Staff Working Paper/Document de travail du personnel — 2020-39

Last updated: September 22, 2020

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

by Amor Aniss Benmoussa[1] Reinhard Ellwanger[1] and Stephen Snudden[2]

[1]International Economic Analysis Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9

[2]Department of Economics
Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 3C5

abenmousa@bankofcanada.ca, rellwanger@bankofcanada.ca, ssnudden@wlu.ca

ISSN 1701-9397

Acknowledgements

We thank Lutz Kilian, Ron Alquist, Christiane Baumeister, Edouard Djeutem, Konrad Zmitrowicz, Pierre Siklos, Wing Chan and Willy Chetwin for detailed discussions, comments and suggestions.

Abstract

We propose a new no-change benchmark to evaluate forecasts of series that are temporally aggregated. The new benchmark is the last high-frequency observation and reflects the null hypothesis that the underlying series, rather than the aggregated series, is unpredictable. Under the random walk null hypothesis, using the last high-frequency observation improves the mean squared prediction errors of the no-change forecast constructed from average monthly or quarterly data by up to 45 percent. We apply this insight to forecasts of the real price of crude oil and show that a new benchmark that relies on monthly closing prices dominates the conventional no-change forecast in terms of forecast accuracy. Although model-based forecasts also improve when models are estimated using closing prices, only the futures-based forecast significantly outperforms the new benchmark. Introducing a more suitable benchmark changes the assessments of different forecasting approaches and of the general predictability of real oil prices.

Bank topic: Econometric and statistical methods; International topics
JEL codes: C1, C53, Q47

1 Introduction

It is common practice to evaluate the performance of a forecast by comparing its accuracy against that of a simple no-change forecast. Implicit in this comparison is the null hypothesis that the series of interest follows a random walk. Improvements over the no-change benchmark permit forecasters to claim that the series is predictable in general and that their specific approach is more accurate than a naive forecast.

The above approach is appealing when the target variable is thought to follow a random walk that occurs at the same frequency as the forecasters' data. However, this is not the case for many macroeconomic applications that feature temporally aggregated time series.[1] The random walk hypothesis typically implies that temporally aggregated data do not follow a random walk and are instead predictable by construction. The most famous example of this effect is time-averaging, which converts an underlying random walk into the cumulative sum of a moving-average process (Working, 1960). In this setting, improvements over the conventional no-change forecast are generally uninformative about the predictability of the underlying series or the practical usefulness of a specific forecasting approach.

Under the random walk null hypothesis, the theoretically optimal forecast of all future observations is the value of the last high-frequency observation. Consequently, the benchmark that is used to evaluate forecasts should be based on this high-frequency observation rather than on the last aggregate observation. As we show in this paper, the difference in the forecast accuracy between these two no-change benchmarks can be sizeable. When the underlying series follows a random walk, the theoretical improvements in the one-step-ahead mean squared prediction error (MSPE) are larger than 45 percent when using the last observed value instead of the last observed monthly or quarterly average value. A simple change of the benchmark can thus have quantitatively large effects on assessments of different forecasting models.

We investigate the relevance of these insights in the case of forecasts of the real price of crude oil, a key global macroeconomic indicator that has attracted considerable attention among forecasters. This application suits the theoretical arguments laid out in this paper for two reasons. First, oil prices are highly persistent and known to be at least partially determined by the forward-looking behaviour of market participants, which makes the random walk assumption a plausible null hypothesis for forecast evaluations. Second, although daily prices for several global crude-

[1] As in Rossana and Seater (1995), we use temporal aggregation to refer to averaging and summation.

oil benchmarks are available at least since the 1980s, the standard real series that are used in numerous structural models and forecasting applications for the global oil market are based on monthly average prices (see, for example, Kilian, 2009; Alquist et al., 2013; Kilian and Murphy, 2014; Baumeister and Hamilton, 2019). An increasing number of studies document that model-based forecasts of the real price of crude oil outperform the average-price no-change benchmark (see, for example, Baumeister and Kilian, 2012; Alquist et al., 2013; Baumeister et al., 2014; Baumeister and Kilian, 2014, 2015; Snudden, 2018; Funk, 2018; Garratt et al., 2019). Based on the insights from a simple random walk model, we construct a new benchmark from monthly closing prices and revisit the claim that these models can predict the real price of crude oil. We also investigate the extent to which the use of closing prices can help improve model-based forecasts more generally.

The main empirical result from these exercises is that replacing average prices with closing prices considerably improves traditional forecasting models for the real price of oil. A simple no-change forecast that is based on the last closing price reduces the MSPE of the conventional no-change forecast that is computed from average monthly prices by 40 percent for one-month-ahead forecasts. The directional accuracy for the one-month-ahead forecasts is higher than 70 percent. The gains decrease with the forecast horizon but are still apparent up to the 12-months-ahead forecast. We also show that estimating forecasting models of the price of oil with the monthly closing price rather than the conventional monthly average price yields similarly large improvements in forecast accuracy.

A central message of this paper is that the no-change forecast that is based on the last high-frequency observation should be used to evaluate the predictability of a variable of interest. Since the new benchmark for the real price of oil has a higher forecast accuracy, it becomes more difficult to establish improvements in forecast accuracy. For example, forecasts from several models that are estimated using closing prices significantly outperform the conventional no-change benchmark that is based on average prices, especially at shorter horizons. However, with the exception of long-horizon forecasts that are based on futures prices, the same forecasts do not outperform the new no-change benchmark. The real price of oil is more difficult to predict and, in that respect, is closer to asset prices than is implied by conventional no-change forecasts.

The gains in forecast accuracy from using closing prices exceed the typical gains from introducing new models or predictor variables that have been proposed to forecast the real price of oil. To put a 40 percent reduction in the MSPE ratio into perspective, existing studies that have advocated for the use of forecast combinations often find MSPE improvements of 12 percent or less from the forecast-combination approach (Baumeister and Kilian, 2015; Garratt et al., 2019). Moreover, the gains from using closing prices are consistent over time and are less dependent on the sample period than are gains from introducing new models or new predictors. Although these improvements are remarkable, their magnitudes and patterns over different forecast horizons are consistent with theoretical results on temporal aggregation.[2] Our findings highlight that, empirically, averaging is indeed a first-order issue for forecasts of macroeconomic variables.

Our baseline estimation relies on a series of monthly closing prices that are converted to real terms and then used in no-change forecasts and to estimate several workhorse forecasting models.[3] The setup differs from the typical setup in the literature on temporal aggregation, which often compares the forecasting performance of pure time-series models at various data frequencies (see, for example, Tiao, 1972; Rossana and Seater, 1995). Instead, we estimate models at the same monthly frequency at which they were originally introduced. This setup addresses the challenge posed by the fact that standard price deflators are only available on a monthly frequency as it allows us to use the same deflator for monthly closing and monthly average prices. Moreover, it facilitates the construction of real-time forecasts, which have become the relevant standard (Baumeister and Kilian, 2012). For our empirical exercises, we update a real-time dataset created by Baumeister and Kilian (2012) and use it to estimate models that rely on real-time oil-market information and other economic variables. Varying only the price series and keeping all other modelling aspects constant allows us to focus our analysis on the informational content of closing prices rather than on other model features. Our results show that the forecasting performance of pure time-series models and of models with economic predictors significantly improves when these models are estimated using closing prices instead of the originally proposed average price. This confirms that a simple change of the price series can have a remarkably strong effect on the forecast performance of various models.

To the best of our knowledge, this is the first paper that uses information on daily oil prices in the context of forecasting the real price of crude oil. Closing 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. Interestingly, although our setup is new, the results are broadly consistent with several distinct pieces of

[2] Tiao (1972) derives MSPE improvements at various horizons for the limiting case in which the average is taken over a increasingly large number of observations that follow a general autoregressive integrated moving-average process in first differences.

[3] Maintaining a monthly (or lower) frequency is generally desirable as it allows forecasters to include additional economic predictors that might improve forecast accuracy (Alquist et al., 2013).

evidence provided in earlier studies. For example, Baumeister and Kilian (2014) document that the monthly random walk model forecasts average quarterly prices better than the quarterly 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) warn that averaging commodity prices can introduce spurious predictability in the context of return-forecasting regressions. These results are all indicative that time-averaging introduces a mechanical loss of information about future price levels. A key contribution of this paper is to show that this loss necessitates the re-evaluation of forecastability as comparisons have conventionally been made against the aggregated no-change forecast. We also show how the introduction of a different benchmark helps establish more meaningful forecast comparisons.

