The New Benchmark for Forecasts of the Real Price of Crude Oil*

Amor Aniss Benmoussa† Reinhard Ellwanger‡ Stephen Snudden§¶

Abstract

This paper investigates forecasts of the real price of crude oil through the lens of temporal aggregation. Estimating forecast models of monthly oil prices using disaggregated oil-price data yields economically large and robust forecast improvements over conventional monthly models. The results also hold for U.S. refinery acquisition costs of crude oil, for which pseudo disaggregated data can be imputed. However, while these forecasts provide unprecedented improvements over the conventional monthly no-change benchmark at short horizons, they do not significantly outperform the random-walk forecast from daily prices. The results call into question the existing evidence on oil-price predictability derived from monthly data.

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.

†Bank of Canada, International Economic Analysis Department

‡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

¶We thank Ron Alquist, Christiane Baumeister, Wing Chan, Willy Chetwin, Edouard Djeutem, Lutz Kilian, Pierre Siklos, Luis Uzeda, and Konrad Zmitrowicz, as well as seminar and conference participants at the Bank of Canada, the European Economic Association Annual Congress, the Canadian Economics Association Annual Conference, and the International Association for Energy Economics International Conference for their discussions, comments and suggestions.

1 Introduction

The real price of crude oil is a key global macroeconomic indicator that has attracted considerable attention among forecasters. A growing literature documents that model-based forecasts of the monthly real price of crude oil can outperform a simple no-change benchmark (Ye et al., 2005; Baumeister and Kilian, 2012; Alquist et al., 2013; Baumeister et al., 2014; Baumeister and Kilian, 2014, 2015; Wang et al., 2015; Yin and Yang, 2016; Snudden, 2018; Zhang et al., 2018; Funk, 2018; Garratt et al., 2019). The conventional interpretation of these results is that model-based forecasts are more useful than naive forecasts, and that the real price of oil is predictable relative to the random walk forecast.

However, the real price of oil used in all of these studies is a monthly average price, which is problematic for several reasons.[1] First, the correct no-change forecast implied by the random walk model is the last observed closing price rather than the last observed monthly average price (Ellwanger and Snudden, 2021b). Second, average prices are predictable relative to the monthly no-change forecast *by construction* for a large class of data generating processes, including the random walk (Working, 1960) and arbitrary ARIMA processes (Ellwanger and Snudden, 2021b). Both results call into question the usefulness of comparing model-based forecasts to the conventional no-change benchmark. Moreover, estimating models with temporally aggregated data is known to reduce forecast efficiency (Wei, 1978; Kohn, 1982), which can be remedied by using disaggregated data in forecasts of aggregated variables (see, e.g. Lütkepohl, 1984; Ellwanger and Snudden, 2021b). Based on these insights, the goal of this paper is to study the effects of introducing daily oil-price data in computing and evaluating model-based forecasts of the monthly real price of crude oil.

Our first contribution is to evaluate the performance of different no-changes forecasts across various crude oil price series. Ellwanger and Snudden (2021b) showed that the random walk forecast constructed from daily Brent prices improves the accuracy of the conventional no-change forecast constructed from monthly average prices. We document similar gains from the random walk forecast based on daily WTI prices. The improvement in the one-step-ahead mean squared prediction error (MSPE) and the directional accuracy are larger than 40 percent. The gains decrease with the forecast horizon, but are still apparent up to the 12-months-ahead forecast. Moreover, we show that any averaging over end-of-month prices worsens the forecast accuracy for both the Brent

[1] The standard series that are used in structural models and forecasting applications of real oil prices are based on monthly averages of daily prices (this is the case for the oil-price series used in Kilian, 2009; Alquist et al., 2013; Kilian and Murphy, 2014; Baumeister and Hamilton, 2019, among many others).

and the WTI forecasts. The success of the random walk forecast bears important consequences for forecast evaluations because it raises the bar for model-based forecasts to claim any practical usefulness. For example, we find that the original evidence for the short-run predictability of the real price of oil presented in Baumeister and Kilian (2012) is overturned when their model-based forecasts are compared to the new random-walk benchmark instead of the conventional, average-price no-change forecast.

Our second contribution is to show how the use of disaggregated data can be extended to the forecasts of the U.S. refinery acquisition cost of crude oil (RAC), which is often used in lieu of other crude oil prices because of its longer history (Alquist et al., 2013). Using disaggregated data for the RAC is complicated by the fact that the RAC is published as a monthly survey. To circumvent this problem, we propose a simple method to impute quasi-end-of-period RAC observations from end-of-period WTI prices. Similar to the previous exercise, the random walk forecast based on the imputed end-of-period RAC yields large forecasting gains over the traditional no-change forecast, with over 30 percent improvements in the MSPE at short horizons. This shows that dis-aggregation of averaged series is useful even in cases where end-of-period observations are not directly observable.

Our third contribution is to evaluate the effectiveness of model-based forecasts that rely on disaggregated data. This includes Lütkepohl (1984)'s approach of estimating models for the daily oil prices and aggregating the forecasts to the monthly frequency. We also compute forecasts under period-end price sampling (PEPS), which selectively samples end-of-period real prices and uses them to replace average prices in model estimations and forecasts (Ellwanger and Snudden, 2021b). We find that both Lütkepohl (1984)'s and the PEPS approach can yield large improvements relative to the conventional no-change forecast and to models estimated with averaged data. However, PEPS tends to perform better for most forecast horizons. The results improve on existing MIDAS forecasts of the real price of crude oil, which represents the main attempt to include higher-frequency variables in forecasts of the real price of oil so far (see, e.g. Baumeister et al., 2015).

A practical advantage of PEPS is that it maintains the lower frequency of the target variable, which allows forecasters to incorporate end-of-period observations easily into existing models. To this end, we compare the performance of Bayesian and classically estimated autoregressive (BAR/AR) and vector autoregressive (BVAR/VAR) models from Baumeister and Kilian (2012), which are estimated with averaged data, to the PEPS forecasts for these models. For the empirical application, we update the real-time dataset created by Baumeister and Kilian (2012).[2] Again, the

[2]Our data set improves upon other studies that have updated the Baumeister and Kilian (2012) real-time data

PEPS approach results in forecast improvements of over 40 percent over the original Baumeister and Kilian (2012) specification, and the improvements are remarkably stable over time. However, despite these substantial forecast improvements over the conventional no-change forecast and models estimated with averaged data, the models estimated with disaggregated data generally do not outperform the random-walk forecast.

Finally, we also investigate other practical implications of introducing disaggregated data into oil-price forecasts. For example, we find that the switch from average to end-of-period observations generally reduces the desired number of autoregressive lags in univariate models. We also show that it reduces relative importance of covariates, as alternative indicators of inventories and real economic activity indicators provide very similar forecasts under the PEPS approach. While we find some evidence that forecast combinations can outperform the random walk model at long forecast horizons, our results are consistent with oil prices being unpredictable at short forecast horizons.

Our results highlight the importance of using disaggregated oil prices in the context of forecasting the average real price of crude oil. Although our setup is new, the results are broadly consistent with several distinct pieces of evidence provided in earlier studies.[3] For example, Baumeister and Kilian (2014) document that the monthly average random-walk model forecasts average quarterly prices better than the quarterly average random-walk model. Baumeister and Kilian (2012) show that most forecasting gains from using revised data stem from improved information on prices, while Bork et al. (2019) and Conlon et al. (2022) document that averaging commodity prices can introduce spurious predictability in the context of return-forecasting regressions. These results are all indicative that time-averaging could introduce a mechanical loss of information about future price levels. Herein, we show that the introduction of disaggregated data restores this information and changes the assessment of model-based forecasts of the real-price of crude oil.

2 Real-time Data

The focus of our empirical application is real-time forecasts of the level of the monthly average real price of crude oil, which is the standard approach in the literature (Baumeister and Kilian, 2012; Alquist et al., 2013). In this setting, the forecaster uses the available information at the end of each

set by incorporating all available monthly vintages of the US Energy Information Administration's MER report.

[3]End-of-period prices have previously been used to evaluate the predictive content of oil-futures prices for the end-of-month nominal oil prices by Alquist and Kilian (2010). However, in the subsequent literature on forecasts for the real price of crude oil, average prices have become the standard for estimation and forecast evaluation.

month to form their prediction for the following months. Forecasts are examined for the monthly series of WTI and Brent crude oil prices, as well as the RAC.

