Baumeister, Christiane; Kilian, Lutz

Working Paper

Forecasting the real price of oil in a changing world: A forecast combination approach

CFS Working Paper, No. 2013/11

Provided in Cooperation with:
Center for Financial Studies (CFS), Goethe University Frankfurt

CENTER FOR FINANCIAL STUDIES

CFS WORKING PAPER

No. 2013/11

Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach

Christiane Baumeister, Lutz Kilian

Center for Financial Studies
Goethe University ■ House of Finance
Grüneburgplatz 1 ■ 60323 Frankfurt am Main

Tel: +49 69 798-30050
Fax: +49 69 798-30077
www.ifk-cfs.de ■ info@ifk-cfs.de

CENTER FOR FINANCIAL STUDIES

CFS Working Paper Series

The *Center for Financial Studies*, located in Goethe University's House of Finance in Frankfurt, is an independent non-profit research center, funded by the non-profit- making organisation Gesellschaft für Kapitalmarktforschung e.V. (GfK). The CFS is financed by donations and by contributions of the GfK members, as well as by national and international research grants. The GfK members comprise major players in Ger- many's financial industry. Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. CFS is also a contributor to policy debates and policy analyses, building upon relevant findings in its research areas.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center´s Research Fellow Program or members of one of the Center´s Research Projects.

If you would like to know more about the Center for Financial Studies, please let us know of your interest.

Prof. Michalis Haliassos, Ph.D. | Prof. Dr. Jan Pieter Krahnen | Prof. Dr. Uwe Walz

Center for Financial Studies
Goethe University ■ House of Finance
Grüneburgplatz 1 ■ 60323 Frankfurt am Main

Tel: +49 69 798-30050
Fax: +49 69 798-30077
www.ifk-cfs.de ■ info@ifk-cfs.de

Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach

November 13, 2013

Christiane Baumeister
Bank of Canada

Lutz Kilian
University of Michigan

Abstract: The U.S. Energy Information Administration (EIA) regularly publishes monthly and quarterly forecasts of the price of crude oil for horizons up to two years, which are widely used by practitioners. Traditionally, such out-of-sample forecasts have been largely judgmental, making them difficult to replicate and justify. An alternative is the use of real-time econometric oil price forecasting models. We investigate the merits of constructing combinations of six such models. Forecast combinations have received little attention in the oil price forecasting literature to date. We demonstrate that over the last 20 years suitably constructed real-time forecast combinations would have been systematically more accurate than the no-change forecast at horizons up to 6 quarters or 18 months. MSPE reduction may be as high as 12% and directional accuracy as high as 72%. The gains in accuracy are robust over time. In contrast, the EIA oil price forecasts not only tend to be less accurate than no-change forecasts, but are much less accurate than our preferred forecast combination. Moreover, including EIA forecasts in the forecast combination systematically lowers the accuracy of the combination forecast. We conclude that suitably constructed forecast combinations should replace traditional judgmental forecasts of the price of oil.

JEL Code: Q43, C53, E32
Key Words: Forecast combination; real-time data; model misspecification; structural change; oil price.

Acknowledgements: The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Bank of Canada. Argyn Toktamyssov provided excellent research assistance. We thank Ron Alquist, Olivier Coibion, the associate editor, and the referees for helpful discussions.

Corresponding author: Lutz Kilian, Department of Economics, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA. Email: lkilian@umich.edu.

Christiane Baumeister, International Economic Analysis Department, Bank of Canada, 234 Laurier Avenue West, Ottawa, ON K1A 0G9, Canada. Email: CBaumeister@bankofcanada.ca.

1. Introduction

Since long-term oil contracts were abandoned around 1980, one of the most challenging forecasting problems has been how to forecast the price of crude oil. One of the few regular producers of oil price forecasts has been the U.S. Energy Information Administration (EIA), which constructs monthly and quarterly forecasts of the price of crude oil for horizons up to two years. EIA oil price forecasts help guide natural resource development and investments in infrastructure. They also play an important role in preparing budget and investment plans. Users of oil price forecasts include international organizations, central banks, governments at the state and federal level as well as a range of industries including utilities and automobile manufacturers.

Traditionally, the EIA's short-term oil price forecasts have been largely judgmental, making them difficult to replicate and justify. Nor have these forecasts been particularly successful when compared with naïve no-change forecasts, as documented in Alquist, Kilian, and Vigfusson (2013). Indeed, many pundits have suggested that changes in the price of oil are inherently unforecastable and that attempts to forecast the price of crude oil are pointless. These agnostics view the current price of oil as the best forecast of future oil prices (see, e.g., Davies 2007, Hamilton 2009). In recent years, however, a number of new econometric forecasting models has been introduced in the literature and has been shown to be more accurate at some horizons than the no-change forecast of the real price of oil even after taking account of real-time data constraints (see, e.g., Baumeister and Kilian 2012, 2013a; Baumeister, Kilian and Zhou 2013).[1] There is no methodology available in the literature that does well at all horizons for which the EIA produces oil price forecasts, however.

In this paper, we explore the question of whether one can improve on both the no-change forecast and the EIA's own judgmental oil price forecasts at the horizons of interest to the EIA. These horizons are

[1] In recent years there has been a resurgence in research on the question of how to forecast the price of commodities in general and the price of oil in particular, at least at horizons up to a year. This literature has examined in depth the predictive power of oil futures prices, the predictive content of changes in oil inventories, oil production, macroeconomic fundamentals, product spreads, and exchange rates as well as the forecasting ability of professional and survey forecasts. Other contributors to this literature include Chernenko, Schwarz, and Wright 2004; Knetsch 2007; Sanders, Manfredo, and Boris 2008; Alquist and Kilian 2010; Chen, Rogoff, and Rossi 2010; Reeve and Vigfusson 2011; Baumeister and Kilian 2013b; and Chinn and Coibion 2013).

longer than those typically examined in earlier studies of oil price forecasts. Given that none of the existing oil price forecasting methods does well at all of these horizons, we investigate the merits of constructing forecasts from suitably chosen combinations of six oil price forecasting models that feature prominently in the recent literature. The set of models includes forecasts from vector autoregressive (VAR) models of the global oil market, forecasts based on recent changes in the price index of non-oil industrial raw materials, forecasts based on West Texas Intermediate (WTI) oil futures prices, the no-change forecast, forecasts based on the spread of the U.S. spot price of gasoline relative to the WTI spot price of crude oil, and a time-varying parameter forecasting model that allows the U.S. gasoline spot spread and the U.S. heating oil spot spread to contribute to the oil price forecast with smoothly varying weights.

Each of these models have been shown to produce more accurate oil price forecasts than the no-change forecast at least at some horizons. Some of these models generate low MSPEs at short horizons; others at somewhat longer horizons. We consider combinations of these forecasting models with fixed equal weights as well as combinations with weights that reflect each model's recent forecasting success. Unlike previous studies we give equal weight to the problem of generating quarterly and monthly forecasts. Indeed, the performance of several of the forecasting models included in our study (such as forecasts based on industrial commodity prices and forecasts based on product spreads) has not previously been evaluated at quarterly horizons.

Our objective is to generate monthly and quarterly oil price forecasts that do not require judgment and are available in real time. We restrict attention to forecast horizons between 1 month and 24 months and between 1 quarter and 8 quarters, consistent with the objective of the EIA. All models are estimated recursively, as is standard in this literature, and subject to real-time data constraints.[2] The weights attached to each forecast are constructed in real time as well. Forecast combinations have received little attention in the oil price forecasting literature to date because until recently oil price forecasters had few

[2] See Baumeister and Kilian (2013a) for a comparison of oil price forecasts based on rolling and recursive regression estimates.

models available to them. With the proliferation of suitable models, forecast combinations are a promising way forward for four reasons. First, even the most accurate forecasting models do not work equally well at all times. The Baumeister and Kilian (2012) oil price forecasting model, for example, works well during times when economic fundamentals show persistent variation, as was the case between 2002 and 2011, but less well at other times. Likewise, there is considerable variation over time in the ability of oil futures prices to forecast the price of oil.

Second, previous research has shown that some forecasting models are more accurate at short horizons and others at longer horizons. For example, forecasting models based on economic fundamentals tend to enjoy superior accuracy at horizons up to 3 months, whereas models based on the spread of refined product prices relative to the price of crude oil tend to be most accurate at horizons between 12 and 24 months.

Third, even the forecasting model with the lowest MSPE may potentially be improved upon by incorporating information from other models with higher MSPE. For example, Baumeister and Kilian (2013a) illustrated that simple equal-weighted averages of quarterly forecasts based on oil futures prices and quarterly forecasts based on VAR models of the global oil market are more accurate at short horizons than either model alone. That evidence, while suggestive, is no substitute for a systematic study of the usefulness of forecast combinations among a wider range of oil price forecasting models. Whether forecast combinations will improve on the single most accurate model, is by no means a foregone conclusion. Baumeister and Kilian (2013a), for example, found that one and only one forecast combination among the individual forecasting models they considered improved on its individual components. Which combination will work best cannot be inferred from the MSPE rankings of individual models. We present a simple algorithm that allows a systematic approach to this problem.