Overall, our results highlight that there can be large benefits from using information on daily observations when constructing forecasts for temporally aggregated data. This insight is particularly relevant for policy makers who often publish price indexes and forecasts based on recently observed average data or estimate forecasting models with average prices.[4] The first, easy-to-implement, step to improve forecasts in this setting is to replace lower-frequency averages with the most recent high-frequency observation.

2 Why Use the Last Closing-Price No-Change Forecast?

The effect of temporal aggregation on the accuracy of the no-change benchmark can best be understood in the cases where higher-frequency observations follow a random walk. This setup is useful because it highlights the consequences of averaging in critical cases in which all future prices are unpredictable, which is the implicit null hypothesis in many existing studies. Conveniently, it also allows us to derive analytical expressions for the MSPEs for the different no-change forecasts that cannot be derived for general settings since the true data-generating process is unknown (Inoue and Kilian, 2006). Moreover, many high-frequency series are very persistent and exhibit only a small degree of autocorrelation in changes, suggesting that a random walk is often a reasonable first-order approximation of the behaviour of such series.

Consider a generic series of daily observations that are labeled "prices" that follow a random

[4] See, e.g., the World Bank Commodity Price Data, the Bank of Canada Monetary Policy Report, or the European Central Bank (ECB) staff macroeconomic projections for the euro area.

walk, such that $p_{t,i}$, the price on day i in month t, is given by

$$p_{t,i} = p_{t,i-1} + \epsilon_{t,i}, \quad \text{for } i = 0, 1, 2, ..., n. \tag{1}$$

Here, n is the number of daily prices within a month (about 21 for the usual business-calendar sampling), $\epsilon_{t,i}$ is a mean-zero *iid* error term with variance σ_ϵ^2.[5] We also define $p_{t,0} = p_{t-1,n}$ to transition between months.[6] The average monthly price in month t is given by

$$\overline{p}_t \equiv \frac{1}{n}\sum_{i=1}^{n} p_{t,i}. \tag{2}$$

The forecaster's goal is to predict the average monthly price k periods ahead, $\overline{p}_{t+k}$, given time t information. It is well-known that the forecast that minimizes the MSPE is the conditional expectation $E_t(\overline{p}_{t+k})$. In our setting, the conditional expectation of the average price k periods ahead, given time t information, is

$$\begin{aligned} E_t(\overline{p}_{t+k}) &= E_t\left(\frac{1}{n}\sum_{i=1}^{n} p_{t+k,i}\right) & (3)\\ &= \frac{1}{n}E_t\left(n \cdot p_{t,n} + n \cdot \sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s}\right) & (4)\\ &= p_{t,n}, & (5)\end{aligned}$$

where the last step follows from the assumption that ϵ is mean-zero, *iid*. Thus, in the MSPE sense, the last trading day in period t is the best predictor of the average monthly price in period $t+k$.

Two different effects arise when price averaging is used instead of the last closing price. First, when the last observed average price is used for comparisons, the forecaster's predictions are evaluated against a benchmark with larger expected squared forecasting errors. To see this, first note that the difference between the last daily price and the last average price is given by

$$p_{t,n} - \overline{p}_t = p_{t,n} - \frac{1}{n}(p_{t,n} + p_{t,n-1} + ... + p_{t,1}) = \frac{1}{n}\sum_{i=2}^{n}(i-1)\cdot\epsilon_{t,i}. \tag{6}$$

The expected squared forecast error for the k-month-ahead average price under the last closing-price

[5] Extending the setup to allow for conditional heteroskedasticity in the innovations is straightforward and does not affect the conclusions of this section as long as the process governing the conditional heteroskedasticity is stationary and the unconditional variance exists.

[6] i.e., the last observed price in month $t-1$ is the price on the opening of the trading day on month t.

no-change forecast is given by

$$E\left[\left(\overline{p}_{t+k} - p_{t,n}\right)^2\right] = E\left[\left(\sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \frac{1}{n}\sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s}\right)^2\right] \tag{7}$$

$$= \left((k-1)\cdot n + \frac{(n+1)\cdot(2n+1)}{6n}\right)\sigma_\epsilon^2, \tag{8}$$

while the expected squared forecast error for the k-month-ahead average price under the average no-change forecast is given by

$$E\left[\left(\overline{p}_{t+k} - \overline{p}_{t}\right)^2\right] = E\left[\left(\sum_{j=1}^{k-1}\sum_{i=1}^{n}\epsilon_{t+j,i} + \frac{1}{n}\sum_{s=1}^{n}(n+1-s)\cdot\epsilon_{t+k,s} - \frac{1}{n}\sum_{l=1}^{n-1} l\cdot\epsilon_{t,l}\right)^2\right] \tag{9}$$

$$= \left((k-1)\cdot n + \frac{(n+1)\cdot(2n+1)}{6n} + \frac{(n-1)\cdot(2n-1)}{6n}\right)\sigma_\epsilon^2, \tag{10}$$

which is unambiguously larger. In practice, most forecasters evaluate the accuracy of their forecasts based on a criterion that involves relative MSPEs, such as the MSPE ratio. When the average price is used as a benchmark, the forecaster is more likely to find reductions in the MSPE from the use of alternative forecasts and to conclude that prices are predictable. This is the case even if prices are entirely unpredictable, as in our example.

The second effect arises when average prices are also used to estimate models, as is the case with most existing econometric forecasts for the real prices of oil. It is well-known that adjacent changes in average prices will be mechanically correlated even if the underlying data-generating process is a martingale (Working 1960). In the context of our setting, the change in the average price can be written as

$$\overline{p}_t - \overline{p}_{t-1} = \frac{1}{n}\left[\sum_{i=1}^{n} i\cdot\epsilon_{t-1,i} + \sum_{j=1}^{n}(n+1-j)\cdot\epsilon_{t,j}\right], \tag{11}$$

such that each innovation $\epsilon_{s,k}$ enters both $\Delta\overline{p}_s$ and $\Delta\overline{p}_{s-1}$. Here, any adjacent changes in average prices will be correlated, even if all future changes in the daily price are unpredictable. Hence, an econometric model that is based on average prices and allows for serial correlation might show statistically significant coefficients and exhibit improved forecasts relative to the no-change forecast $\bar{p}_t$, even in the critical case in which prices are unpredictable. Both effects can be remedied by using closing prices to construct both no-change forecasts and model-based forecasts.

A notable feature of the forecast improvements of the last-closing-price forecast over the last-average-price forecast is that all forecasting gains occur at the one-month-ahead prediction. Because of the improvements at the initial step, forecasts that are based on the last closing price will still be better for other horizons, but it is important to note that there is no *additional* improvement in forecastability beyond the first period. For example, the percent improvement in the terms of the MSPE ratio—the difference between the expected forecasting errors, normalized by the average-price benchmark—is

$$\frac{E\left[\left(\overline{p}_{t+k}-\overline{p}_t\right)^2\right]-E\left[\left(\overline{p}_{t+k}-p_{t,n}\right)^2\right]}{E\left[\left(\overline{p}_{t+k}-\overline{p}_t\right)^2\right]}=\frac{(n-1)\cdot(2n-1)}{6(k-1)\cdot n^2+(n+1)\cdot(2n+1)+(n-1)\cdot(2n-1)}. \tag{12}$$

Because the forecasting error increases with the forecasting horizon, but the forecasting gains remain constant, the relative gains, while still measurable for a larger k, decrease to zero asymptotically. Equation 12 also highlights that, in our setting, the theoretical MSPE ratio of the forecast from the last closing price relative to the forecast from the last average price has a closed-form expression. In fact, this ratio only depends on the forecast horizon, k, and on the number of high-frequency observations, n, that are averaged over a given low-frequency time interval. The typical crude-oil forecast is based on monthly average data, with $n = 21$. The theoretical MSPE ratios of the monthly closing-price no-change forecast—relative to the monthly average no-change forecast—that are derived using Equation 12 are 0.54, 0.88, 0.95, 0.97, 0.99 at horizons of 1, 3, 6, 12 and 24 months, respectively. The relative MSPE ratios are lowest for the one-month-ahead forecast, with improvements of 46 percent. However, this number quickly decreases with the forecasting horizon, falling to 12 percent for the three-months-ahead forecast and to a mere 1 percent improvement at the 24-months-ahead forecast.[7] As documented in the empirical application to oil prices below, this pattern is strongly reflected in the actual data; thus, lending support for the intuition provided in this section.