Crude oil prices and the RAC are obtained from the U.S. Energy Information Administration (EIA). For Brent and WTI, the EIA reports the average monthly price as simple averages of the daily closing price. We construct a series of end-of-month prices by taking the closing price on the last trading day of each month, which represents the last observed price contained in the forecaster's monthly information set. For WTI and Brent crude oil prices, both the end-of-month price and the monthly average price are observed in real time.[4]

Disaggregating the RAC is complicated by the fact that no daily price series exists. For our empirical analysis, we impute two distinct series of artificial end-of-period RAC observations by applying the growth rate of the end-of-month WTI (Brent) oil price over the monthly average WTI (Brent) oil price to the monthly RAC price.[5] Thus,

$$p_{t,n}^{RAC} = p_{t,a}^{RAC} \cdot (p_{t,n}^{x} / p_{t,a}^{x}), \tag{1}$$

where $p_{t,n}$ stands for the end-of-period observation, $p_{t,a}$ stands for the monthly average. The superscript RAC denotes the RAC nominal price series and $x \in \{WTI, Brent\}$ is the price series used for imputation of the pseudo-end-of-period calculation.

Nominal monthly average prices, end-of-month prices and the RAC are all converted to real prices using the seasonally adjusted U.S. consumer price index. The price index is the seasonally adjusted U.S. consumer price index obtained from the FRASER database of the Federal Reserve Bank of St. Louis and the real-time database of the Philadelphia Federal Reserve.

A contribution of this paper is to extend the real-time data of Baumeister and Kilian (2012) using historical data vintages from the EIA's Monthly Energy Review and Short-Term Energy Outlook. Other than in earlier updates of this dataset provided in Garratt et al. (2019), we collect all monthly vintages after 2010M12, which allows us to include revisions for the entire history of each vintage.[6] Real-time data vintages start in 1991M1 and contain historical data starting from 1973M1. Real-time data on U.S. crude-oil inventories, U.S. petroleum inventories and OECD petroleum inventories are obtained from historical releases in the EIA's Monthly Energy Review or the International Data Browser. The real-time version of Kilian (2009)'s real economic activity

[4] See Appendix A1 for details.

[5] This is akin to the standard practice of nowcasting the average nominal RAC via the growth rate of WTI prices (see, e.g. Baumeister and Kilian, 2012).

[6] See Appendix A1 for details.

index is computed using the corrected formula (Kilian, 2019).

All nowcasts follow Baumeister and Kilian (2012). For example, the monthly average RAC is nowcasted using the month-over-month growth rate of the monthly WTI series. Any missing real-time observations for the consumer price index are nowcasted using the average historical growth rate. To nowcast the imputed end-of-period observations for the RAC, we use the historical vintage of the nowcasted RAC prices at the time of the forecast. As discussed in Section 4, none of the main results are affected by different nowcasting choices.

For estimation of econometric models, prices for Brent and WTI are backcasted to 1973M1. The series of WTI and Brent prices provided by the EIA begin in 1986M1 and 1987M5, respectively. Prices are backcasted to 1983M4 using the growth rate of the nearby WTI futures contract. All monthly prices before 1983M4 are backcasted using the growth rate of the RAC. Our main findings are robust to estimating models with data starting in 1983 or 1986.

All forecasts are computed out-of-sample, beginning in 1992M1. Models are estimated with an expanding window using data beginning in 1973M1. To be consistent with the existing literature, the MSPE ratio and the success ratio for directional accuracy are reported relative to the monthly average no-change forecast. However, we also test for statistically significant improvements vis-à-vis the random-walk no-change forecast following Ellwanger and Snudden (2021b). Statistical tests for the null hypothesis of equal predictability using the MSPE ratio follow Diebold and Mariano (1995), while the null hypothesis that the success ratios are drawn at random are assessed using the test of Pesaran and Timmermann (2009).[7] Both tests account for serial dependence in the forecasts.

3 Real-time Forecasts

The effects of introducing end-of-month prices on real-time forecasts of the real price of crude oil are investigated in several exercises. First, we compare no-change forecasts based on end-of-month data to the standard average-price no-change forecast for alternative oil price series. Second, we compare the performance of two forecasting approaches that rely on disaggregated data in univariate time-series models. Third, we replicate the econometric models proposed in Baumeister and Kilian (2012) and compare the forecasts from the original models to forecasts from models estimated with

[7] With real-time data and the iterative out-of-sample forecasting setup, the assumptions of the Diebold-Mariano test are not fulfilled in our setup (Kilian, 2015). As is standard, the ratios are still reported with this warning in mind.

period-end prices. Fourth, we examine how the use of end-of-month observations affects the choice of lag length and covariates in forecasting models. Finally, we examine whether oil futures prices and forecast combinations can improve the accuracy of short-term forecasts based on disaggregated data.

3.1 Comparison of No-Change Benchmarks

The literature on forecasts of the real crude oil price has focused on the average real price of crude oil. Throughout the paper, we maintain the goal of forecasting this average real price, $R^x_{t,a} = p^x_{t,a}/CPI_t$, where $p^x_{t,a}$ is the monthly average of daily nominal closing prices, denoted by subscript a, CPI_t is the U.S. consumer price index, and x is the crude oil price series, such as WTI or Brent. It has been standard practice to compare the performance of a forecast of this real price h steps ahead, $\hat{R}^x_{t+h|t,a}$, to that of the average-price no-change forecast:

$$\hat{R}^{x,no-change}_{t+h|t,a} = R^x_{t,a}. \tag{2}$$

However, Ellwanger and Snudden (2021b) show that for many macroeconomic series, the forecast performance of the conventional no-change forecast in equation (2) is dominated by the random walk no-change forecast that is based on end-of-period observations:

$$\hat{R}^{x,random-walk}_{t+h|t,a} = R^x_{t,n}, \tag{3}$$

where $R^x_{t,n} = p^x_{t,n}/CPI_t$ and n denotes the last disaggregated observation within a given period, the closing price on the last trading day of the month. The forecast in Equation (3) corresponds to a random walk model for daily prices (rather than for monthly averages prices) and yields large MSPE gains over the average-price no-change forecast under the null hypothesis that *all* future prices are unpredictable (Ellwanger and Snudden, 2021b). In this section, we investigate the performance and robustness of the random-walk no-change forecast across different crude-oil markers and against alternative measures of average end-of-month prices.

The first two columns of Table 1 report the MSPEs and success ratios for the random-walk forecasts relative to the monthly average no-change forecasts. The random-walk forecasts are significantly more accurate than the monthly average no-change forecasts, with accuracy-improvements of over 40 percent at the one-month-ahead-horizon. The result holds for both Brent and WTI. The improvements in the success ratio are largest for the one-step-ahead prediction, but remain

Table 1. Random-Walk vs Average-Price No-Change Forecasts of the Real Price of Crude Oil

Price	Brent	WTI	RAC	RAC	WTI	WTI	WTI
Sampling	EoM	EoM	Brent EoM	WTI EoM	EoM 2 Day Avg.	EoM 1 Week Avg.	EoM 2 Week Avg.
Horizon				MSPE Ratio			
1	**0.58 (0.000)**	**0.59 (0.000)**	**0.70 (0.003)**	**0.70 (0.001)**	**0.60 (0.000)**	**0.67 (0.000)**	**0.78 (0.001)**
3	**0.92 (0.017)**	**0.89 (0.012)**	**0.91 (0.031)**	**0.89 (0.010)**	**0.89 (0.017)**	**0.92 (0.025)**	0.95 (0.071)
6	0.98 (0.123)	**0.95 (0.029)**	0.98 (0.131)	**0.95 (0.027)**	**0.96 (0.044)**	0.98 (0.155)	0.99 (0.297)
12	0.98 (0.170)	**0.96 (0.028)**	0.99 (0.191)	**0.96 (0.018)**	0.98 (0.116)	1.00 (0.473)	1.01 (0.714)
24	1.00 (0.443)	0.99 (0.189)	1.00 (0.422)	0.98 (0.123)	1.00 (0.440)	1.01 (0.838)	1.01 (0.936)
				Success Ratio			
1	**0.72 (0.000)**	**0.71 (0.000)**	**0.70 (0.000)**	**0.71 (0.000)**	**0.70 (0.000)**	**0.70 (0.000)**	**0.67 (0.000)**
3	**0.57 (0.005)**	**0.61 (0.000)**	**0.60 (0.000)**	**0.61 (0.000)**	**0.61 (0.000)**	**0.59 (0.000)**	**0.57 (0.006)**
6	**0.56 (0.013)**	0.55 (0.057)	**0.58 (0.001)**	**0.58 (0.006)**	**0.56 (0.039)**	**0.55 (0.043)**	**0.55 (0.047)**
12	**0.59 (0.000)**	**0.57 (0.007)**	**0.58 (0.001)**	**0.57 (0.009)**	**0.58 (0.003)**	**0.56 (0.017)**	0.52 (0.322)
24	**0.56 (0.015)**	0.52 (0.292)	**0.55 (0.027)**	0.54 (0.153)	0.51 (0.415)	0.49 (0.737)	0.47 (0.880)

Notes: Random walk versus monthly average no-change forecasts for alternative crude oil prices benchmarks, 1992M1–2021M1. The random walk forecast is the no-change forecast computed from the end-of-month closing price. *EoM* refers to end-of-month observations. *Brent EoM* and *WTI EoM* stand for RAC closing prices which are imputed using Brent and WTI prices, respectively. *Avg.* stands for average. Brackets report the p-values for serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) for equal MSPEs relative to the monthly average no-change forecast.

