# Working Papers



# Forecasting the Brent Oil Price: Addressing Time-Variation in Forecast Performance

Cristiana Belu Manescu
Ine Van Robays

CESIFO WORKING PAPER NO. 6242
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
DECEMBER 2016



***ISSN 2364-1428***

# Forecasting the Brent Oil Price: Addressing Time-Variation in Forecast Performance


## Abstract

This paper explores a range of different forecast methods for Brent oil prices and analyses their performance relative to oil futures and the random walk over the period 1995Q1 - 2015Q2, including periods of stable, upwardly trending and rapidly dropping oil prices. None of the individual methods considered outperforms either benchmark consistently over time or across forecast horizons. To address this instability, we propose a forecast combination for predicting quarterly real Brent oil prices. A four-model combination - consisting of futures, risk-adjusted futures, a Bayesian VAR and a DSGE model of the oil market - predicts oil prices more accurately compared to all methods evaluated up to 11 quarters ahead and generates forecasts whose performance is robust over time. The improvements in forecast accuracy and stability are noticeable in terms of both point forecasts – with MSPE gains of 23% relative to futures at the 11 quarter-ahead horizon and a directional accuracy of 70% – and density forecasts – with CRPS gains of 50% relative to futures and logarithmic score gains of 90%, both at the 7-quarter ahead horizon.



JEL-Codes: Q430, C530, E370.

Keywords: Brent oil prices, real-time, combining forecasts, time-variation, and density forecasts.



*Cristiana Belu Manescu*
*European Commission*
*170 Rue de la Loi*
*Belgium – 1040 Bruxelles*
*cristiana.manescu@ec.europa.eu*

*Ine Van Robays**
*European Central Bank*
*Kaiserstraße 29*
*Germany – 60311 Frankfurt am Main*
*ine.van-robays@ecb.int*

*corresponding author



April 2016

The views expressed in this paper are solely those of the authors and cannot be attributed to the European Central Bank or the Eurosystem. Without implicating, we would like to thank Lorenzo Cappello and Nina Buchmann for outstanding research assistance; Galo Nuño for his valuable input and comments; participants at the 22nd SNDE conference in New York, 2014, and the 1st IAAE conference in London, 2014, for their comments; one anonymous referee of the ECB Working Paper Series for his suggestions; as well as two anonymous referees and an associate editor of the International Journal of Forecasting for improving the quality of this paper. Useful comments were also received from Lutz Kilian, Christiane Baumeister and Kirstin Hubrich. We would also like to thank Lutz Kilian for providing us with the real-time dataset as well as some of the codes developed in Baumeister and Kilian (2012), Michele Lenza for the BVAR codes developed in Giannone et al (2015), Galo Nuño for the DSGE codes developed in Nakov and Nuño (2014) and Francesco Ravazzolo for the codes used to compute the density forecasts evaluation statistics.

## 1. Introduction

Given the importance of oil prices in driving inflation and economic activity, accurately forecasting oil prices is key. Designing models that outperform the random walk forecast for oil, however, far from straightforward (see e.g. Alquist et al. 2013). Several recent studies have proposed methods to improve oil price projections, ranging from simple forecasting rules to complex estimated models or model combinations which improve the oil price forecast to some extent (e.g. Alquist et al. 2013, Baumeister and Kilian 2012, 2014, 2015a).

However, this recent literature has three main limitations. First, only few studies have touched upon forecasting the Brent oil price.[3] Most studies have focused on the refiner`s acquisition cost of imported crude oil (RAC) or the West Texas Intermediate (WTI) price (e.g. Pagano and Pisani 2009, Baumeister and Kilian 2012, 2015a and related papers). While the WTI is particularly relevant for the North-American oil market, in other parts of the world, Brent crude oil has traditionally been the more representative oil price benchmark. Moreover, Brent oil has recently strengthened its prominence as global benchmark as the WTI price increasingly reflected US-specific dynamics. Forecasting Brent oil prices is as such not the same as forecasting WTI or RAC oil prices.

Second, it is not clear how sensitive the forecast accuracy of recently proposed methods is to the evaluation period considered. Having stability in forecast performance is crucial and seems particularly relevant for oil price forecasting as oil price dynamics substantially change over time – varying between being stable, trending upwards, falling substantially and displaying high volatility at the same time (see Figure 1). The forecast accuracy of individual forecast approaches might be highly time-dependent as these typically only capture a specific oil price behavior. While sample-variation is acknowledged as an important issue in oil price forecasting in recent work (Baumeister and Kilian 2014, Baumeister et al. 2014), no systematic evaluation has been carried out so far to document the magnitude of this instability by class of models and over different forecast horizons for Brent oil prices.

Third, the recent oil price forecasting literature has focused exclusively on comparing models based on their point forecast, disregarding other type of useful information such as the level of uncertainty associated to the point forecasts. In the macroeconomics literature, several contributions (e.g. Jore et al. 2010, Clark 2011, Clark and Ravazzolo, 2014) have focused on density forecasts and shown that prediction of higher moments are useful to discriminate among models. In practice, several central banks (e.g. Bank of England, Norges Bank) have recently started to use density forecasts in communicating the uncertainty around their macroeconomic projections. However, a systematic evaluation of how standard oil price forecasting models perform in terms of density forecasts has not yet been conducted.

In this regard, this paper contributes to the literature in three ways. First, this paper provides a systematic evaluation of a range of forecasting models for the Brent oil price and rigorously documents

[3] One example is Baumeister and Kilian (2016).

the time-variation in the performance of these individual forecasts. We show that relying on forecast statistics without considering the performance over time can hide important information on a model`s forecasting accuracy. As such, we explicitly use forecast stability as one of the selection criteria for choosing the preferred model for forecasting Brent oil prices.

Second, based on finding considerable instability in forecast performance for the individual models, this paper proposes a forecast combination model for Brent oil prices. In addition to offering gains in forecast accuracy, the model combination renders a projection whose performance is remarkably robust over time. In this we follow Baumeister and Kilian (2015a) and Baumeister et al. (2014) who have proposed similar approaches for the RAC and the WTI price. Importantly, we use a different set of models than employed in that literature in order to better capture Brent price dynamics and circumvent some data limitations that prevent us from using the same models. Even though our focus is forecasting Brent oil prices, the model combination proposed in this paper also offers accurate projections for WTI oil prices.

Finally, this paper proposes a comparison of the oil price forecasting models in terms of their density forecasts accuracy. We find that our proposed model combination offers significant accuracy gains relative to all individual models when considering the full probability distribution of the forecasts. Interestingly, when looking at density forecasts, a combination with time-varying weights significantly outperforms an equal-weight strategy, a result that confirms earlier findings in the macroeconomic literature.

We evaluate the out-of-sample performance of nine different forecast methods for Brent oil prices which are estimated recursively over the period 1995Q1 – 2015Q2 using a real-time dataset which is based on an augmented version of the real-time database first created by Baumeister and Kilian (2012). We choose the set of forecast models based on recent advances in the oil price forecasting literature, although one of the approaches is new. Various forecast evaluation tools are employed: the mean squared prediction error (MSPE), the forecast bias, the success ratio that measures directional accuracy and recursive MSPE ratios calculated over a rolling window to demonstrate the stability in forecast performance. Given the focus on stability, we show most of the results over the full evaluation sample as well as over different subsamples. We aim at improving the quarterly forecast in both the short and long term (up to 11 quarters ahead) which is longer than the horizons looked at in previous studies for the WTI or RAC price. We choose futures prices as the primary benchmark against which the models are assessed in contrast to earlier studies since this is a widely used tool for oil price forecasting by central banks, international institutions and other policy makers. In a large part of the evaluation period and over several forecast horizons, oil futures are a tougher benchmark to beat compared to the random walk. We also report the forecast performance relative to the random walk to maintain the comparability of our results to the literature.

Several interesting results emerge from this analysis. Concerning the forecast performance of the individual models, the results show that all individual forecasting methods suffer from significant time variation in their performance as no single forecasting approach consistently outperforms the futures or the random walk over the sample period or across horizons. Of the nine models evaluated, four methods however stand out in predicting Brent crude oil prices, each over a specific period and/or forecast horizon: (i) oil futures, (ii) risk-adjusted futures that aim to correct the futures` forecast error for a time-varying risk premium, (iii) a Bayesian VAR (BVAR) model of the oil market including data on oil fundamentals and global economic activity, and (iv) a DSGE model that has a detailed structure of the supply side of the oil market. Compared to the no-change, the BVAR improves the forecast in the short and medium term (with gains up to 6% in MSPE terms), the futures and risk-adjusted futures are more accurate for medium-term projections (with gains up to 15% and 23% respectively), and the DSGE for medium to longer-term forecasts (with gains up to 20%). Similar gains are found relative to the futures benchmark. Concerning the sample-dependence of the forecast performance, the DSGE and the risk-adjusted futures generated large forecast errors in the period in which oil prices were relatively stable (1995-2001), and the futures when prices were trending upwards (2002-2007), whereas these approaches outperformed the no-change in the other subsamples including the more recent period characterized by high oil price volatility induced by the global financial crisis and the mid-2014 oil price drop.

Due to this instability in performance across periods and forecast horizons, it is clear that relying on an individual forecast model will not generate reliable oil price projections. However, these differences in forecast performance can be exploited by combining the projections into a single forecast which might be more accurate and stable over time, see also Baumeister et al. (2014), Baumeister and Kilian (2015a).

We propose an equal-weighted four-model combination for Brent oil price forecasting consisting of the futures, the risk-adjusted futures, the BVAR model and the DSGE model. Compared to individual forecasts, it is clear that combining these projections offers gains in accuracy and also generates a forecast whose performance is remarkably more robust over time, even over the recent period including the financial crisis and the large mid-2014 oil price drop. On average across forecast horizons, the four-model combination offers MSPE gains of 14%, with an improvement of about 10% in the short run which increases to a noticeable 23% at the 11 quarter-ahead forecast horizon. In addition, in all subsamples considered and for most forecast horizons, the four-model combination predicts Brent oil prices more accurately than the futures; a stability in forecast accuracy no other individual approach considered in this paper can offer. Moreover, the model combination has a lower forecast bias and predicts the direction of the oil price change more accurately than both benchmarks at almost all forecast horizons with an average directional accuracy of 70%. These conclusions are robust to using WTI oil prices instead of Brent prices and using different weighting schemes for the forecast combination.

Finally, our analysis also shows that taking the full probability distribution of the models' forecasts into account can be a useful additional tool to discriminate among models. For example, the individual models considered differ in terms of the level of uncertainty associated to their point forecasts, with the no-change, the BVAR and the non-oil index showing higher levels of uncertainty compared to VAR in levels and VAR with GDP. In terms of forecast accuracy of the density forecasts, the BVAR outperforms all the other models on the full sample. The no-change, VAR in differences and the DSGE model at longer horizons also show relatively good forecast accuracy, while the futures, adjusted-futures and the non-oil index have the lowest accuracy especially at longer-horizons and especially according to the logarithmic score. More importantly, the four-model combination (with time-varying weights) outperforms all individual models – including the BVAR i.e. the best performing individual model – also when taking the entire probability distribution of the forecast into account. Relative to futures, the four-model combination shows substantial gains at the 1-quarter ahead horizon by 30% and 65% in terms of CRPS and logarithmic score, respectively, and even larger at the 7-quarter horizon by 49% and 89%, respectively. Relative to the random walk, the improvements of the four-model combination are of 21% and 46% at the 1-quarter ahead in terms of CRPS and logarithmic score, respectively, and of 50% and 39% at the 7-quarters ahead.

The remainder of this paper is structured as follows. The next section describes the different forecasting methods for quarterly Brent crude oil prices. Section 3 shows the results of the forecast evaluation exercise of the individual projections, first on average over the complete forecast evaluation period and then over different subsamples to illustrate the instability in forecast performance over time. Section 4 discusses the choice of the forecast combination and evaluates its performance relative to that of the individual projections. In section 5, we discuss the robustness of the results, while in Section 6 we present the analysis of the density forecast evaluation. Section 7 concludes.

## 2. Evaluation of the forecasting performance of selected oil price models

We evaluate the forecasting performance of nine different forecasting approaches in detail, i.e. the futures, the risk-adjusted futures, the non-oil commodity price index, four structural VAR models based on the Baumeister and Kilian (2012) model, a DSGE model based on Nakov and Nuño (2014) and the random walk. The forecast evaluation is done out-of sample using data as available in real-time based on an augmented version of the proprietary real-time data base first created by Baumeister and Kilian (2012). The forecast variable of interest is the level of the Brent real oil price. We only consider methods that predict Brent crude oil prices directly.[4] The choice of the set of forecast models is based on recent advances in the oil price forecasting literature.[5] Of course, this list of models is not exhaustive. Most of the approaches we look at have been shown to have certain power to forecast oil prices. As we only focus on Brent oil prices, we modify these approaches to directly forecast real Brent oil. In addition, we

[4] Although other studies that focus on forecasting WTI or RAC have been also touching upon Brent price forecasting, this has often been done indirectly by including the spread between the WTI or RAC with the Brent, and forecasting this spread using a random walk (e.g. Baumeister and Kilian 2014). However, we do not follow this approach as (1) Baumeister and Kilian (2014) note that the model with the spread and the model including the Brent price directly do not differ much which is confirmed by our own evaluation and (2) the spread of the Brent with the RAC and WTI has not been stable over time which might cause a random walk forecast of the spread to not always be the best forecast.

[5] Recently, there has been an increase in the literature proposing models or techniques to improve oil price forecasting. Alquist et al. (2013) give a comprehensive overview of the forecasting performance of diverse models to forecast nominal and real oil prices, including regressions-based predictors and no-change forecasts based on macroeconomic indicators. A distinct class of models introduced by Baumeister, Kilian and Zhou (2013) exploits the information contained in the so-called product spread, i.e. the difference between refined product market prices and the crude oil purchase price for refineries, and is found to work well for the RAC and the WTI. Also for the RAC and the WTI, Baumeister, Kilian and Lee (2014) and Baumeister and Kilian (2015a) propose a variety of model combinations with good forecasting accuracy. For the Brent oil price, however, the literature is, to our knowledge, very limited. For example, Baumeister and Kilian (2012) analyses the accuracy of several Brent oil models, but only in the short-term (up to 4 quarters ahead).

include a model new to the oil price forecasting literature that differs in several aspects from the ones considered so far, i.e. the oil market DSGE model of Nakov and Nuño (2014) which has been chosen because of its more detailed structure of the supply side of the oil market.