Fourth, one can think of forecast combinations as providing insurance not only against possible model misspecification, but also against smooth structural change. Such structural change may arise from changes in market structure, in the structure of the global economy, or in the accessibility of crude oil, for example.

Our analysis focuses on both the U.S. refiners' acquisition cost for crude oil imports, which is commonly viewed as a proxy for the global price of oil, and the spot price of West Texas Intermediate (WTI) crude oil frequently cited in the press. There are two problems with modeling WTI prices. One is that the WTI spot price was subject to government regulation until the early 1980s and hence not representative for the market price of oil. The other problem is that the WTI price has suffered from structural instability since 2011, when restrictions on U.S. crude oil exports and transportation bottlenecks prevented arbitrage between the WTI price in the United States and the price of Brent crude oil in the United Kingdom. As a consequence, generating WTI forecasts in some cases requires suitable modifications of the baseline forecasting model (see Baumeister and Kilian 2013a).[3]

Although the accuracy of the forecast combinations remains remarkably robust across alternative specifications, our results indicate that inverse MSPE weights based on recursive or rolling windows of past data generate less accurate forecasts than constant equal weights. We also assess the contribution of each model to the accuracy of a given equal-weighted forecast combination and demonstrate that in practice two of the six models in the forecast combination provide little value added and may be dropped. The four models that appear essential are the oil market VAR model, the model based on non-oil commodity prices, the model based on oil futures spreads, and the time-varying product spread model. In contrast, allowing the combination to assign positive weight to the no-change forecast actually increases the MSPE at all horizons up to 18 months (or 6 quarters), providing a powerful argument against the agnostic position that the real price of oil is unforecastable. The same applies to the gasoline spread model. Dropping these two models further improves accuracy at horizons up to 18 months (or up to 6 quarters), but at the expense of somewhat lower accuracy at longer horizons.

[3] We do not report results for the Brent price of crude oil. The reason is that there are no suitable data available for applying some of the forecasting models considered in this paper. For example, there do not exist long enough time series for Brent futures prices at longer maturities, and no suitable spot price data are available for the Rotterdam gasoline and heating oil markets. Without access to these data there is no reason to expect forecast combinations to replicate the successes reported in this paper for other oil price measures. We note, however, that the Brent price has remained stable in relation to the U.S. refiners' acquisition cost for oil imports even in recent years, so to some extent our results for the refiners' acquisition cost are expected to be representative for the Brent price and could be mapped into forecasts of the Brent price, as discussed in Baumeister and Kilian (2013a).

Our evidence shows that over the last twenty years suitably constructed real-time forecast combinations would have been more accurate than the no-change forecast at every horizon up to 18 months (or up to 6 quarters). The preferred forecast combination works equally well for monthly and for quarterly forecasts, and its performance has been quite robust over time. For example, relative to the no-change forecast, quarterly forecast combinations may reduce the MSPE by up to 12%. They also have statistically significant directional accuracy as high as 72%. Using monthly forecast combinations, the MSPE reductions may be as high as 13% and directional accuracy as high as 65%. The overall pattern of results is similar in monthly and quarterly data. Forecast combinations help improve accuracy both for the real U.S. refiners' acquisition cost for crude oil imports and for the real WTI price.

By comparison the judgmental EIA forecasts, to the extent that consistent time series for these forecasts can be constructed, are considerably less accurate. We show in particular that the quarterly EIA forecasts not only tend to be less accurate than the no-change forecast, but are much less accurate than the preferred forecast combination. Moreover, including these EIA forecasts in the forecast combination systematically increases the MSPE of the combination forecast. We conclude that suitably constructed forecast combinations should replace traditional judgmental forecasts of the real price of oil.

The remainder of the paper is organized as follows. In section 2, we briefly review the six forecasting models to be combined in Section 3 and the data employed in constructing the forecasts. In section 3, we investigate how to choose the weights used in combining monthly oil price forecasts. Section 4 formally evaluates whether the forecast accuracy may be improved further by dropping one or more of the models under consideration. We also examine how the gains in accuracy of this combination forecast accumulate over time and show that the results are systematic rather than being driven by outliers. Section 5 extends the analysis to forecasts of the real price of oil at quarterly horizons. In section 6 we evaluate the EIA forecasts and compare them to the no-change forecast. We also examine how adding the EIA forecasts to the forecast combination affects the forecast accuracy. We conclude in section 7.

2. Six Approaches to Forecasting the Real Price of Oil

In this section we briefly review the forecasting models to be combined in section 3. Our focus is on forecasts of the real price of oil at horizons between 1 and 24 months. The maximum forecast horizon is dictated by the needs of the EIA. We also follow the EIA in focusing on monthly averages of the price of oil. Each of the models below has been shown to generate more accurate real-time out-of-sample forecasts than the no-change forecast at least at some forecast horizons, although not all models have been examined at horizons beyond 12 months. All models are estimated subject to real-time data constraints using a suitably extended and updated version of the real-time database developed for Baumeister and Kilian (2012).

2.1. Forecast based on a VAR Model of the Global Oil Market

The first model is a reduced-form VAR model that includes the key variables relevant to the determination of the real price of oil in global markets:

$$B(L)y_t = \nu + u_t$$

where $y_t = \left[\Delta prod_t, rea_t, r_t^{oil}, \Delta inv_t\right]'$ refers to a vector including the percent change in global crude oil production, a measure of global real activity, the log of the U.S. refiners' acquisition cost for crude oil imports deflated by the log of the U.S. CPI, and the change in global crude oil inventories, ν denotes the intercept, $B(L)$ denotes the autoregressive lag order polynomial, and u_t is a white noise innovation. The inventory data are constructed by multiplying U.S. crude oil inventories by the ratio of OECD petroleum inventories to U.S. petroleum inventories. Petroleum inventories are defined to include both stocks of crude oil and stocks of refined products. The inventory data are from the EIA's Monthly Energy Review, which also provides data on global oil production and the refiners' acquisition cost. The global real activity index is constructed from data on global dry cargo ocean shipping freight rates as described in Kilian (2009).

This VAR model may be viewed as the reduced-form representation of the structural global oil market model developed in Kilian and Murphy (2013). Its forecast accuracy has been examined extensively in the literature (see, e.g., Alquist et al. 2013; Baumeister and Kilian 2012, 2013a). Throughout the paper we estimate the unrestricted VAR model with 12 autoregressive lags by the method of least squares.[4] Forecasts $\hat{r}_{t+h|h}$ of the log of the real price of oil are constructed iteratively from the estimated VAR model conditional on the most recent data and converted to levels, resulting in the forecast

$$\hat{R}^{oil}_{t+h|t} = \exp(\hat{r}^{oil,VAR}_{t+h|t}) \tag{1}$$

Forecasts for the real WTI price are constructed from the same VAR model by assuming that the most recent spread between the log WTI price and the log of the U.S. refiner's acquisition cost remains unchanged in the future. By rescaling the forecasts of the U.S. refiners' acquisition cost in this manner, we allow the relationship between the two oil price measures to evolve as a random walk. This approach has been shown to be more accurate than the simpler approach of replacing the U.S. real refiners' acquisition cost in the VAR model by the real WTI price (see Baumeister and Kilian 2013a). The WTI spot price data are from the FRED database of the Federal Reserve Bank of St. Louis, which also provides the real-time U.S. CPI data used to deflate the two measures of the nominal price of oil.

2.2 Forecast based on the Price of Non-Oil Industrial Raw Materials

Much of the empirical success of VAR forecasting models of the real price of oil can be traced to the use of measures of global real economic activity that help capture fluctuations in the demand for industrial commodities. A much simpler forecasting method – based on the same intuition that there are broad-based predictable shifts in the demand for globally traded commodities – exploits real-time information from recent cumulative changes in non-oil industrial commodity price indices. As discussed in Baumeister and Kilian (2012), a forecast of the real price of oil may be constructed as follows:

[4] Similar results would be obtained by imposing standard Bayesian priors in estimation.

$$R^{oil}_{t+h|t} = R^{oil}_t \left(1 + \pi_t^{h,industrial\ raw\ materials} - E_t(\pi_t^h)\right), \qquad (2)$$

where $\pi_t^{h,industrial\ raw\ materials}$ stands for the percent change of an index of the spot price of industrial raw materials (other than oil) over the preceding h months. This index is available in real time from the Commodity Research Bureau. The term $E_t(\pi_t^h)$ is the expected inflation rate over the next h periods. In practice, this expectation is proxied by recursively constructed averages of past U.S. inflation data, starting in July 1986.[5]

2.3. No-Change Forecast

Baumeister and Kilian (2012) made the point that forecasting models based on economic fundamentals such as the VAR model above perform best during times of persistent and hence predictable fluctuations in economic fundamentals. That is why the VAR forecasting model does particularly well before, during and after the Great Recession of 2008. In contrast, during other times the model is only about as accurate as the no-change forecast. Indeed, there are indications that, since 2011, we have once again entered a period during which VAR models offer at best minimal gains relative to the no-change forecast and that this situation will persist as long as the world economy stagnates. This observation raises the question of whether we would be better off at times if we replaced the VAR forecast by the no-change forecast or at least downweighted the VAR forecast relative to the no-change forecast. This line of reasoning suggests that we want to allow for the forecast combination to put positive weight on the no-change forecast:

$$\hat{R}^{oil}_{t+h|t} = R^{oil}_t, \qquad (3)$$

where R^{oil}_t denotes the real price of oil in levels (as opposed to logs).