Altogether, the results in this section demonstrate that improvements over the conventional no-change forecast are not necessarily indicative that prices are predictable. Although the derivations are limited to the pure random walk setting, it can be shown that the usefulness of closing prices for forecasting extends to more general processes. As long as daily prices have a persistent component, the last closing price will contain information that can lead to substantial gains in forecast efficiency

[7] See table A1, which also shows the theoretical MSPE ratios for quarterly data ($n = 63$).

relative to the average no-change forecast (Tiao, 1972). This insight is relevant because oil prices are indeed very persistent and practically undistinguishable from non-stationary processes.[8] We therefore expect that the intuition that is highlighted by the random walk setting will carry over to the empirical exercises presented in the remainder of the paper.

3 Data and Real-Time Forecast Method

The focus of our empirical application is real-time forecasts of monthly averages of oil prices, an approach that has become the standard in the literature. In the typical setup of existing studies, the forecaster uses the available information at the end of each month to form their prediction for the following months. The last observed price contained in the forecaster's information set is the closing price of the final trading day of the month, the monthly closing price. Similar to the existing literature, we treat this closing price as being observed in real time.[9] However, while the existing literature uses closing prices only in calculations of monthly average prices, we argue that the closing price by itself provides superior predictive information.

The empirical results in the main section are computed for the monthly price of WTI crude oil.[10] We obtained monthly average prices of WTI crude oil from the EIA. These average prices are simple averages of the daily closing prices that are also provided by the EIA. We construct a series of monthly closing prices by taking the closing price on the last trading day of each month. Both the monthly average prices and the monthly closing prices are deflated to real prices using the seasonally adjusted U.S. consumer price index described below.

To investigate how the no-change forecast that is based on the monthly closing price compares against the alternatives, we also replicate several model-based forecasts that are proposed in the literature. Univariate methods that are shown to outperform the no-change forecast include autoregressive (AR) and autoregressive integrated moving-average models (ARIMA) (Alquist et al., 2013; Baumeister and Kilian, 2012) and autoregressive fractionally integrated models (ARFI) (Snudden,

[8] The measure of persistence used in Alquist and Kilian (2010) —the sum of the autoregressive coefficients from autoregressive models—yields values of about 0.98 for both the real monthly average prices and the real monthly closing prices. The sample period for these estimates is 1992M1 to 2018M12, and the models are estimated with 12 lags.

[9] For the principal global oil-price benchmarks—West Texas Intermediate (WTI) and Brent—nominal prices are available on a daily frequency or higher. For example, daily data, including closing prices for spot prices of WTI and Brent are each updated by the Energy Information Administration (EIA) between 7:30 and 8:30 a.m. This justifies treating the EIA spot prices as being observed in real time.

[10] In the robustness analysis, we show that similar results are obtained for other crude-oil benchmarks and data frequencies.

2018); multivariate models include vector autoregressions (VAR) and Bayesian VARs (BVAR) (Alquist et al., 2013; Baumeister and Kilian, 2012). In addition to the log of the real price of crude oil, the BVAR and VAR models include the percentage change in global crude-oil production, the global real economic activity indicator of Kilian (2009), and the change in global crude-oil inventories.

Following the previous literature, all models are estimated with log prices and converted to actual price levels, such that

$$\hat{\bar{R}}^X_{t+h|t} = exp\left(\hat{\bar{r}}^X_{t+h|t}\right), \tag{13}$$

where $X \in \{AR, VAR, ARIMA, AFRI, BVAR\}$, $\hat{\bar{R}}^X_{t+h|t}$ is the forecast of the monthly average real price of crude oil in levels, $\hat{\bar{r}}^X_{t+h|t}$ is the equivalent in logs, and exp is the exponential function. The only exception is an additional estimation of univariate AR models that use the percent change in real prices. In this case, forecasts for the net growth rate, $\bar{g}^{oil}_t = (\bar{R}^{oil}_t/\bar{R}^{oil}_{t-1} - 1)$, are mapped into levels via

$$\hat{\bar{R}}^{oil}_t = (1 + \hat{\bar{g}}^{oil}_t)\hat{\bar{R}}^{oil}_{t-1}, \tag{14}$$

where $\hat{\bar{R}}^{oil}_{t-1}$ uses historical monthly average data up to T and recursively estimated forecasts thereafter.

An additional, estimation-free forecast is computed from oil futures using the methodology of Alquist et al. (2013):

$$\hat{\bar{R}}^F_{t+h|t} = \bar{R}^{oil}_t\left(1 + \bar{f}^h_t - \bar{s}_t - E_t\left(\pi_{t+h}\right)\right), \tag{15}$$

where $\bar{s}_t$ is the monthly average of the log nominal price of crude, $\bar{R}^{oil}_t$ is the monthly average of the real price of crude, $\bar{f}^h_t$ is the log of the monthly average of the daily nominal price of crude-oil futures for maturity h at time t, and $E_t\left(\pi_{t+h}\right)$ is the expected U.S. inflation rate.[11] Finally, we also consider an equal-weighted forecast combination of all model-based forecasts and a futures-based forecast to examine the potential gains from pooling these forecasts (Baumeister et al., 2014; Baumeister and Kilian, 2015; Garratt et al., 2019).

All of our forecasts are implemented using real-time data. Vintages start in 1991M1 and contain historical data starting from 1973M1. 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

[11] The expected U.S. inflation rate is computed as the historical average for the CPI inflation rate from 1986M7. Since the nominal price of oil is generally much more variable than CPI inflation, the exact choice of the (expected) deflator has no direct bearing on our results.

the real-time database of the Philadelphia Federal Reserve. Any missing real-time observations for the price index are nowcasted using the average historical growth rate, as in Baumeister and Kilian (2012). 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 and are nowcasted following Baumeister and Kilian (2012).[12] The real-time version of the real economic activity index is computed using the corrected formula (Kilian, 2019). As discussed in section 6, none of the main results are affected by the use of different oil-price series or nowcasting choices.

Our forecasting exercise follows the setting of Baumeister and Kilian (2012). Our forecasts are computed out-of-sample and in a recursive fashion, for the sample period 1992M1 to 2018M12. We use two common criteria for the forecast evaluations, namely the MSPE ratio and the success ratio for directional accuracy.[13] To be as consistent as possible with the existing literature, both measures are constructed against the no-change forecasts that are derived from monthly average prices. However, we also highlight statistically significant improvements vis-à-vis the new benchmark; i.e., the no-change forecast that is constructed from the monthly closing price. Statistical tests for the null hypothesis of equal predictability using the MSPE ratio are based on Diebold-Mariano (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).[14] Both tests account for serial dependence in the forecasts.

The results are robust to the choice of the lag length in the model-based forecasts that rely on autoregressive components. For the BVAR and VAR, we report the results for a lag length of 12 months, one of the benchmark specifications in the existing literature (Baumeister and Kilian, 2012). For the univariate models, we report the results from the lag-length that produces the best forecasting results.[15]

[12] We update the real-time data of Baumeister and Kilian (2012) using historical vintages from the EIA and following the methodology described in Baumeister and Kilian (2012). See section A.2 for details.

[13] For the monthly closing price, the directional forecasts are based on the difference between the monthly closing price and the average monthly price.

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

[15] The number of lags searched is from 1 to 24.

4 Forecast Performance

Table 1 reports the end-of-sample MSPEs and the success ratios for the monthly closing price no-change forecasts and also the alternative forecasts. The forecasts that use the monthly closing prices strongly outperforms the forecasts that use the monthly average price no-change in both the mean squared prediction and the directional accuracy, especially at shorter horizons. For one-month-ahead forecasts, the monthly closing price results in 40 percent improvements in the MSPE and the success ratios, both of which are significant at the 1 percent level. These improvements decrease with increases in the forecasting horizon but are still statistically significant at the 5 percent significance level for the 12-months-ahead forecast. The decay in the relative forecasting performance is both qualitatively and quantitatively very similar to the theoretical prediction of the pure random walk model for daily oil prices. For example, the theoretical improvements for the 1-, 3- and 12-month horizons are 46, 12 and 3 percent, respectively, while the empirical counterparts are 39, 11 and 4 percent, respectively. These results strongly corroborate the idea that the forecasting gains are realized at the one-step-ahead prediction and become relatively less important for longer-horizon forecasts.