### 2.1 Futures price

In many international policy institutions, the assumptions on the future development of oil prices are based on the futures prices of Brent crude oil. Following Baumeister and Kilian (2012), futures-based forecasts of the real Brent oil price *h*-periods ahead (denoted by $\hat{R}_{t,t+h}$) are generated as follows:

$$\hat{R}_{t,t+h} = R_t(1 + f_{t,t+h} - s_t - \hat{\pi}_{t,t+h}),$$

where $R_t$ denotes the current level of the real Brent oil price, $f_{t,t+h}$ the log of the current oil futures price with maturity *h*, $s_t$ the corresponding log of Brent nominal spot price, and $\hat{\pi}_{t,t+h}$ the forecast level of US CPI inflation, which equals the current level of US CPI inflation over the last h months.

Futures prices are a simple forecasting tool and might provide some indication of market expectations of future oil prices. It is well-known however that oil futures prices are not identical to the expected price of oil (see e.g. Alquist and Kilian 2010, Singleton 2014, Hamilton and Wu 2014, Baumeister and Kilian 2015b). Futures prices of risky assets such as oil can deviate from the expected spot price due to a risk premium component. Unlike purely financial assets, storable commodities also carry additional costs such as storage costs and benefits such as holding physical inventories for consumption, typically referred to as the convenience yield. In normal times, the risk and the convenience yield component are typically larger than the cost component, which causes oil futures to generally underpredict actual oil prices.[6] Several studies have shown that futures prices do not deliver accurate forecasts for oil prices, see for example Alquist and Kilian (2010), Alquist et al. (2013) and Reeve and Vigfusson (2011). We will refer to this model as the '*futures*'.

### 2.2 Risk-adjusted futures of Pagano and Pisani (2009)

The second model we consider is the risk-adjusted futures model. This is a statistical model constructed by Pagano and Pisani (2009) that corrects the futures-based projection for a time varying risk premium. The idea is that the forecast error generated by the futures, which varies over time, is to some extent predictable.[7] More specifically, the realized futures-based forecast errors $fe_{t+h}^{(h)}$ at time *t+h* are related to the US business cycle as follows:

$$fe_{t+h}^{(h)} = a^{(h)} + \beta^{(h)} UCap_{t-1} + \varepsilon_{t+h}^{(h)}$$

[6] The situation when the futures contract of oil is traded at a lower price than the spot contract in the oil market is referred to as "backwardation". If the reverse occurs, in case spot prices are higher than futures prices as storage costs are for example very high, the oil market is in "contango".

[7] Although Pagano and Pisani (2009) originally proposed this approach for WTI oil prices, we apply it to Brent oil prices as it can be expected that US capacity utilization of manufacturing is a proxy for the global risk premium and therefore would also work to adjust the forecast error made by the Brent futures.

with *UCap* the degree of capacity utilization in US manufacturing as a proxy for the US business cycle, which is a proxy for the time-varying risk premium that creates a wedge between the futures forecast and the realized oil price. Based on this relationship, we first generate the forecast for the nominal Brent oil price that corrects the futures for the time-varying risk premium:

$$\hat{P}_{t,t+h} = \left( F_{t,t+h} - a^{(h)} - \beta^{(h)} UCap_{t-1} \right)$$

where $\hat{P}_{t,t+h}$ is the nominal risk-adjusted futures price for Brent oil h-period ahead, and $F_{t,t+h}$ is the nominal futures price h-period ahead. These nominal forecasts are then converted to real forecasts using the same formula as for real futures prices, except that the log of current oil futures is replaced with the log of risk-adjusted oil futures as defined above. We will refer to this model as the '*risk-adjusted futures*'.

### 2.3 Non-oil commodity price index

The third forecasting approach exploits information in non-oil commodity prices. Alquist et al. (2013) show that percentage changes in the industrial raw materials index predict nominal WTI prices well in the short run, which is confirmed by Baumeister and Kilian (2012) who demonstrate that it also works for the US refiners' acquisition cost of imported crude oil. The forecasting approach is as follows:

$$\hat{R}_{t,t+h} = R_t * \left( 1 + \Delta P_{t,h}^{non-oil} - \hat{\pi}_{t,t+h} \right)$$

with $\Delta P_{t,h}^{non-oil}$ the percentage change in the Commodity Research Bureau (CBR) industrial raw material nominal price index over the past *h* months, with *h* the forecast horizon. The intuition behind this approach is that changes in non-oil commodity prices such as metals reflect fluctuations in global economic activity which also affects oil prices. We will refer to this forecasting approach as the '*non-oil index*'.

### 2.4 VAR model for Brent oil prices based on Baumeister and Kilian (2012)

The fourth forecasting approach is the reduced form VAR model of Baumeister and Kilian (2012) applied to Brent oil prices which has the following representation:

$$Y_t = c + A(L) Y_{t-1} + u_t$$

This model is a four variable monthly VAR with $Y_t$ the vector of endogenous variables comprising (1) the log-real price of Brent oil deflated by the U.S. consumer price index, (2) the percentage change in global crude oil production, (3) the Kilian (2009) index of global real economic activity, measured in deviation from trend, and (4) a proxy for the change in above-ground global crude oil inventories. The vector *c* contains the intercepts and monthly seasonal dummy variables, $A(L)$ is a matrix polynomial in the lag operator *L* and $u_t$ is a vector of reduced form error terms. Besides including Brent oil prices instead of

RAC or WTI prices, the exact same specification is used as in Baumeister and Kilian (2012). We will refer to this forecasting approach as the '*VAR in levels*'.

### 2.5 VAR model for Brent oil prices in log differences

In the 'VAR in levels' model the oil price is specified in log-levels under the assumption that it is stationary. However, since it is not clear whether oil prices in levels are stationary or not, we consider an alternative VAR model which imposes a unit root on the real price of oil. Thus, the fifth forecasting model we consider is similar to the VAR model just described, but assumes that oil prices are integrated of order one so that the first difference of the logarithms is stationary. Hence, oil prices are included in first differences of the logarithm in the model instead of in log-levels.

We will refer to this model as the '*VAR in differences'*.

### 2.6 VAR model for Brent oil prices with GDP as index of real economic activity

To capture oil price movements driven by changes in economic activity, the VAR model of Baumeister and Kilian (2012) includes a global economic activity measure constructed from data on global dry cargo ocean shipping freight rates as described in Kilian (2009). This index is designed to capture shifts in the demand for industrial commodities in global business markets. In addition to the advantages of this index outlined in Kilian (2009), its timely availability avoids making any assumptions on its recent evolution when forecasting in real time. However, as noted by Kilian (2009), the index is not free from drawbacks.[8]

We consider the VAR model with oil prices in log levels *('VAR in levels'*) in which the global economic activity measure is replaced by a monthly OECD measure of global GDP, comprising OECD member countries plus six major non-OECD member countries (Brazil, China, India, Indonesia, Russian Federation and South-Africa). As this global GDP series was not available over the entire forecast evaluation period but only created recently, we assume the series to have a three-month lag in availability and nowcast it using the average growth rate of monthly GDP up to that point in time. As such, this GDP measure is only available in pseudo-real time. The GDP series is expressed in the first difference of the logarithm to make it stationary. We will refer to this forecasting model as the '*VAR with GDP*'.

### 2.7 VAR model using Bayesian prior information

Bayesian estimation is well-known to the oil price forecasting literature (e.g. Baumeister and Kilian 2012 and 2014) and has been used as an efficient way to estimate densely parametrized VAR models such as those presented above. Bayesian VARs have been shown to outperform their classical counterparts in certain instances, with larger gains for specifications with longer lags (e.g. Baumeister and Kilian, 2012).

Similar to Baumeister and Kilian (2012) and related work, we apply the Bayesian shrinkage method of Giannone et al. (2015) to the 'VAR with GDP' model (owing to its superior accuracy compared to other

[8] Due to the procyclicality of shipbuilding, the freight rate index can be expected to lag increases in economic activity as spare capacity in shipping will absorb some of the impact of higher economic growth. Moreover, as strong economic growth will encourage shipbuilding and reduce freight rates, it might even indicate a decline in economic activity when in reality, economic activity is growing. These drawbacks might have important implications for forecasting.

VAR specifications, as we show later). Specifically, we treat also the coefficients that govern the tightness of the prior (hyperparameters) as random parameters, and we conduct inference on them as for the other coefficients of the model. In particular, our results are obtained by fixing the hyperparameters at the mode of their posterior distribution.[9] The prior distributions for the other coefficients in the model are in the Normal-Inverse Wishart class, and adopt the specification most commonly used in the literature (i.e. the 'Minnesota', the 'sum-of-coefficients' and the 'dummy-initial-observation').[10]

Our prior choice differs from Baumeister and Kilian (2012) – which consider only the Minesota prior – as it also includes two priors that allow for cointegration. While cointegration may appear inconsistent with our VAR specification, this should not be an issue when using the specific estimation method of Giannone et al. (2015) for two reasons. First, the estimates of the hyperparameters are not "dogmatic" and can deviate from the prior beliefs based on the information embedded in the likelihood; this should reduce by a lot the risk of specifying priors that are completely at odds with the data. Second, these priors that allow for cointegration have been shown to provide a very good out-of-sample forecast accuracy compared to flat-priors VARs in several empirical applications. The increase in out-of-sample accuracy is due to the ability of bayesian shrinkage methods to control for overfitting in the context of VARs, which are highly parameterized models. In particular, in contrast to flat-prior VARs, the priors on the sum-of-coefficients are shown to reduce the importance of the deterministic component in explaining the variation in the time series – a sort of deterministic overfitting – and therefore increase the out-of-sample accuracy (see Giannone et al. 2014 and the references therein). Indeed, we have also found that this specification outperforms a number of alternative prior specifications and data-transforms, including the Baumeister and Kilian (2012) prior specification. This analysis is reported in the appendix. We will refer to this model as the '*BVAR'* model.

### 2.8 DSGE model of the oil market based on Nakov and Nuño (2014)

Nakov and Nuño (2014) have constructed a general equilibrium model of the global oil market consisting of an oil-importing and two oil-exporting regions in which the oil price, oil production and oil consumption are jointly determined as equilibrium outcomes of optimizing decisions. On the supply side, they approximate the industrial structure of the oil market by modeling a dominant oil-exporting firm that internalizes the behavioral responses of its competitive fringe producers and the oil consumers. Real oil prices are assumed to follow a trend over the long term as the productivity growth in the oil sector is lower compared to that of the global economy. Their model is calibrated such that the dominant firm with monopolistic power matches Saudi Arabian data. Given the rich structure of the oil supply dynamics in this DSGE model which captures specific historic supply-driven episodes well, we include this model in set of forecasting methods that this paper evaluates. For more details on the model, we refer to Nakov and Nuño (2014).

---

[9] We assume the informative priors of Giannone et al. (2015) for the hyperparameters. Notice that, in case of flat hyperpriors, the posterior distribution of the hyperparameters is proportional to the marginal likelihood.

[10] Only oil prices are included in levels, while the other variables are transformed to stationarity. The Minnesotta prior is centered on one for the oil price and at zero for the other variables.

The calibration of the DSGE model is the same as in that paper, which replicates the main stylized facts of the global oil market for the period from 1973 to 2009.[11] To generate forecasts, we use this calibrated version of the model – not an estimated one – and we proceed in two steps. First, the model is solved by first order perturbation methods, resulting in a linear state space model:

$$\text{State equation:} \qquad Y_{t+1} = AY_t + w_{t+1}, w_t \sim N(0, Q), \qquad (1)$$

where the value of the coefficients of matrices $A$ and $Q$ are a nonlinear function of the structural parameters of the DSGE. The state variable $Y_t$ is the vector of endogenous variables which follow an autoregressive process; $w_{t+1}$ is the vector of shocks following a white noise process. There are four shocks in the model: a demand shock (a TFP shock to oil consuming countries), an efficiency shock (a shock to the preferences of households in consuming countries), a supply shock (a shock to the oil production function in oil producers other than Saudi Arabia) and a Saudi Arabia-specific supply shock (a shock to the oil production function in Saudi Arabia).

Second, the model is embedded into a Kalman filter.

$$\text{Measurement equation:} \qquad X_t = BY_t + v_t, v_t \sim N(0, R) \qquad (2)$$

which relates the unobservable state $Y_t$ to the observable variables $X_t$. $v_t$ denotes the measurement errors which follow a white noise and are uncorrelated with $w_t$. Matrix $B$ is the mapping from state variables into observed ones. We consider five observable variables that capture the main facts in the oil market: (i) the growth rate of real Brent oil prices, (ii) the growth rate of world oil production, (iii) the growth rate of Saudi Arabia production, (iv) the growth rate of non-Saudi Arabia production and (v) the growth rate of OECD global GDP.

To generate the forecasts of the price of oil, we first employ both the measurement and the state equations (1) and (2) of the Kalman filter to estimate the current state $Y_t$.[12] Then we project the future state $Y_{t+1}$ using the state-space equation (1) with zero shocks ($w_{t+1}$). Knowing $Y_{t+1}$, we compute the implicit forecasts of the observables at *t+1* ($X_{t+1}$) – among which is the forecast of the oil price growth rate – based on relation (2). Forecasts at longer horizons are obtained iteratively based on relation (2). Finally, the level of the real oil price is then recovered. This process is repeated recursively at each point in the evaluation sample, using real-time data for the five observables. We will refer to this forecasting model as the '*DSGE*'.

### 2.9 Random walk

Finally, we include a random walk without drift which sets the *h*-period real oil price forecast equal to the real oil price of today:

[11] Such as non-Saudi Arabia growth rate of oil production, production share and average spare capacity of Saudi Arabia or global GDP growth. The long-run growth rate of the oil price is set to the differential growth rate between world GDP and global oil production and the short-run price elasticity of oil demand is set to 0.05.

12 This computes the initial value $Y_0$ and the path of structural shocks $w_t$ that maximizes the likelihood given the path of observed variables $X_t$.

$$\hat{R}_{t,t+h} = R_t$$

Following Baumeister and Kilian (2014), we use the last monthly observation for the real price of oil for the quarterly projection as this has been shown to deliver a more accurate prediction. This is the conventional benchmark to which other models` forecasting performance is typically compared in the literature, as a better forecasting performance relative to the no-change forecast indicates that oil prices can be forecasted to some extent. As this method projects oil prices to remain constant, we will refer to this approach as the '*no-change'*.