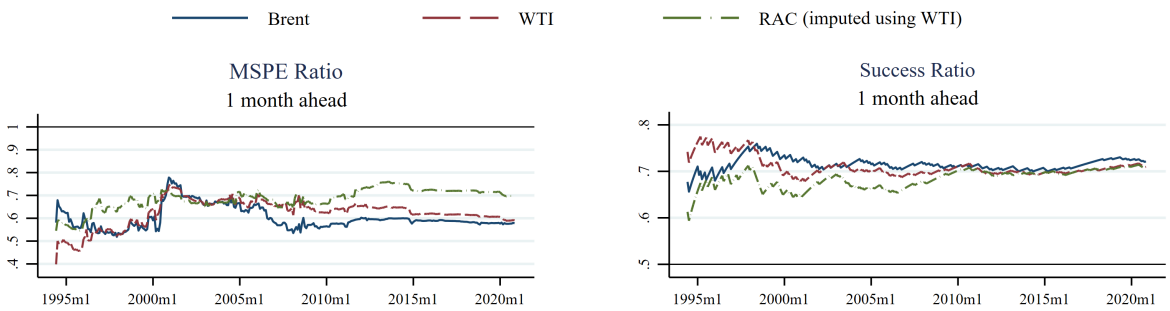
substantial and often significant for longer-term forecasts. For the MSPEs, the gains decrease as the forecasting horizon increases but are still statistically significant at the 5 percent significance level for up 12-months-ahead forecast. The decay in the relative forecasting performance is quantitatively very similar to the theoretical prediction of the pure random walk model for daily oil prices (Ellwanger and Snudden, 2021b). These results hence corroborate the idea that crude oil prices are closely approximated by a random walk at the daily frequency.

The performance of the quasi end-of-period random walk forecast for the RAC is also very good, with accuracy-improvements of 30 percent at the one-step-ahead prediction. For all forecast horizons, the MSPE and success ratios of the imputed end-of period price of RAC do at least as well as the monthly average no-change forecast, and often significantly better. This is true whether the RAC closing prices are imputed using WTI or Brent prices. However, the imputation using the WTI price produces lower MSPE ratios beyond the one-month-ahead horizon. These results highlight that the even in cases where the end-of-period price is not directly observed, attempts to dis-aggregate such series can yield first-order forecast gains.

Price-averaging could be potentially useful for forecasters as it tends to reduce the variance of the underlying series. This is particularly true when the series is affected by random errors, which could occur from measurement error or from market-microstructure dynamics. Table 1 reports the

no-change forecasts that are computed from averaging daily closing prices over the last two trading days, the last week of trading, and the last two weeks of trading before the end of the month. The results indicate that there is no benefit from averaging, as it only leads to a deterioration in the forecast performance. Moreover, this worsening increases with the length of the averaging window. Thus, in practice, the benefits from using end-of-month prices clearly outweigh the concerns over increasing the volatility relative to average price.[8]

Figure 1. Recursively Estimated Random Walk Forecasts vs Average-Price No-Change Forecasts



Notes: Relative forecast performance of random walk forecasts versus monthly average no-change forecasts of the real price of crude oil, 1992M1–2021M1. The first 30 months are omitted to reduce starting-point effects. The RAC end-of-month price is imputed using daily WTI crude oil prices. The left panel shows the evolution of the MSPE ratio of the one-month-ahead random walk forecast relative to the average-price no-change forecast. The right panel shows the evolution of the directional accuracy as measured by the success ratio.

How stable are these results over time? The performance of previously proposed forecast methods for the real price of crude oil are often very sensitive to the sample period (Baumeister et al., 2014; Baumeister and Kilian, 2015; Snudden, 2018; Garratt et al., 2019; Baumeister et al., 2020). By contrast, as Figure 1 documents, the forecast improvements of the end-of-month relative to the monthly average no-change forecast at the one-month-ahead horizon are robust across all potential sample end dates between 1995 and 2021. The improvements in the MSPE and success ratio are statistically significant at the 5 percent level for 100 percent of the sample, which is remarkable for oil-price forecasts. As such, end-of-period prices seem to restore a mechanical loss of information in average prices which is largely independent of the oil-price behavior over specific episodes.

The good performance of the random walk forecast is generally consistent with the idea that daily crude oil prices are well approximated by a random walk. In line with theoretical results for forecasts of temporally aggregated non-stationary series, averaging worsens the forecast performance of the no-change forecast (see, e.g. Working, 1960; Wei, 1978). Our results suggest that the

[8]The empirical variances of the monthly closing prices of crude oil are statistically indistinguishable from the variances of the monthly average prices of crude oil (see Appendix A1). This suggests that the improvements from using closing prices is driven by a smaller bias, again consistent with a random walk model for daily prices.

random-walk no-change forecast, computed from the last observed closing price, should be used as a benchmark for forecast evaluations of the real price of crude oil. This new benchmark produces large forecast gains over the conventional, average-price no-change forecast and thus raises the bar for model-based forecast to claim any usefulness over naive forecasts. The practical importance of this effect is highlighted throughout the remainder of this paper.

3.2 Disaggregated Data for Univariate Models

A common feature of existing forecasts of the real price of crude oil is that models are estimated with monthly average prices. However, estimating models with temporally aggregating data can lead to a loss of information and reduce forecasting efficiency (Wei, 1978; Kohn, 1982). This section explores whether and how forecasts relying on daily oil-price data can help restore this information and thus improve traditional oil-price forecasts based on monthly data. To the best of our knowledge, this paper presents the first application of models estimated with daily crude oil prices to forecast the average real price of crude oil.

One approach to deal with aggregation is to estimate models and compute forecasts for disaggregated data first, and to aggregate the forecasts to the desired lower frequency afterwards (Lütkepohl, 1984). This approach avoids aggregation and the associated information loss altogether. In our context, this implies estimating models for daily oil prices and taking the monthly average over the forecasted daily prices.

Ellwanger and Snudden (2021b) propose an alternative approach that includes information from daily prices but maintains the lower frequency of the dependent variable: period-end price sampling (PEPS). In our context, this method uses end-of-month closing prices instead of monthly average prices in econometric models. The model-implied forecasts for end-of-period prices are then taken to predict average prices. The PEPS approach is potentially useful when daily prices behave similar to a random walk, in which case expectations of future average and end-of-period observations coincide. It often reduces the number of estimated parameters and might therefore produce better out-of-sample forecasts than models estimated for daily data.

Another approach that relies on disaggregated data is mixed-data sampling (MIDAS, see Ghysels et al., 2007). MIDAS allows forecasters to forecast aggregated variables with higher frequency, disaggregated predictors. The MIDAS approach has been applied to forecasts of the real price of crude oil with some success by Zhang and Wang (2019) and Degiannakis and Filis (2018). However, while the approach has been shown to result in forecast gains, the gains are usually not so large as

to be able to improve upon forecast based on aggregated prices (Baumeister et al., 2015). Given that this method is well documented in previous papers and their forecast performance is worse than that of the random-walk no-change forecast, we do not replicate them here.

This section compares forecasts from three different types of univariate models of the real price of WTI crude oil. Specifically, we consider autoregressive models, AR(p), autoregressive moving average models, ARMA(p,q), and autoregressive models with autoregressive conditional heteroskedasticity in the innovations, AR-ARCH(p,m). The models are estimated using monthly average data (as is standard in the existing literature), daily data (Lütkepohl (1984)'s "disaggregated" approach), and end-of-month data (the PEPS approach).