Table 1. Real-Time Forecasts, Real Price of WTI Crude Oil, 1992–2018

Months Ahead	Last WTI Close Price	BVAR(12)	VAR(12)	AR(12)	AR(12) %Δ	ARFI(1)	ARMA(1,1)	Futures Curve	Model Averaging
					MSPE Ratios				
1	0.61***	0.97	1.01	0.94	0.95*	0.93	0.92	1.00	0.91
3	0.89**	1.00	1.00	0.97	0.99	0.96	0.95	0.97	0.92
6	0.95**	1.05	1.04	1.00	1.04	0.99	0.95	0.96	0.95
12	0.96**	1.10	1.11	1.00	1.10	1.00	0.94	**0.85****	0.94
24	0.99	1.08	1.06	1.02	1.19	1.04	0.96	**0.82***	0.91
					Success Ratios				
1	0.71***	0.51	0.54**	0.52	0.49	0.52	0.53	0.47	0.49
3	0.60***	0.53	0.54*	0.49	0.57**	0.49	0.50	0.49	0.52
6	0.56**	0.53	0.55*	0.48	0.54	0.49	0.46	0.53	0.51
12	0.59***	0.49	0.56**	0.53	0.50	0.51	0.50	0.61***	0.51
24	0.53	0.51	0.55	0.57	0.47	0.55	0.56	**0.62*****	0.55

Note: Recursive, dynamic, out-of-sample-forecasts 1992M1–2018M12 for West Texas Intermediate crude converted into real levels. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing price. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, 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.

It is well-known that both the bias and the variance of a forecast contribute to its MSPE. Since one might expect average prices to have a lower variance than end-of-month prices, a natural question is why averaging is not more useful in reducing the MSPE. Again, the key to this answer lies in the near-random-walk behaviour of daily prices. It can be shown that in the case of an exact random walk and for typical sample sizes, the variances of the average price and the last closing price are almost indistinguishable. Likewise, over our evaluation period, the empirical variance of the real monthly average price is 144.1 relative to a variance of 146.5 for the real last closing price. The p-value for the test of the difference in the variance is 0.44. On the other hand, the empirical bias of the closing price tends to be much smaller than that of the average price, which indicates that a reduction in the bias drives our results.

The results for the monthly closing price stand in stark contrast to those for the econometric models, which generally do not improve the traditional no-change benchmark. The only exception is for the VAR(12) model, which features statistically significant improvements in the directional accuracy over all but the 24-month horizon. However, the MSPE ratios indicate that, according to this metric, both the BVAR and the VAR models perform poorly. As shown below, this result stems from a deterioration in the performance of these models during the 2010-2014 period.[16]

More strikingly, none of the individual estimation-based forecasts improve upon the monthly closing-price benchmark in a statistically significant way. Only the estimation-free forecast provided by the futures curve yields large and significant improvements in both the MSPE and the directional accuracy of the 12- and 24-months-ahead forecasts. As highlighted by the bold-faced values, only the futures-based forecast beats the closing-price benchmark at the 5 percent significance level and only at horizons of 12 and 24 months. While model averaging yields generally favorable results with MSPE reductions of 5 to 10 percent at various horizons, these improvements are not statistically significantly better than the average price no-change forecast.

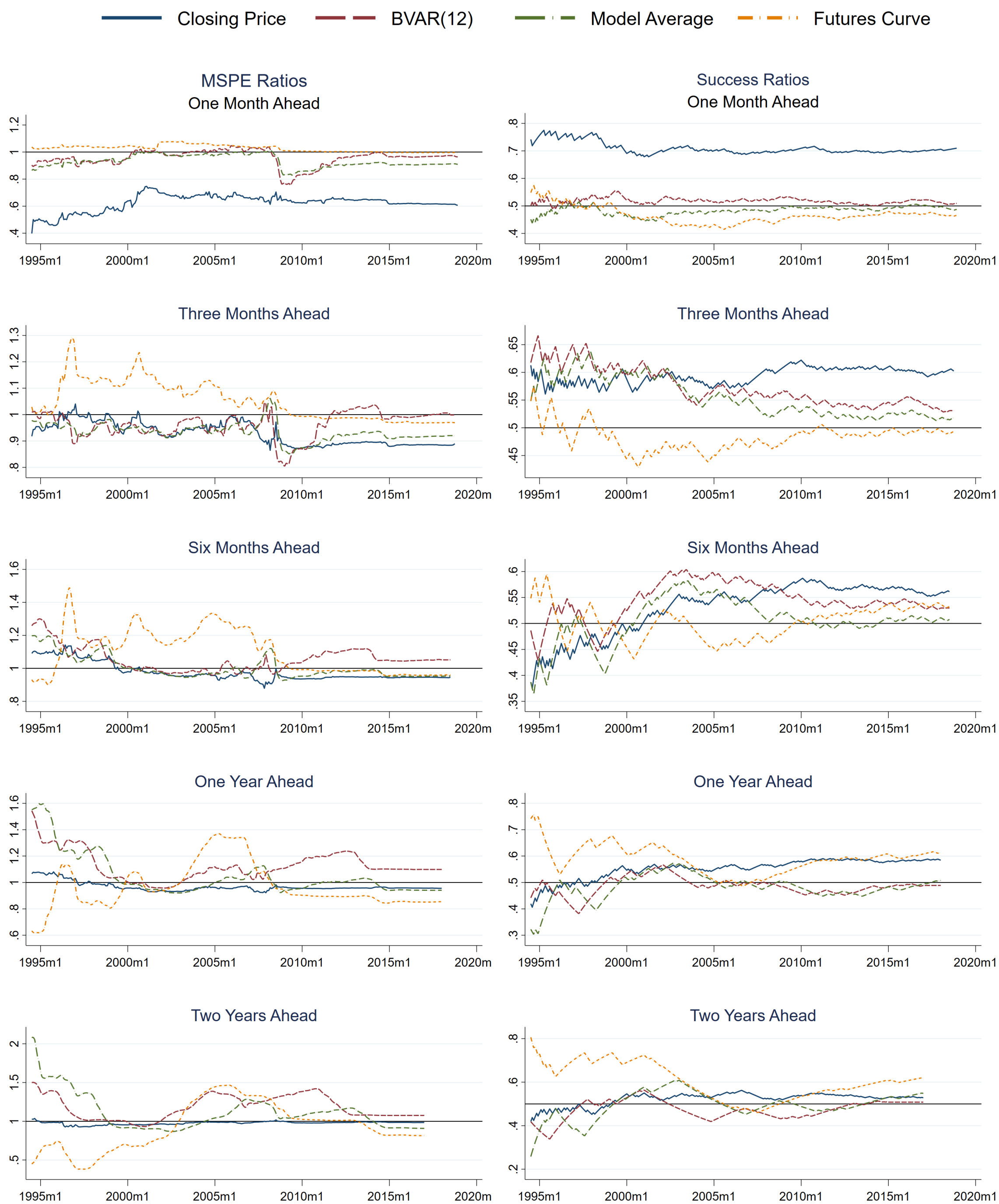
The forecast performance of oil-price-prediction models are often sensitive to the sample period.[17] To investigate the robustness of these models' performance over time, the evolution of the recursively updated MSPE ratio and the directional accuracy are reported in Figure 1.[18] The forecasting improvements of the monthly closing prices are extremely stable for shorter horizons. By contrast, the performance of the other forecasts is more variable. For example, at the one-month-ahead horizon, the closing price significantly outperforms the average price no-change forecast in the MSPE precision for 100 percent of the sample, while the BVAR forecast produces MSPE ratios

[16]Garratt et al. (2019) and Baumeister et al. (2020) have recently documented a similar result.

[17]See, for example, Baumeister et al. (2014); Baumeister and Kilian (2015); Snudden (2018); Garratt et al. (2019).

[18]To conserve space, the graphs focus on the monthly closing-price benchmark, BVAR(12), futures-based and model-average.