2.4. Forecast based on Oil Futures Prices

Yet another approach is to exploit information from oil futures markets. Many practitioners rely on the price of oil futures contracts in generating forecasts of the nominal price of oil. This forecast can then be

[5] Undoubtedly, the inflation forecast could be refined further, but there is little loss in generality in our approach, given that fluctuations in the nominal price of oil dominate the evolution of the real price of oil.

converted to a forecast for the real price of oil by subtracting expected inflation. This approach is embodied in the forecasting model

$$R_{t+h|t}^{oil} = R_t^{oil}\left(1 + f_t^h - s_t - E(\pi_t^h)\right), \tag{4}$$

where R_t denotes the current level of the real price of oil, f_t^h is the log of the current WTI oil futures price for maturity h, s_t is the corresponding WTI spot price, and $E_t(\pi_t^h)$ is again the expected inflation rate over the next *h* periods. Both f_t^h and s_t are available in real time. The oil futures prices are obtained from Bloomberg. Although forecasts based on (4) are not significantly more accurate than no-change forecast at horizons of 1, 3, or 6 months (and sometimes less accurate), especially in recent years the accuracy of futures-based forecasts at horizons of 9 and 12 months has improved. In this paper, we use the monthly WTI oil futures price data up to a horizon of 18 months, which is the maximum horizon for which the construction of continuous monthly time series is feasible, given our evaluation period. This means that for horizons beyond 18 months the futures-based forecast receives zero weight in the forecast combinations we construct in sections 3 through 5.

2.5. Spread between the Spot Prices of Gasoline and Crude Oil

Another promising class of oil price forecasting models involves the use of product spreads. Many market practitioners believe that a rising spread between the price of gasoline and the price of crude oil signals upward pressures on the price of crude oil. For example, Goldman Sachs in April 2013 cut its oil price forecast, citing significant pressure on product spreads, which it interpreted as an indication of reduced demand for products (see Strumpf 2013). Such a forecasting model was derived in Baumeister et al. (2013):

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

where s_t^{gas} is the log of the nominal U.S. spot price of gasoline, s_t is the log of the spot price of WTI crude oil as defined earlier, and $\hat{\beta}$ is obtained from estimating the model

$$\Delta s_{t+h|t} = \beta\left[s_t^{gas} - s_t\right] + \varepsilon_{t+h}$$

recursively by the method of least squares. It can be demonstrated that imposing an intercept of zero, as shown in (5), greatly enhances the out-of-sample accuracy of this model. This gasoline spread model greatly improves on the accuracy of a no-change forecast especially at horizon beyond one year, making it a natural complement to models based on economic fundamentals which are most accurate at shorter horizons. The gasoline spot price data are readily available in real time from the EIA. For further details see Baumeister et al. (2013).

2.6. Time-Varying Parameter Model of the Gasoline and Heating Oil Spreads

The simplicity of the forecast based on the gasoline spread is appealing, yet there are reasons to be wary. One concern is that the price of crude oil is likely to be determined by the refined product in highest demand. According to Verleger (2011), traditionally, in the United States this product has been gasoline, but more recently it has been heating oil (which is nearly equivalent to diesel fuel), suggesting a forecasting model that allows for both a gasoline spread and a heating oil spread with time-varying coefficients. Another concern is that crude oil supply shocks, local capacity constraints in refining, changes in environmental regulations, or other market turmoil may all temporarily undermine the predictive power of product spreads. These considerations motivate the following generalization of model (5), introduced in Baumeister et al. (2013).

We first recursively estimate the time-varying regression model

$$\Delta s_{t+h|t} = \beta_{1t}\left[s_t^{gas} - s_t\right] + \beta_{2t}\left[s_t^{heat} - s_t\right] + \varepsilon_{t+h}$$

where the additional variable s_t^{heat} is the log of the nominal U.S. spot price of heating oil. The product prices are from the EIA. For details on the data sources the reader is referred to Baumeister et al. (2013). In estimating the model, we postulate that $\varepsilon_{t+h} \sim NID(0, \sigma^2)$, while the time-varying coefficients $\theta_t = \left[\beta_{1t} \quad \beta_{2t}\right]'$ evolve according to a random walk as $\theta_t = \theta_{t-1} + \xi_t$, and ξ_t is independent Gaussian white noise with variance Q. The intercept has again been restricted to zero, following Baumeister et al. (2013)

who show that this restriction greatly improves the out-of-sample accuracy. This state-space model is estimated using a Gibbs sampling algorithm. The conditional posterior of θ_t is normal, and its mean and variance can be derived via standard Kalman filter recursions (see Kim and Nelson 1999). Conditional on an estimate of θ_t, the conditional posterior distribution of σ^2 is inverse Gamma and that of Q is inverse Wishart.

Given the TVP estimates, we then construct the TVP model forecast:

$$\hat{R}^{oil}_{t+h|t} = R^{oil}_t \exp\left\{\hat{\beta}_{1t}\left[s^{gas}_t - s_t\right] + \hat{\beta}_{2t}\left[s^{heat}_t - s_t\right] - E_t(\pi^h_t)\right\} \tag{6}$$

by Monte Carlo integration as the mean of the forecasts simulated based on 1,000 Gibbs iterations conditional on the most recent data. Our forecasts take into account that the model parameters continue to drift over the forecast horizon according to their law of motion. The first 30 observations of the initial estimation period are used as a training sample to calibrate the priors and to initialize the Kalman filter.

This TVP product spread model has been shown to be systematically more accurate than the no-change forecast, especially at horizons beyond one year. At some horizons it produces forecasts even more accurate than the gasoline spread forecast (5). Hence, there is reason to believe that this approach may have additional predictive information not already contained in the simpler gasoline spread model.

3. Baseline Results

Knowing ex post that one or the other forecasting method would have been more accurate is not of much use to applied forecasters. The challenge is to be able to detect in real time when one model should be downweighted compared to another. A natural approach to measuring the real-time forecast accuracy of competing models is to construct inverse MSPE weights based on the recent forecasting performance of each model. These weights may then be used to construct a suitable weighted average of the forecasting models in question. This forecast combination approach has a long tradition in econometrics (see, e.g., Diebold and Pauly 1987; Stock and Watson 2004). The smaller the MSPE of a model is at date t, the larger the weight that this model receives in the combination forecast

$$\hat{R}^{oil}_{t+h|t} = \sum_{k=1}^{6} \omega_{k,t} \hat{R}^{oil,k}_{t+h|t}, \qquad \omega_{k,t} = \frac{m_{k,t}^{-1}}{\sum_{j=1}^{6} m_{j,t}^{-1}},$$

where $m_{k,t}$ is the recursive MSPE of model k in period t. In practice, the MSPE estimates must be initialized. We proceed by assigning equal weight to each model when entering the evaluation period. For subsequent periods, we then recursively update the MSPE of each model. The advantage of inverse MSPE weights is that they allow the forecast combination to adjust according to the recent MSPE of each model. An alternative much simpler approach is to impose equal weights of 1/6 to each forecast throughout the sample. The latter approach still provides some insurance against forecast breakdowns and forecasting model misspecification, but does not explicitly allow for structural change.

All forecasting models are evaluated on the same evaluation period of 1992.1-2012.9. Using such a long evaluation periods reduces the odds of spurious fits. The initial estimation period ends in 1991.12. Some forecasting models such as the VAR models are estimated on data back to 1973.2. For other forecasting models, the estimation period starts much later, reflecting the availability of the data. For example, monthly spot prices for gasoline and heating oil are available only starting in 1986. It is important to stress again that our data are in many cases subject to real-time data constraints. Where appropriate we rely on an updated and extended version of the real-time data base developed at the Bank of Canada for the purpose of forecasting oil prices (see Baumeister and Kilian 2012, 2013a,b). We use the real price of oil in the March 2013 vintage up to September 2012 as a proxy for the ex-post revised data, against which all forecasts are evaluated.

