

Real-time forecast combinations for the oil price

Anthony Garratt^{1,2} | Shaun P. Vahey^{1,2} | Yunyi Zhang¹

¹Department of Finance, Warwick Business School, University of Warwick, Coventry, UK

²National Institute of Economic and Social Research (NIESR), London, UK

Correspondence

Yunyi Zhang, Department of Finance, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK.
Email: Yunyi.Zhang@warwick.ac.uk

Summary

Baumeister and Kilian (*Journal of Business and Economic Statistics*, 2015, 33(3), 338–351) combine forecasts from six empirical models to predict real oil prices. In this paper, we broadly reproduce their main economic findings, employing their preferred measures of the real oil price and other real-time variables. Mindful of the importance of Brent crude oil as a global price benchmark, we extend consideration to the North Sea-based measure and update the evaluation sample to 2017:12. We model the oil price futures curve using a factor-based Nelson–Siegel specification estimated in real time to fill in missing values for oil price futures in the raw data. We find that the combined forecasts for Brent are as effective as for other oil price measures. The extended sample using the oil price measures adopted by Baumeister and Kilian yields similar results to those reported in their paper. Also, the futures-based model improves forecast accuracy at longer horizons.

1 | INTRODUCTION

Notable recent features of the three real oil prices measures illustrated in Figure 1 include: from 2011, the divergence between Brent crude, the US Refiners' Acquisition Cost for imports (RAC) and the West Texas Intermediate (WTI); the relative convergence during 2014; and the lower conditional means of all these measures from 2014.

Baumeister and Kilian (2015) compare the forecasting performance of six econometric models for the real oil price, individually and in combination relative to a no-change benchmark model. Their analysis is restricted to a sample ending in 2012:09—excluding much of the more recent data plotted in Figure 1—and neglects the Brent crude oil price. Arguably, the Brent measure represents an increasingly important benchmark for the world oil price; see discussions by (among others), Morana (2001), Alberola, Chevallier, and Chèze (2008), and Baumeister and Kilian (2016).

In this paper, we consider three extensions to their analysis: first, the robustness of the results reported by Baumeister and Kilian (2015) to utilizing the real Brent measure (as well as the WTI and RAC measures); second, the sensitivity of their findings to a longer evaluation sample, ending in 2017:12, rather than 2012:9; third, the consideration of futures-based forecasts at longer forecast horizons. We find evidence of similar forecast accuracy across real oil price measures and over the extended evaluation sample, confirming the general findings of Baumeister and Kilian, but with stronger forecasting performance at longer horizons. This last feature of our results arises from our real-time factor-based estimation of the oil price futures curve using the specification of Nelson and Siegel (1987).

We provide a multivariate data set vital for subsequent real-time research on the oil market. The database includes real-time measurements by data vintage for variables similar to those described by Baumeister and Kilian (2012), updated so that 2018:06 represents the last time series observation for all variables. We provide detailed data descriptions in the documentation, together with the real-time data, on <https://www.niesr.ac.uk/real-time-forecast-combinations-oil-price>.