Figure 1. Evolution of Real-Time Forecast Criteria, Real Price of WTI Crude Oil, 1992–2018



Note: Dynamic, recursive, out-of-sample forecasts 1992M1–2018M12. The forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. Success ratios are calculated to quantify the accuracy of the forecast direction and represent the fraction of times the forecast correctly predicts the direction of the change in the real price of oil. All forecast criteria are evaluated in the levels of the real price of oil. The first 30 months are dropped to allow for the law of large numbers to begin working.

of less than one for about 70 percent of the sample. In many cases, the peak performance of the model-based forecasts coincides with the financial crisis of 2008 but deteriorates thereafter. By contrast, the relative forecast performance of the monthly closing price is very stable and avoids much of the variability in the performance around times of large price movements. This lends further support to the idea that averaging leads to a loss of information that is mechanical and largely independent of the price behaviour that occurs over an episode.

5 Estimating Models with Closing Prices

The results presented so far strongly suggest that the monthly closing-price no-change forecast should be used as the benchmark for forecast evaluations. An additional insight from our illustrative theoretical setting is that estimating models by using average prices introduces mechanical correlation into any adjacent price changes. Even though many existing models are dynamic and allow for autocorrelations, estimating the respective coefficients introduces an additional level of estimation uncertainty that will, in practice, diminish the out-of-sample forecasting performance of models that use small samples. Therefore, a natural question is whether forecasting using econometric models and futures prices can be improved by using closing prices when estimating these models.

To answer this question, we re-estimate all previously discussed models using real monthly closing prices instead of average monthly prices. The corresponding forecast is

$$\hat{R}^X_{t+h|t,n} = exp\left(\hat{r}^X_{t+h|t,n}\right), \tag{16}$$

where $X \in \{AR, VAR, ARIMA, AFRI, BVAR\}$, $\hat{R}^X_{t+h|t,n}$ is the forecast of the monthly closing real price of crude oil in levels, and $\hat{r}^X_{t+h|t,n}$ is the equivalent in logs. For the AR model that is estimated in growth rates, the forecasts for the net growth rates in closing prices, $g^{oil}_{t,n} = (R^{oil}_{t,n}/R^{oil}_{t-1,n} - 1)$, are mapped into the price levels via

$$\hat{R}^{oil}_{t,n} = (1 + \hat{g}^{oil}_{t,n})\hat{R}^{oil}_{t-1,n}, \tag{17}$$

where $\hat{R}^{oil}_{t-1,n}$ uses historical monthly closing data up to T and uses recursively estimated forecasts thereafter.

For the futures-based model, we use the closing contract prices on the last trading day of the month instead of the monthly average contract price:

$$\hat{R}^{F}_{t+h|t,n} = R^{oil}_{t,n}\left(1 + f^{h}_{t,n} - s_{t,n} - E_t\left(\pi_{t+h}\right)\right), \tag{18}$$

where $s_{t,n}$ is the log of the monthly nominal closing price of crude, $R^{oil}_{t,n}$ is the monthly closing real price of crude, and $f^{h}_{t,n}$ is the log of the monthly closing nominal price of crude-oil futures for maturity h at time t. All other aspects of the forecasts are kept identical to the exercise in the previous section.[19]

Table 2. Real-Time Forecasts, Real Price of WTI Crude Oil, End-of-Month Closing Prices, 1992–2018

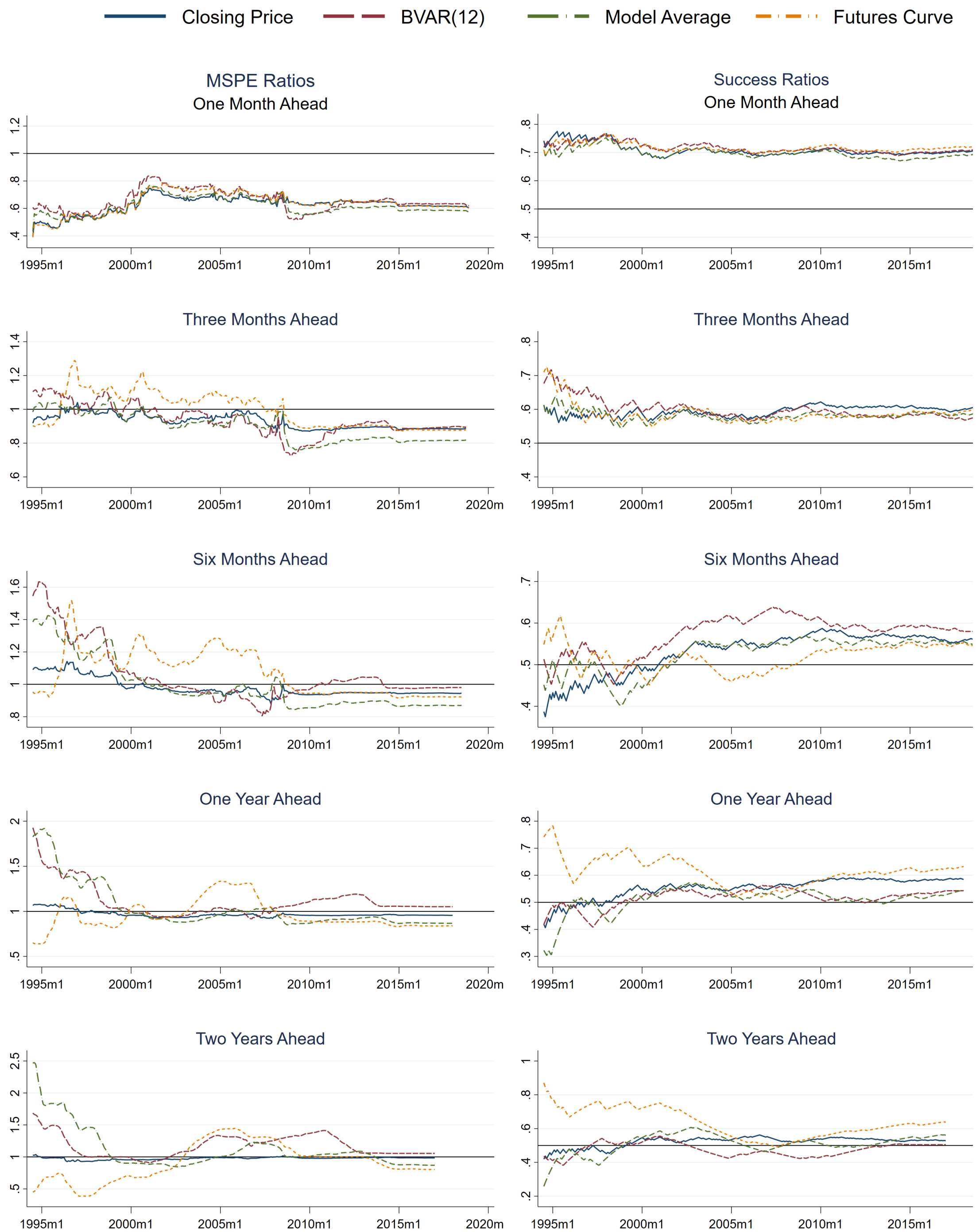
Months Ahead	Last WTI Close Price	BVAR(12)	VAR(12)	AR(2)	AR(1) %Δ	ARFI(2)	ARMA(1,1)	Last Future Price Curve	Model Averaging
					MSPE Ratios				
1	0.61***	0.62**	0.73*	0.57***	0.62***	0.57***	0.57***	0.60***	0.58***
3	0.89**	0.90	0.91	0.85	0.93*	0.86	0.86	0.88**	0.82*
6	0.95**	0.98	0.97	0.89	1.04	0.89	0.90	0.92*	0.87
12	0.96**	1.05	1.06	0.90	1.15	0.90	0.91	**0.84****	0.87*
24	0.99	1.06	1.06	0.95	1.48	0.95	0.97	**0.80***	0.87
					Success Ratios				
1	0.71***	0.71***	0.66***	0.71***	0.73***	0.72***	0.71***	0.73***	0.70***
3	0.60***	0.58**	0.56**	0.57**	0.62***	0.60***	0.60***	0.60***	0.59***
6	0.56**	0.58**	0.58**	0.52	0.58**	0.51	0.52	0.55	0.55
12	0.59***	0.54	0.58**	0.54	0.59**	0.51	0.53	**0.63*****	0.54
24	0.53	0.51	**0.57***	0.58	0.56	**0.55**	0.54	**0.64*****	0.56

Note: Recursive, dynamic, out-of-sample-forecasts from 1992M1–2018M12 for West Texas Intermediate crude converted to real levels. The models are estimated using the monthly closing price beginning in 1973M1. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table 2 shows that substantial gains in the model-based forecast performance are achieved by estimating models using the monthly closing prices instead of the average monthly price. At the one-month horizon, all model-based forecasts produce large improvements of about 40 percent over the average no-change forecast. The improvements in the MSPE ratios are statistically significant at the 10 percent level for the VAR model, at the 5 percent level for the BVAR model, and at the 1 percent level for all other models. The success ratios also show large improvements in the directional accuracy for all horizons, most of which are statistically significant at the 5 percent

[19]Monthly closing prices are available from the EIA, starting in January 1986. To construct a series that overlaps with the monthly average data, we backcast the closing prices until January 1973, using daily futures prices and WTI prices as described in Appendix A.2. The results are robust to estimating models using a sample that starts in January 1986.