Below we assess the accuracy of various forecast combinations based on their recursive MSPE over the evaluation period (expressed as a ratio relative to the MSPE of the no-change forecast). MSPE ratios below 1 mean that the forecast in question is more accurate than the no-change forecast. We also examine the directional accuracy of the forecast combinations. Under the null hypothesis of no directional accuracy, the model should be no more successful at predicting the direction of change in the price of oil than would be tossing a fair coin with success probability 0.5, so any success ratio higher than

0.5 indicates an improvement over the no-change forecast. Tests of the null of no directional accuracy are conducted using the test of Pesaran and Timmermann (2009). There are to our knowledge no tests for the statistical significance of MSPE reductions for estimated forecast combinations. Standard tests in the literature are based on the premise that we compare the same two models at each point in time. Given that the model weights used in the combination evolve over time that premise appears violated.

It is useful to start with the evidence for the real U.S. refiners' acquisition cost for crude oil imports. The first column of Table 1 shows results for the equal-weighted combination of all six forecasting models. Each model receives weight 1/6 throughout. There is clear evidence of reductions in the MSPE relative to the no-change forecast up to a horizon of 18 months. The highest reduction is 9%. At horizons of 21 and 24 months, the equal-weighted forecast combination has an MSPE about as high as the no-change forecast. The equal-weighted forecast combination also has significant directional accuracy at horizons as long as 18 months. The highest success ratio is 65%. It is worth pointing out that some of the models included in the forecast combination achieve even larger gains at some horizons, but none produce consistently large accuracy gains across all horizons up to 18 months. This pattern of results is not specific to the real U.S. refiner's acquisition cost for crude oil imports. The sixth column in Table 1 illustrates that similar, if somewhat weaker, results hold when forecasting the real WTI price. The reductions in the recursive MSPE are as high as 9% and the success ratios are as high as 61% and often statistically significant.

An obvious question is whether we can improve on these results by weighting each model based on its recent forecasting success. The second column shows that the forecast combination based on recursively estimated inverse MSPE weights again systematically reduces the recursive MSPE at all horizons but horizon 21; in the latter case, it is almost as accurate as the no-change forecast. Overall the results are similar to those in column (1), but there is a slight reduction in accuracy. For example, the reductions in the recursive MSPE tend to be slightly lower. They biggest reduction is 9%. The highest success ratio drops to 63%, and the degree of statistical significance is lower than in column (1). A similar slight reduction of the forecasting accuracy can be observed in column (7) for the real WTI price.

Recursively constructed forecast combination weights may not be optimal when dealing with smooth structural change in the global oil market. In the latter case, a more natural approach would be to use rolling windows in estimating the weights. An obvious question concerns the length of these windows. The more pronounced the structural change (or the less stable the individual forecasting models), the shorter the window length should be. At the same time, the window cannot be too short without the estimates of the weights becoming too noisy. In Table 1 we experimented with three window lengths: 36, 24 and 12 months. Columns (3) through (5) of Table 1 relate to the real U.S. refiner's acquisition cost for crude oil imports. They show that, while the results remain remarkably robust across specifications, the use of rolling weights tends to undermine forecast accuracy. The shorter is the window length, the less accurate are the results. The same pattern can be seen for the real WTI price in columns (8) through (10).

Overall, neither recursively constructed weights nor weights based on rolling windows of data can be recommended. The reason for this perhaps unexpected result is that these weights must be constructed based on the fit of the model estimated $h+6$ periods ago, where h denotes the forecast horizon in months, and the additional delay of 6 months refers to the delays in the availability of reliable data. Although one could instead construct weights based on the nowcasts for data that are not released yet and based on preliminary data to the extent available, we found that this strategy does not improve forecast accuracy and cannot be recommended. It can be shown that if forecasters had access to the ex-post revised data in constructing forecast combination weights, the accuracy of the combination of real-time forecasts would increase considerably at all horizons. This point is moot, however, because this information is not available in real time, making equal-weighted forecast combinations the preferred choice in practice.

4. Sensitivity Analysis

Our baseline combination involves six forecasting models. A question of practical interest is whether all of these models are required or whether the set of models may be reduced further. This question may be addressed by recomputing the accuracy of the forecast combination, having eliminated one model at a time from the equal-weighted forecast combination, as shown in Table 2. In each case, the weights are accordingly set to 1/5. Our point of departure is the result in column (1) of Table 1. Evidence that dropping one of the six models systematically lowers the recursive MSPE ratio would be an indication that this model ought to be eliminated from the forecast combination, if we care about the MSPE outcomes. We illustrate this approach in Table 2 for the case of the real U.S. refiners' acquisition cost for oil imports.

The first column of Table 2 shows that leaving out the VAR model all else equal raises the MSPE ratio compared with column (1) in Table 1 at short horizons, reaffirming our decision to include this model in the forecast combination, while lowering it at longer horizons. The second and third columns of Table 2 provide evidence that the futures-based forecast helps improve the accuracy of the forecast combination at intermediate horizons, while the forecast based on the non-oil commodity price model contributes at short horizons. In sharp contrast, the fourth column shows that the recursive MSPE may be lowered at all but the last two forecast horizons, if we eliminate the no-change model from the forecast combination. This result contradicts the rationale we provided in section 2 for including the no-change forecast in the forecast combination. A similar systematic improvement in the recursive MSPE can be observed after dropping the gasoline spread model in the fifth column, but not for the TVP product spread model in the last column, which contributes to the overall forecast accuracy especially at longer horizons. This evidence suggests that, at horizons up to 18 months, the gasoline spread model adds nothing beyond the predictive information in the TVP product spread model and should be eliminated from the forecast combination.

Similar results (not shown) also hold for the real WTI price, the main difference being that the VAR model contributes at horizons as long as 21 months, while the commodity price model is useful only at short horizons. As in Table 2, the oil futures spread model contributes to the accuracy of the forecast combination mainly at medium-term horizons, whereas the TVP product spread model contributes at medium and longer horizons.

Table 3 takes the analysis a step further by eliminating both the no-change forecast and the gasoline spread model from the equal-weighted forecast combination. The first column of Table 3 confirms that eliminating both models further improves the accuracy of the forecast combination at most horizons up to 18 months, at the cost of worsening it at longer horizons. A forecast combination consisting only of the VAR model, the commodity price model, the oil futures spread model, and the TVP product spread model, has lower recursive MSPE than the no-change forecast at all horizons from 1 month to 18 months. The reductions in the MSPE range from 4% to 13%. The improvements in directional accuracy are statistically significant at all but one horizon and range from 55% to 65%, depending on the horizon. Similar results hold for the real WTI price, as shown in the second column of Table 3, albeit with somewhat lower directional accuracy. The highest success ratio is still 62%, but the directional accuracy is statistically significant at only three of the first seven horizons considered.

Because all of these results were achieved in real time, the evidence in Tables 1 and 3 shows that there is a practical alternative to the construction of judgmental forecasts of the real price of oil. Whether the models in Table 3 are preferred relative to columns (1) and (6) in Table 1, depends on how much we care about forecasting beyond horizons of 18 months. Either way, an important question is whether the recursive MSPE reductions shown in Tables 1 and 3 are driven by one or two unusual episodes in the data or whether they are more systematic. Figure 1 addresses this question by plotting the recursive MSPE ratio at each horizon for the evaluation period since 1997. We disregard the earlier MSPE ratios for being based on too short a recursive evaluation period to be considered reliable. For illustrative purposes we focus on the real U.S. refiners' acquisition cost for crude oil imports and the method in column (1) of Table 1. Very similar results apply when using the method in column (1) of Table 3. The plot shows the evolution of the recursive MSPE ratio over time. The last entry on the right corresponds to the entry for horizon 24 in column (1) of Table 1.

Figure 1 illustrates that at horizons of 1 and 3 the equal-weighted forecast combination has been consistently more accurate than the no-change forecast throughout the entire evaluation period since 1997. Similar results hold at horizons of 12, 15 and 18 months, but at horizons of 6 and 9 months there is evidence of the no-change forecast having been slightly more accurate early in the sample during 1997-2001. This pattern is consistent with the underlying regression estimates becoming more reliable as the estimation window lengthens. It may also reflect the fact that the usefulness of economic fundamentals as predictors tends to be concentrated at shorter horizons, whereas that of product spreads tends to be most pronounced at horizons of one year and beyond, effectively limiting the sources of predictive information available at horizons of 6 and 9 months. Be that as it may, since 2001, the forecast combination has consistently been more accurate than the no-change forecast at all horizons between 1 month and 18 months.

The accuracy of the equal-weighted forecast combination at horizons of 21 months and 24 months is more erratic. One the one hand, relatively large recursive MSPE reductions occurred during 2003-2009, at the time of the Great Surge in oil prices driven by a booming world economy (see, e.g., Kilian and Murphy 2013). On the other hand, there are times when the no-change forecast has been more accurate. In the latter case, however, the recursive MSPE differences have rarely exceeded 5%. Thus, the analysis in Figure 1 lends further credence to the overall results reported in Tables 1 and 3.