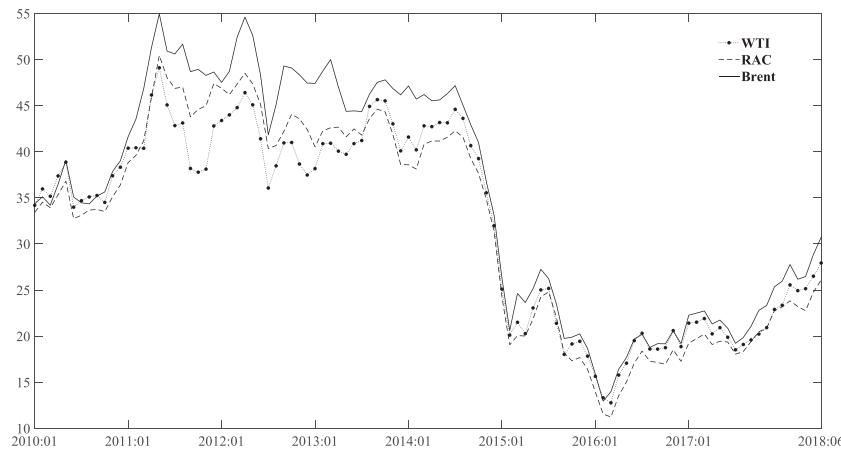


FIGURE 1 Real oil price measures

The remainder of this paper is structured as follows. The next section summarizes the real-time oil market data set, together with the forecast combination methods deployed by Baumeister and Kilian (2015), also considered by Baumeister, Kilian, and Lee (2014). The subsequent section describes the results and the final section concludes.

2 | REAL-TIME DATA AND MODEL SPACE

When compiling the monthly real-time data for the oil market, we broadly followed the nowcast and backcast methods described in Baumeister and Kilian (2012). The main differences between our approach and Baumeister and Kilian's are (1) the inclusion of Brent crude prices; (2) the extended monthly sample with the last observation of 2018:06; and (3) the use of crude oil price futures data from 1991:12 to 2018:06 for longer horizon forecasting.

We collected real-time data for the US consumer price index (CPI), the real-world economic activity index of Kilian (2009), and the following variables provided by the Energy Information Association (EIA): the US RAC for imports, world crude oil production, US crude oil inventories, US petroleum inventories, and OECD petroleum inventories. The EIA publications provided real-time measurements over a variable window, up to 3 years prior to the most recent observation.

Following the conventional terminology in the real-time macroeconomic forecasting literature, we define a “vintage” of data as the historical time series observed by forecasters at a specific point in time (sometimes known as the “vintage date”). For example, the 2018:06 vintage includes observations only available at the end of June 2017. There are 319 vintages in total in the database, summarized in the documentation available together with the data from <https://www.niesr.ac.uk/real-time-forecast-combinations-oil-price>.

Following Baumeister and Kilian (2015), we used a point forecast combination methodology to mitigate issues of model misspecification. They combined point forecasts from six specifications using equal weights and inverse mean squared predictive error (MSPE) weights. The six specifications include: an unrestricted global oil market vector autoregression (VAR), a commodity price model, an oil price futures spread model, a gasoline spread model, a time-varying parameter (TVP) product spread, and a no-change benchmark model.

Baumeister and Kilian (2015) forecast the nominal crude oil price deflated by the US CPI, based on information at time t for the period $t + h$, $\hat{R}_{t+h|t}$, where h is the forecast horizon. We examine point forecast combinations based on the six different specifications:

$$\hat{R}_{t+h|t} = \sum_{k=1}^6 w_{k,t} \hat{R}_{t+h|t}^k, \quad (1)$$

where the weights, $w_{k,t}$, are assigned to model k at time t . Equal weights, $w_{k,t} = \frac{1}{6}$ and rolling and recursively estimated MSPE-based weights are used, where the latter are defined as

$$w_{k,t} = \frac{m_{k,t}^{-1}}{\sum_{j=1}^6 m_{j,t}^{-1}},$$

where $m_{k,t}^{-1}$ is the inverse MSPE of model k calculated with respect to observed outcomes available at time t .

With $\hat{R}_{t+h|t}^k$ denoting the forecast from the k th specification, the six models are described below.

1. An unrestricted global oil market vector autoregression (VAR):

$$\hat{R}_{t+h|t}^1 = \exp \left(\hat{r}_{t+h|t}^{\text{VAR}} \right), \quad (2)$$

where $\hat{r}_{t+h|t}^{\text{VAR}}$ is the forecast of logged real crude oil prices and the VAR has four variables: the percentage change in global crude oil production, the business cycle index of global real economic activity (rea) of Kilian (2009), the log of the RAC oil price deflated by the log of CPI, and the change in global crude oil inventories. Following Baumeister and Kilian (2015), the VAR lag length is 12 months. The WTI (Brent) forecasts are constructed as the forecasts of RAC scaled by the ratio of WTI (Brent) to RAC. We explore robustness to an alternative activity measure suggested by Hamilton (2018) in the subsequent section.