For the disaggregated and the PEPS approach, the current-month CPI index is used to deflate the end-of-the month observation. To deflate intra-month observations under the disaggregated approach, a daily real-time measure of the CPI is interpolated using a cubic spline, but the results are robust to linear interpolation. We follow standard parameterizations for the lag length.[9] For models estimated with monthly data, the models are AR(2), ARMA(1,1), and AR-ARCH(2,1). Since there are about 22 business days in a month, the corresponding models for the daily data are AR(44), ARMA(22,22), and ARCH(44,1). However, similar to Lütkepohl (1984) we found that ARMA models with long lags and in particular the ARMA(22,22) model resulted in ill-conditioned optimization problems. We therefore consider an ARMA(1,1) model for daily data instead. None of our qualitative results is affect by this choice.

The models are estimated in a recursive fashion, starting in 1973M1 for the monthly models and in 1986M1 for the disaggregated approach. The evaluation period is 1992M1-2021M1. The models are estimated with prices in log real levels and forecasts are converted to real levels,

$$\hat{R}^{x}_{t+h|t,j} = exp\left(\hat{\hat{r}}^{x}_{t+h|t,j}\right), \tag{4}$$

where $j \in \{a, n\}$ for oil series x, $\hat{\hat{r}}^{x}_{t+h|t,j}$ is forecast of the real price in logs, and exp is the exponential function. Forecasts for daily prices are averaged to monthly forecasts after transforming to levels.

The first three columns of Table 2 report the forecast performance of the univariate models when estimated using the standard monthly average price series. Most of these models can outperform the average price no-change forecasts, especially at shorter forecast horizons. However, the improvements appear relatively modest (below 10 percent) and, for the most part, not statis-

[9]The relationship between temporal aggregation and the lag length of forecasting models is further investigated in Section 3.4.

Table 2. Forecasts of the Real Price of WTI Crude Oil Based on Univariate Models

Model	AR(2)	ARMA(1,1)	AR-ARCH (2,1)	AR(44)	ARMA(1,1)	AR-ARCH (44,1)	AR(2)	ARMA(1,1)	AR-ARCH (2,1)
Method	Monthly Average Prices			Daily Prices			Period-End Price Sampling		
Horizon					MSPE Ratio				
1	0.90 (0.156)	0.91 (0.131)	0.93 (0.071)	0.60 (0.000)	0.61 (0.000)	0.62 (0.000)	0.56 (0.001)	0.56 (0.001)	0.63 (0.000)
3	0.95 (0.333)	0.95 (0.256)	0.97 (0.213)	0.94 (0.219)	0.98 (0.429)	0.98 (0.307)	0.84 (0.113)	0.84 (0.088)	0.87 (0.021)
6	0.96 (0.406)	0.96 (0.335)	0.99 (0.382)	1.05 (0.754)	1.08 (0.653)	1.15 (0.876)	0.89 (0.215)	0.89 (0.180)	0.89 (0.054)
12	0.98 (0.463)	0.95 (0.336)	0.99 (0.386)	1.12 (0.830)	1.18 (0.773)	1.33 (0.948)	0.90 (0.250)	0.89 (0.190)	0.85 (0.020)
24	1.05 (0.589)	0.96 (0.409)	1.03 (0.695)	1.18 (0.835)	1.33 (0.891)	1.67 (0.960)	0.95 (0.385)	0.92 (0.308)	**0.82 (0.015)**
					Success Ratio				
1	0.52 (0.315)	0.53 (0.107)	0.53 (0.305)	0.72 (0.000)	0.68 (0.000)	0.72 (0.000)	0.71 (0.000)	0.72 (0.000)	0.70 (0.000)
3	0.49 (0.638)	0.52 (0.093)	0.50 (0.802)	0.55 (0.075)	0.53 (0.145)	0.57 (0.041)	0.58 (0.002)	0.59 (0.000)	0.61 (0.000)
6	0.49 (0.652)	0.48 (0.554)	0.46 (0.971)	0.48 (0.567)	0.50 (0.399)	0.52 (0.457)	0.53 (0.276)	0.54 (0.096)	0.59 (0.012)
12	0.54 (0.250)	0.53 (0.213)	0.52 (0.577)	0.50 (0.399)	0.53 (0.198)	0.50 (0.633)	0.54 (0.261)	0.56 (0.064)	0.58 (0.055)
24	0.59 (0.063)	0.54 (0.190)	0.53 (0.479)	0.57 (0.065)	0.56 (0.070)	0.46 (0.925)	0.61 (0.035)	0.56 (0.086)	0.60 (0.042)

Note: Real-time, out-of-sample forecasts for the real price of WTI crude oil, 1992M1–2021M1. Models estimated with "Daily Prices" are aggregated to the monthly frequency following (Lütkepohl, 1984); Models estimated with "Period-End Price Sampling" follow Ellwanger and Snudden (2021b). The brackets show the p-values serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs relative to the monthly average-price no-change and of Pesaran and Timmermann (2009) for the null of no directional accuracy. Boldface values indicate improvements over the random-walk no-change forecast at the five percent level.

tically significant. The results from the disaggregated approach, shown in the next three columns, show about 40 percent improvements at the 1-step ahead prediction for all three models. The improvements are significant at the 0.1 percent significance level and apply to both the MSPE ratios and the success ratios. However, for longer forecast horizons, the performance of the models deteriorates quickly. For 6-months-ahead and longer, the MSPE ratios indicate that forecasts from the disaggregated approach are worse than the average-price no-change forecast. The worsening performance of the MSPE at longer horizons could be because the daily WTI sample starts later, in 1986M1. Moreover, the AR and AR-ARCH models are heavily parameterized, which could weigh on their out-of-sample performance.

The forecasts based on the PEPS approach are evaluated in the final 3 columns of Table 2. For all horizons and models, the forecasts outperform the conventional no-change benchmark. At the one-step ahead, the PEPS approach results in the largest improvements in forecast accuracy of up to 44 percent, which are significant at the 0.1 percent level. Although the forecast performance tends to decrease with the forecast horizons, the performance of the AR-ARCH model remains significantly better than that of the average-price no-change forecast at longer horizons. The success ratios are also significant at the ten percent level for all horizons for the ARMA and AR-ARCH forecasts.

Overall, these results suggest that forecast approaches employing disaggregated methods significantly outperform models estimated with averaged data. It can be shown that this increase

in performance is very stable over time. For the disaggregated and PEPS approach, the one-step ahead forecast gains are relatively stable and remain significant for all ending dates of the forecast evaluation sample. However, when using average prices, the forecast performance depends on the choice of evaluation period.

Finally, Table 2 also shows the importance of the choice of the benchmark to evaluate forecast models of the real price of oil. Whereas the conventional interpretation of these results would be to advocate for the use of model-based forecasts and to declare the price of oil predictable, the absence of boldface values (for all but one long-run forecast) suggest that none of the forecast models/methods significantly improve upon the random-walk no-change forecast at short and medium horizons. This further highlights that improvements over the average-price no-change benchmark are, by themselves, not suitable to provide evidence in favor of model-based approaches.

3.3 Disaggregated Data for Multivariate Models: Replication of Baumeister and Kilian (2012)

A practical advantage of using the period-end price sampling approach proposed by Ellwanger and Snudden (2021b) is that it is easily extended to multivariate models that include lower-frequency predictor variables. Existing multivariate forecast models of the real price of crude oil can thus be adapted by replacing the monthly average with end-of-month prices. We illustrate this point by investigating the performance of the Bayesian and classically estimated autoregressive (BAR/AR) and vector autoregressive (BVAR/VAR) model-based forecasts of Baumeister and Kilian (2012) for both averaged and end-of-period prices.

Following Baumeister and Kilian (2012), the BVAR/VAR models include four oil-market variables: oil prices, the growth rate of global oil production, a proxy for the change in global crude oil inventories, and the REA as a measure of global economic activity. The estimation (including priors for the Bayesian models) follows Baumeister and Kilian (2012). Accordingly, we consider models with 12 autoregressive lags, and all models are estimated with prices in log real levels and forecasts are converted to real price levels.[10] Our exhibits focus on forecasts of WTI crude oil prices (similar results are obtained for Brent prices) and the RAC.

Table 3 reports the forecast performances of the (B)AR and (B)VAR models for the real price of WTI crude oil for the full sample going to 2021. In the updated sample, the models estimated with average prices do not significantly improve upon the monthly average-price no-change forecast.

[10] Similar results are obtained for models with 24 lags.