5. Extensions to Quarterly Horizons

As mentioned earlier, the EIA forecasts not only the monthly price of oil, but also quarterly averages. The construction of quarterly forecasts has been studied in depth in Baumeister and Kilian (2013a), who showed that the most accurate forecasts of the quarterly real price of oil are typically obtained by aggregating forecasts from models estimated at monthly frequency to quarterly frequency. For example, the average of the January, February and March forecasts generated in December of the preceding year would constitute the forecast for the first quarter of the subsequent year.

There are two ways of proceeding. One method is first to construct forecast combinations of the forecasts generated each month for the monthly horizons 1 through 24 (similar to the results shown in Tables 1 and 3) and then to aggregate the resulting monthly forecasts by quarter. This approach has the advantage that the monthly combination forecasts are fully consistent with the quarterly combination forecasts. An alternative method is to aggregate the monthly forecasts for each individual forecasting method first and then to construct forecast combinations on the resulting quarterly forecasts. In the case of equal-weighted forecast combinations, this distinction becomes moot. Table 4 shows results for quarterly equal-weighted forecast combinations constructed this manner. The benchmark is again the no-change forecast based on the most recent monthly real price of oil in each quarter.

Given the construction of the quarterly real price of oil as the average of the monthly prices, it is not possible to infer from the results in Tables 1 and 3 how accurate the combination forecasts for the quarterly data will be. The MSPE of the latter also depends on the unknown covariance between the monthly forecasts. Table 4 shows that, nevertheless, the quarterly equal-weighted forecast combination performs quite well for the real U.S. refiners' acquisition cost for oil imports. Column (1) of Table 4 shows systematic reductions in the recursive MSPE at all horizons up to six quarters ranging from 5% to 8%. The directional accuracy ranges at these horizons ranges from 59% to 69% and is mostly highly statistically significant. Similar results are also obtained for the real WTI price in column (4).

Eliminating the no-change forecast and the gasoline spread forecasting model as in Table 3 further improves the accuracy of the quarterly forecasts at horizons up to six quarters. The MSPE reductions in column (2) may be as high as 12% and the success ratios as high as 72% and statistically significant throughout. Similar results hold for the real WTI price in column (5) of Table 4.

It is useful to contrast these results with those from the TVP spread model. Baumeister et al. (2013) established that this model is more accurate than other forecasting methods at forecasting the real price of oil at horizons between 12 and 24 months. Table 4 shows the corresponding results at quarterly horizons. It is evident that this model is more accurate than any of the forecast combinations especially at horizons of 7 and 8 quarters. The advantage of the forecast combination is that it provides insurance against any one model breaking down unexpectedly. As with all insurance, there is a cost involved. In this case, the cost is a potential loss of accuracy compared with the single-most accurate model at any one horizon. There is also a benefit, however. In this example, the benefit is that the forecast combination is more accurate than the TVP spread model at horizons of one and two quarters. Moreover, while the TVP spread model has high directional accuracy, much of that directional accuracy is not statistically significant, whereas the directional accuracy of the forecast combination usually is.

As in the case of the monthly forecasts, it is useful to examine the evolution of the recursive MSPE ratios. The results are fully in line with the earlier evidence. Figure 2 confirms that at horizons of 1, 4, 5, and 6 quarters, the equal-weighted forecast combination has been consistently more accurate than the no-change forecast since 1997. At horizons of 2 and 3 quarters it has been consistently more accurate since 2000 and 2011, respectively. The performance at horizons of 7 and 8 quarters mirrors the results for the 21-month and 24-month horizons in Figure 1.

6. The U.S. Energy Information Administration's Forecasts

So far we have not compared the accuracy of our preferred forecast combination directly with that of the EIA's judgmental oil price forecasts. The reason is that these forecasts are not available for all horizons and data frequencies considered in our analysis. In this section, we evaluate the EIA oil price forecasts to the extent that they are available for our evaluation period. The data source for the EIA oil price forecasts is the EIA's *Short-Term Economic Outlook*. Most importantly for our purposes, this publication provides quarterly forecasts of the U.S. refiners' acquisition cost for imports for horizons up to 6 quarters. Given the irregular pattern of the reports, however, consistent time series of quarterly forecasts dating back to 1991 can only be obtained for horizons from one to four quarters ahead.[6] This quarterly time series allows us to gauge more directly the accuracy of the EIA's judgmental forecasts and to compare their accuracy to that of alternative forecasts. This question is interesting also because of the possibility that the EIA may have early access to oil market data allowing it to generate more accurate real-time forecasts than econometricians. We investigate this question below.

For the period 1991 to 1996, the *Short-Term Energy Outlook* is issued every quarter. From 1997

[6] The corresponding monthly EIA forecasts of the U.S refiners' acquisition cost are available only from 2004 onwards. They are reported for horizons up to 24 months. The EIA also reports monthly and quarterly WTI forecasts, but again the time series are short. Given the well-known sensitivity of forecast rankings to using small evaluation periods, we focus on the quarterly EIA forecasts of the RAC that are available for the same long evaluation period already considered for our forecast combinations .

onwards, the publication is released each month but only reports quarterly forecasts. The monthly publications are usually issued within the first two weeks of each month. The EIA updates its quarterly forecasts in each monthly report incorporating new information as it becomes available. To make the forecast comparison meaningful it is important to match the information set of the EIA as closely as possible with our information set. Our default convention (timing convention (a) in Table 5) uses the end-of-quarter issues of the *Short-Term Energy Outlook* (i.e., March, June, September and December). The oil price reported for the current quarter is taken as the nowcast and the oil prices reported for the subsequent quarters represent the forecasts. The corresponding real oil price forecasts are obtained by adjusting the nominal EIA oil price forecasts for expected inflation exactly as already shown for the forecasts based on oil futures prices.

The first column of Table 5 shows that this EIA forecast tends to be less accurate than the no-change forecast used as the default benchmark throughout the paper. For example, at the one-quarter horizon, the EIA forecast has an MSPE that is 62% higher than the no-change forecast, while the preferred forecast combination in Table 4 has 9% lower MSPE than the benchmark. At higher horizons, the extent of the losses from using the EIA forecast diminishes, but even at the four-quarter horizon, the EIA forecast is only about as accurate as the random walk, while the forecast combination is 11% more accurate. Nor does the EIA forecast have any statistically significant directional accuracy, in sharp contrast to our preferred forecast combination.

Under timing convention (a), the nowcast and forecasts effectively rely on information for the first two months of the current quarter given that the *Short-Term Energy Outlook* is released within the first two weeks of the third month of the quarter. While this timing accurately reflects the real-time availability of these forecasts, it implies that the EIA forecast has a slight informational disadvantage compared with model-based forecasts. One way of gauging how much of a difference this fact makes is to evaluate the EIA forecast under a different timing convention (timing convention (b) in Table 5). The proposal is to rely instead on the issue of the *Short-Term Energy Outlook* that appears in the first month of the following quarter (i.e., April, July, October and January). Under this convention, the price reported for the previous quarter (e.g., the Q1 price in the April issue) is considered the nowcast, and the price quoted for the current quarter (e.g., Q2 in the April issue) is the 1-quarter-ahead forecast.

Although this alternative convention gives the EIA forecast an informational advantage of up to one month relative to model-based oil price forecasts, the second column of Table 5 shows that even with that informational advantage the EIA forecast is never significantly more accurate than the no-change forecast with MSPE ratios as high as 1.12 at some horizons. There is some directional accuracy, but lower than for the preferred forecast combination and usually not statistically significant. This evidence suggests that the EIA's forecasts are systematically less accurate than our forecast combinations at all four horizons, even controlling for any informational delays associated with the timing of the releases of the EIA forecasts.

These results are clearly even less favorable to the EIA forecast than those reported in Alquist et al. (2013) for selected quarterly horizons. The reason is that in Alquist et al. (2013) the EIA forecasts were compared with the EIA's own quarterly nowcast, as shown in the third column of Table 5 under timing convention (a), rather than the no-change forecast from the real-time data base. As subsequently shown in Baumeister and Kilian (2013a) no-change forecasts based on quarterly nowcasts are distinctly less accurate than quarterly no-change forecasts based on the most recent monthly value of the price of oil, which explains the much higher MSPE ratio of the EIA forecast in column (1) of Table 5. It is the latter result that provides an accurate representation of the forecasting ability of the EIA forecast.

To summarize, the EIA forecast appears systematically less accurate compared with the preferred forecast combination discussed in section 5. This does not rule out that adding the judgmental EIA forecast to the forecast combination could further improve the accuracy of the forecast combination, of course. This possibility is investigated in the last two columns of Table 5, which demonstrates that adding the EIA forecasts to the quarterly forecast combination would have increased the MSPE ratio by as much as 0.05 at horizons of 1 and 2 quarters and marginally at the 3-quarter horizon, while having essentially no effect on the MSPE at the 4 quarter horizon. Thus, there is no gain from incorporating the EIA forecasts into the forecast combination. We conclude that suitable forecast combinations provide a viable alternative to the judgmental oil price forecasts currently reported in the *Short-Term Energy Outlook*.