2. A commodity price-based model:

$$\hat{R}_{t+h|t}^2 = R_t^{\text{oil}} \left(1 + \pi_t^{h,\text{raw}} - E_t \left(\pi_{t+h}^h \right) \right), \quad (3)$$

where $\pi_t^{h,\text{raw}}$ is the difference between the log price of non-oil industrial raw materials at t and $t - h$, and R_t^{oil} is the real oil price measure. Following Baumeister and Kilian (2015), $E_t(\pi_{t+h}^h)$ is expected US inflation, based on the historical average for CPI inflation from 1986:07.

3. A futures-based model:

$$\hat{R}_{t+h|t}^3 = R_t^{\text{oil}} \left(1 + f_t^h - s_t - E_t \left(\pi_{t+h}^h \right) \right), \quad (4)$$

where s_t is the log of monthly WTI spot price, and f_t^h is the log of oil price futures for maturity h observed at t . WTI and RAC forecasts are based on WTI futures; Brent forecasts are based on Brent futures. The monthly oil futures prices for WTI are the average of daily futures closed prices collected from Bloomberg. There are missing values in the Bloomberg source for our evaluation sample, for monthly WTI oil futures at horizons greater than 17 months, and for Brent futures at horizons beyond 8 months. Baumeister and Kilian (2015, p. 341) gave zero weight to the futures-based forecasts at long horizons due to the missing values. To avoid dropping the futures-based forecasts in the combinations at these horizons, we fill in missing data by estimating a real-time factor-based model for crude oil price futures. Following Hevia, Petrella, and Sola (2016) and Garratt and Petrella (2018), we assume that futures prices are a function of two factors—the level and slope—and impose Nelson and Siegel's (1987) parametric restrictions to the loadings. A VAR(1) is assumed for the dynamics, estimation exploits the Kalman filter, and we use the estimated model to fill in the missing values. See the database documentation for further details.

4. A gasoline spread based model:

$$\hat{R}_{t+h|t}^4 = R_t^{\text{oil}} \exp \left\{ \hat{\beta} \left[s_t^{\text{gas}} - s_t \right] - E_t \left(\pi_{t+h}^h \right) \right\}, \quad (5)$$

where s_t^{gas} is the log spot price of gasoline and $\hat{\beta}$ is estimated from the regression $\Delta s_{t+h} = \beta[s_t^{\text{gas}} - s_t] + \varepsilon_{t+h}$, employing ordinary least squares. We denote the h -step-ahead log difference of spot WTI prices as $\Delta s_{t+h} = s_{t+h} - s_t$.

5. A time-varying parameter (TVP) product spread model:

$$\hat{R}_{t+h|t}^5 = R_t^{\text{oil}} \exp \left\{ \hat{\delta}_{1t} \left[s_t^{\text{gas}} - s_t \right] + \hat{\delta}_{2t} \left[s_t^{\text{heat}} - s_t \right] - E_t \left(\pi_{t+h}^h \right) \right\}. \quad (6)$$

The parameters $\hat{\delta}_{1t}$ and $\hat{\delta}_{2t}$ are estimated from

$$\Delta s_{t+h} = \delta_{1t} \left[s_t^{\text{gas}} - s_t \right] + \delta_{2t} \left[s_t^{\text{heat}} - s_t \right] + e_{t+h},$$

where s_t^{heat} is the log spot price of heating oil with the error term $e_{t+h} \sim \text{n.i.d.}(0, \sigma^2)$. The TVP model Bayesian estimation of gasoline and heating oil spreads employs an independent Normal-Wishart prior and the Gibbs sampler.

6. No change forecast, from the random walk model:

$$\hat{R}_{t+h|t}^6 = R_t^{\text{oil}}. \quad (7)$$

The no-change forecast is included in the forecast combinations, and is used as the benchmark.

Each specification uses a different sample, following Baumeister and Kilian (2015), to maximize the number of observations for parameter estimation. See Baumeister et al. (2018) for further details on spread models.