## 3. Forecast evaluation of various methods for real Brent oil price forecasting

### 3.1 Set-up of the forecast evaluation exercise

The period considered for the forecast evaluation is 1995Q1 – 2015Q2 and we forecast up to 11 quarters ahead.[13] The forecasts based on the VAR models use data back to 1973M1 for the estimation of the model parameters. The evaluation of the forecast is done out of sample using real-time data. The real-time data set is the one developed by Baumeister and Kilian (2012, 2014) and used in subsequent studies, extended with more recent observations and complemented with additional time series as necessary. More specifically, the Baumeister and Kilian (2012, 2014) dataset includes: spot Brent and WTI oil prices, an index of bulk dry cargo ocean shipping freight rates, world crude oil production, US crude oil inventories, US petroleum inventories, OECD petroleum inventories, the US consumer price index for all urban consumers. In addition, a real-time data set for Brent and WTI futures prices, US capacity in manufacturing industries, industrial raw material price index, the OECD measure of global GDP and Saudi Arabia oil production has also been constructed. For a more detailed description of the dataset we refer to the appendix. Following the literature standard, the models are estimated at monthly frequency and re-estimated recursively at each point in time in the evaluation exercise, except for the DSGE model whose parameters are calibrated.[14] The forecast evaluation is done on quarterly forecasts of the real price of oil which are obtained as a simple average of monthly forecasts during the quarter.[15]

As forecast evaluation tools we employ the mean squared prediction error (MSPE), the forecast bias, the success ratio that measures directional accuracy and recursive MSPE ratios calculated over a rolling

[13] As futures prices with maturity 34 to 36 months ahead are not available for either WTI or Brent during an important part of the 1990s, we have excluded the 12-quarter ahead horizon from the analysis. The evaluation window starts only in 1995 because of lack of sufficient data for futures at the start of the 1990s which is needed to estimate the regressions for the risk-adjusted futures.

[14] These parameters are calibrated using data available over the complete sample. This could give the DSGE model a slight advantage in the real-time out-of-sample evaluation exercise, although the calibrated parameters refer to long-term trends which are based on theory and therefore can be assumed to not change over time.

[15] Following Baumeister and Kilian (2014), quarterly forecasts for the first quarter are calculated as the average of the values predicted for January, February and March, as of December of the last year.

window to demonstrate the stability in forecast performance. Each time, the forecast accuracy is compared relative to the futures and the no-change. It is important to note that we cannot fully rely on the standard significance tests to evaluate whether the differences in MSPE are also statistically significant as the forecasts are generated in real time. Standard tests require the underlying data to be stationary which is not the case given that several of the data series used are subject to revisions, see Baumeister and Kilian (2012) and the references therein.

### 3.2 Average forecast performance over the total sample

Table 1 shows the MSPE of the different models over the complete forecast evaluation period 1995Q1 – 2015Q2. A value lower than one means that the model outperforms the futures forecast or the random walk on average and vice versa. None of the models succeeds in forecasting real Brent oil prices more accurately than the futures across all forecast horizons (see Panel A). Compared to the futures, the four VAR-based models only offer gains for forecasting in the very short term (up to 2 quarters-ahead with gains up to 22% for the BVAR) which is also the case for the non-oil index. The risk-adjusted futures are more accurate for medium-term projections (with MSPE gains up to 10%) and the DSGE for medium to longer-term forecasts (with gains up to 13%). If the forecast performance is compared to the no-change similar results are found (see Panel B of Table 1).

Table 2 shows the bias and the success ratio over the complete sample.[16] With the exception of the DSGE, the bias for all forecasting methods considered is negative over most forecast horizons implying that, on average, the models tend to underpredict actual Brent oil prices. The VAR in differences has the smallest absolute forecast error over short-term forecast horizons and the DSGE model over medium- to long term horizons (see Panel A of Table 2). Panel B of Table 2 shows the success ratio that measures the directional accuracy. All models predict the direction of the oil price change better than a random guess up to 5 quarters ahead, and in the long term, the DSGE model has the highest directional accuracy of 75% of all forecasting methods considered.

### 3.3 Forecast performance of individual methods over time

However, these statistics - which are calculated over the complete forecast evaluation period - hide important information on the performance of these individual models. This is not only true for the forecast accuracy as measured by the MSPE but also for the forecast bias and the success ratio.

To illustrate the time variation in forecast accuracy, Figure 2 shows the recursive MSPE ratios of the nine forecasting approaches calculated over a rolling window of six years relative to the monthly no-change forecast over different forecast horizons over the period 2000Q1 – 2015Q2.[17] The no-change is chosen as the main reference to also show the variation in the accuracy of futures-based projections. Each chart

[16] For the reason that the quarterly no-change forecast is calculated as the last monthly observation for the real price of oil following Baumeister and Kilian (2014), the success ratio for the no-change can differ from 50% as we are comparing a monthly price forecast for the random walk to the average realized quarterly oil price.

[17] To better illustrate the time-variation in the forecast performance, we show the recursive MSPE as an average over a rolling window instead of over a recursive window as the latter would less clearly reflect instability in forecast accuracy as the MSPE over the full evaluation period would for example remain dependent on the forecast performance at start of the evaluation period. We start in 2000Q1 as we take the rolling mean squared error over a rolling window of six years to allow for enough observations to have a reliable MSPE estimate.

is shown on two different scales for reasons of clarity as some models generate large MSPEs throughout the sample. As the middle part of the sample is over-represented as it contributes to both earlier and later windows, the MSPE results are also shown in Table 3 for three non-overlapping time periods: 1995Q1 – 2001Q4, 2002Q1 – 2007Q4 and 2008Q1 – 2015Q2. Although the choice of these periods is arbitrary, the first period corresponds more to a period in which oil price were relatively stable, the second to one in which oil prices trended upwards and the third one to a period in which oil prices displayed significant volatility (see also Figure 1). Similar tables for the forecast bias and the success ratio can be found in the appendix.

When evaluating the forecasting performance of all individual methods over time, it is remarkable that none of the individual models considered manages to consistently outperform the futures-based or the no-change projection neither over time nor across forecast horizons. Each individual projection seems to capture only a specific behavior of oil prices.

Except during the period in which oil prices were trending upwards (2002-2007), it is difficult to consistently improve upon the futures-based oil price forecast, see Panel A of Table 3. In the second half of the 1990s, even though most of the models outperformed the futures-based forecast in the short term, they did not manage to do so at longer-term horizons. Also in the more recent period characterized by substantial oil price volatility, it is difficult to beat the futures-based forecast beyond the very short term, with the notable exception of the DSGE model and the risk-adjusted futures. The relatively good performance of oil futures is an interesting finding given the apparent consensus in the literature that futures prices are generally inaccurate predictors for future oil spot prices, see e.g. Alquist and Kilian (2010), Alquist et al. (2013) and Reeve and Vigfusson (2011).

Also relative to the no-change, it is clear that none of the models considered manages to consistently outperform (see Figure 2 and Panel B of Table 3). Baumeister and Kilian (2012), for example, show that the VAR in levels for RAC and WTI prices significantly beats the random walk forecast up to a horizon of one year over the period 1992-2010 on average. Panel B in Figure 2 however shows that although the model performed relatively well for Brent oil forecasting in the period prior to the global financial crisis, its performance steadily deteriorated since then. Particularly during the large drop in oil prices since mid-2014, this and the other VAR-based models made large forecast errors in predicting oil prices over short-term forecast horizons.[18] Only the futures and the risk-adjusted futures provided gains in forecasting the large oil price drop over the 3-quarter ahead forecast horizon. The variation in performance of the VAR-based models might be explained by the observation that they seem to forecast oil prices relatively well in times when economic fundamentals show persistent variation but less well at other times as noted by Baumeister and Kilian (2015a). A similar variation in forecast performance is found for the non-oil price index-based forecast.

The DSGE model, in contrast, generated very large forecast errors in the 1990s, but outperformed the random walk at almost every forecasting horizon from the early 2000s on with large gains in forecast

[18] Repeating the forecast evaluation for the WTI oil prices, which is one of the oil price measures used in Baumeister and Kilian (2012) instead of the Brent price, we find the same conclusions although our sample period starts slightly later, see the robustness section.

accuracy especially at longer term horizons (see Panel B of Table 3 and e.g. Panel D in Figure 2). Also the risk-adjusted futures projections are subject to significant time variation in their performance, both across time periods and forecast horizons (see e.g. Panel A and D of Figure 2). Compared to the no-change forecast, the BVAR model generally fares well in predicting Brent oil prices and delivering a relatively stable forecast performance across time and forecast horizons, although similar to the other VAR-based models it also generated larger forecast error over short-term forecast horizons more recently when oil prices fell steeply. Compared to oil futures, however, the forecast accuracy of the BVAR is noticeably less accurate and stable (see Panel A of Table 3).[19]

### 3.4 Implications of the results: time variation in performance and forecast combination

These results have two important implications. First, all the forecasting methods considered suffer from significant time variation in their performance which implies that not taking into account this performance instability might lead to choosing an approach that, although it performed well in the past, will not outperform anymore in the future. For the reason that forecasting accuracy is sample-dependent, only using performance statistics calculated over the complete forecast evaluation period may cancel out important information about a model`s forecasting properties. A second implication is that the differences in model performance over time and across forecast horizons suggests that combining these forecasts might offer significant gains in forecast accuracy. A forecast combination can exploit the fact that different forecasting methods perform better during specific periods and over specific forecasting horizons, and provide some insurance against structural breaks in the evolution of the variable to be forecasted if the data generating process is unknown (Hendry and Clements 2002). For oil price forecasting, combining individual forecasts has already been exploited by Baumeister et al. (2014) and Baumeister and Kilian (2015a) and for WTI and RAC oil prices. In the next section, we show that a forecast model combination also works well for Brent oil prices.

## 4. Predicting Brent oil prices using a forecast combination to address performance instability

A requirement of a successful forecast combination is that the individual forecasts differ sufficiently in their forecasting properties (de Menezes, Bunn and Taylor 2000). In this respect, based on our analysis so far, four individual forecast methods stand out; the futures, the risk-adjusted futures, the BVAR and the DSGE. First, concerning the forecast horizon, the forecasts based on the BVAR offer gains in forecasting accuracy in the short run on average, the futures and the risk-adjusted futures at medium term horizons and the DSGE model in the medium to long term (see Table 1). Second, concerning the performance over time, these models perform differently in different periods over the sample (see Figure 2). The DSGE and the risk-adjusted futures generated large forecast errors in the period in which oil prices were relatively stable (1995-2001), and the futures when prices were trending upwards (2002-2007), whereas these approaches outperformed the no-change in the other subsamples including the

[19] The results of this forecast evaluation do not depend on a few outliers. Looking at the squared forecast errors of the different models over time for different forecast horizons, for example, show that the forecast methods display the same forecast properties as is displayed in the subsample results.

more recent period characterized by high oil price volatility induced by the global financial crisis and the mid-2014 oil price drop, see Panel B of Table 3.

### 4.1 Choosing a forecast combination for Brent oil price forecasting

Given many possible combinations, the preferred forecast combination will crucially depend on the loss function of the forecaster, i.e. the relative importance of factors such as accuracy, stability in forecast performance and avoidance of large forecast errors. Also the benchmark relative to which the forecast needs to be improved is important. In this paper, the interest lies mainly in improving the Brent price forecast relative to the futures-based projection used as the baseline forecast in many policy institutions. In addition to finding a combination whose projection is more accurate, the aim is also to find a combination whose performance is stable over time.

We therefore use two evaluation criteria when choosing the forecast combination for Brent oil prices. A first measure is the MSPE gain of the model combination relative to the futures benchmark (‘*MSPE improvement’*) and the second one is the percentage times that a specific model outperforms compared to the futures (*‘outperformance’)*. The latter measure is included to take the stability of the forecast performance into account. The MSPE calculated over the complete forecast evaluation period can mask important variation in performance over time as small forecast errors over one specific time period can more than offset large errors over another. The percentage outperformance measure is less subject to this as it does not depend on the magnitude of the forecast errors and gives an equal weighting to every point forecast over the evaluation period.

We set up the evaluation exercise as follows: starting from the futures benchmark, we evaluate whether adding one of the other eight individual models offers gains in both accuracy and stability over time.[20] If it does so, we add it to the model combination and evaluate whether including additional models can improve the Brent oil price forecast further.

Figure 3 summarizes the results of the forecast evaluation of different forecast model combinations over the full sample 1995Q1-2015Q2. All model combinations considered are listed along the X-axis. The ‘*outperformance*’ measure is represented by the triangles measured on the left axis and shows the percentage times that a specific model outperforms compared to the futures; a value of 0.60 for example indicates that a specific model combination outperforms the futures 60% of the time over the complete evaluation period. The ‘*MSPE improvement*’, represented by the squares and measured on the right axis shows the relative MSPE gain of the model combination relative to the futures benchmark. A value higher than zero indicates that the model combination considered is more accurate in Brent oil price forecasting compared to the futures; a value of 0.10 for example means that the specific model combination has an MSPE 10% lower than the futures. The two measures are averages across forecast horizons. For each combination, we choose to weigh each individual forecast equally as the literature has shown that this simple equal-weighting scheme often performs the best (e.g. Stock and Watson

[20] Although the choice of starting with the futures is arbitrary, we end up choosing the same forecast combination when starting from a different model. These results are available from the authors upon request.

2004, Aiolfi and Timmermann 2006, Baumeister and Kilian 2015a). We also confirm this later on when testing the sensitivity of the results to different combination weights in the robustness section.

Looking at the '*MSPE improvement*' (squares), the first part of Figure 3 ('2m') shows that combining the futures forecast with other individual projections only offers gains in forecast accuracy when the futures are combined with the no-change, the BVAR, the DSGE or the risk-adjusted futures as the MSPE improvement is larger than zero. The percentage '*outperformance*' of each of the two-model combinations (triangles) in contrast, is always higher than 50% indicating that combining two models already provides more stability in forecast accuracy over time than when using the futures only. The combination including the DSGE has the highest gains in forecast accuracy of the two-model combinations considered (MSPE improvement of 15%) and as such, we include it in the combination and we evaluate whether its forecast performance can be improved further by adding a third projection.

The second part of Figure 3 ('3m') displays the forecasting performance when the remaining individual projections are added to the two-model combination consisting of the futures and the DSGE. All three-model combinations evaluated have a lower MSPE relative to the futures projection and in addition offer more stability in performance over time. Particularly the three-model combination including the risk-adjusted futures outperforms the futures close to 70% of the time with an MSPE improvement of 16% on average.