7. Conclusion

One of the challenges faced by producers of short-term oil price forecasts such as the EIA is how to generate real-time forecasts of the price of oil that are more accurate than the no-change forecast. We showed how to construct such forecasts in a timely manner without relying on judgment. Our analysis relied on forecast combinations of several forecasting models that by themselves are superior to the no-change forecast at least at some horizons. These models utilize as predictors lags of the real price of oil, current oil spot prices and oil futures prices, current spot prices in the market for refined products, and current and lagged data on economic fundamentals such as oil production, global real activity, other industrial commodity prices, and changes in crude oil stocks. The potential advantage of such data-based forecast combinations is that the resulting oil price forecasts may be more robust across forecast horizons and over time than even the best individual forecasting models. They may provide some insurance not only against possible model misspecification, but also against smooth structural change. Both problems are potentially important concerns when forecasting oil prices.

The most accurate forecasts are obtained based on constant equal weights for all forecasting models. There is no evidence that choosing the weights based on recent forecasting performance improves the forecast accuracy. We demonstrated that combinations of forecasts from VAR models of the global oil market, forecasts based on non-oil industrial commodity prices, forecasts based on oil futures prices, and forecasts from TVP product spread models, in particular, are systematically more accurate than the no-change forecast at all horizons from 1 month to 18 months. Depending on the horizon, our forecast combinations lower the MSPE by as much as 13% relative to the no-change forecast, and they have directional accuracy as high as 65%. Similar results are obtained for measures of the global price of oil such as the U.S. refiners' acquisition cost, and the WTI price of crude oil. These results are also remarkably robust over time.

While much of our analysis focuses on forecasting the monthly real price of oil, we showed that our results can be generalized to the problem of forecasting the quarterly real price of oil. In the latter, there are systematic improvements in accuracy at horizons up to six quarters with MSPE reductions as high as 12% and statistically significant directional accuracy as high as 72%. An important question is whether the recursive MSPE reductions are driven by one or two unusual episodes in the data or whether they are more systematic. We found that suitably constructed combination forecasts have been more accurate than the no-change forecast since about 2001 and in many cases since 1997.

Although we do not pursue this extension in this paper, we note that it would be straightforward to extend our analysis to the problem of forecasting the nominal price of oil. This would only involve the additional step of scaling our forecast of the real price of oil by the expected inflation rate over the relevant forecast horizon. We also provided an algorithm for deciding which forecasting models to include in the forecast combination. We used this algorithm to establish that not all of the forecasting models included in the baseline forecast combination contribute toward lower MSPEs, allowing us to eliminate some models from consideration.

References

Alquist, R., and Kilian, L. (2010), "What Do We Learn from the Price of Crude Oil Futures?" *Journal of Applied Econometrics*, 25, 539-573.

Alquist, R., Kilian, L., and R.J. Vigfusson (2013), "Forecasting the Price of Oil," forthcoming in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2, Amsterdam: North-Holland.

Baumeister, C., and L. Kilian (2012), "Real-Time Forecasts of the Real Price of Oil," *Journal of Business and Economic Statistics*, 30, 326-336.

Baumeister, C., and L. Kilian (2013a), "What Central Bankers Need to Know about Forecasting Oil Prices," forthcoming: *International Economic Review*.

Baumeister, C., and L. Kilian (2013b), "Real-Time Analysis of Oil Price Risks using Forecast

Scenarios," mimeo, University of Michigan.

Baumeister, C., Kilian, L., and X. Zhou (2013), "Are Product Spreads Useful for Forecasting? An Empirical Evaluation of the Verleger Hypothesis," mimeo, University of Michigan.

Chen, Y-C., Rogoff, K.S., and B. Rossi (2010), "Can Exchange Rates Forecast Commodity Prices?" *Quarterly Journal of Economics*, 125, 1145-1194.

Chernenko, S.V., Schwarz, K.B. and J.H. Wright (2004), "The Information Content of Forward and Futures Prices," International Finance Discussion Paper No. 808, Board of Governors of the Federal Reserve System.

Chinn, M., and O. Coibion (2013), "The Predictive Content of Commodity Futures," forthcoming: *Journal of Futures Markets*.

Davies, P. (2007), "What's the Value of an Energy Economist?" Speech presented at the International Association for Energy Economics, Wellington, New Zealand.

Diebold, F.X., and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X., and P. Pauly (1987), "Structural Change and the Combination of Forecasts," *Journal of Forecasting*, 6, 21-40.

Hamilton, J.D. (2009), "Understanding Crude Oil Prices," *Energy Journal*, 30, 179-206.

Kilian, L. (2009), "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 99, 1053-1069.

Kilian, L., and Murphy, D.P. (2013), "The Role of Inventories and Speculative Trading in the Global Market for Crude Oil," forthcoming: *Journal of Applied Econometrics*

Kim, C.J., and C.R. Nelson (1999), *State Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge, MA: MIT Press.

Knetsch, T.A. (2007), "Forecasting the Price of Oil via Convenience Yield Predictions," *Journal of Forecasting*, 26, 527-549.

Pesaran, M.H., and A. Timmermann (2009), "Testing Dependence Among Serially Correlated Multicategory Variables," *Journal of the American Statistical Association*, 104, 325-337.

Reeve, T.A., and R.J. Vigfusson (2011), "Evaluating the Forecasting Performance of Commodity Futures Prices," International Finance Discussion Paper No. 1025, Board of Governors of the Federal Reserve System.

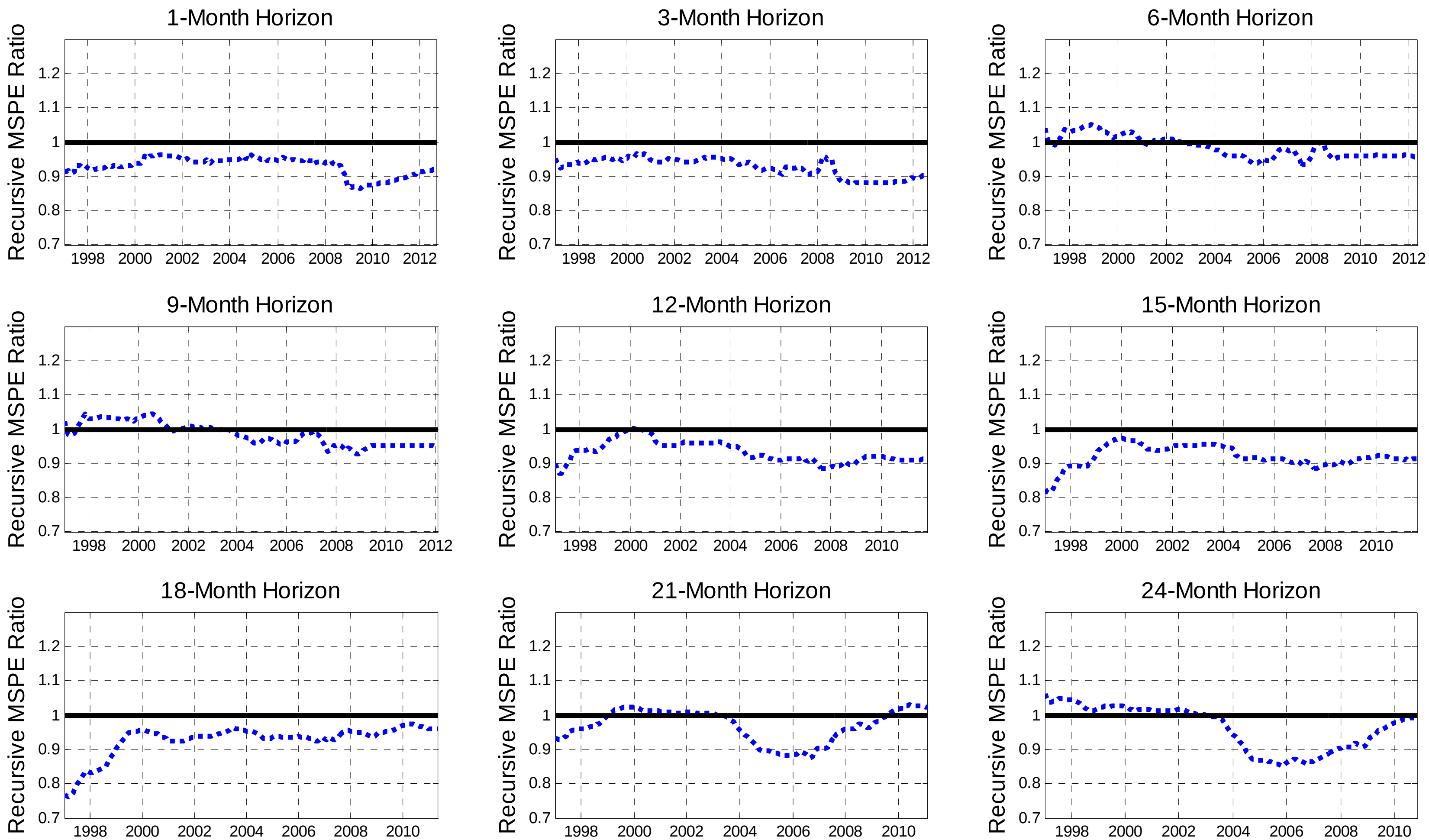
Sanders, D.R., Manfredo, M.R., and K. Boris (2008), "Evaluating Information in Multiple Horizon Forecasts: The DOE's Energy Price Forecasts," *Energy Economics*, 31, 189-196.