Figure 2. Evolution of Real-Time Forecast Criteria for the Real Price of WTI, Using End-of-Month Closing Prices, 1992–2018



Note: Dynamic, recursive, out-of-sample forecasts 1992M1–2018M12. Models are estimated using the monthly closing price starting in 1986M1. Forecast criteria reported include the recursive MSPE expressed as a ratio relative to the monthly average no-change forecast. Success ratios are calculated to quantify the accuracy of the forecast direction and represent the fraction of times the forecast correctly predicts the direction of change in the real price of oil. All forecast criteria are evaluated in the levels of the real price of oil. The first 30 months are dropped to allow for the law of large numbers to begin working.

level. These results stand in stark contrast to the forecasts that are derived from the same models but based on average prices (Table 1), suggesting that closing prices should also be used to estimate these models even if the forecaster's goal is to predict average prices.

Interestingly, for the univariate models, the number of autoregressive coefficients is smaller in the models that produce the best forecasts. For example, the AR model that is estimated in levels produces the best forecasts with 2 lags when estimated using the monthly closing price, whereas 12 lags are commonly found to be preferred when the model is estimated using the monthly average price. This is consistent with Tiao (1972) and Rossana and Seater (1995), who show that time-averaging can have important effects on the dynamics of economic time series.

The evolution of the recursively updated forecasting criteria is reported in Figure 2. At the one-month horizon, all models now outperform the average monthly price no-change forecast. The improvements in the MSPE ratio and the directional accuracy are statistically significant at the 5 percent level for 100 percent of the sample. In 93 percent of the sample, model averaging produces lower MSPE ratios compared to the monthly closing price at the three-month horizon. In over 80 percent of the sample, the futures model produces superior criteria at the one-and two-year horizons. This documents that relative to the conventional benchmark, model-based forecasts become much more stable when the models are estimated using end-of-the-month closing prices.

Similar to the previous exercise, the bold-faced numbers in Table 2 highlight the improvements of the forecasts of the models that are estimated using the closing prices as opposed to those that use the last-closing-price no-change benchmark. Although model-based forecasts are considerably improved and, in some cases, show lower MSPE ratios and better directional accuracy than the new benchmark, these improvements are rarely statistically significant and consistent across both criteria. This shows that the choice of the benchmark matters and that improvements over the conventional benchmark are not necessarily indicative that oil prices are predictable. As before, only the futures-based forecasts for 1 and 2 years are both economically and statistically more accurate than the new benchmark.

6 Extensions and Robustness

The results presented in the main section of the paper are remarkably robust to a series of alternative modelling choices, such as using different oil-price benchmarks, using ex-post revised data instead of real-time data, using nominal prices instead of real prices, and using quarterly instead of monthly data. This robustness corroborates that averaging leads to a mechanical loss of information that is relevant for forecasting average prices.

As a first robustness check, the forecasting exercises from the main section are repeated for the real price of Brent crude oil and the real refiner acquisition import costs (RAC) instead of the real price of WTI. The results are presented in Tables A3 and A4. Compared to the forecasts for WTI prices, the performance of the monthly closing-price no-change forecast relative to the monthly average no-change forecast is even better for Brent prices. For the one-step-ahead real-time forecasts, monthly closing prices result in over 40 percent improvements in the MSPE ratio and directional accuracy. Model-based forecasts for Brent crude that are estimated using monthly closing prices are of similar quality to those for WTI, suggesting that the monthly closing price is also the appropriate benchmark for forecasting the real price of Brent crude oil.

Closing-price forecasts for the RAC are complicated by the fact that no daily price series exists for the RAC. However, there exists a straightforward solution that resembles the standard practice of nowcasting the average RAC. The RAC closing price is nowcasted by applying the growth rate of the WTI closing price over the previous month's average WTI price to the previous month's RAC price.[20] In terms of its forecast performance, the nowcasted closing price of the RAC is remarkably similar to that documented for the other oil benchmarks. As shown in Table A5, at the one-month horizon, the MSPE and success ratio improve by a statistically significant 30 percent relative to the average-price no-change forecast. The relative improvements for all other horizons are almost identical to those found in the case of real prices of WTI. Similar to the baseline results, the forecasts that are based on the monthly closing prices of futures contracts are more accurate than the no-change forecasts. The improvements are statistically significant at the 5 percent level for both the evaluation criteria and for all horizons relative to the average no-change forecast but only at the one- and two-year horizons in the case of the last-closing-price benchmark.

Using ex-post revised data instead of real-time data does not affect our main results. For the ex-post revised data, the gains from the closing-price no-change forecast are almost identical to our baseline results (Table A6) and are still unable to significantly outperform the closing-price benchmark.[21] Moreover, the results are robust to alternative methods of deflating the closing prices using the CPI deflator. Affixing the monthly index level to the closing price of the month

[20]The computations are performed for nominal prices. The imputed closing prices for the RAC are converted into real terms using the same deflater that is used to deflate the other oil-price series.

[21]Consistent with Baumeister and Kilian (2012), we also find that the forecast performance of the BVAR and VAR models for all crude-price series is robust to alternative methods of nowcasting the CPI, the oil-inventory and oil-production series.

outperforms alternative daily interpolations. These results are expected as the potential revisions and deviations in the CPI deflator are very small compared to the fluctuations in nominal oil prices.

The finding that the forecasts of workhorse models that use economic predictors are unable to significantly outperform the new no-change benchmark is also robust to proposed modifications in the predictor variables. Consistent with Snudden (2018), we find that the most accurate VAR forecasts originate from models that are estimated using the growth rate of global industrial production and U.S. petroleum inventories.[22] Still, the improvements from these modifications are not statistically significantly more accurate than the new benchmark. While we cannot not rule out that additional improvements in forecast performance are possible by using high-frequency observations in combination with recently proposed forecasting strategies, this further corroborates the idea that standard econometric models are less suitable to forecasting oil prices than previously thought.

The main results not only hold for real prices but also for nominal prices. As shown in Table A8, the relative forecast gains of using the nominal closing price vis-à-vis the average nominal price are very similar to those documented for real oil prices. This result is again consistent with the idea that fluctuations in nominal oil prices rather than fluctuations in inflation drive our results. For the case of nominal oil prices, the use of the last closing price as a no-change benchmark has been previously proposed by Alquist and Kilian (2010). The authors also document that futures prices did not outperform the last-closing-price benchmark before 2010. Our results are based on a decade of additional data in which futures prices have become increasingly useful in forecasting both real and nominal prices. This example highlights again that gains from alternative forecasts tend to be much more dependent on the sample period than on the results from using closing prices.

The closing-price no-change forecast remains superior to the average-price no-change forecast for quarterly and annual data. The use of lower-frequency data and forecasts are of primary interest to policy makers (Baumeister and Kilian, 2014, 2015). Table A7 shows the results for forecasts of the quarterly average real price of crude oil, where all models are estimated using average quarterly variables. Consistent with the theoretical prediction in section 2, the improvements in the MSPE ratio at the one-step-ahead prediction are even larger at the quarterly than the monthly frequency. The last closing price significantly outperforms the quarterly average-price no-change forecast at the one-quarter horizon for the MSPE ratio and up to one year ahead for the success ratio. Again,

[22]The good forecast performance of global industrial production has also been documented in a comparison of different economic activity measures in a pseudo real-time setting by Baumeister et al. (2020).

only the futures curve can outperform the closing-price no-change forecasts but only at longer forecasting horizons.