3 | RESULTS

Baumeister and Kilian (2015) evaluate the forecasts for the WTI and RAC real oil price measures from 1992:01 to 2012:09 using the 2013:03 data vintage as the target variable. Here we focus on the broader replication with the evaluation extended to 2017:12 for the monthly real Brent measure and also consider combinations that include the futures-based model. Results for the RAC and WTI measures (for the extended evaluation sample), the shorter evaluation time frame (for Brent), and the sensitivity of the inclusion of futures-based forecasts at longer horizons are presented in Supporting Information Appendix A (available online).

Table 1 reports the MSPE and success ratios of the point combination forecasts for various forecasting horizons (shown in the first column), evaluated on observations from 1992:01 to 2017:12, with the 2018:06 vintage data as the target real Brent price. The upper panel presents the end-evaluation MSPE ratios, relative to the no-change forecast. If the MSPE ratio is below 1, the combination forecast is more accurate than the benchmark. The lower panel presents end-evaluation success ratios. These describe the directional accuracy, with a success ratio higher than 0.5 indicating an improvement. The results for point forecast combinations with equal, recursive MSPE and rolling MSPE weights (with window sizes of 36, 24, and 12 months, respectively) are reported in the columns.

Echoing the WTI and RAC results on the shorter evaluation period reported by Baumeister and Kilian (2015), we find evidence of improvements in forecast accuracy from combinations for the Brent measure. The second column of Table 1 displays MSPE and success ratios consistent with improved accuracy (relative to the benchmark) for the equal weight combination at all forecast horizons from 1 to 24 months. The results using MSPE weights are similar to those for equal weights, regardless of whether the combinations are recursive (third column) or rolling (remaining columns).

Table 2 compares the accuracy of the equal weight combinations with and without futures-based forecasts for the 1992:01–2017:12 evaluation sample at horizons beyond 8 months for Brent. At shorter horizons there are no missing values

TABLE 1 Forecast accuracy for Brent, evaluation 1992:01 to 2017:12

MH	Real Brent price		Rolling weights based on windows of length		
	Equal weight	Recursive weights	36	24	12
<i>Recursive MSPE ratios</i>					
1	0.941** (0.030)	0.925** (0.003)	0.926** (0.003)	0.930** (0.005)	0.946** (0.042)
3	0.935** (0.005)	0.935** (0.004)	0.936** (0.004)	0.939** (0.005)	0.941** (0.005)
6	0.978* (0.093)	0.984 (0.147)	0.990 (0.241)	0.989 (0.204)	0.986 (0.177)
9	0.961** (0.007)	0.967** (0.017)	0.971** (0.023)	0.969** (0.024)	0.966** (0.026)
12	0.929** (0.000)	0.937** (0.000)	0.940** (0.000)	0.936** (0.000)	0.920** (0.000)
15	0.923** (0.000)	0.936** (0.000)	0.943** (0.000)	0.947** (0.000)	0.938** (0.000)
18	0.937** (0.000)	0.956** (0.003)	0.968** (0.028)	0.979* (0.100)	0.981 (0.194)
21	0.955** (0.002)	0.979* (0.076)	0.989 (0.234)	0.994 (0.349)	0.994 (0.391)
24	0.950** (0.002)	0.980* (0.094)	0.979* (0.096)	0.981 (0.107)	1.012 (0.730)
<i>Success ratios</i>					
1	0.526* (0.069)	0.545** (0.027)	0.548** (0.026)	0.538* (0.053)	0.551** (0.014)
3	0.561** (0.015)	0.548** (0.042)	0.568** (0.007)	0.568** (0.007)	0.574** (0.004)
6	0.534 (0.112)	0.521 (0.218)	0.524 (0.189)	0.528 (0.128)	0.547** (0.036)
9	0.569** (0.003)	0.566** (0.005)	0.549** (0.028)	0.539* (0.053)	0.556** (0.009)
12	0.605** (0.000)	0.598** (0.000)	0.601** (0.000)	0.598** (0.000)	0.628** (0.000)
15	0.621** (0.000)	0.621** (0.000)	0.634** (0.000)	0.614** (0.000)	0.631** (0.000)
18	0.586** (0.000)	0.586** (0.000)	0.559** (0.002)	0.549** (0.009)	0.539** (0.020)
21	0.558** (0.004)	0.538** (0.017)	0.555** (0.004)	0.524* (0.070)	0.521* (0.085)
24	0.540* (0.055)	0.488 (0.553)	0.536* (0.065)	0.509 (0.305)	0.512 (0.250)