Stock, J.H., and M.W. Watson (2004), "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting*, 23, 405-430.

Strumpf, D. (2013), "Goldman Cuts the Near-Term Brent Crude Forecast to $100 a Barrel," *Wall Street Journal*, April 23.

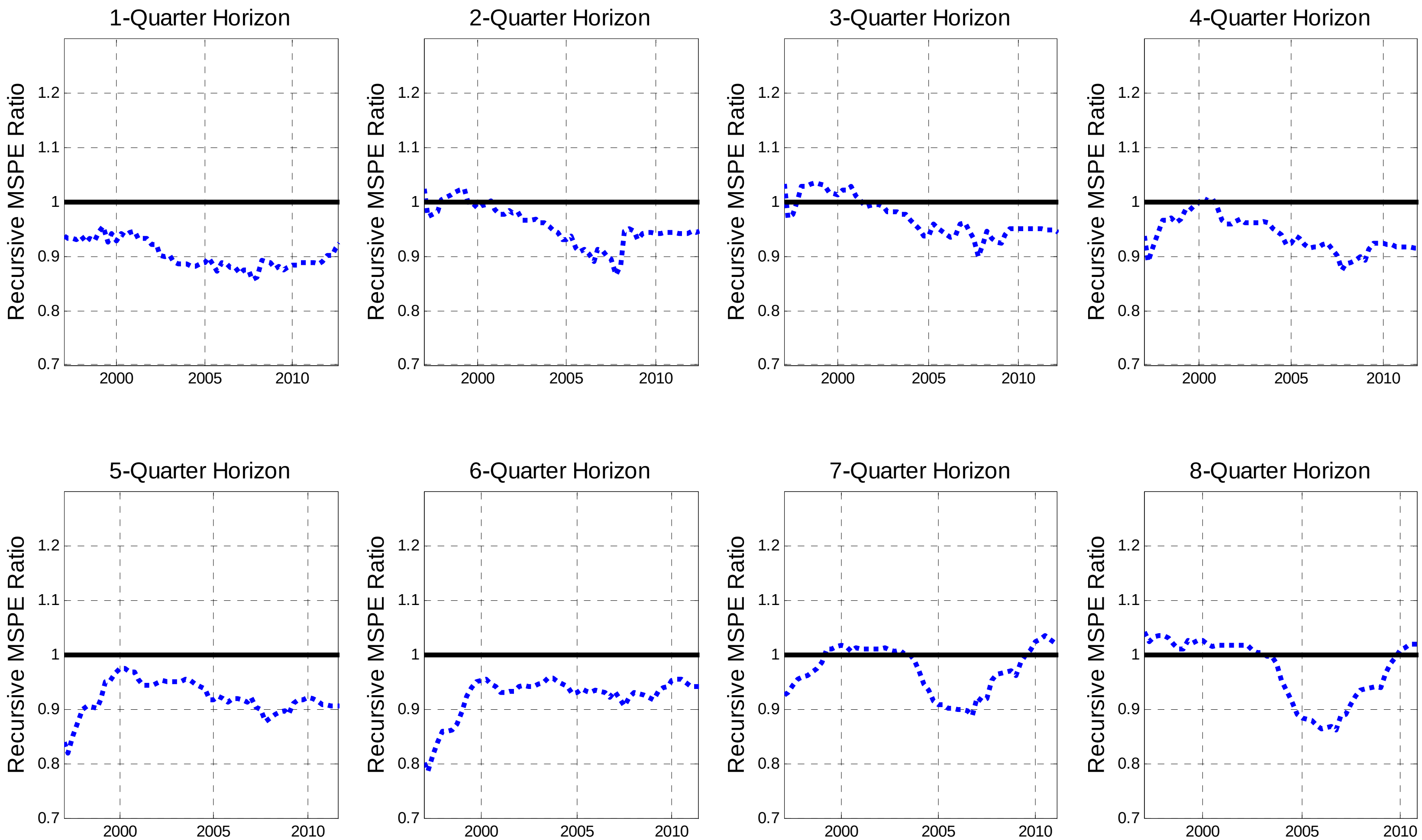
Verleger, P.K. (2011), "The Margin, Currency, and the Price of Oil," *Business Economics*, 46, 71-82.

Figure 1: Real-Time Recursive MSPE Ratio Relative to No-Change Forecast for Refiner's Acquisition Cost for Oil Imports Equal-Weighted Combination of All Six Forecasting Models



NOTES: Results based on the forecast combination shown in Table 3. A ratio below 1 indicates an improvement relative to the no-change forecast. The plot shows the evolution of the recursive MSPE ratio over time for the forecast evaluation period since 1997. This increases the reliability of the MSPE estimates and allows the MSPE ratio to stabilize.

Figure 2: Real-Time Recursive MSPE Ratio Relative to No-Change Forecast for Refiner's Acquisition Cost for Oil Imports Equal-Weighted Combination of All Six Forecasting Models



NOTES: Results based on the forecast combination shown in Table 4. A ratio below 1 indicates an improvement relative to the no-change forecast. The plot shows the evolution of the recursive MSPE ratio over time for the forecast evaluation period since 1997. This increases the reliability of the MSPE estimates and allows the MSPE ratio to stabilize.

Table 1: Real-Time Forecast Accuracy of Baseline Forecast Combination Based on All Six Forecasting Models

	Real U.S. Refiners' Acquisition Cost for Oil Imports					Real WTI Price				
	Equal Weights	Recursive Weights	Rolling Weights Based on Windows of Length			Equal Weights	Recursive Weights	Rolling Weights Based on Windows of Length		
			36	24	12			36	24	12
Monthly Horizon			Recursive MSPE Ratios					Recursive MSPE Ratios		
1	**0.922**	**0.927**	**0.931**	**0.929**	**0.925**	**0.911**	**0.917**	**0.918**	**0.919**	**0.918**
3	**0.906**	**0.912**	**0.916**	**0.919**	**0.919**	**0.906**	**0.914**	**0.918**	**0.915**	**0.926**
6	**0.957**	**0.964**	**0.966**	**0.976**	**0.971**	**0.962**	**0.968**	**0.972**	**0.979**	**0.980**
9	**0.948**	**0.955**	**0.954**	**0.961**	**0.966**	**0.952**	**0.958**	**0.954**	**0.961**	**0.968**
12	**0.912**	**0.918**	**0.917**	**0.916**	**0.922**	**0.920**	**0.925**	**0.921**	**0.916**	**0.918**
15	**0.913**	**0.922**	**0.929**	**0.927**	**0.946**	**0.922**	**0.934**	**0.940**	**0.936**	**0.929**
18	**0.962**	**0.979**	1.002	1.002	1.023	**0.963**	**0.986**	1.009	1.013	1.035
21	1.025	1.030	1.051	1.058	1.110	1.023	1.032	1.054	1.065	1.092
24	**0.992**	**0.987**	**0.991**	**1.000**	1.043	**0.984**	**0.987**	**0.994**	1.009	1.055
			Success Ratios					Success Ratios		
1	**0.570***	**0.588****	**0.562***	**0.558****	**0.566****	**0.517**	**0.512**	**0.521**	**0.517**	**0.517**
3	**0.588***	**0.592***	**0.588***	**0.592***	**0.592***	**0.567****	**0.576***	**0.576***	**0.567****	**0.563**
6	**0.556**	**0.535**	**0.539**	**0.522**	0.491	**0.543**	**0.517**	**0.526**	**0.517**	0.496
9	**0.575**	**0.575**	**0.562**	**0.558**	**0.527**	**0.562**	**0.562**	**0.571**	**0.544**	**0.527**
12	**0.614***	**0.614***	**0.609***	**0.600***	**0.627***	**0.605***	**0.596***	**0.586****	**0.600***	**0.600***
15	**0.645***	**0.626***	**0.630***	**0.640***	**0.612***	**0.612***	**0.608***	**0.617***	**0.612***	**0.608***
18	**0.611***	**0.553***	**0.539**	**0.519**	**0.519**	**0.572****	**0.548***	**0.548**	**0.534**	**0.534**
21	**0.550**	**0.564**	**0.569**	**0.564**	0.460	**0.550**	**0.574**	**0.555**	**0.564**	0.460
24	**0.566**	**0.531**	**0.556**	**0.566**	**0.520**	**0.556**	**0.520**	**0.526**	**0.551**	**0.536**

NOTES: The models are described in the text. Boldface indicates improvements relative to the no-change forecast. * denotes significance at the 5% level and ** at the 10% level based on the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy. The statistical significance of the MSPE reductions cannot be assessed because none of the currently available tests of equal predictive accuracy applies in this setting.