7 Conclusion

Academics and policy makers have invested significant effort in analyzing whether macroeconomic variables are predictable and which models should be used to forecast such series. In this context, the practice of averaging raw high-frequency observations to construct lower-frequency series can lead to a significant loss of information about future averages. For the real price of oil, a key global macroeconomic indicator, replacing monthly average prices with monthly closing prices improves the MSPEs and the directional accuracy of existing models by up to 40 percent.

The loss of information has important implications for conducting forecast comparisons with a no-change forecast. The common approach of comparing forecasts against a no-change forecast that is based on the last average at the forecaster's data frequency is unsuitable for making inferences about the data-generating process. Instead, forecasts should be compared to an alternative benchmark: the no-change forecast that is constructed from the last higher-frequency observation. The new benchmark reflects the hypothesis that the underlying data, rather than aggregated observations, are generated by a random walk. When the underlying series are highly persistent, as is the case for daily oil prices, the new benchmark dominates the conventional benchmark in terms of forecast accuracy.

These findings have two broader implications. First, the introduction of a new benchmark can raise the bar for model-based forecasts to claim improvements over the no-change forecast. We show that this is indeed the case for the real price of crude oil. Forecasts that are generated from several popular models often outperform the conventional no-change benchmark, especially when these models are estimated using closing prices instead of average prices. However, they generally do not improve upon the new benchmark. Only the futures-based forecast provides better forecasts than the monthly closing-price benchmark and only for horizons of one year and beyond. This result suggests that real oil prices are more difficult to predict and, in this sense, closer to asset prices than implied by the previous literature.

The second implication concerns policy makers and applied forecasters who use aggregate data to construct their forecasts. Our results highlight that incorporating information from high-frequency observations can yield large gains even in the context of the simple models many practitioners prefer. Such gains are likely to occur in any setting where forecasters work with aggregated data and the underlying series are very persistent. This includes the prices of other storable commodities that are known to be highly persistent and are commonly published as averaged data (Deaton and Laroque, 1992).[23] In this environment, one would expect forecasts from econometric models to beat the conventional no-change forecasts that are based on aggregated data, even if the underlying data used to obtain the aggregated series is entirely or approximately unpredictable. For policy makers and applied forecasters, the easiest way to improve traditional forecasts for such series—particularly for short- and medium-term horizons—is to rely on the last higher-frequency observation rather than on lower-frequency averages.

[23] Examples of popular commodity prices and price indexes that are published as average data are the IMF Primary Commodity Prices, the World Bank Pink Sheet, and the Bank of Canada Commodity Price Index and its components.

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. 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). Do exchange rates really help forecasting commodity prices? *Mimeo.*

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *The Review of Economic Studies*, 59(1):1–23.

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

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.

Inoue, A. and Kilian, L. (2006). On the selection of forecasting models. *Journal of Econometrics*, 130(2):273–306.

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.

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.

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

A Appendix (Not for publication)

A.1 Theoretical MSPE Ratios Under the Random Walk Assumption

Table A1. Theoretical MSPE Ratios for Monthly and Quarterly Data Frequency

	Monthly				
Horizon	1	3	6	12	24
MSPE ratio	0.54	0.88	0.95	0.97	0.99
	Quarterly				
Horizon	1	2	4	8	
MSPE ratio	0.51	0.80	0.91	0.96	

Note: The mean square prediction error (MSPE) ratios are computed under the assumption that daily prices follow a simple random walk. The calculations use 21 daily observations for monthly and 63 daily observations for quarterly data.

A.2 Construction of Real-Time Data

This section describes the construction of the real-time data used in the empirical exercises. To remain consistent with the existing literature, we use monthly real-time vintages from Baumeister and Kilian (2012) from 1991M12 to 2010M12 and extend the vintages using the same methodology and data sources for the period of 2011M01 to 2019M06. This approach is similar to Garratt et al. (2019), who recently made their real-time dataset available online. Our methodology differs primarily in that our update uses the excel files that contain the full historical releases of the EIA's Monthly Energy Review. The full coverage of the historical data ensures that all historical revisions and all significant digits are included after 2010M12, which was not possible for parts of the sample in Garratt et al. (2019).

The data coverage of all of the real-time vintages starts in 1973M1. 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's Monthly Energy Review (MER). For the construction of the Real Economic Activity index proposed by Kilian (2009), we use Kilian's freight-rate data until 1986 and, thereafter, we use monthly averages of the Baltic Dry Index obtained from Haver Analytics. 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.

Crude-oil price data for WTI and Brent is available in real time and is not subject to historical revisions and are obtained from the EIA. The monthly average series are backcasted using historical data for monthly average prices from the International Monetary Fund's International Financial Statistics database. Daily closing prices for WTI and Brent are obtained from EIA Petroleum and other liquids. Daily prices are updated 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 not subject to revisions, which justifies treating the data as observed in real time. The monthly closing price is the closing price on the last trading day of the month.

The series of daily WTI and Brent prices provided by the EIA begin in 1986M1 and 1987M5, respectively. In Table 2 of section 5, the models are estimated with a series of monthly closing prices that are backcasted to 1973M1. 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. Values for observations prior to 1983M4 are backcasted using the end-of-the-month WTI price. The monthly closing price of Brent is backcasted using the growth rate in the monthly closing price of WTI. The backcasted Brent crude-price series is used to construct the log real monthly closing-price series for the estimation of the models in Table A4. For horizons of less then one year, the findings are robust to estimating the models using monthly closing spot prices beginning in 1986M1. The backcasted series improves the estimation-based forecasts primarily at one- and two-year horizons.

Table A2. Descriptive Statistics of Oil Price Data, 1973–2018

Monthly Average	Date Range	Mean	Std. Dev.	Min	Max
U.S. Refiner Acquisition Cost, Imported	1973m1 - 2018m12	3.00	0.51	1.74	4.07
Brent	1973m1 - 2018m12	3.06	0.52	1.79	4.11
West Texas Intermediate	1973m1 - 2018m12	3.06	0.47	1.93	4.12
Monthly Closing Price	**Date Range**	**Mean**	**Std. Dev.**	**Min**	**Max**
Brent	1973m1 - 2018m12	3.03	0.51	1.80	4.15
West Texas Intermediate	1973m1 - 2018m12	3.06	0.47	1.94	4.16

Note: All series are the log level of real prices. 2019M6 data vintage. Nominal crude-oil price data obtained from the Energy Information Administration and the consumer price index is obtained from Philadelphia Federal Reserve.

A.2.1 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 methodology proposed by Baumeister and Kilian (2012):

- Missing observations for global crude-oil production, 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.

A.3 Forecast Robustness

Table A3. Real-Time Forecasts, Real Price of Brent Crude Oil, 1973–2018

Months Ahead	Last Brent Close Price	BVAR(12)	VAR(12)	AR(12)	AR(12) %Δ	ARFI(1)	ARMA(1,1)	Futures Curve	Model Averaging
					MSPE Ratios				
1	0.57***	1.02	1.08	0.96	0.97*	0.97	0.95	1.05	0.95
3	0.91**	1.05	1.08	0.99	1.00	1.00	0.98	1.06	0.97
6	0.97*	1.07	1.10	1.01	1.03	1.03	0.98	1.03	0.98
12	0.98	1.09	1.12	1.01	1.08	1.09	1.00	0.91*	0.97
24	1.00	1.09	1.11	1.00	1.19	1.09	1.01	**0.81****	0.92
					Success Ratios				
1	0.73***	0.47	0.52	0.48	0.51	0.47	0.48	0.52	0.49
3	0.58***	0.46	0.50	0.49	0.56*	0.48	0.48	0.47	0.49
6	0.57***	0.45	0.48	0.52	0.48	0.52	0.51	0.50	0.50
12	0.60***	0.44	0.51	0.54	0.53	0.54	0.51	0.61***	0.52
24	0.56**	0.51	0.54	0.60*	0.50	**0.59***	**0.57**	0.61***	0.57