Table 3. Forecasts of the Real Price of WTI Crude Oil from Key Models of Baumeister and Kilian (2012) Estimated with Average Prices and End-of-Month Prices

Method	Monthly Average Prices				Period-End Price Sampling			
Model	BVAR	VAR	AR	BAR	BVAR	VAR	AR	BAR
Horizon	MSPE Ratio							
1	0.97 (0.374)	1.05 (0.633)	0.93 (0.239)	0.92 (0.147)	0.62 (0.004)	0.74 (0.083)	0.58 (0.002)	0.57 (0.001)
3	1.04 (0.630)	1.05 (0.630)	0.98 (0.440)	0.97 (0.359)	0.94 (0.299)	0.91 (0.344)	0.87 (0.126)	0.86 (0.081)
6	1.08 (0.847)	1.15 (0.916)	1.02 (0.565)	1.00 (0.503)	1.04 (0.657)	0.99 (0.464)	0.94 (0.271)	0.92 (0.168)
12	1.15 (0.918)	1.28 (0.987)	1.05 (0.632)	1.00 (0.500)	1.13 (0.813)	1.11 (0.793)	0.96 (0.362)	0.92 (0.190)
24	1.12 (0.776)	1.34 (0.946)	1.13 (0.728)	0.98 (0.455)	1.07 (0.694)	1.03 (0.570)	0.99 (0.470)	0.91 (0.237)
	Success Ratio							
1	0.46 (0.885)	0.45 (0.952)	0.50 (0.519)	0.48 (0.730)	0.70 (0.000)	0.69 (0.000)	0.72 (0.000)	0.72 (0.000)
3	0.53 (0.208)	0.50 (0.414)	0.48 (0.745)	0.51 (0.274)	0.60 (0.001)	0.58 (0.005)	0.57 (0.008)	0.61 (0.000)
6	0.54 (0.188)	0.53 (0.152)	0.49 (0.641)	0.49 (0.594)	0.60 (0.004)	0.61 (0.001)	0.54 (0.191)	0.53 (0.258)
12	0.55 (0.164)	0.54 (0.139)	0.55 (0.175)	0.53 (0.288)	0.57 (0.064)	0.61 (0.002)	0.55 (0.196)	0.53 (0.222)
24	0.51 (0.377)	0.53 (0.145)	0.59 (0.079)	0.58 (0.077)	0.54 (0.227)	0.59 (0.019)	0.59 (0.060)	0.59 (0.048)

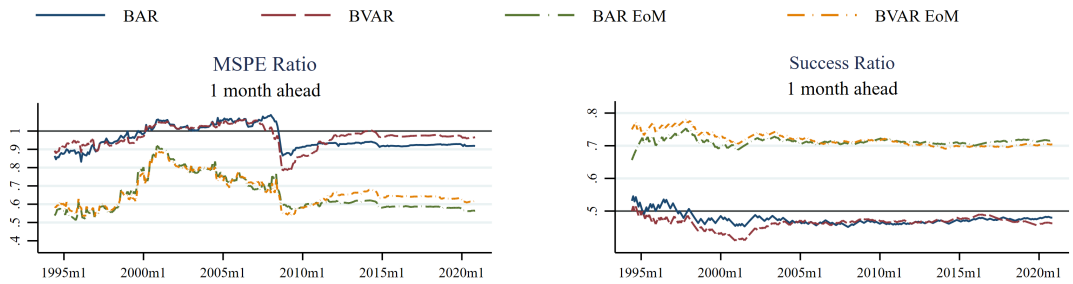
Note: Real-time, out-of-sample-forecasts of the real price of WTI crude oil, 1992M1–2021M1. All models are estimated with 12 autoregressive lags. The brackets show the p-values serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs relative to the monthly average-price no-change and of Pesaran and Timmermann (2009) for the null of no directional accuracy. Boldface values indicate improvements over the random-walk no-change forecast at the five percent level.

This deterioration in forecast accuracy is consistent with evidence from other recent studies (Funk, 2018; Snudden, 2018; Baumeister et al., 2020). The multivariate models have larger MSPEs than the average-price no-change forecast at all forecast horizons and exhibit only small, insignificant improvements in terms of directional accuracy. The (B)AR models tend to perform better, although significant improvements can only be found in directional accuracy at the 10 percent level at the 24-month horizon. Figure 2 depicts the evolution of the forecast criteria over time and shows that deterioration in the MSPE ratios, particularly for the BVAR, occurs mostly during the years 2010 to 2014.

By contrast, the models estimated with end-of-period prices outperform the same models estimated with monthly average prices in almost all cases, and significantly so at short horizons. As shown in Figure 2, the improvements in forecast-accuracy at the 1-step ahead prediction forecast are of the order of 40 percent and robust to all forecast evaluation sample ending dates between 1995 and 2021. Even during the peak performance of the BVAR model around 2008, forecast improvements as large as 40 percent could have been realized by estimating the model with end-of-period prices rather than monthly average prices.

The RAC forecasts exhibit a similar pattern. Figure 3 shows that estimating models with the quasi-end-of-month observations instead of the average RAC yield forecast improvements of over 30 percent, which are again robust for all sample end dates between 1995 and 2021. This highlights

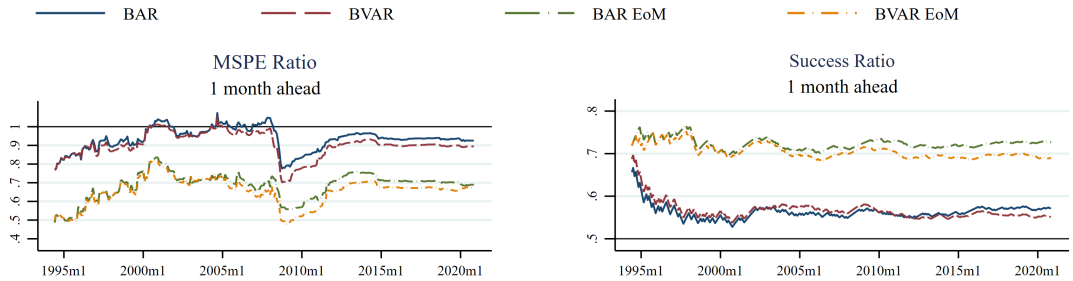
Figure 2. Recursive Forecasts of the Real Price of WTI crude oil from Baumeister and Kilian (2012)'s Models Estimated with End-of-Month and Average Prices



Note: Dynamic, recursive, out-of-sample forecasts of the real price of WTI crude oil, 1992M1–2021M1. Models are estimated with 12 lags. "EoM" refers to models estimated with quasi-end-of-month prices. Forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. Success ratios represent the fraction of times the forecast correctly predicts the direction of change in the real price of oil. The first 30 months are dropped to reduce starting-point effects.

that even in the case where the end-of-periods observations are not available, approximating such observations with simple imputation methods can produce large and robust forecast gains in the context of multivariate models.

Figure 3. Recursive Forecasts of the RAC from Baumeister and Kilian (2012)'s Models Estimated with End-of-Month and Average Prices



Note: Dynamic, recursive, out-of-sample forecasts of the RAC, 1992M1–2021M1. Models are estimated with 12 lags. "EoM" refers to models estimated with quasi-end-of-month prices. Forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. Success ratios represent the fraction of times the forecast correctly predicts the direction of change in the real price of oil. The first 30 months are dropped to reduce starting-point effects.

Our results suggest that the PEPS approach is a simple yet effective way of improving forecast accuracy for a wide variety of forecast approaches. However, as can be noted by the absence of boldfaced values, even with PEPS none of the model-based forecasts significantly improve upon the random walk no-change benchmark. This again highlights how important it is to compare model-based forecasts against the higher standard of the random-walk no-change forecast.

To further illustrate this point, Table 4 shows the performance of the multivariate models preferred by Baumeister and Kilian (2012) over their original sample period covering 1992M1 to

2010M6. Using the updated REA slightly deteriorates the forecast performance of the (B)VARs reported in Baumeister and Kilian (2012), but our replication still supports their conclusion that model-based forecasts of the real price of WTI crude oil and the RAC are more accurate than the average-price no-change forecast at shorter horizons.