Table 2: Real-Time Recursive MSPE Ratios of "Leave-One-Out" Forecast Combinations with Equal Weights

	Real U.S. Refiners' Acquisition Cost for Oil Imports					
Model Left Out:	VAR	Oil Futures Spread	Commodity Prices	No Change	Gasoline Spread	TVP Product Spread
Monthly Horizon			Changes in Recursive MSPE Ratios			
1	**0.041**	-0.011	**0.013**	-0.012	-0.009	-0.007
3	**0.031**	-0.007	**0.029**	-0.016	-0.011	-0.010
6	**0.012**	**0.005**	-0.001	-0.007	-0.001	**0.008**
9	-0.006	**0.014**	-0.005	-0.008	0	**0.024**
12	-0.006	**0.022**	0	-0.015	-0.001	**0.022**
15	-0.014	**0.029**	-0.001	-0.015	-0.002	**0.021**
18	-0.023	**0.034**	-0.007	-0.005	0	**0.023**
21	-0.014	0	-0.037	**0.012**	**0.015**	**0.052**
24	-0.002	0	-0.058	**0.005**	**0.020**	**0.085**

NOTES: The models are described in the text. Boldface indicates increases relative to the MSPE ratio in column (1) of Table 1. Increases mean that the model left out would have improved forecast accuracy if included, whereas decreases mean that it would have worsened forecast accuracy. The statistical significance of the MSPE reductions cannot be assessed because none of the currently available tests of equal predictive accuracy applies in this setting.

Table 3: Real-Time Forecast Accuracy of Forecast Combination with Equal Weights after Dropping the No-Change Forecast and Gasoline Spread Forecast

	Real U.S. Refiners' Acquisition Cost for Oil Imports	Real WTI Price
Monthly Horizon	Recursive MSPE Ratios	
1	**0.897**	**0.880**
3	**0.874**	**0.873**
6	**0.949**	**0.956**
9	**0.939**	**0.943**
12	**0.892**	**0.902**
15	**0.893**	**0.906**
18	**0.957**	**0.959**
21	1.065	1.064
24	1.029	1.017
	Success Ratios	
1	**0.554***	**0.517**
3	**0.609***	**0.592***
6	**0.556**	**0.543**
9	**0.580****	**0.562**
12	**0.609***	**0.605***
15	**0.650***	**0.617***
18	**0.601***	**0.577****
21	**0.550**	**0.550**
24	**0.561**	**0.551**

NOTES: The models are described in the text. Boldface indicates improvements relative to the no-change forecast. * denotes significance at the 5% level and ** at the 10% level based on the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy. The statistical significance of the MSPE reductions cannot be assessed because none of the currently available tests of equal predictive accuracy applies in this setting.

Table 4: Real-Time Forecast Accuracy of Equal-Weighted Forecast Combinations at Quarterly Horizons

	Real U.S. Refiners' Acquisition Cost for Oil Imports			Real WTI Price		
	All Six Models	Four Models	TVP Spread Model Only	All Six Models	Four Models	TVP Spread Model Only
Quarterly Horizon			Recursive MSPE Ratios			
1	**0.929**	**0.906**	1.020	**0.914**	**0.882**	1.034
2	**0.947**	**0.928**	1.016	**0.947**	**0.929**	1.009
3	**0.942**	**0.928**	**0.905**	**0.948**	**0.935**	**0.906**
4	**0.916**	**0.894**	**0.854**	**0.917**	**0.895**	**0.851**
5	**0.907**	**0.883**	**0.871**	**0.915**	**0.894**	**0.889**
6	**0.942**	**0.929**	**0.900**	**0.946**	**0.935**	**0.913**
7	1.024	1.054	**0.929**	1.025	1.057	**0.976**
8	1.029	1.057	**0.883**	1.048	1.048	**0.936**
			Success Ratios			
1	**0.688***	**0.688***	**0.713***	**0.688***	**0.700***	**0.700***
2	**0.654***	**0.628***	**0.615****	**0.667***	**0.654***	**0.641***
3	**0.684***	**0.658***	**0.553**	**0.658***	**0.645***	**0.592**
4	**0.676***	**0.716***	**0.635**	**0.649***	**0.676***	**0.649**
5	**0.611***	**0.625***	**0.681**	**0.611***	**0.611***	**0.611**
6	**0.586**	**0.614***	**0.657**	**0.571**	**0.600****	**0.643**
7	**0.500**	**0.544**	**0.691**	**0.500**	**0.559**	**0.677**
8	**0.500**	0.470	**0.652**	**0.515**	0.485	**0.652**

NOTES: The four models are obtained by dropping the no-change forecast and gasoline-spread forecast from the set of models to be combined. The TVP spread model is included as benchmark. Boldface indicates improvements relative to the no-change forecast. * denotes significance at the 5% level and ** at the 10% level based on the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy. The statistical significance of the MSPE reductions cannot be assessed because none of the currently available tests of equal predictive accuracy applies in this setting.

Table 5: Real-Time Forecast Accuracy of EIA Forecasts of the Real U.S. Refiners' Acquisition Cost for Imports

Quarterly Horizon	Timing Convention (a) MSPE Ratio Relative to No-Change Forecast	Timing Convention (b) MSPE Ratio Relative to No-Change Forecast	Timing Convention (a) MSPE Ratio Relative to EIA Quarterly Nowcast	Timing Convention (a) Four Models with EIA Forecast Added Change in MSPE Ratio Relative to No-Change Forecast	Timing Convention (a) All Six Models with EIA Forecast Added Change in MSPE Ratio Relative to No-Change Forecast
1	1.618	**0.996**	**0.926**	**0.054**	**0.037**
2	1.291	1.122	1.113	**0.046**	**0.031**
3	1.086	1.046	1.049	**0.012**	**0.007**
4	1.004	**0.957**	**0.948**	**0.003**	-0.002
	Success Ratio	Success Ratio	Success Ratio		
1	0.374	**0.602***	0.374		
2	0.463	**0.573**	0.463		
3	0.469	**0.556**	0.469		
4	**0.588**	**0.638***	**0.588**		

NOTES: The timing convention (a) for the EIA forecasts mimics a user of the EIA forecast who downloads the forecast at the end of each quarter at the same point in time when the model-based forecasts are constructed. The timing convention (b) allows the EIA an informational advantage of up to one month relative to the model-based forecasts. The comparison with the EIA quarterly nowcast is equivalent to the exercise for the nominal price of oil reported in Alquist et al. (2013). In the first three columns, boldface indicates improvements relative to the no-change forecast. * denotes significance at the 5% level and ** at the 10% level. The statistical significance of the success ratios is assessed based on the Pesaran and Timmermann (2009) test for the null hypothesis of no directional accuracy. The statistical significance of the MSPE reductions is assessed based on the Diebold and Mariano (1995) test. In the last two columns, boldface indicates an increase in the MSPE resulting from the inclusion of the EIA forecast in the forecast combination. The statistical significance of the MSPE increases cannot be assessed because none of the currently available tests of equal predictive accuracy applies in this setting.

CFS WORKING PAPER SERIES

No.	Author(s)	Title
2013/10	**Christiane Baumeister** **Lutz Kilian**	Do Oil Price Increases Cause Higher Food Prices?
2013/09	**Christiane Baumeister** **Lutz Kilian**	Are Product Spreads Useful for Forecasting? Empirical Evaluation of the Verleger Hypothesis
2013/08	**Hugh H. Kim** **Raimond Maurer** **Olivia S. Mitchell**	Time is Money: Life Cycle Rational Inertia and Delegation of Investment Management
2013/07	**Andreas Hubener** **Raimond Maurer** **Olivia S. Mitchell**	How Family Status and Social Security Claiming Options Shape Optimal Life Cycle Portfolios
2013/06	**Günter Coenen** **Anders Warne**	Risks to Price Stability, the Zero Lower Bound and Forward Guidance: A Real-Time Assessment
2013/05	**Stefan Mittnik**	VaR-implied Tail-correlation Matrices
2013/04	**Hans-Joachim Dübel**	The Capital Structure of Banks and Practice of Bank Restructuring
2013/03	**Daniele Pianeselli** **Andrea Zaghini**	The Cost of Firms' Debt Financing
2013/02	**Otmar Issing**	A New Paradigm for Monetary Policy?

All CFS Working Papers can be downloaded at *www.ifk-cfs.de/publications/working-papers*.