Note. MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, p -values of a Harvey et al. (1997) small-sample adjustment of the Diebold and Mariano (1995) test are reported in parentheses after recursive MSPE ratios. We also report p -values for the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy in parentheses after success ratios. Asterisks denote significance at the *10% and **5% levels.

TABLE 2 Forecast accuracy for Brent, equal weight combinations, excluding and including futures-based forecasts (FUTURES), evaluation 1992:01 to 2017:12

MH	Recursive MSPE ratios		Success ratios	
	Excluding FUTURES	Including FUTURES	Excluding FUTURES	Including FUTURES
9	0.973* (0.096)	0.961** (0.007)	0.549** (0.030)	0.569** (0.003)
10	0.965** (0.033)	0.951** (0.001)	0.578** (0.002)	0.594** (0.000)
11	0.952** (0.006)	0.936** (0.000)	0.609** (0.000)	0.629** (0.000)
12	0.948** (0.003)	0.929** (0.000)	0.601** (0.000)	0.605** (0.000)
13	0.951** (0.003)	0.926** (0.000)	0.553** (0.029)	0.610** (0.000)
14	0.951** (0.003)	0.923** (0.000)	0.582** (0.002)	0.609** (0.000)
15	0.955** (0.005)	0.923** (0.000)	0.597** (0.000)	0.621** (0.000)
16	0.960** (0.012)	0.926** (0.000)	0.569** (0.007)	0.620** (0.000)
17	0.971** (0.047)	0.933** (0.000)	0.568** (0.005)	0.584** (0.000)
18	0.978 (0.106)	0.937** (0.000)	0.576** (0.001)	0.586** (0.000)
19	0.983 (0.161)	0.941** (0.000)	0.554** (0.011)	0.571** (0.001)
20	0.995 (0.372)	0.949** (0.001)	0.570** (0.004)	0.563** (0.002)
21	1.004 (0.591)	0.955** (0.002)	0.548** (0.047)	0.558** (0.004)
22	1.002 (0.553)	0.952** (0.001)	0.522 (0.240)	0.553** (0.013)
23	1.000 (0.494)	0.949** (0.001)	0.531 (0.158)	0.559** (0.009)
24	1.002 (0.547)	0.950** (0.002)	0.519 (0.284)	0.540* (0.055)

Note. MH represents monthly forecast horizons. Boldface indicates improvements relative to the no-change forecast. As a rough guide, *p*-values of a Harvey et al. (1997) small-sample adjustment of the Diebold and Mariano (1995) test are reported in parentheses after recursive MSPE ratios. We also report *p*-values for the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy in brackets after success ratios. Asterisks denote significance at the *10% and **5% levels.