Note: Recursive, dynamic, out-of-sample, real-time forecasts 1992M1–2018M12 for Brent crude converted into real levels. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table A4. Real-Time Forecasts, Real Price of Brent Crude Oil, End-of-Month Closing Prices, 1973–2018

Months Ahead	Last Brent Close Price	BVAR(12)	VAR(12)	AR(2)	AR(1) %Δ	ARFI(2)	ARMA(1,1)	Last Futures Price Curve	Model Averaging
					MSPE Ratios				
1	0.57***	0.68**	0.83	0.58***	0.59***	0.58***	0.57***	0.64***	0.57***
3	0.91**	1.00	1.06	0.91	0.96	0.92	0.91	0.97	0.89
6	0.97*	1.08	1.12	0.96	1.07	0.96	0.96	0.99	0.95
12	0.98	1.13	1.19	0.98	1.21	0.98	0.99	**0.90****	0.96
24	1.00	1.07	1.09	0.98	1.54	0.99	1.01	**0.80****	0.91
					Success Ratios				
1	0.73***	0.66***	0.65***	0.72***	0.71***	0.72***	0.72***	0.71***	0.71***
3	0.58***	0.58***	0.53	0.53	0.57*	0.53	0.52	0.54*	0.56**
6	0.57***	0.58**	0.59**	0.55	0.56	0.53	0.52	0.56*	0.56**
12	0.60***	0.54	0.57**	0.52	0.55	0.51	0.52	**0.62*****	0.54
24	0.56**	0.53	0.57*	**0.59***	0.55	0.56	0.55	0.63***	0.58*

Note: Recursive, dynamic, out-of-sample, real-time forecasts 1992M1–2018M12 for Brent crude converted into real levels. Models are estimated using the monthly closing-price beginning in 1973M1. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table A5. Real-Time Forecasts, Real Price of U.S. Refiner Import Acquisition Cost of Crude Oil, Nowcasted Using Growth in the Average Price of WTI, 1973–2018

Months Ahead	Last WTI Close Price	BVAR(12)	VAR(12)	AR(12)	AR(12) %Δ	ARFI(1)	ARMA(1,1)	Last Futures Price Curve	Model Averaging
					MSPE Ratios				
1	0.71***	0.89	0.92	0.94	0.95	0.90	0.94	0.70***	0.80*
3	0.89**	0.97	0.96	0.95	0.97	0.92	0.93	0.86**	0.86
6	0.95**	1.08	1.08	1.00	1.03	0.97	0.95	0.90**	0.94
12	0.96**	1.15	1.18	1.03	1.09	0.98	0.95	**0.82*****	0.94
24	0.98	1.05	1.03	1.03	1.18	1.01	0.93	**0.77****	0.86
					Success Ratios				
1	0.71***	0.54**	0.57***	0.56**	0.56**	0.55**	0.57***	0.72***	0.60***
3	0.62***	0.54*	0.54*	0.54	0.57**	0.54**	0.56***	0.61***	0.56*
6	0.59***	0.54	0.56**	0.51	0.53	0.51	0.51	0.56**	0.55
12	0.58***	0.52	0.61***	0.58	0.50	0.55	0.54	**0.64*****	0.61**
24	0.54	0.50	0.58**	0.59*	0.46	0.54	0.56	**0.65*****	0.61**

Note: Recursive, dynamic, out-of-sample, real-time forecasts 1992M1–2018M12 for U.S. refiner imported acquisition cost of crude oil, nowcasted using the growth rate in the average monthly WTI price, converted into real levels. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table A6. Ex-post Revised Data Forecasts, Real Price of WTI Crude Oil, 1973–2018

Months Ahead	Last WTI Close Price	BVAR(12)	VAR(12)	AR(12)	AR(12) %Δ	ARFI(1)	ARMA(1,1)	Futures Curve	Model Averaging
					MSPE Ratios				
1	0.61***	0.94	0.97	0.98	0.95*	0.91	0.91	1.00	0.90
3	0.89**	1.01	0.97	1.00	0.99	0.94	0.95	0.97	0.91
6	0.95**	1.09	1.03	1.08	1.04	0.95	0.95	0.96	0.95
12	0.96**	1.16	1.10	1.17	1.10	0.94	0.94	**0.85****	0.93
24	0.99	1.12	0.97	1.08	1.19	0.97	0.96	**0.81***	0.86
					Success Ratios				
1	0.71***	0.54*	0.52	0.52	0.51	0.51	0.53	0.47	0.49
3	0.60***	0.54	0.57**	0.57**	0.57**	0.50	0.49	0.49	0.53
6	0.56**	0.52	0.56**	0.56**	0.54	0.46	0.47	0.53	0.50
12	0.59***	0.49	0.56**	0.57**	0.51	0.50	0.50	0.61***	0.54
24	0.53	0.52	0.57*	0.57**	0.48	0.57	0.56	**0.62*****	0.56

Note: Recursive, dynamic, out-of-sample, ex-post forecasts from 1992M1–2018M12 for West Texas Intermediate crude converted into real levels. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the monthly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table A7. Quarterly Real-Time Forecasts, Real Price of WTI Crude Oil, 1973–2018

Quarters Ahead	Last WTI Close Price	BVAR(4)	VAR(4)	AR(8)	AR(8) %Δ	ARFI(2)	ARMA(1,1)	Futures Curve	Model Averaging
					MSPE Ratios				
1	0.58**	0.99	1.00	0.94	0.97	0.94	0.93	0.61**	0.84
2	0.90	1.08	1.12	0.98	1.03	0.98	0.97	0.88*	0.94
4	0.96	1.11	1.15	0.97	1.06	0.97	0.97	0.87*	0.94
8	1.02	1.07	1.14	0.96	1.18	1.00	0.99	0.83	0.92
					Success Ratios				
1	0.73***	0.53	0.55	0.57*	0.60***	0.54	0.49	0.69***	0.57*
2	0.65***	0.53	0.53	0.55	0.52	0.47	0.46	0.51	0.55
4	0.63***	0.51	0.51	0.51	0.55	0.52	0.52	0.58**	0.51
8	0.49	0.51	0.53	0.55	0.56	0.56	0.55	**0.57****	0.57

Note: Recursive, dynamic, out-of-sample, real-time forecasts 1992Q1–2018Q4 for West Texas Intermediate crude converted into real levels. Bold values indicate significant improvements over the quarterly closing-price no-change forecast at the 5 percent level. *Model Averaging* refers to equal-weighted model-based forecast combinations excluding the quarterly closing-price forecast. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.

Table A8. Real-Time Forecasts for the Nominal Price of WTI Crude Oil, 1973–2018

Months Ahead	Last Daily Close Price	Last Week Close Ave	Last Two Weeks Close Ave	Futures Curve Last Daily Close Price	Futures Curve Monthly Ave	Futures Curve Last Day of Contract	Futures Curve Contract Ave
				MSPE Ratios			
1	0.59***	0.74***	0.81***	0.59***	0.99	1.07	1.45
3	0.89**	0.94**	0.95**	0.88**	0.97	0.97	1.09
6	0.94**	0.99	0.99	0.92*	0.96	0.98	1.00
12	0.95**	1.00	1.00	**0.84***	**0.86**	**0.87**	**0.86**
24	0.98	1.01	1.01	**0.78**	**0.79**	**0.80**	**0.80**
				Success Ratios			
1	0.71***	0.70***	0.63***	0.72***	0.47	0.54**	0.36
3	0.61***	0.59***	0.57***	0.60***	0.54	0.52	0.49
6	0.57**	0.56**	0.54	0.57*	0.53	0.52	0.51
12	0.60***	0.56**	0.52	0.59**	0.59**	0.59**	0.58**
24	0.53	0.49	0.48	**0.63***	**0.64***	**0.64***	0.63***

Note: Recursive, dynamic, out-of-sample-forecasts 1992M1–2018M12 for West Texas Intermediate crude using nominal prices. Bold values indicate significant improvements over the monthly closing-price no-change forecast at the 5 percent level. ***, **, and * denote significance at the 1, 5, and 10 percent level, respectively, of serial dependence robust tests of Pesaran and Timmermann (2009) for the null of no directional accuracy and Diebold and Mariano (1995) of equal MSPEs relative to the monthly average no-change forecast.