Departing from this three-model combination including the futures, the DSGE and the risk-adjusted futures, the picture is less clear when it comes to possible four-model combinations as several trade-offs emerge, see part 3 of Figure 3 ('4m'). The four-model combination including the BVAR is the best performing of the four-model combinations considered in terms of MSPE gains and stability, although it does not outperform the two- or three-model combination on average. The two- and three-model combinations however have larger forecast errors in the short run certainly when looking at the earlier subsamples, as shown in Table 4. As such, the four-model combination including the BVAR generally offers a more consistent improvement in forecast accuracy across horizons and given that the loss in overall average forecast accuracy is small, we prefer the four-model combination consisting of the futures, the DSGE, the risk-adjusted futures and the BVAR.

Extending this four-model combination further by including a fifth model overall reduces the MSPE gains, see part 4 of Figure 3 ('5m'). In addition, Table 4 indicates that adding the VAR with GDP given its largest MSPE gains on average renders the projection more unstable especially in the more recent period. Finally, including all models into the forecast combination clearly worsens forecast accuracy, see part 5 of Figure 3 ('9m') which indicates that the optimal number of models in the forecast combination is limited.

### 4.2 Performance of the four-model combination

When comparing the performance of this four-model combination with the individual forecasts, it is clear that combining several projections offers noticeable gains in forecast accuracy and generates a forecast whose performance is remarkably more robust over time even when including the period of the global financial crisis and the mid-2014 oil price drop in the evaluation sample. The forecast statistics of the four-model combination are shown in Tables 1-3 and Figure 2.

Concerning forecast accuracy, the four-model combination outperforms the futures-based forecast at every forecast horizon on average over the total sample, with increasing accuracy gains at medium- to longer-term horizons (see Panel A of Table 1). On average across forecast horizons, the four-model combination offers MSPE gains of 14%, with an improvement of about 10% in the short run which increases to a noticeable 23% at the 11 quarter-ahead forecast horizon. In addition, the performance of the four-model combination is remarkably stable over time. In all subsamples considered and for most forecast horizons, the four-model combination predicts Brent oil prices more accurately than the futures; a stability in forecast accuracy no other individual approach considered in this paper can offer (see Panel A of Table 3). This stability in outperformance relative to the futures can also be seen in the Figure 2 as the rolling MSPE of the four-model combination lies below that of the futures over most of the forecast evaluation sample for all forecast horizons shown, including during periods of relatively stable oil prices, upwardly trending prices and during episodes of high oil price volatility such as following the global financial crisis and the mid-2014 oil price drop. The negative bias inherent in the forecast combination projections is smaller than that of the futures and the no-change at every forecast horizon considered (see Panel A of Table 2). In addition, the four-model combination predicts the direction of the change in Brent oil prices better compared to a random guess and the futures with a directional accuracy of 70% on average across forecast horizons (see Panel B of Table 2). Also taking the no-change as the reference point, the model combination`s performance is remarkably better than the individual models offering MSPE gains of 17% on average across forecast horizons with the gains generally increasing with forecast horizon and reaching 27% at the 7 quarter-ahead prediction horizon (see Panel B in Table 1). The outperformance of the four-model combination relative to the no-change is somewhat more subject to time-variation than when compared to the futures (see Panel B in Table 3) although it can offer clear gains in times when oil prices display large shifts as observed over the past few years (see Table 3 and Figure 2).

## 5 Robustness of the results

The results are robust to deleting a specific model from the proposed four-model combination, using different combination weights and predicting WTI instead of Brent oil prices.

First, as the choice of the combination might be sensitive to the ordering in which the models are chosen it is useful to check whether deleting a specific model from the combination can improve the forecast accuracy of the four-model combination we propose.[21] Table 5 lists the gains and losses in MSPE of deleting individual forecasts from the four-model combination. Only excluding the futures might offer some marginal gains in forecast accuracy over the entire forecast evaluation period on

[21] See Baumeister, Kilian and Lee (2014) and Baumeister and Kilian (2015a) who have been using this approach.

average, but this comes at the cost of losing stability in forecast performance. As shown in Panel B of Table 5, the forecast performance of the combination excluding the futures worsens across all horizons over the 1995-2001 evaluation period and most of the horizons over the 2008-2015 sample. In sum, the forecast accuracy of the combination cannot be consistently improved by deleting one of the models from the combination.

Second, there are several ways to choose the weights of the individual forecasts in the forecast combination besides equal weighting, see Timmermann (2006) for a general overview and Baumeister, et al. (2014) and Baumeister and Kilian (2015a) in the case of oil price forecast combinations. We explore the sensitivity of the forecasting performance to a multitude of different weighting schemes: (i) an ordinary least-squares estimate of the weights obtained by regressing actual oil price realizations on the individual forecasts, with and without an intercept in the regression to account for possible biases, (ii) performance-based weights generated as the inverse MSPE ratio which are calculated as the ratio of the inverse MSPE of the individual projections and the sum of inverse MSPE across all models, (iii) decaying weights which give more importance to recent forecasting performance[22], (iv) a trimming approach which discards the worst performing model from the equally-weighted combination, and finally (v) an outperforming weighting scheme which sets the weights equal to the percentage of times that an individual forecast has the minimum MSPE. We calculate the performance-based weights over a one year rolling window to allow for structural breaks in the weights.[23] Table 6, which shows the MSPE of the four-model combination using different combination weights relative to the equal-weighted combination, clearly demonstrates that using equal weights delivers the most accurate prediction. This is in line with the literature that shows that equal weighting often dominates other weighting schemes (Stock and Watson 2004, Aiolfi and Timmermann 2006).

Third, repeating the forecast evaluation exercise for WTI instead of Brent oil prices is interesting as the results become better comparable to the literature that has been focusing on WTI prices and as it would shed light on whether the benefits the four-model combination offers are mainly due to the choice of oil price benchmark. The results in Table 7 show that the benefits of the specific four-model combination generally also hold for forecasting quarterly real WTI crude oil prices although its gains in terms of stability in forecast accuracy are less high than when forecasting Brent oil prices. More specifically, on average over the complete evaluation period, the four-model combination outperforms for WTI oil price forecasting outperforms the futures and the no-change over all forecast horizons considered with the exception on the one quarter-ahead horizons, with MSPE gains of 19% at the 11 quarter-ahead forecast horizon relative to the futures and 22% at the 6 quarter-ahead forecast horizon relative to the no-change. Similar gains are found relative to the no-change. In terms of the magnitude, the accuracy gains offered by the four-model combination are slightly larger compared to those generated by the model combinations proposed by Baumeister and Kilian (2015a) for WTI oil price forecasting, although a

[22] The weights are assumed to decay geometrically by multiplying the squared error by 1/n where n = 1, …T is the total number of observations. From these rescaled weights, we calculate the MSPE of the individual forecast relative to the total MSPE and use the inverse as the combination weights.

[23] The results remain the same when we use a longer rolling window of three years for the weights. These results are available upon request.

different evaluation period is considered. Compared to when forecasting Brent prices, however, the gains in forecast accuracy offered by the four-model combination are less stable over time.

## 6 Density forecast evaluation

Real-time quarterly density forecasts for the individual models are generated by recursive bootstrap simulation methods on the basis of the in-sample monthly forecast errors (see e.g. Baumeister and Kilian, 2014). This method is applied to all models with the exception of the BVAR for which predictive densities are directly generated by the Bayesian estimation method. For the other models, forecast errors are generated differently depending on whether models are estimated/calibrated (i.e. VAR in levels, VAR in differences, VAR with GDP, DSGE and adjusted-futures) and those that are not (i.e. no-change, futures and non-oil index). In the case of estimated or calibrated models, the in-sample forecast errors are generated recursively over each estimation sample. Special care has been taken to correct for the fact that forecast errors at horizons larger than 1 are autocorrelated due to the fact that forecasts are generated iteratively. To recover the i.i.d residuals from which the random draws are generated, an MA(h-1) process is first fitted to the forecast errors at horizon h (see Alquist et al. 2013).[24] For models that are not estimated, forecast errors are recursively calculated based on historical data since 1973M1 or as available in the case of non-oil index and futures.[25] To be consistent with the point forecast evaluation conducted in the paper, all forecast errors are generated at monthly frequency. Density forecasts at monthly frequency are obtained by extracting 1000 random draws from the in-sample monthly forecast errors at each horizon and adding them to the monthly mean forecast. Quarterly density forecasts are then obtained as the quarterly average of the monthly densities. The forecast accuracy of density forecasts is assessed on the basis of the average logarithmic score and the average CRPS statistics computed over the same sample period – 1995Q1 – 2015Q2 – as the point forecast evaluation.[26]

Three combination schemes are used to generate the real-time density forecasts for the model combination: a constant weight scheme which assigns equal-weights to the four models in the combination and two time-varying weights schemes proposed by Billio et al. (2013). These time-varying schemes are compatible with an incomplete model set (meaning that all models can be individually

[24] Note that an MA process could not be fitted to the estimated residuals starting at horizon 6 for VAR in levels and VAR with GDP and at horizon 8 for DSGE and VAR in differences. Therefore, density forecasts for these models were not generated for all horizons.

[25] Non-oil index data is available since August 1980, whereas futures data for Brent oil prices starts in January 1990. Note that this implies a much shorter series of historical forecast errors for futures (and necessarily also for risk-adjusted futures) than for the rest of the models, which may impact the estimated density forecasts of both futures and adjusted-futures.

[26] The average CRPS is computed as $\sum_{t=1}^{T}(\int_{-\infty}^{\infty}(F(z_t) - 1\{y_t \leq z_t\})^2 dz)/T = \sum_{t=1}^{T}(E_p|Y_t - y_t| - 0.5E_f|Y_t - Y_t'|)/T$, where F() is the cdf of $f(y_y|I_{t-1})$ and $Y_t$ and $Y_t'$ are random draws from $f(y_y|I_{t-1})$. The average logarithmic score is computed as $\sum_{t=1}^{T}\ln(f(y_t|\, I_{t-1}))/T$ where $f(y_y|I_{t-1})$ is a predictive density for $y_t$constructed using information up to time $t-1$. The lower the CRPS and the higher the logarithmic score, the higher the accuracy of the predictive density.

misspecified). For both schemes, weight dynamics are driven by the past performance of the density forecasts, and one of them also considers learning mechanisms. For more details we refer to the Billio et al. (2013) paper and also to the Matlab implementation described in Casarin et al. (2013).

Looking first at the individual models, we begin by inspecting the density forecasts illustrated at the 3-quarters ahead horizon in Figure 4 (density forecasts for other horizons are available from the authors upon request). Several observations are noteworthy. First, all models generally display asymmetric probability bands throughout the sample. Most models (most notably no-change, VAR in differences, DSGE and BVAR) estimate a wider price range above the median than below it, as illustrated by the upper 95% probability band which is wider than the equivalent lower band throughout the sample period. This suggests that there is larger room for oil prices turning higher than the median forecast rather than below it. This pattern is generally preserved at longer horizons. In contrast, for the recent part of the sample since 2010, futures and adjusted-futures tend to estimate instead a larger price range below the median forecast than above it at horizons up to 4-quarters ahead. Starting with the 5-quarter ahead horizon, however, the probability bands around the median forecast become more similar for both futures and adjusted-futures.

Second, the level of uncertainty around the median oil price forecast differs quite significantly across models. To measure uncertainty we construct a metric that is invariant to the shifts in the oil price level, namely the ratio between the 90% probability price range and the median oil price forecast, which is displayed in Figure 5 for the three-quarter ahead horizon. The figure illustrates that the no-change, the non-oil index or the BVAR display much higher levels of uncertainty (with ratios higher than 0.6) than the VAR in levels or the VAR with GDP (ratios around 0.4). A ratio of 0.4 means that the price range between the 95% and the 5% quantiles is 40% of the median price forecast, implying that with 90% probability the price is forecast to lie within the range given by the median forecast +/-20% if the estimated probability distribution is symmetric. Moreover, the models also differ in terms of how the level of uncertainty varies over time. Whereas for certain models (VAR in levels, VAR with GDP and DSGE) the level of uncertainty remains broadly constant over time, it increases slightly for others (VAR in differences and BVAR) while it decreases for the remaining ones (no-change, futures, non-oil index and adjusted futures). More interestingly, in the case of futures, adjusted-futures and non-oil index we observe large jumps in the level of uncertainty at certain points during the sample, which correspond to periods of large drops in oil prices, such as during 2008 and 1998. This indicates that during periods of falling oil prices, these forecasts are accompanied by a sharp increase in the levels of uncertainty.

In terms of density evaluation statistics (see Table 8), the BVAR appears to outperform all other models according to both the logarithmic score and the CRPS over the whole sample period. The no-change, VAR in differences and the DSGE model at longer horizons also show relatively good forecast accuracy which is similar to that of the BVAR. In contrast, the futures, adjusted-futures and the non-oil index show the lowest accuracy especially at longer-horizons and especially according to the logarithmic score.

Turning to the four-model combination, the evaluation statistics in Table 8 indicate that of the three combination schemes considered, the combination with time-varying weights without learning outperforms the alternatives and is slightly better than the one with time-varying weights and learning.

Thus, when considering the uncertainty around the point forecast, allowing the combination weights to vary over time outperforms equal-weight strategies, a result that confirms existing evidence in macroeconomics (e.g. Jore et al., 2010). Moreover, the best performing four-model combination outperforms all individual models by a large margin especially at longer horizons in terms of both accuracy measures. For example, it outperforms the BVAR at the 1-quarter ahead horizon by 21% and 16% in terms of CRPS and logarithmic score, respectively, and at the 7-quarter horizon by 45% and 35%, respectively. As regards uncertainty, the level of uncertainty in the case of the four model combination fluctuates between 0.1 and 0.6 at the three-quarter ahead horizon, showing big jumps at particular episodes of large drops in oil prices, such as the peak corresponding to the forecast for 2009Q3 done in 2008Q4.