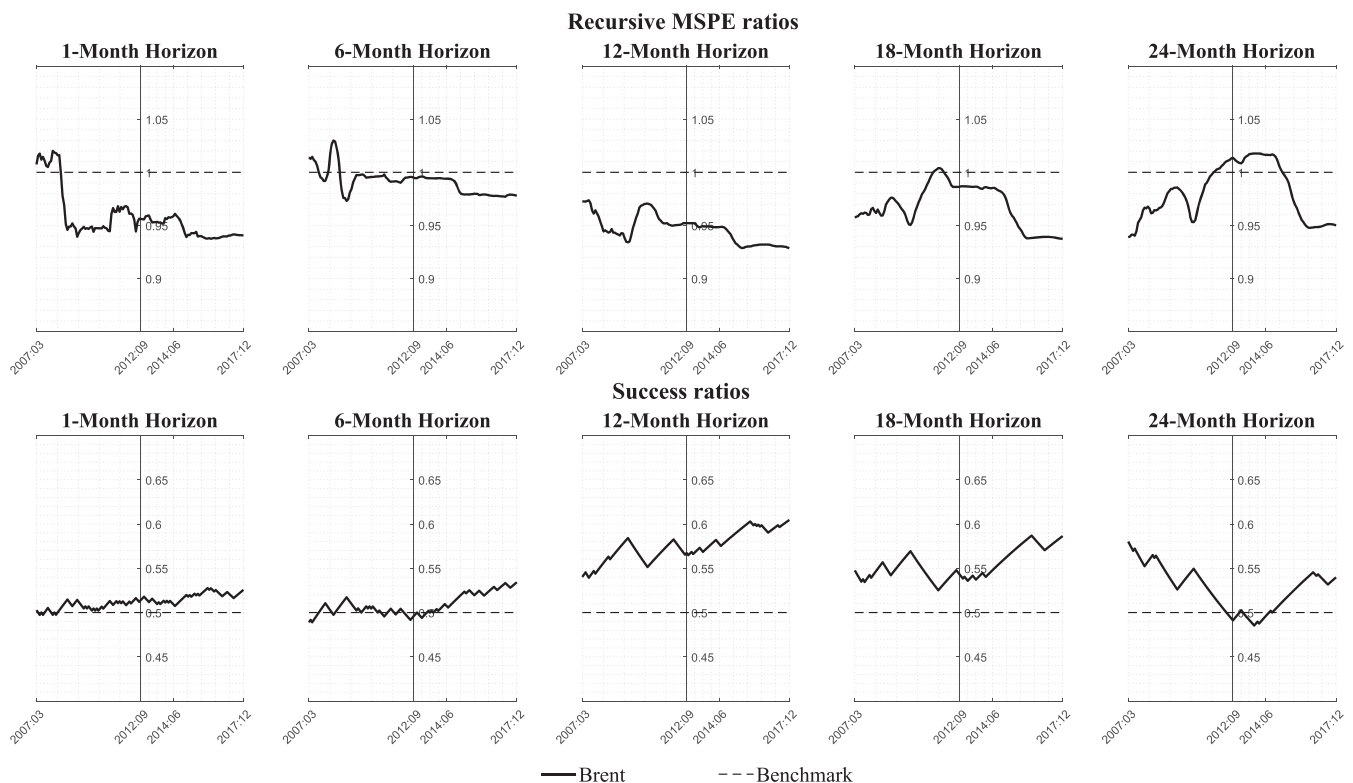


FIGURE 2 Recursive forecast accuracy of equal weighted combinations

on futures for Brent. Further details are provided in the Supporting Information Appendix. The inclusion of futures-based forecasts reduces MSPE ratios and raises the success ratios. As a rough guide, we note that the Harvey, Leybourne, and Newbold (1997) small-sample adjustment of the Diebold and Mariano (1995) test and the Pesaran and Timmermann (2009) test both imply stronger support. These results have to be taken with a grain of salt because the Diebold and Mariano (1995) test does not, strictly speaking, apply in our context (see ; Kilian, 2015).

Digging a little deeper into the real-time properties of the forecast combinations, Figure 2 plots the recursive MSPE ratios (top panel) and the recursive success ratios (bottom panel) of the equal weight combinations for selected horizons (1, 6, 12, 18, and 24 months), with end evaluation dates from 2007:03 to 2017:12. The 2012:09 end evaluation considered by Baumeister and Kilian (2015) sits in the middle of the x-axis for each cell. The equal weight combination is preferred if the line lies below 1 for the upper panel and above 0.5 for the lower panel. The recursive MSPE and success ratios consistently indicate that equal weight combinations dominate the benchmark before and after 2012:09 (with the exception of the 24-month horizon case for MSPE between 2012 and 2015).