Table 4. Replication of Baumeister and Kilian (2012) and Comparison to Random-Walk No-Change Forecast

	WTI			RAC		
Model	RW (EoM)	BVAR	VAR	RW (EoM)	BVAR	VAR
Horizon	MSPE Ratio					
1	0.64 (0.005)	0.86 (0.201)	0.85 (0.237)	0.69 (0.011)	0.79 (0.147)	0.75 (0.145)
3	0.88 (0.047)	0.90 (0.224)	0.86 (0.281)	0.88 (0.044)	0.87 (0.194)	0.81 (0.223)
6	0.94 (0.035)	1.05 (0.753)	1.05 (0.654)	0.94 (0.030)	1.07 (0.824)	1.05 (0.646)
12	0.96 (0.046)	1.17 (0.957)	1.16 (0.944)	0.95 (0.027)	1.20 (0.975)	1.17 (0.944)
	Success Ratio					
1	0.71 (0.000)	0.47 (0.896)	0.51 (0.332)	0.70 (0.000)	0.57 (0.031)	0.57 (0.011)
3	0.62 (0.001)	0.58 (0.091)	0.56 (0.101)	0.64 (0.000)	0.57 (0.047)	0.58 (0.019)
6	0.58 (0.028)	0.56 (0.326)	0.53 (0.367)	0.60 (0.008)	0.55 (0.217)	0.53 (0.263)
12	0.58 (0.031)	0.57 (0.191)	0.59 (0.037)	0.58 (0.049)	0.53 (0.411)	0.61 (0.010)

Note: Replication of Baumeister and Kilian (2012)'s real-time, out-of-sample-forecasts of WTI crude oil prices and the RAC for the original sample period, 1992M1–2010M6. The models are estimated with 12 autoregressive lags. *RW(EoM)* stands for the random walk forecasts based on end-of-month prices. The brackets show the p-values serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs relative to the monthly average-price no-change and of Pesaran and Timmermann (2009) for the null of no directional accuracy. Boldface values indicate improvements over the random-walk no-change forecast at the five percent level.

The standard interpretation of this short-run predictability is that model-based forecasts are preferable over naive forecasts and that oil prices behave differently from the prices of other assets such as stocks and exchange rates (Baumeister and Kilian, 2012). However, short-run predictability relative to the conventional no-change forecast can also be a feature arising from temporally aggregating time-series that are originally unpredictable (Ellwanger and Snudden, 2021b). In fact, the direct comparison with the new benchmark, also displayed in Table 4, shows that the original Baumeister and Kilian (2012) models performed no better (and often considerably worse) than the random walk forecast from daily prices at forecast horizons up to one year. This shows that contrary to common wisdom, the original results of Baumeister and Kilian (2012) do not provide evidence in favor of model-based approaches, nor against the idea that oil prices have a strong asset-price component. Instead, they are generally consistent with a random walk model for oil prices.

3.4 Period-End Price Sampling and Covariances

Given that averaging alters the dynamics and serial correlation of a series (see, e.g. Tiao, 1972; Rossana and Seater, 1995), it seems plausible that end-of-period sampling affects established properties of model features. Consequently, this section investigates how the use of PEPS affects the lag length selection and the predictive content of additional regressors in forecast models of the real price of oil.

Table 5. Percent of Sample End-Points at which the Forecasts from the AR(2) Outperforms the Forecasts from the AR(12) Model

Method	Horizon	1	3	6	12	24
EoM Prices	MSPE Ratio	94%	91%	90%	100%	100%
	Success Ratio	74%	100%	46%	100%	82%
Average Prices	MSPE Ratio	92%	48%	46%	49%	72%
	Success Ratio	96%	63%	71%	25%	100%

Note: Dynamic, recursive, out-of-sample forecasts of WTI crude oil prices, 1992M1–2021M1. The first 30 months dropped to reduce starting-point effects. The table displays the percent of sample end-points at which the forecasts from the AR(2) outperforms the forecasts from the AR(12) model. *EoM Prices* stands for autoregressive model estimated with end-of-month prices, *Average Prices* for models estimated average prices.

To tease out the effect of PEPS on forecast performance arising from changes in serial correlation, we focus on univariate AR models. The relative performance of recursively estimated models with 2 and 12 autoregressive lags is illustrated for forecasts of the real price of WTI crude oil. Because the relative forecasting performance of models changes over time, Table 5 reports the percent of sample end-points which favors the AR(2) model. Previous studies advocate for the use of 12 or 24 lags when estimating autoregressive models (Alquist and Kilian, 2010; Baumeister and Kilian, 2012). However, without aggregation, the number of autoregressive lags may be lower. As can be seen in Table 5, the MSPE criterion and, albeit to a somewhat lesser extent, the success ratio, heavily favor the AR(2) model over the AR(12) model under the PEPS approach. By contrast, under the conventional estimation with average price, there is no clear evidence for either a longer or a shorter lag length. While the AR(2) performs well at the 1-step-ahead and the 24-month-ahead prediction, there is no clear favorite for medium term forecasts, with AR(12) models often outperforming the AR(2). Overall, these results suggest that the PEPS reduces the number of optimal autoregressive lags in forecasting models.

To explore how co-variances are affected by PEPS, we also compare the forecast performance of VAR models estimated with alternative measures of economic activity and oil inventories. For real activity indexes, we consider four alternative measures to the REA: the composite leading indicator

of the OECD and six major non-OECD countries (CLI); the Global Economic Conditions (GCON) indicator from Baumeister et al. (2020); and World Industrial Production (WIP). For inventories, we consider the unscaled U.S. petroleum inventories as an alternative to the standard proxy of global crude oil inventories, which is obtained by scaling U.S. crude oil inventories by the ratio of OECD petroleum stocks to US petroleum stocks (Kilian and Murphy, 2014; Baumeister and Kilian, 2012). The VAR models are estimated with 2 lags, which produces more accurate short-horizon forecasts than models with longer lag lengths.

Table 6. VAR-Model Forecasts of the Real Price of WTI Crude Oil with Different Predictors under the PEPS approach

Activity Index	REA	WIP	CLI	GCON	REA	WIP	CLI	GCON
Inventories	Global Crude Oil				U.S. Petroleum			
Horizon	MSPE Ratio							
1	0.62 (0.004)	0.55 (0.002)	0.56 (0.002)	0.57 (0.005)	0.62 (0.003)	0.54 (0.002)	0.56 (0.001)	0.55 (0.004)
3	0.95 (0.304)	0.81 (0.112)	0.82 (0.102)	0.82 (0.145)	0.96 (0.337)	0.82 (0.109)	0.83 (0.101)	0.83 (0.139)
6	1.11 (0.773)	0.88 (0.217)	0.86 (0.153)	0.86 (0.220)	1.12 (0.782)	0.88 (0.214)	0.86 (0.153)	0.87 (0.222)
12	1.30 (0.905)	0.90 (0.248)	0.86 (0.142)	0.88 (0.221)	1.30 (0.908)	0.91 (0.268)	0.86 (0.158)	0.89 (0.241)
24	1.25 (0.875)	0.95 (0.375)	0.90 (0.270)	0.95 (0.394)	1.25 (0.872)	0.96 (0.409)	0.91 (0.290)	0.97 (0.421)
	Success Ratio							
1	0.69 (0.000)	0.71 (0.000)	0.71 (0.000)	0.70 (0.000)	0.71 (0.000)	0.72 (0.000)	0.70 (0.000)	0.70 (0.000)
3	0.57 (0.012)	0.58 (0.003)	0.61 (0.000)	0.59 (0.003)	0.56 (0.017)	0.58 (0.003)	0.61 (0.000)	0.58 (0.003)
6	0.57 (0.076)	0.55 (0.094)	0.56 (0.077)	0.55 (0.112)	0.56 (0.114)	0.54 (0.168)	0.57 (0.050)	0.55 (0.163)
12	0.55 (0.205)	0.54 (0.215)	0.53 (0.261)	0.55 (0.181)	0.55 (0.207)	0.54 (0.238)	0.53 (0.324)	0.55 (0.182)
24	0.55 (0.239)	0.60 (0.045)	0.60 (0.047)	0.61 (0.030)	0.56 (0.189)	0.60 (0.045)	0.60 (0.047)	0.61 (0.034)

Note: Real-time, out-of-sample-forecasts of the real price of WTI crude oil, 1992M1–2021M1. All models are estimated with two lags and end-of-month oil-price data. The table presents results for VAR models with different combinations of measures of global economic activity and changes in oil inventories. The brackets show the p-values serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs relative to the monthly average-price no-change and of Pesaran and Timmermann (2009) for the null of no directional accuracy. Boldface values indicate improvements over the random-walk no-change forecast at the five percent level.