Finally, Figure 6 illustrates the evolution of the models' weights over time. The figure highlights that weights vary considerably across horizons and over time. At shorter horizons, e.g. at the three-quarters ahead, the combination loads generally evenly in the four models – slightly more on the futures and adjusted-futures, with weight around 0.3 each –, except for specific episodes when the BVAR and especially the DSGE load more significantly. Such episodes appear to correspond to periods of sustained oil price increases, such as the second half of 1996, during 1999-2000 and between 2002 and 2006. During the 2002 and 2006 oil price run-up, for example, the weight of the DSGE model increases from nearly 0 to 0.6 to the detriment of futures and adjusted-futures. In contrast, at longer horizons – illustrated in Figure 7 by the seven-quarter ahead horizon – the combination loads mostly on futures, adjusted-futures and DSGE, while the BVAR contributes only about 0.1 on average throughout the sample and without large variations (this may indicate that the BVAR is better suitable for short-term forecasting). Moreover, the fluctuation in the weights of these three models is much more pronounced than at shorter horizons. In particular, the loading on the DSGE appears as the reverse of that on futures, while the loading on the adjusted futures mirrors that of futures but with less fluctuation. Similar to the picture at the short-horizon, the loading on the DSGE increases significantly during periods of sustained oil price rises, in contrast to futures whose loading increases more during periods of stable or falling oil prices. For example, during the 2002-2006 episode, the combination loads increasingly more on the DSGE, peaking at 0.92 for the oil price forecast of 2005Q4 to the detriment of futures, whose contribution drops to nearly 0. In contrast, during the 1997-1998 period of oil price decreases, the weight on futures increases reaching a maximum of 0.81 for the forecast corresponding to 1999Q2 to the detriment of the DSGE whose weight drops to zero.

## 7 Conclusions

This paper reviews the forecasting performance of a set of methods for predicting quarterly real Brent oil prices. The aim is to find a tool offering a consistent improvement relative to futures-based predictions up to 11 quarters ahead that also delivers a stable performance over time given continuous changes in oil price dynamics.

The forecast approaches evaluated are the random walk, the futures, the risk-adjusted futures based on Pagano and Pisani (2009), a non-oil commodity price index, a DSGE model of the oil market based on Nakov and Nuño (2014) and four alternatives of the Baumeister and Kilian (2012) oil market VAR model: one with Brent oil prices in log levels, one with Brent prices in first differences, one in which the Kilian (2009) index of economic activity is replaced with a monthly global GDP measure and finally, one in which the VAR with prices in levels is estimated using Bayesian techniques.

Remarkably, none of the individual forecast approaches manages to consistently predict Brent oil prices better than the futures or random walk over time or across forecast horizons. Nevertheless, this performance instability can be exploited by combining projections of different models into one forecast that delivers a more accurate forecast and stable performance when oil price dynamics change. In this paper, we propose an equally-weighted four-model forecast combination for quarterly Brent oil price forecasting consisting of Brent oil futures, risk-adjusted futures, a BVAR and DSGE model of the oil market. This combination predicts Brent oil prices more accurately in real time and out-of-sample compared to the futures and the random walk up to 11 quarters ahead, with MSPE gains reaching 23% at the 11 quarter-ahead horizon relative to the futures. In addition, the four-model combination generates a forecast whose performance is remarkably robust over time, outperforming the futures on average in periods of stable, upwardly trending or rapidly falling oil prices. Moreover, the model combination reduces the forecast bias and predicts the direction of oil price changes more accurately than a random guess with a directional accuracy of 70% on average. Similar gains in forecast performance are found using this four-model combination for predicting WTI oil prices. In addition, the four-model combination (with time-varying weights) outperforms all individual models also when taking the full probability distribution of the forecasts into account.

## References


Aiolfi, M. and A. Timmermann (2006), “Persistence in Forecasting Performance and Conditional Combination Strategies”, Journal of Econometrics, 35 (1-2), 31-53.

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

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

Billio, M., Casarin R., Ravazzolo F. and H.K.van Dijk (2013), "Time-varying combinations of predictive densities using nonlinear filtering," Journal of Econometrics, 177, 213-232.

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 (2014), “What Central Bankers Need to Know about Forecasting Oil Prices,” International Economic Review, 55, 869-889.

Baumeister, C. and L. Kilian (2015a), “Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach,” Journal of Business and Economic Statistics, 33(3), July 2015, 338-351.

Baumeister, C. and L. Kilian (2015b), “A General Approach to Recovering Market Expectations from Futures Prices With an Application to Crude Oil,” mimeo.

Baumeister, C. and L. Kilian (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, March 2016.

Baumeister, C., L. Kilian and X. Zhou (2013), “Are Product Spreads Useful for Forecasting? An Empirical Evaluation of the Verleger Hypothesis,” forthcoming Macroeconomic Dynamics.

Baumeister, C., L. Kilian and T. Lee (2014), “Are there gains from pooling real-time oil price forecasts?” Energy Economics, 46, 33-43.

Casarin, R., S. Grassi, F. Ravazzolo and H.K. van Dijk (2013) "Parallel sequential Monte Carlo for efficient density combination: the DeCo Matlab Toolbox", Working Paper No. 08, Ca' Foscari University of Venice, Department of Economics.

Clark, T. (2011), "Real-time density forecasts from BVARs with stochastic volatility", Journal of Business and Economic Statistics, 29, 327-341.

Clark, T. and F. Ravazzolo (2014), "The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility", Journal of Applied Econometrics, 30 (4), 551-575.

de Menezes, L.M., D.W. Bunn and J.W. Taylor, (2000), "Review of Guidelines for the Use of Combined Forecasts", European Journal of Operational Research, 120, 190–204.

Giannone, D., M. Lenza, and G. Primiceri (2014), "Priors for the long run," mimeo.

Giannone, D., M. Lenza, and G. Primiceri (2015), "Prior Selection for Vector Autoregressions," Review of Economics and Statistics, 97(2), 436-451.

Hamilton, J. and C. Wu (2014), "Risk Premia in Crude Oil Futures Prices", Journal of International Money and Finance, 42, 9-1-37.

Hendry, D. and M.P. Clements (2002), "Pooling of Forecasts", Econometrics Journal, 5, 1-26.

Jore, A.S., J. Mitchell and S.P. Vahey (2010), "Combining forecast densities from VARs with uncertain instabilities", Journal of Applied Econometrics, 25 (4), 621-634.

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.

Nakov, A. and G. Nuño (2014), "Saudi Arabia and the Oil Market", Economic Journal, 132, 1333-1362.

Pagano P. and M. Pisani (2009), "Risk-Adjusted Forecasts of Oil Prices", The B.E. Journal of Macroeconomics, 9(1), 1-28.

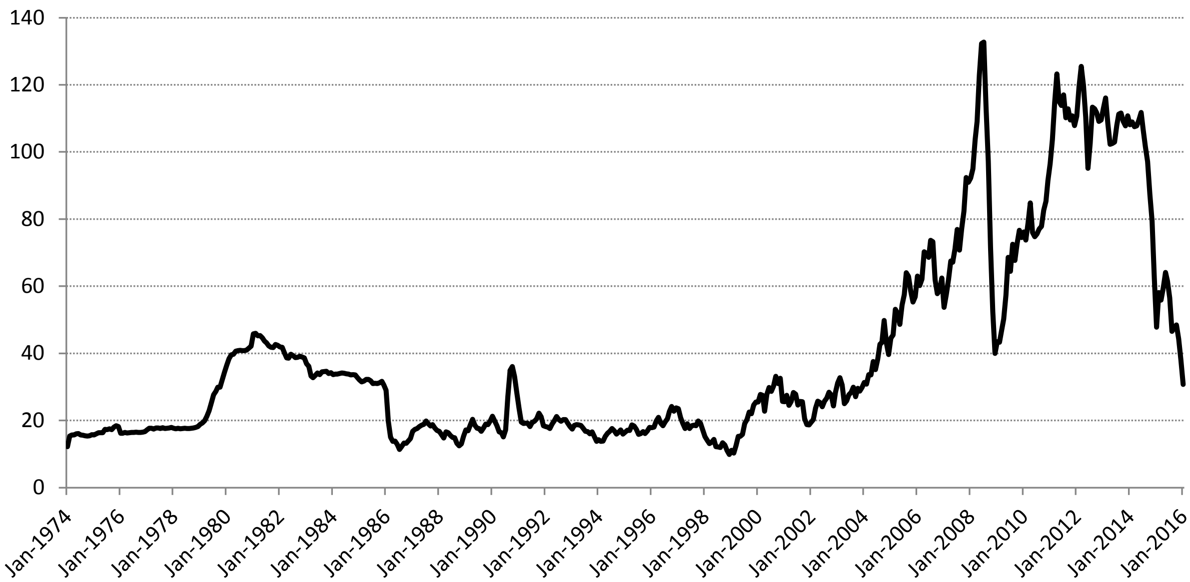
Reeve T.A. and R.J. Vigfusson (2011), "Evaluating the Forecasting Performance of Commodity Futures Prices," International Finance Discussion Papers 1025, Board of Governors of the Federal Reserve System (U.S.).

Singleton, K.J. (2014), "Investor Flows and the 2008 Boom/Bust in Oil Prices", Management Science, 60, 300–318.

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.

Timmermann, A. (2006), "Forecast Combinations", Handbook of Economic Forecasting, 1, 135–196.

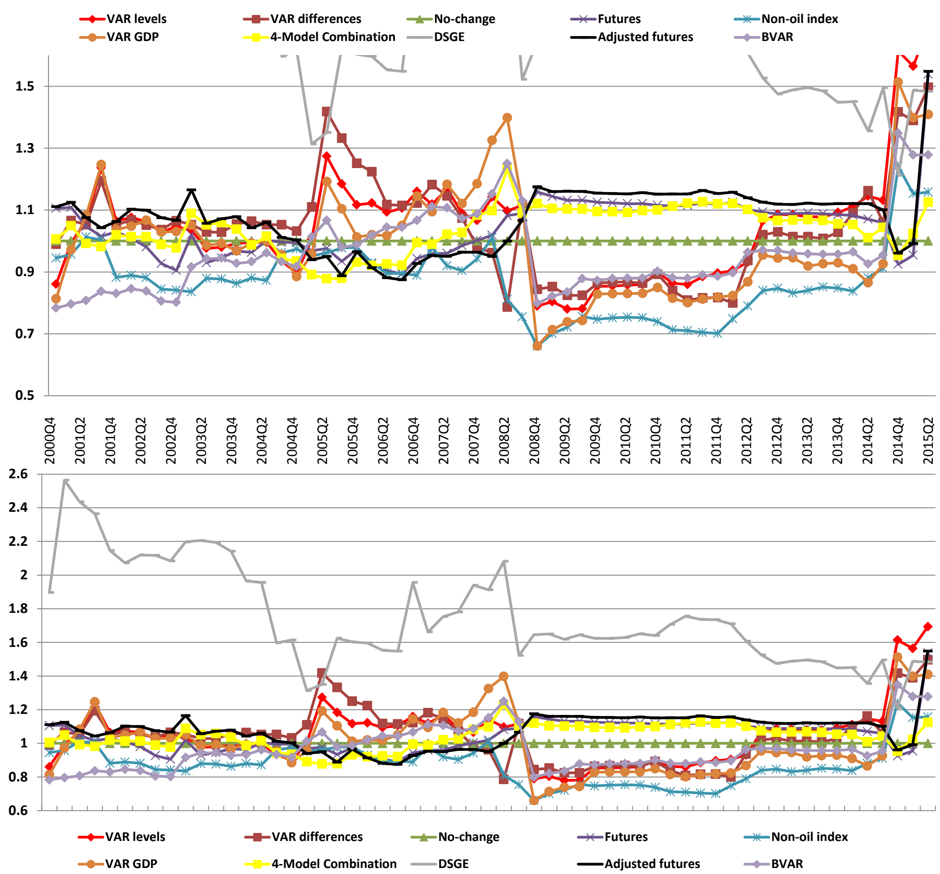
**Figure 1. Real Brent oil price over time**



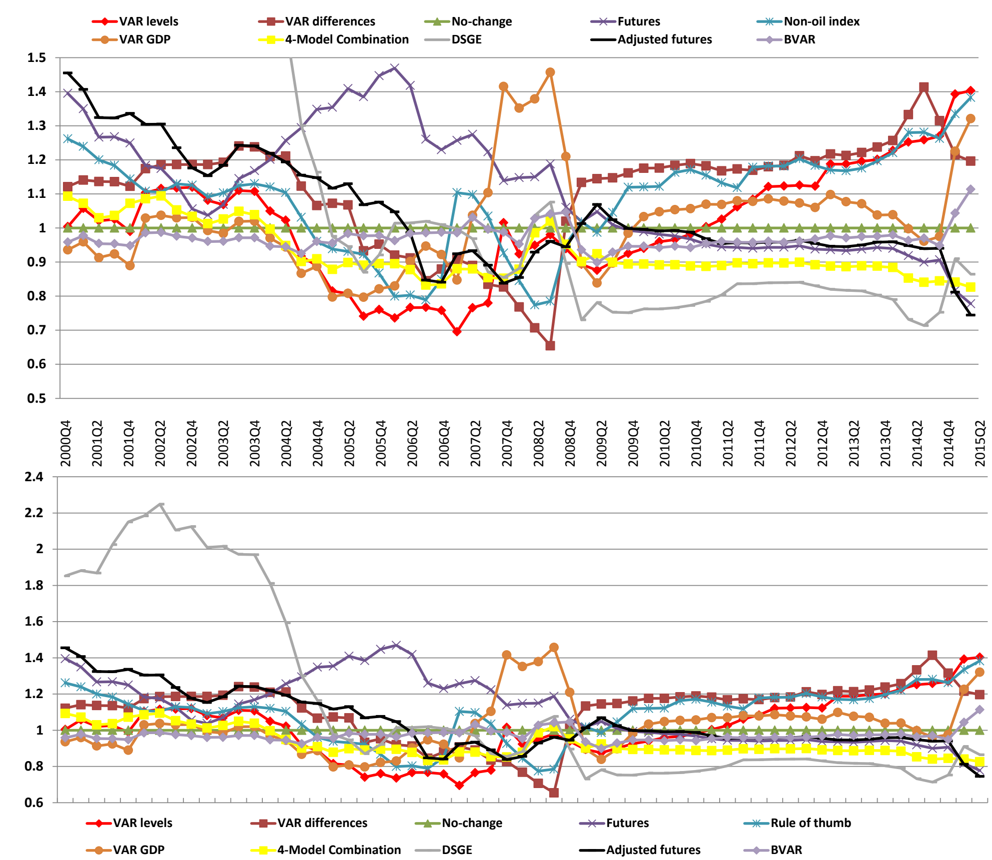

Source: Energy Information Administration and FRED economic data
Notes: Monthly data, the nominal Brent oil price in USD per barrel has been backcasted with the growth rate of the refiner`s acquisition cost of imported crude oil up to May 1987 and has been made real using US CPI, index 1982-84=100.

