332)**	***0.54 (0.046)***
BVAR(12)	**0.55 (0.016)**	**0.55 (0.045)**	**0.54 (0.140)**	**0.51 (0.351)**	**0.55 (0.075)**	**0.54 (0.095)**	**0.54 (0.060)**	0.50 (0.403)	0.49 (0.593)	0.50 (0.377)
AR(24)	**0.56 (0.017)**	**0.55 (0.033)**	**0.54 (0.099)**	0.50 (0.419)	**0.53 (0.244)**	**0.54 (0.089)**	**0.52 (0.230)**	**0.55 (0.057)**	0.47 (0.721)	0.48 (0.608)
VAR(24)	**0.59 (0.001)**	**0.55 (0.027)**	**0.55 (0.077)**	**0.51 (0.328)**	**0.54 (0.143)**	**0.54 (0.096)**	**0.53 (0.166)**	**0.53 (0.177)**	0.48 (0.628)	0.48 (0.668)
BAR(24)	**0.53 (0.151)**	**0.52 (0.218)**	**0.53 (0.166)**	0.48 (0.637)	0.49 (0.549)	**0.51 (0.350)**	**0.51 (0.304)**	0.49 (0.543)	0.45 (0.866)	0.46 (0.824)
BVAR(24)	**0.56 (0.009)**	***0.59 (0.004)***	**0.57 (0.029)**	**0.53 (0.264)**	**0.56 (0.092)**	**0.53 (0.123)**	***0.55 (0.029)***	**0.56 (0.069)**	0.50 (0.492)	**0.51 (0.318)**
Futures	0.46 (0.393)	**0.53 (0.150)**	**0.57 (0.048)**	***0.66 (0.000)***	***0.66 (0.000)***	**0.51 (0.300)**	0.49 (0.433)	**0.51 (0.272)**	***0.58 (0.006)***	***0.61 (0.000)***
CRB Index	**0.55 (0.035)**	**0.59 (0.008)**	***0.61 (0.005)***	**0.56 (0.112)**	**0.58 (0.053)**	**0.54 (0.105)**	**0.55 (0.064)**	***0.57 (0.042)***	**0.55 (0.149)**	**0.55 (0.152)**

Note: Replication of Baumeister and Kilian (2012)'s real-time, out-of-sample forecasts of the monthly average real price of RAC crude oil, 1992M1–2021M1. Brackets report the p-values for the serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and of Diebold and Mariano (1995) for equal MSFEs relative to the corresponding no-change forecast. Boldface values indicate gains relative to the respective no-change forecast. Statistically significant improvements over the end-of-month no-change forecast at the five percent significance level are italicized.

However, apart from the ARMA(1,1) at the 12-month horizon, none of the model-based forecasts outperform the end-of-month no-change forecast in both mean-squared precision and directional accuracy. This is true for both the WTI and RAC forecasts. As in the original sample, only the futures-based forecast significantly outperforms the end-of-month no-change forecast in mean-squared precision and directional accuracy, and only at the one-year horizon. These results suggest that our qualitative conclusions are robust to alternative sample end points.

4 Conclusion

Our study revisits the seminal findings of Baumeister and Kilian (2012) by replicating their real-time forecasts of the real price of crude oil and correcting a key methodological error. Our results suggest that the findings of Baumeister and Kilian (2012) stem from the fact that their no-change forecast is a monthly average price. Correcting for the error changes the directional accuracy assessment of alternative forecast approaches. Moreover, the inability of forecasts to consistently outperform the end-of-month no-change forecast – the traditional random walk forecast – questions the interpretation that the real price of crude oil is predictable in population at short horizons (Alquist et al., 2013).

Our replication study highlights the broader implications of a pervasive error in macroeconomic forecasting. These implications could only have been identified through replications. The results challenge the current consensus of the oil price forecasting literature and show the need to revisit conclusions of subsequent studies in this area. An open question is how they extend to the forecast of other primary commodity prices and macroeconomic series, which are often expressed as monthly averages. Our findings emphasize the need to reassess conclusions in all studies forecasting period-average data and motivate further replication efforts.

Acknowledgements

We thank Ron Alquist, Christiane Baumeister, and Lutz Kilian, as well as seminar and conference participants at the European Economic Association Annual Congress, the Canadian Economics Association Annual Conference, the International Association for Energy Economics International Conference, and the International Symposium on Forecasting for their discussions, comments and suggestions.

Data Availability Statement

The data and programs that support the findings of this study are openly provided by Benyo et al. (2025) at OPENICPSR, https://doi.org/10.3886/E218641V4, 2025-6-17.

References

Alquist, R., Kilian, L., and Vigfusson, R. (2013). Forecasting the price of oil. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, chapter 8, pages 427–507. Elsevier, 1 edition.

Baumeister, C. and Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, 30(2):326–336.

Baumeister, C. and Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics*, 33(3):338–351.

Baumeister, C., Kilian, L., and Lee, T. K. (2014). Are there gains from pooling real-time oil price forecasts? *Energy Economics*, 46:S33–S43.

Baumeister, C., Korobilis, D., and Lee, T. K. (2022). Energy markets and global economic conditions. *The Review of Economics and Statistics*, 104(4):828–844.

Benyo, E., Ellwanger, R., and Snudden, S. (2025). ECIN replication package for "A reappraisal of real-time forecasts of the real price of oil". Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2025-2-9. https://doi.org/10.3886/E218641V1.

Cowles, A. (1960). A revision of previous conclusions regarding stock price behavior. *Econometrica: Journal of the Econometric Society*, 28(4):909–915.

Cowles, A. and Jones, H. E. (1937). Some a posteriori probabilities in stock market action. *Econometrica*, 5(3):280–294.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Ellwanger, R. and Snudden, S. (2023). Forecasts of the real price of oil revisited: Do they beat the random walk? *Journal of Banking and Finance*, 154:1–8.

Filippidis, M., Filis, G., and Magkonis, G. (2024). Evaluating oil price forecasts: A meta-analysis. *The Energy Journal*, 45(2):71–89.

Funk, C. (2018). Forecasting the real price of oil-time-variation and forecast combination. *Energy Economics*, 76:288–302.

Garratt, A., Vahey, S. P., and Zhang, Y. (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics*, 34(3):456–462.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3):1053–69.

Kilian, L. (2015). Comment on Francis X. Diebold's "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests". *Journal of Business and Economic Statistics*, 33(1):13–17.

Kilian, L. (2019). Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economics Letters*, 178:106–110.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.

Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: an extension of hamilton's results. *Journal of political Economy*, 97(3):740–744.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Snudden, S. (2018). Targeted growth rates for long-horizon crude oil price forecasts. *International Journal of Forecasting*, 34(1):1–16.

Weiss, A. A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, 26(3):271–281.

Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica*, 28(4):916–918.