Both Snudden (2018) and Funk (2018) find evidence that in VAR models, including changes in U.S. petroleum inventories provides superior forecasts to including changes in Kilian and Murphy (2014)'s proxy for global crude oil inventories. However, in the updated sample and with period-end price sampling, the forecast accuracy is consistently very similar across both inventory measures (Table 6). Regarding the economic activity measures, recent studies have advocated to use the WIP in VAR models (Funk, 2018; Snudden, 2018; Baumeister et al., 2020). Table 6 shows that similar to the case of oil inventories, the forecast performance of models estimated with different economic activity is almost indistinguishable across the GCON, CLI and WIP measures of global economic activity. Only the corrected measure of the REA stands out, as models estimated with

the REA perform worse at both short and longer horizons. Taken together, these exercises suggest that the choice of alternative predictor variables might become less relevant when end-of-month prices are used in forecast models.

Overall, the performance of models estimated with the PEPS approach is impressive relative to models estimated with average prices. However, none of the forecast models explored can outperform the random-walk no-change benchmark at the 5 percent level at any forecast horizon. The general message from this section is that established findings for models estimated with average prices need to be re-evaluated when information on disaggregated oil prices is incorporated into oil-price forecasts.

3.5 Forecast Combinations

Despite the large forecast gains from using the PEPS approach, the individual models considered so far are unable to outperform the random-walk no-change forecast at horizons less than two years. However, Ellwanger and Snudden (2021a) show that futures-based forecasts based on end-of-period futures prices can outperform the random-walk no-change at horizons of one year and beyond. Further, forecast combination approaches can often improve oil-price forecasts, as has been documented in previous studies that examined models estimated with average prices (Baumeister et al., 2014; Baumeister and Kilian, 2015; Garratt et al., 2019). In this section, we investigate whether pooling model-based and futures-based forecasts computed with end-of-period data can further improve forecast accuracy.

Several model-based forecast combinations are considered. First, we compute an equal-weighted combination of four VAR models, estimated with the four different global activity indexes: REA, CLI, WIP, and GCON. Second, we also consider an equal-combination of the corresponding BVAR forecasts. Third, we consider an equal-weighted combination of four univariate models: AR, ARMA, AR-ARCH and BAR.

We then consider three further forecast combinations involving a larger set of models and information from oil futures prices. First, we consider an equal weighted forecast combination across all multivariate and univariate models described in the previous paragraph. We then add the futures-based forecast to the equal weighted forecast with the models. Finally, given the previously documented success of the futures-based forecasts, we also consider a forecast combination that gives half the weight to the futures-based forecasts and half the weight to an equal-weighted combination of all model-based forecasts.

All models are estimated with end-of-period observations following the PEPS approach described in the previous sections. Consistent with our previous results, the (B)VAR models are estimated with two lags.[11] The univariate models are estimated with the same lag length as in the previous sections, (B)AR(2), ARMA(1,1) and AR-ARCH(2,1). Futures-based forecasts are constructed using direct forecasts of the end-of-month futures prices following Ellwanger and Snudden (2021a).

Table 7. Futures-based Forecasts and Forecast Combinations for the Real Price of WTI Crude Oil

Forecasts	EoM	Futures	4 VARs	4 BVARs	4 Univar.	All 12 Models	Models and Futures	1/2 Models, 1/2 Futures
Horizon				MSPE Ratio				
1	0.59 (0.000)	0.59 (0.000)	0.55 (0.001)	0.55 (0.001)	0.56 (0.000)	0.55 (0.001)	0.55 (0.001)	0.55 (0.000)
3	0.89 (0.012)	0.87 (0.015)	0.83 (0.086)	0.83 (0.084)	0.84 (0.066)	0.83 (0.071)	0.83 (0.061)	0.83 (0.025)
6	0.95 (0.029)	0.91 (0.046)	0.89 (0.143)	0.90 (0.117)	0.89 (0.134)	0.88 (0.123)	0.88 (0.100)	0.87 (0.034)
12	0.96 (0.028)	**0.83 (0.011)**	0.92 (0.242)	0.93 (0.172)	0.87 (0.122)	0.90 (0.171)	0.88 (0.120)	**0.82 (0.013)**
24	0.99 (0.189)	**0.79 (0.034)**	0.95 (0.387)	0.93 (0.292)	0.87 (0.161)	0.91 (0.291)	0.88 (0.218)	**0.78 (0.036)**
				Success Ratio				
1	0.71 (0.000)	0.72 (0.000)	0.71 (0.000)	0.71 (0.000)	0.71 (0.000)	0.71 (0.000)	0.70 (0.000)	0.71 (0.000)
3	0.61 (0.000)	0.60 (0.001)	0.60 (0.000)	0.61 (0.000)	0.60 (0.000)	0.59 (0.001)	0.60 (0.000)	0.62 (0.000)
6	0.55 (0.057)	0.56 (0.048)	0.58 (0.025)	0.57 (0.039)	0.52 (0.311)	0.58 (0.019)	0.58 (0.028)	0.56 (0.060)
12	0.57 (0.007)	**0.65 (0.000)**	0.58 (0.053)	0.56 (0.088)	0.54 (0.230)	0.55 (0.145)	0.57 (0.065)	0.58 (0.035)
24	0.52 (0.292)	**0.64 (0.000)**	0.63 (0.015)	**0.58 (0.077)**	0.60 (0.041)	0.62 (0.018)	0.62 (0.017)	**0.59 (0.033)**

Note: Real-time, out-of-sample-forecasts of the real price of WTI crude oil, 1992M1–2021M1. "All 12 Models" is an equal-weighted forecast of the 4 univariate (Univar.), the 4 BVAR, and the 4 VAR forecasts. The brackets show the p-values serial-dependence-robust tests of Diebold and Mariano (1995) for the null hypothesis of equal MSPEs relative to the monthly average-price no-change and of Pesaran and Timmermann (2009) for the null of no directional accuracy. Boldface values indicate improvements over the random-walk no-change forecast at the five percent level

For reference, the first two columns of Table 7 replicate the performance of the random walk no-change forecast and Ellwanger and Snudden (2021a)'s futures-based forecasts, respectively. Despite the exceptional forecast accuracy of the futures-based forecast relative to the monthly average no-change forecast, it only significantly outperform the random-walk no-change forecast at the one- and two-year horizon.

The combination of the four VAR, four BVAR, and four univariate models is displayed in columns 4 to 6, respectively. Although the model-based forecast combinations perform better than individual models and generally exhibit lower MSPEs than the random walk no-change forecast, the differences are relatively small and not statistically significant. Only the two-year-ahead forecast for the combination of the four BVARs significantly improve upon the random-walk no-change forecast for the success ratio. Similar results are obtained for the equal-weighted combination of

[11] The results in this section are qualitatively unchanged for (B)VAR models with longer lags.

all 12 models and the equal-weighted combination of all 12 models and the futures-based forecast, displayed in columns 7 and 8.

Finally, the last column of Table 7 show that a forecast combination that puts a larger weight on the futures-based forecast performs best. Its forecast is significantly better than the random walk no-change at longer forecasts horizons, with MSPE reductions of up to 21% at the 2-year-ahead forecast. Still, much of these gains in performance seems to be driven by the futures-based forecast, and the combination actually deteriorates the success ratio of the futures-based forecast at longer horizons.

Taken together, we find that combining forecasts and incorporating information from futures can improve the accuracy of individual forecasts. The results mirror existing findings for models estimated with average prices (Baumeister et al., 2014; Baumeister and Kilian, 2015). However, the gains from pooling model-based forecasts and including futures prices occur at longer-term forecast horizons. The inability of futures and models to significantly outperform the random-walk no-change forecasts at shorter forecast horizons suggests again that oil prices are closer to asset prices than assumed in the previous literature.

4 Extensions and Robustness

The results presented in the main section of the paper are remarkably robust to alternative modelling choices. It can be shown that using ex-post revised data instead of real-time data does not affect our main results.[12] Moreover, our results are robust to alternative methods of deflating the end-of-period and daily prices using the CPI deflator, as well as alternative methods of nowcasting the CPI. This result is expected, as fluctuations in the CPI deflator are generally small compared to the fluctuations in nominal oil prices.

The main results not only hold for real prices, but also for monthly average nominal prices. The relative forecast gains of using the nominal closing price vis-à-vis the average nominal price are similar to those documented for real oil prices. This result is consistent with the idea that fluctuations in nominal oil prices, rather than fluctuations in inflation, drive our results.