## Figure 2. Recursive MSPE over 6-year rolling window of different Brent crude oil forecasting methods

### PANEL A: 1 quarter-ahead forecast horizon



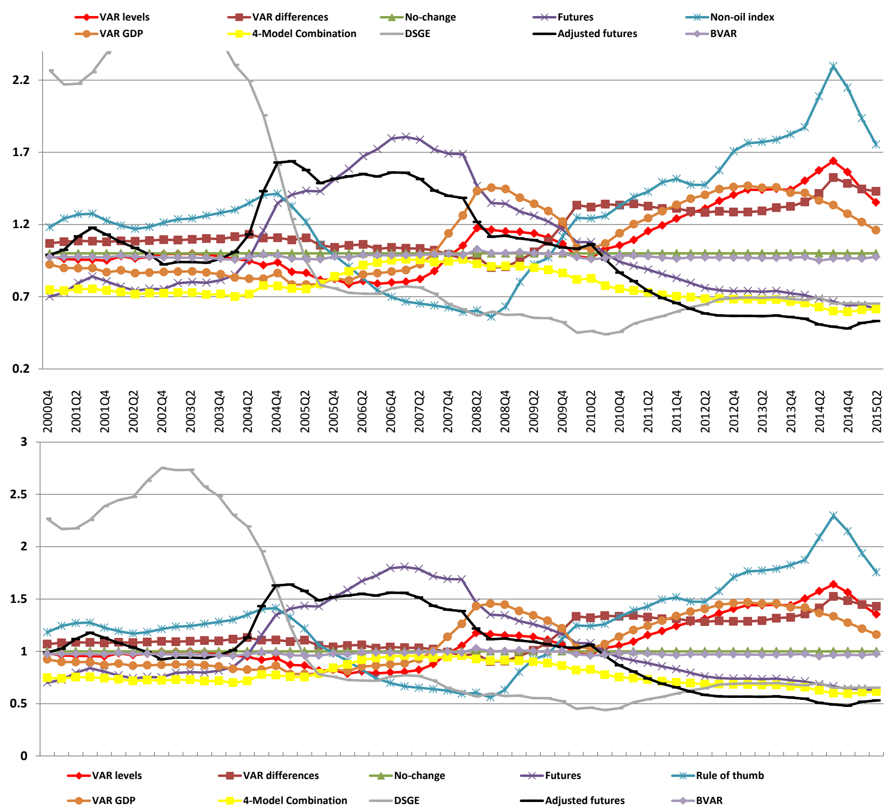

### PANEL B: 3 quarter-ahead forecast horizon



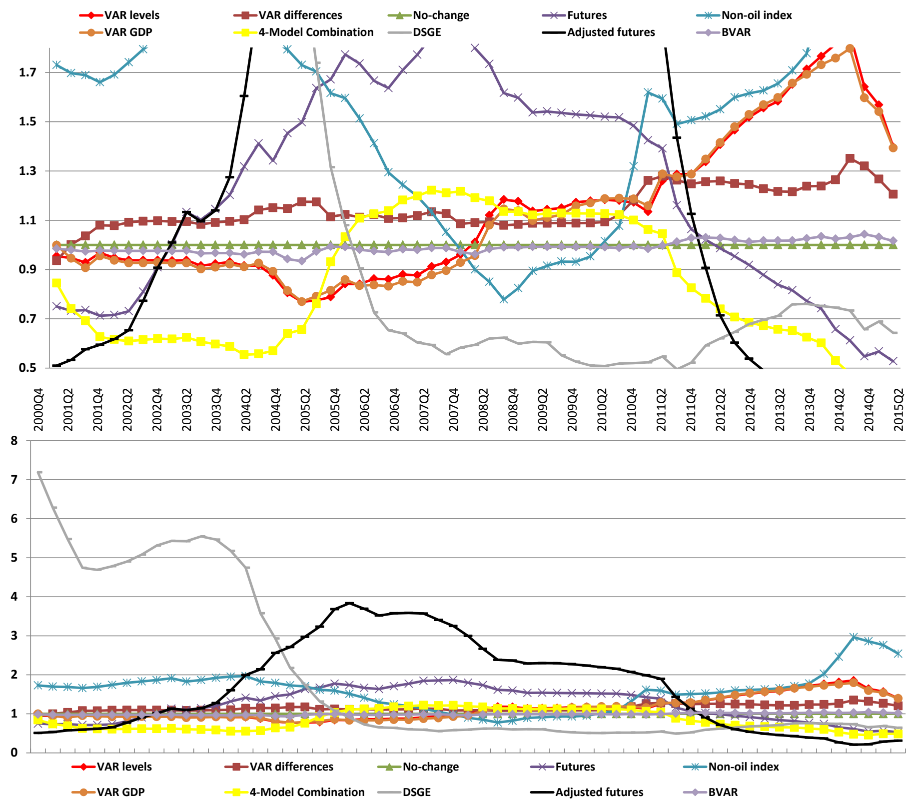

Notes: The lines represent the recursive mean squared prediction errors (MSPE) over a rolling window of six years relative to the no-change forecast over the period 2000Q4 – 2015Q2 for the different forecasting methods including the forecast combination. Per panel, the same chart is shown twice on two different scales due to the large forecast errors of some models. A value lower than one means that the method outperforms the random walk forecast over the past six years on average. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

**Figure 2 continued. Recursive MSPE over 6-year rolling window of different Brent crude oil forecasting methods**

**PANEL C: 7 quarter-ahead forecast horizon**

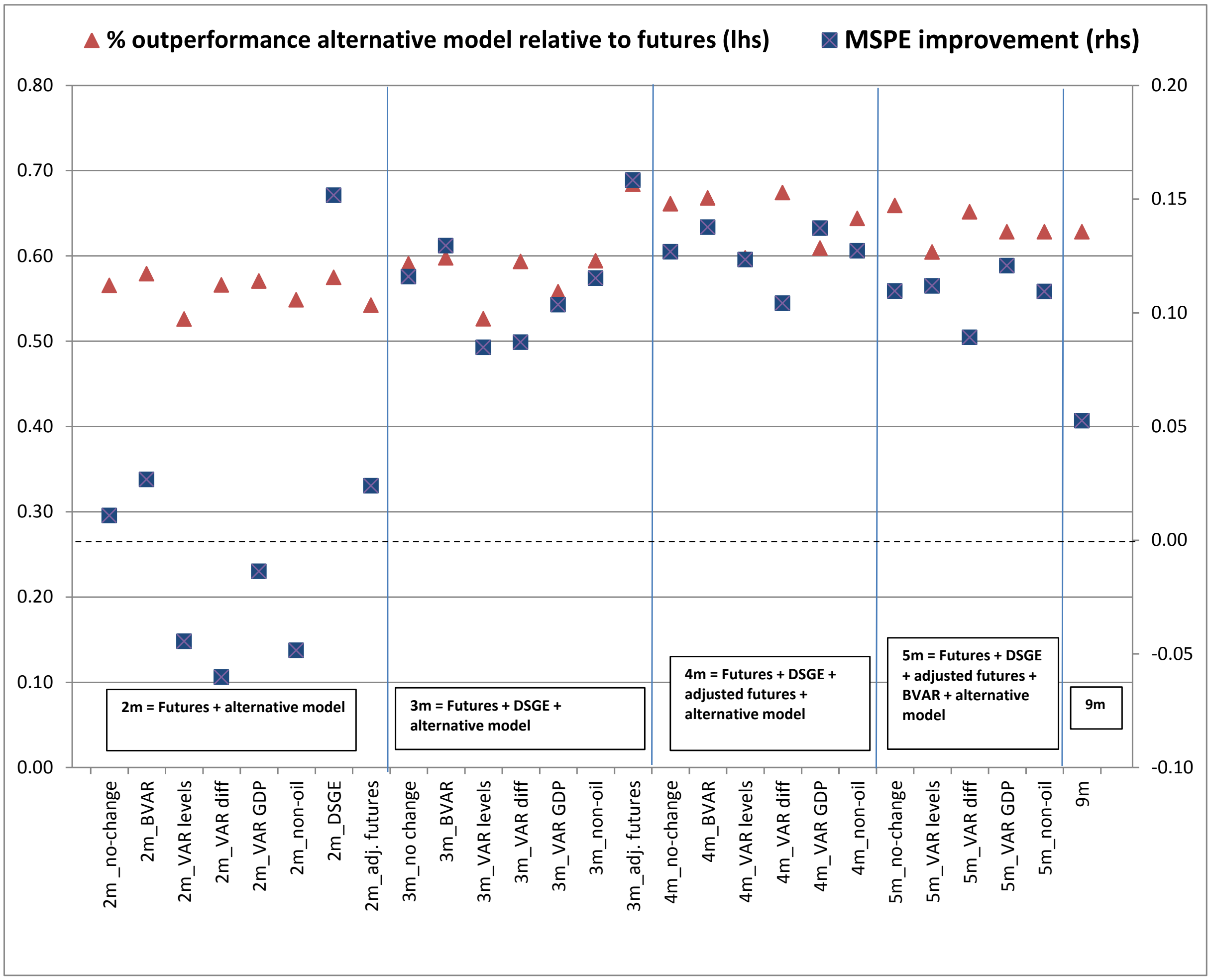



**PANEL D: 11 quarter-ahead forecast horizon**



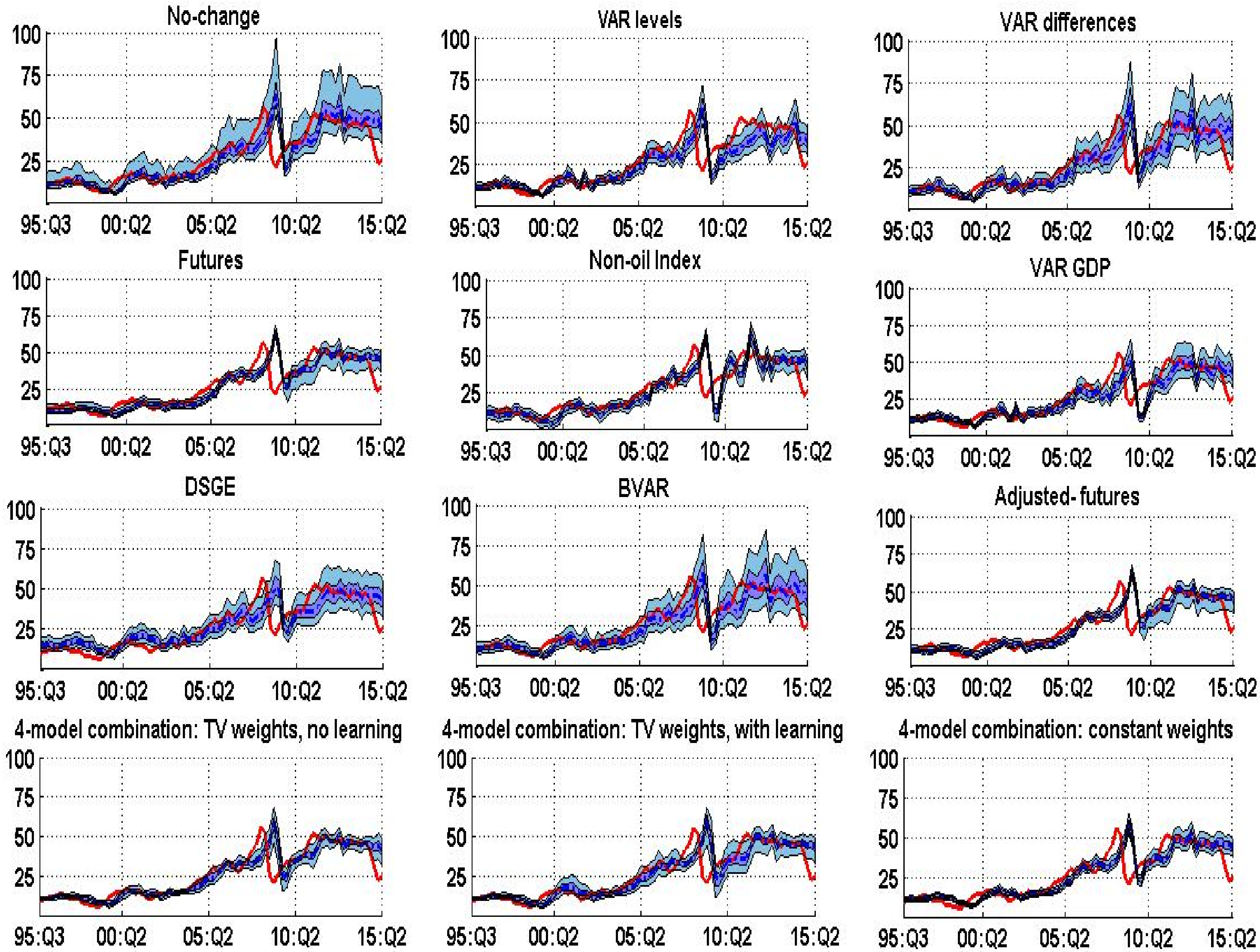

Notes: The lines represent the recursive mean squared prediction errors (MSPE) over a rolling window of six years relative to the no-change forecast over the period 2000Q4 – 2015Q2 for the different forecasting methods including the forecast combination. Per panel, the same chart is shown twice on two different scales due to the large forecast errors of some models. A value lower than one means that the method outperforms the random walk forecast over the past six years on average. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