As a further robustness check, we replicated our analysis with the 24-month cumulative growth rate of real shipping rates proposed by Hamilton (2018) as an alternative measure of global activity. Tables A-5, A-6 and A-7 in the Supporting Information Appendix correspond to Table 1 using this activity measure but not the real; the online database documentation describes the data details. We found strong evidence of improvements in forecast accuracy from the combinations regardless. See Kilian and Zhou (2018) and Kilian (2018) for discussions of various measures of activity.

4 | CONCLUSIONS

In this paper, we have broadly replicated the findings of Baumeister and Kilian using point forecast combinations to predict the real oil price in real time, based on an extended evaluation sample from 1992:01 to 2017:12. We found the accuracy of their point forecast combinations to be robust across different measures of the oil price and over various evaluation samples. We also confirmed that including futures-based information improves longer horizon forecasts. Subsequent researchers will find the real-time data set for this study particularly helpful when investigating new candidate models and methods for both point and density forecast combinations.

ACKNOWLEDGMENTS

We thank Lutz Kilian, Christiane Baumeister, Beili Zhu, Badi Baltagi, and an anonymous referee for helpful comments. Data, documentation and a not-for-publication Appendix are available from <https://www.niesr.ac.uk/real-time-forecast-combinations-oil-price>.

REFERENCES

- Alberola, E., Chevallier, J., & Chèze, B. (2008). Price drivers and structural breaks in European carbon prices 2005–2007. *Energy Policy*, 36(2), 787–797. <https://doi.org/10.1016/j.enpol.2007.10.029>
- Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business and Economic Statistics*, 30(2), 326–336. <https://doi.org/10.1080/107350015.2011.648859>
- Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business and Economic Statistics*, 33(3), 338–351. <https://doi.org/10.1080/07350015.2014.949342>
- Baumeister, C., & Kilian, L. (2016). Understanding the decline in the price of oil since June 2014. *Journal of the Association of Environmental and Resource Economists*, 3(1), 131–158. <https://doi.org/10.1086/684160>
- Baumeister, C., Kilian, L., & Lee, T. K. (2014). Are there gains from pooling real-time oil price forecasts? *Energy Economics*, 46, S33–S43. <https://doi.org/10.1016/j.eneco.2014.08.008>
- Baumeister, C., Kilian, L., & Zhou, X. (2018). Are product spreads useful for forecasting oil prices? An empirical evaluation of the Verleger hypothesis. *Macroeconomic Dynamics*, 22(3), 562–580. <https://doi.org/10.1017/S1365100516000237>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253–263. <https://doi.org/10.2307/1392185>
- Garratt, A., & Petrella, I. (2018). The informational content of commodity prices for probabilistic inflation forecasts. Unpublished manuscript.
- Hamilton, J. D. (2018). Measuring global economic activity. Department of Economics, 0508 University of California, San Diego 9500 Gilman Drive La Jolla, CA 92093-0508 (*Working paper*) Retrieved from <http://econweb.ucsd.edu/~jhamilto/REA.pdf>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hevia, C., Petrella, I., & Sola, M. (2016). Risk premia and seasonality in commodity futures. *Journal of Applied Econometrics*, 33(6), 853–873. <https://doi.org/10.1002/jae.2631>

- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3), 1053–1069. <https://doi.org/10.1257/aer.99.3.1053>
- 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. <https://doi.org/10.1080/07350015.2014.969430>
- Kilian, L. (2018). Measuring global economic activity: Reply. (*Working paper*). Ann Arbor, MI: University of Michigan Retrieved from http://www-personal.umich.edu/~lkilian/kilian_correction.pdf
- Kilian, L., & Zhou, X. (2018). Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance*, 88, 54–78. <https://doi.org/10.1016/j.jimonfin.2018.07.001>
- Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, 23(3), 325–338. [https://doi.org/10.1016/S0140-9883\(00\)00075-X](https://doi.org/10.1016/S0140-9883(00)00075-X)
- Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, 60(4), 473–489.
- Pesaran, M. H., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485), 325–337.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Garratt A, Vahey SP, Zhang Y. Real-time forecast combinations for the oil price. *J Appl Econ*. 2019;34:456–462. <https://doi.org/10.1002/jae.2673>