The closing-price no-change forecast is also superior to the average-price no-change forecast for quarterly and annual data, which are of primary interest to policymakers (Baumeister and Kilian, 2014, 2015). Consistent with the theoretical predictions in Ellwanger and Snudden (2021b), the

[12] For brevity, we do not provide additional exhibits for this section. Tables and figures are available upon request.

improvements in the MSPE ratio at the one-step-ahead prediction are even larger at the quarterly and annual frequency compared to the monthly frequency.

5 Conclusion

Model-based forecasts of the real price of crude oil have traditionally been computed from monthly average data. We have shown how the use of daily oil-price data can provide large improvements over the conventional practice. Univariate models that are estimated with daily or end-of-month observations yield forecast improvements of up to 40% over traditional forecasts. Similar forecasting gains are obtained from classical and Bayesian VARs that are estimated with end-of-month observations. The results hold for both WTI and Brent crude oil prices as well as for U.S. refinery acquisition costs of crude oil, for which pseudo disaggregated data can be imputed.

Our forecasts provide unprecedented improvements over the conventional monthly no-change forecast computed from average data. However, a key insight from our paper is that this standard forecast comparison can be misleading. The random-walk benchmark computed from end-of-month prices is shown to raise the bar for model-based forecasts to claim practical usefulness.

Despite the impressive gains relative to the conventional benchmark, models estimated with disaggregated data generally do not significantly outperform the random-walk forecast. We have also shown that the new benchmark overturns influential results of the previous literature that was thought to favor model-based forecasts over naive forecasts. Overall, our results demonstrate that information from daily oil prices is crucial not only for implementing accurate model-based forecasts of the real price of oil, but also for properly evaluating the merit of new and existing forecast approaches.

References

Alquist, R. and Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25(4):539–573.

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., Guérin, P., and Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? The midas touch at work. *International Journal of Forecasting*, 31(2):238–252.

Baumeister, C. and Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, 109(5):1873–1910.

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. (2014). What central bankers need to know about forecasting oil prices. *International Economic Review*, 55(3):869–889.

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. (2020). Energy markets and global economic conditions. *NBER Working Paper No. 27001.*

Bork, L., Rovira Kaltwasser, P., and Sercu, P. (2019). Can exchange rates forecast commodity prices? A reconsideration. *Mimeo.*

Conlon, T., Cotter, J., and Eyiah-Donkor, E. (2022). The illusion of oil return predictability: The choice of data matters! *Journal of Banking & Finance*, 134:106331.

Degiannakis, S. and Filis, G. (2018). Forecasting oil prices: High-frequency financial data are indeed useful. *Energy Economics*, 76:388–402.

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. (2021a). Futures prices are useful predictors of the spot price of crude oil. *LCERPA Working Paper 2021-4*.

Ellwanger, R. and Snudden, S. (2021b). Predictability of aggregated time series. *LCERPA Working Paper No. 2021-5*.

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.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric reviews*, 26(1):53–90.

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.

Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3):454–478.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, 18(3):337–349.

Lütkepohl, H. (1984). Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business & Economic Statistics*, 2(3):201–214.

OECD (2022). Composite leading indicator (CLI). doi: 10.1787/4a174487-en.

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

Rossana, R. J. and Seater, J. J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, 13(4):441–451.

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

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59(3):525–531.

Wang, Y., Liu, L., Diao, X., and Wu, C. (2015). Forecasting the real prices of crude oil under economic and statistical constraints. *Energy Economics*, 51:599–608.

Wei, W. W. (1978). Some consequences of temporal aggregation in seasonal time series models. In *Seasonal analysis of economic time series*, pages 433–448. NBER.

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

Ye, M., Zyren, J., and Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, 21(3):491–501.

Yin, L. and Yang, Q. (2016). Predicting the oil prices: do technical indicators help? *Energy Economics*, 56:338–350.

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

Zhang, Y.-J. and Wang, J.-L. (2019). Do high-frequency stock market data help forecast crude oil prices? Evidence from the midas models. *Energy Economics*, 78:192–201.

Online Appendix (Not for publication)

A1 Real-Time Data

This section describes the construction of the real-time data used in the empirical exercises.

Oil market variables: The real-time data for the nominal U.S. refiner acquisition cost for crude-oil imports, world crude-oil production, U.S. crude-oil inventories, U.S. petroleum inventories, and OECD petroleum inventories are obtained from the EIA. Real-time vintages start in 1973M1 and use the vintages of Baumeister and Kilian (2012) from 1991M12 to 2010M12. Vintages from 2011M01 onwards are collected using the same data sources. [13] The updated real-time data is similar to Garratt et al. (2019) but differs primarily in that all historical releases, and thus data revisions, are included.

Other economic variables: For the construction of the Real Economic Activity index we use Kilian (2009)'s freight-rate data until 1984M12 and, thereafter, monthly values of the Baltic Dry Index obtained from Bloomberg. The index is computed using the corrected formula (Kilian, 2019). As in Baumeister and Kilian (2012), the real-time data for the monthly seasonally adjusted U.S. consumer price index for all urban consumers is obtained from the Economic Indicators published by the Council of Economic Advisers from the FRASER database of the Federal Reserve Bank of St. Louis and from the macroeconomic real-time database of the Federal Reserve Bank of Philadelphia. The real-time composite leading indicator (CLI) of the OECD and six major non-OECD countries is obtained from the Organization for Economic Co-Operation and Development (OECD, 2022). CLI is available with a one-month delay until the 2006M05 vintage; vintages before that date use the 2006M05 vintage and assume a one-month delay. Pseudo real-time data is constructed using the 2021M6 vintages of the Global Economic Conditions (GCON) indicator (Baumeister et al., 2020) and World Industrial Production (WIP) of the OECD and 6 major other countries (Baumeister and Hamilton, 2019), assuming a one and two month data delay, respectively.

Crude oil prices: Crude-oil price data for WTI and Brent are obtained from the EIA. Daily closing prices for WTI and Brent are used to calculate both the monthly average and end-of-month prices and are obtained from EIA Petroleum and other liquids. The monthly closing price is the closing price on the last trading day of the month. Prices are updated daily by the EIA in "Today in Energy." The data for the previous day is updated between 7:30 a.m. and 8:30 a.m. EST and is

[13] This real-time database is updated monthly and is publicly available, along with further documentation on the corresponding author's website.

not subject to revisions, which justifies treating both the end-of-month and monthly average data as observed in real time.

The series of daily WTI and Brent prices provided by the EIA begin in 1986M1 and 1987M5, respectively. The monthly closing WTI price is backcasted from 1986M1 to 1983M4 using the growth rate in the monthly closing price of the one-month-ahead futures contract. All prices before 1983M4 are backcasted using the growth rate of the RAC. Our main findings are robust to estimating the models using end-of-month prices beginning in 1986M1 or 1983M4.

Table A1. Descriptive Statistics for the Real Prices of Crude Oil

Monthly Average	Date Range	Mean	Std. Dev.	Min	Max
U.S. Refiner Acquisition Cost, Imported	1973m1 - 2021m1	24.03	12.31	5.95	60.85
Brent	1973m1 - 2021m1	23.50	10.67	6.08	61.56
West Texas Intermediate	1973m1 - 2021m1	22.63	11.09	5.71	58.34
End of Month	**Date Range**	**Mean**	**Std. Dev.**	**Min**	**Max**
U.S. Refiner Acquisition Cost, Imported	1973m1 - 2021m1	24.07	12.37	5.76	63.64
Brent	1973m1 - 2021m1	22.27	10.23	5.02	64.36
West Texas Intermediate	1973m1 - 2021m1	21.39	10.51	5.13	60.68

Note: All series are the level of real prices from 1973M1–2021M5. 2021M5 data vintage. The end-of-month price for the U.S. refiner acquisition cost (imported) is imputed using the WTI price. Nominal crude-oil price data obtained from the Energy Information Administration and the consumer price index is obtained from the Federal Reserve of Philadelphia.

Nowcasting

The CPI series, the global crude-oil market variables used in the VARs and the refiners import price of crude oil are subject to historical revisions and are reported with a lag. For each vintage, observations that are missing due to reporting lags are nowcasted following the standard approaches (Baumeister and Kilian, 2012; Funk, 2018; Baumeister et al., 2020):

- Missing observations for global crude-oil production, WIP, GCON, CLI, and the U.S. CPI, and U.S. crude-oil inventories are nowcasted by extending the series by using the average of the historical growth rate at the respective point in time.
- Missing observations for the ratio of OECD petroleum inventories to U.S. petroleum inventories are kept constant at the last available value for this ratio.
- Monthly nominal U.S. crude oil imported acquisition cost by refiners are extrapolated with the growth rate of the monthly average of the nominal WTI price.