**Figure 3. Forecast evaluation of different forecast combinations relative to futures**



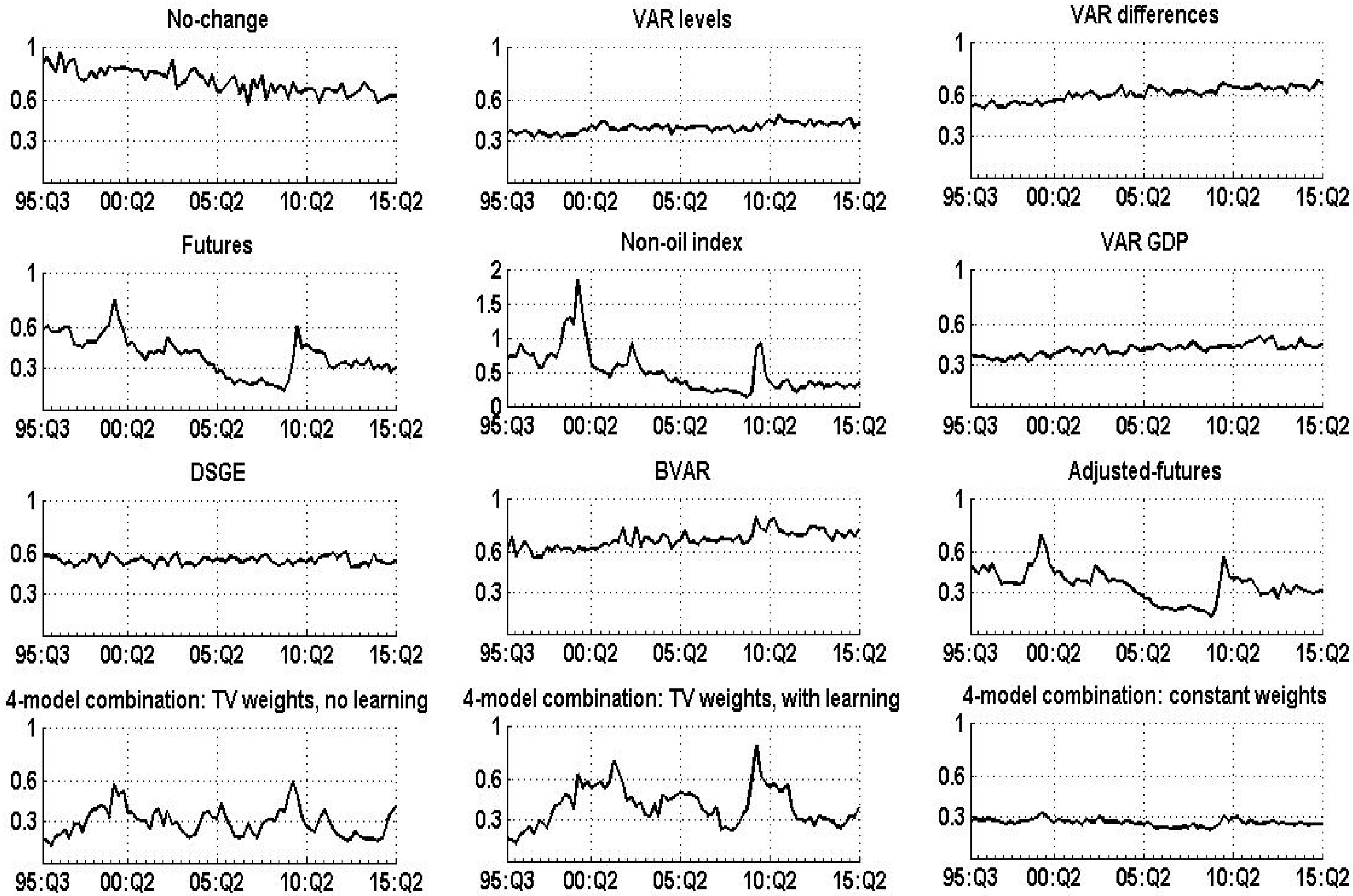

Notes: this chart shows the average forecast performance over the sample 1995Q1 – 2015Q2 of different forecasting models relative to the futures.The alternative models whose performance is compared to the futures are listed on the X-axis; 2m, 3m, 4m,5m and 9m indicate that the alternative model is a forecast combination consisting of respectively 2,3,4,5 and 9 individual models. The triangles represent the percentage times that the alternative model outperforms the futures (shown on the left Y-axis); a value of 0.60 for example indicates that a specific model combination outperforms the futures 60% of the time over the complete evaluation period. The squares indicate the MSPE improvement of the alternative model relative to the futures, which has been averaged over the different forecast horizons (shown on the right Y-axis). A MSPE improvement higher than zero indicates that the alternative model outperforms the futures in average MSPE sense; a value of 0.10 means that the alternative model has an MSPE 10% lower than that of the futures.

**Figure 4. Predictive densities of quarterly real Brent oil price at the 3-quarter-ahead horizon**




Notes: The bloack lines outside the light blue shaded area represents the 5% and 95% quantiles, while the black lines bordering the dark blue area represent the 25% and 75% quantiles. The blue dotted line represents the median of the predictive density, while the red line represents the outcome price.

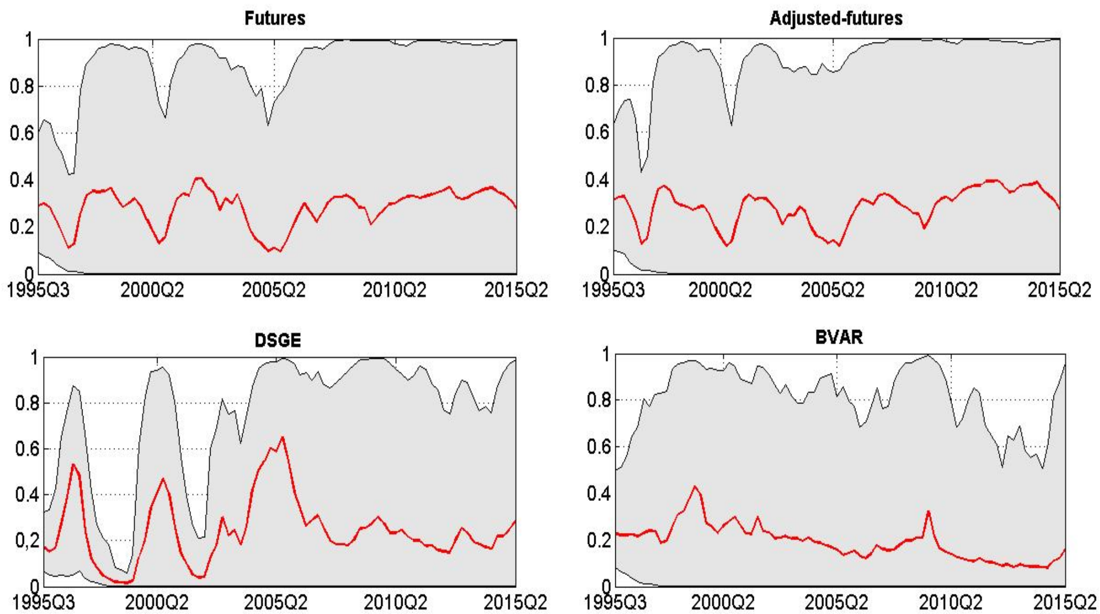
**Figure 5. A measure of the uncertainty level around the median forecast densities at the 3-quarter ahead horizon for real Brent oil price**



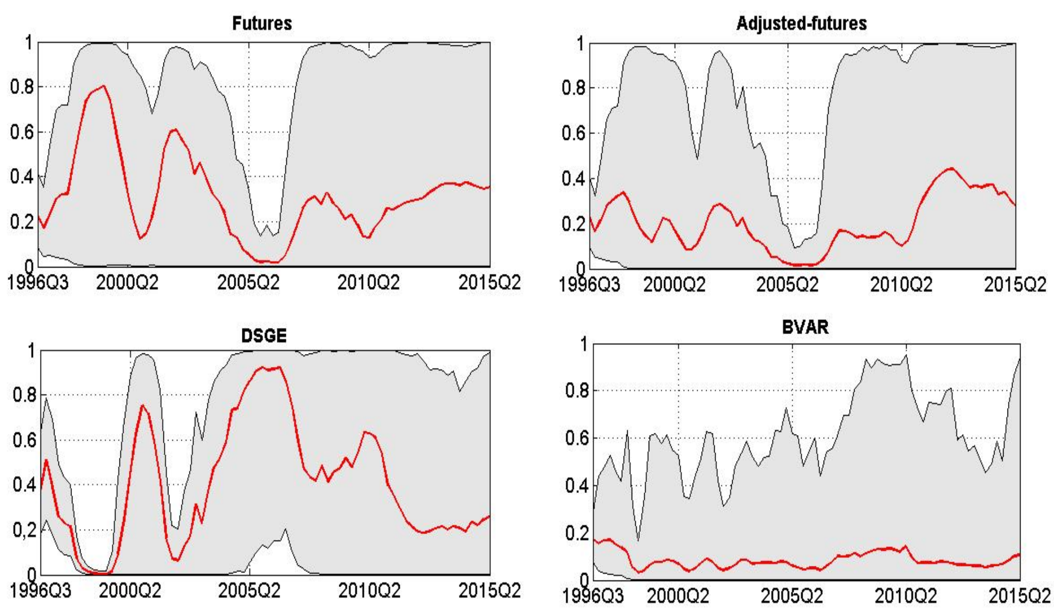

Notes: The level of uncertainty around the median forecast is computed as the ratio between the 90% probability ranage – i.e. the differenbce between the 95% and 5% quantiles – and the median price forecast. The higher the ratio, the higher the level of uncertainty. A ratio of 1 indicates that with 90% probability the outcome price is forecast to lie within the range given by the median forecast +/-50%, if the probability distribution is symmetric.

**Figure 6. Time-varying weights without learning of the 4-model combination**

**Panel A: 3-quarter ahead horizon**




**Panel B: 7-quarter ahead horizon**




Notes: Average filtered time-varying weights without learning (solid line) with 2.5% and 97.5% quantiles (grey area). Note that the quantiles are obtained using different draws from the predictive densities.

## Table 1. MSPE of real quarterly Brent oil price forecasts: complete forecast evaluation period

### PANEL A: relative to the futures forecast

| Horizon (Quarters) | 1995-2015 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | VAR levels | VAR differences | No-change | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination |
| 1 | **0.86** | **0.83** | **0.80** | **0.69** | **0.77** | 1.30 | **0.78** | 1.02 | **0.89** |
| 2 | 1.03 | 1.06 | **0.96** | 1.01 | **0.91** | **0.99** | **0.90** | 1.03 | **0.93** |
| 3 | 1.11 | 1.21 | 1.04 | 1.18 | 1.06 | **0.91** | 1.02 | 1.01 | **0.93** |
| 4 | 1.23 | 1.36 | 1.11 | 1.17 | 1.15 | **0.96** | 1.08 | **0.97** | **0.93** |
| 5 | 1.38 | 1.51 | 1.16 | 1.22 | 1.26 | **0.95** | 1.14 | **0.94** | **0.91** |
| 6 | 1.48 | 1.56 | 1.17 | 1.39 | 1.37 | **0.94** | 1.17 | **0.90** | **0.88** |
| 7 | 1.43 | 1.44 | 1.15 | 1.62 | 1.38 | **0.92** | 1.13 | **0.90** | **0.83** |
| 8 | 1.37 | 1.34 | 1.09 | 1.76 | 1.32 | **0.89** | 1.08 | **0.95** | **0.81** |
| 9 | 1.32 | 1.28 | 1.08 | 1.94 | 1.27 | **0.90** | 1.07 | 1.04 | **0.80** |
| 10 | 1.29 | 1.23 | 1.05 | 1.90 | 1.26 | **0.88** | 1.05 | 1.16 | **0.80** |
| 11 | 1.24 | 1.13 | **0.99** | 1.76 | 1.23 | **0.87** | **0.99** | 1.23 | **0.77** |

### PANEL B: relative to the no-change forecast

| Horizon (Quarters) | 1995-2015 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination |
| 1 | 1.06 | 1.04 | 1.24 | **0.86** | **0.95** | 1.61 | **0.97** | 1.27 | 1.11 |
| 2 | 1.07 | 1.10 | 1.04 | 1.05 | **0.94** | 1.03 | **0.94** | 1.06 | **0.96** |
| 3 | 1.07 | 1.16 | **0.96** | 1.13 | 1.01 | **0.87** | **0.98** | **0.97** | **0.89** |
| 4 | 1.11 | 1.23 | **0.90** | 1.06 | 1.04 | **0.87** | **0.98** | **0.88** | **0.84** |
| 5 | 1.19 | 1.30 | **0.86** | 1.06 | 1.09 | **0.82** | **0.98** | **0.81** | **0.79** |
| 6 | 1.27 | 1.34 | **0.85** | 1.18 | 1.17 | **0.81** | **1.00** | **0.77** | **0.75** |
| 7 | 1.24 | 1.25 | **0.87** | 1.41 | 1.20 | **0.80** | **0.98** | **0.78** | **0.73** |
| 8 | 1.26 | 1.23 | **0.92** | 1.61 | 1.21 | **0.82** | **0.99** | **0.87** | **0.74** |
| 9 | 1.22 | 1.19 | **0.93** | 1.80 | 1.18 | **0.83** | **0.99** | **0.96** | **0.74** |
| 10 | 1.23 | 1.17 | **0.95** | 1.81 | 1.20 | **0.84** | **1.00** | 1.10 | **0.76** |
| 11 | 1.26 | 1.15 | 1.01 | 1.78 | 1.24 | **0.88** | **1.00** | 1.24 | **0.78** |

Notes: The tables show the mean squared prediction errors (MSPE) relative to the futures (Panel A) and no-change forecast (Panel B) over the total sample period 1995Q1 – 2015Q2 for various individual forecasting methods and the forecast combination (in the grey column). A value lower than one - indicated in bold - means that the method outperforms the benchmark on average over the sample period. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

## Table 2. Bias and success ratio of real quarterly Brent oil price forecasts: complete forecast evaluation period

### PANEL A: Forecast bias

| Horizon (Quarters) | 1995-2015 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-change | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | **-0.22** | **-0.66** | **-0.02** | **-0.07** | **-0.16** | **-0.71** | **0.07** | **-0.25** | **0.10** | **-0.04** |
| 2 | **-0.38** | **-1.59** | **0.01** | **-0.73** | **-0.11** | **-1.56** | **0.25** | **-0.33** | **-0.10** | **-0.23** |
| 3 | **-0.56** | **-2.18** | **0.01** | **-1.32** | **-0.03** | **-2.19** | **0.35** | **-0.56** | **-0.30** | **-0.46** |
| 4 | **-0.95** | **-2.88** | **-0.31** | **-2.15** | **-0.22** | **-3.18** | **0.36** | **-1.01** | **-0.77** | **-0.89** |
| 5 | **-1.42** | **-3.74** | **-0.77** | **-2.99** | **-0.46** | **-4.28** | **0.37** | **-1.46** | **-1.22** | **-1.33** |
| 6 | **-1.88** | **-4.47** | **-1.14** | **-3.80** | **-0.65** | **-5.05** | **0.40** | **-1.87** | **-1.75** | **-1.76** |
| 7 | **-2.36** | **-5.37** | **-1.82** | **-4.62** | **-0.81** | **-5.73** | **0.46** | **-2.41** | **-2.55** | **-2.28** |
| 8 | **-2.85** | **-6.22** | **-2.52** | **-5.37** | **-0.88** | **-6.39** | **0.52** | **-2.91** | **-3.40** | **-2.79** |
| 9 | **-3.28** | **-6.94** | **-3.04** | **-6.10** | **-0.78** | **-6.95** | **0.63** | **-3.33** | **-4.17** | **-3.24** |
| 10 | **-3.78** | **-7.59** | **-3.55** | **-6.82** | **-0.87** | **-7.51** | **0.67** | **-3.83** | **-4.94** | **-3.73** |
| 11 | **-4.33** | **-8.30** | **-4.21** | **-7.59** | **-1.30** | **-8.14** | **0.70** | **-4.41** | **-5.54** | **-4.21** |

### PANEL B: Success ratio

| Horizon (Quarters) | 1995-2015 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-change | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | **0.60** | **0.70** | **0.70** | **0.60** | **0.66** | **0.66** | **0.54** | **0.66** | **0.66** | **0.70** |
| 2 | **0.63** | **0.60** | **0.62** | **0.58** | **0.62** | **0.60** | **0.58** | **0.63** | **0.69** | **0.62** |
| 3 | **0.63** | **0.54** | **0.60** | **0.63** | **0.66** | **0.63** | **0.64** | **0.61** | **0.64** | **0.64** |
| 4 | **0.63** | **0.63** | **0.59** | **0.68** | **0.65** | **0.63** | **0.68** | **0.57** | **0.70** | **0.73** |
| 5 | **0.55** | **0.58** | **0.53** | **0.64** | **0.59** | **0.59** | **0.69** | **0.53** | **0.67** | **0.76** |
| 6 | **0.52** | **0.60** | **0.49** | **0.65** | **0.57** | **0.58** | **0.75** | **0.49** | **0.62** | **0.78** |
| 7 | **0.51** | **0.59** | **0.46** | **0.59** | **0.53** | **0.57** | **0.76** | **0.50** | **0.55** | **0.76** |
| 8 | **0.49** | **0.56** | **0.45** | **0.60** | **0.49** | **0.53** | **0.75** | **0.48** | **0.57** | **0.72** |
| 9 | **0.47** | **0.61** | **0.45** | **0.59** | **0.38** | **0.59** | **0.74** | **0.47** | **0.61** | **0.70** |
| 10 | **0.49** | **0.56** | **0.44** | **0.62** | **0.36** | **0.56** | **0.71** | **0.49** | **0.62** | **0.68** |
| 11 | **0.53** | **0.57** | **0.47** | **0.57** | **0.33** | **0.54** | **0.75** | **0.53** | **0.53** | **0.61** |

Notes: The tables show the forecast bias (Panel A) and the success ratio (Panel B) over the total sample period 1995Q1 – 2015Q2 for various forecasting methods including the forecast combination (in the grey column). A negative number for the bias indicates and underprediction of actual oil prices on average and vice versa. The success ratio gives the probability of predicting the correct change in the oil price. A value higher than 0.5 denotes a higher directional accuracy than at random. For the reason that the quarterly no-change forecast is calculated as the last monthly observation for the real price of oil following Baumeister and Kilian (2014), the success ratio for the no-change can differ from 0.5 as we are comparing a monthly price forecast to an average realized quarterly oil price.The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

## Table 3. MSPE of real quarterly Brent oil price forecasts: different subsamples

### PANEL A: relative to the futures-based forecast

| Horizon (Quarters) | 1995-2001 | | | | | | | | | 2002-2007 | | | | | | | | | 2008-2015 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VAR levels | VAR differences | No-change | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination | VAR levels | VAR differences | No-change | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination | VAR levels | VAR differences | No-change | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination |
| 1 | 1.04 | 1.03 | **0.97** | **0.86** | 1.02 | 2.14 | **0.82** | 1.02 | **0.99** | 1.06 | **0.98** | **1.00** | **0.94** | 1.18 | 1.94 | 1.08 | **0.96** | 1.08 | **0.82** | **0.81** | **0.77** | **0.65** | **0.70** | 1.18 | **0.74** | 1.03 | **0.86** |
| 2 | 1.06 | 1.09 | **0.89** | **0.93** | 1.02 | 1.67 | **0.86** | 1.05 | **0.92** | **0.83** | **0.65** | **0.88** | **0.65** | 1.07 | 1.14 | **0.78** | **0.85** | **0.83** | 1.05 | 1.10 | **0.98** | 1.05 | **0.89** | **0.94** | **0.92** | 1.04 | **0.94** |
| 3 | **0.80** | **0.90** | **0.80** | **0.93** | **0.71** | 1.82 | **0.76** | 1.07 | **0.86** | **0.89** | **0.73** | **0.88** | **0.81** | 1.24 | **0.75** | **0.87** | **0.74** | **0.75** | 1.16 | 1.30 | 1.08 | 1.24 | 1.05 | **0.88** | 1.05 | 1.04 | **0.96** |
| 4 | **0.84** | **0.97** | **0.89** | 1.09 | **0.74** | 2.10 | **0.86** | 1.17 | **0.89** | **0.58** | **0.56** | **0.71** | **0.48** | **0.91** | **0.54** | **0.68** | **0.62** | **0.61** | 1.38 | 1.54 | 1.20 | 1.31 | 1.23 | **0.94** | 1.18 | 1.02 | **1.00** |
| 5 | 1.04 | 1.19 | 1.09 | 1.34 | **0.90** | 2.19 | 1.06 | 1.24 | **0.91** | **0.60** | **0.60** | **0.65** | **0.41** | **0.83** | **0.45** | **0.63** | **0.63** | **0.56** | 1.65 | 1.81 | 1.31 | 1.45 | 1.43 | **0.97** | 1.29 | **1.00** | 1.02 |
| 6 | 1.15 | 1.26 | 1.16 | 1.40 | 1.01 | 2.50 | 1.14 | 1.25 | **0.91** | **0.55** | **0.54** | **0.60** | **0.33** | **0.70** | **0.42** | **0.58** | **0.67** | **0.55** | 1.93 | 2.04 | 1.42 | 1.84 | 1.70 | **0.99** | 1.42 | **0.95** | 1.02 |
| 7 | 1.18 | 1.34 | 1.24 | 1.52 | 1.08 | 2.96 | 1.21 | 1.40 | **0.92** | **0.58** | **0.59** | **0.59** | **0.37** | **0.67** | **0.38** | **0.59** | **0.83** | **0.56** | 1.96 | 1.95 | 1.47 | 2.36 | 1.84 | **0.99** | 1.44 | **0.88** | **0.99** |
| 8 | 1.32 | 1.51 | 1.40 | 1.86 | 1.26 | 3.64 | 1.36 | 1.38 | **0.94** | **0.58** | **0.63** | **0.57** | **0.40** | **0.61** | **0.37** | **0.57** | 1.08 | **0.58** | 1.91 | 1.80 | 1.41 | 2.66 | 1.81 | **0.96** | 1.39 | **0.83** | **0.95** |
| 9 | 1.46 | 1.67 | 1.56 | 2.29 | 1.40 | 4.67 | 1.51 | 1.23 | **0.96** | **0.52** | **0.59** | **0.53** | **0.40** | **0.53** | **0.35** | **0.52** | 1.34 | **0.60** | 1.92 | 1.78 | 1.46 | 3.09 | 1.83 | **0.99** | 1.45 | **0.79** | **0.95** |
| 10 | 1.47 | 1.67 | 1.55 | 2.51 | 1.44 | 5.65 | 1.51 | 1.04 | **0.93** | **0.52** | **0.58** | **0.54** | **0.46** | **0.50** | **0.35** | **0.52** | 1.61 | **0.64** | 1.99 | 1.79 | 1.49 | 3.18 | 1.95 | 1.01 | 1.51 | **0.75** | **0.93** |
| 11 | 1.32 | 1.51 | 1.40 | 2.36 | 1.31 | 6.55 | 1.36 | **0.86** | **0.86** | **0.51** | **0.58** | **0.54** | **0.53** | **0.50** | **0.31** | **0.52** | 1.74 | **0.65** | 2.07 | 1.73 | 1.47 | 3.11 | 2.05 | 1.06 | 1.50 | **0.67** | **0.90** |

### PANEL B: relative to the no-change forecast

| Horizon (Quarters) | 1995-2001 | | | | | | | | | 2002-2007 | | | | | | | | | 2008-2015 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination | VAR levels | VAR differences | Futures | Non-oil index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-model combination |
| 1 | 1.07 | 1.07 | 1.03 | **0.89** | 1.06 | 2.21 | **0.85** | 1.05 | 1.02 | 1.07 | **0.98** | 1.00 | **0.94** | 1.18 | 1.94 | 1.09 | **0.96** | 1.08 | 1.06 | 1.04 | 1.29 | **0.85** | **0.91** | 1.53 | **0.96** | 1.33 | 1.11 |
| 2 | 1.19 | 1.22 | 1.12 | 1.04 | 1.15 | 1.88 | **0.96** | 1.18 | 1.03 | **0.95** | **0.74** | 1.14 | **0.74** | 1.21 | 1.30 | **0.88** | **0.97** | **0.95** | 1.08 | 1.13 | 1.02 | 1.08 | **0.91** | **0.97** | **0.94** | 1.07 | **0.96** |
| 3 | **0.99** | 1.12 | 1.25 | 1.17 | **0.89** | 2.28 | **0.95** | 1.34 | 1.08 | 1.02 | **0.83** | 1.14 | **0.93** | 1.42 | **0.85** | **0.99** | **0.84** | **0.85** | 1.08 | 1.20 | **0.93** | 1.15 | **0.97** | **0.81** | **0.98** | **0.96** | **0.89** |
| 4 | **0.95** | 1.09 | 1.12 | 1.22 | **0.83** | 2.36 | **0.96** | 1.31 | **1.00** | **0.82** | **0.79** | 1.41 | **0.67** | 1.28 | **0.76** | **0.95** | **0.88** | **0.85** | 1.15 | 1.29 | **0.83** | 1.09 | 1.03 | **0.79** | **0.98** | **0.85** | **0.83** |
| 5 | **0.95** | 1.09 | **0.91** | 1.23 | **0.82** | 2.00 | **0.97** | 1.14 | **0.84** | **0.92** | **0.92** | 1.54 | **0.63** | 1.28 | **0.69** | **0.98** | **0.96** | **0.86** | 1.25 | 1.38 | **0.76** | 1.10 | 1.09 | **0.74** | **0.99** | **0.76** | **0.77** |
| 6 | **0.99** | 1.09 | **0.87** | 1.21 | **0.87** | 2.17 | **0.99** | 1.08 | **0.79** | **0.93** | **0.91** | 1.67 | **0.55** | 1.18 | **0.70** | **0.96** | 1.12 | **0.91** | 1.36 | 1.44 | **0.70** | 1.30 | 1.20 | **0.70** | 1.00 | **0.67** | **0.72** |
| 7 | **0.95** | 1.08 | **0.81** | 1.23 | **0.87** | 2.39 | **0.97** | 1.13 | **0.74** | **0.98** | **0.99** | 1.69 | **0.62** | 1.14 | **0.65** | **1.00** | 1.40 | **0.94** | 1.34 | 1.33 | **0.68** | 1.61 | 1.25 | **0.68** | **0.98** | **0.60** | **0.67** |
| 8 | **0.94** | 1.08 | **0.71** | 1.33 | **0.90** | 2.59 | **0.97** | **0.98** | **0.67** | 1.00 | 1.10 | 1.74 | **0.70** | 1.07 | **0.65** | **1.00** | 1.88 | 1.01 | 1.36 | 1.28 | **0.71** | 1.89 | 1.28 | **0.68** | **0.99** | **0.59** | **0.67** |
| 9 | **0.94** | 1.07 | **0.64** | 1.47 | **0.90** | 3.00 | **0.97** | **0.79** | **0.62** | **0.98** | 1.10 | 1.88 | **0.76** | **1.00** | **0.66** | **0.98** | 2.52 | 1.12 | 1.32 | 1.22 | **0.69** | 2.12 | 1.26 | **0.68** | **0.99** | **0.54** | **0.65** |
| 10 | **0.95** | 1.08 | **0.64** | 1.61 | **0.92** | 3.64 | **0.97** | **0.67** | **0.60** | **0.97** | 1.09 | 1.87 | **0.87** | **0.94** | **0.65** | **0.96** | 3.02 | 1.19 | 1.34 | 1.20 | **0.67** | 2.14 | 1.31 | **0.68** | 1.01 | **0.50** | **0.63** |
| 11 | **0.95** | 1.08 | **0.72** | 1.69 | **0.94** | 4.69 | **0.98** | **0.62** | **0.62** | **0.96** | 1.09 | 1.87 | **0.98** | **0.93** | **0.58** | **0.97** | 3.25 | 1.22 | 1.40 | 1.18 | **0.68** | 2.11 | 1.39 | **0.72** | 1.02 | **0.46** | **0.61** |

Notes: The tables show the mean squared prediction errors (MSPE) relative to the futures (Panel A) and no-change forecast (Panel B) over different sample periods for various forecasting methods including the forecast combination (in the grey column). A value lower than one – indicated in bold - means that the method outperforms the benchmark on average over the sample period indicated at the top of each table. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

**Table 4. Comparison of forecast performance over time of different model combinations relative to the futures**

**Panel A: total sample 1995 - 2015**

| **1995 - 2015** | 2 models | 3 Models | 4 models | 5 models |
|---|---|---|---|---|
| **Horizon (Quarters)** | ***+DSGE*** | ***+ adj. futures*** | ***+BVAR*** | ***+ VAR GDP*** |
| 1 | 1.03 | 1.00 | **0.89** | **0.81** |
| 2 | **0.96** | **0.97** | **0.93** | **0.89** |
| 3 | **0.91** | **0.93** | **0.93** | **0.93** |
| 4 | **0.91** | **0.91** | **0.93** | **0.94** |
| 5 | **0.88** | **0.86** | **0.91** | **0.94** |
| 6 | **0.85** | **0.81** | **0.88** | **0.93** |
| 7 | **0.81** | **0.77** | **0.83** | **0.89** |
| 8 | **0.78** | **0.75** | **0.81** | **0.86** |
| 9 | **0.76** | **0.75** | **0.80** | **0.85** |
| 10 | **0.74** | **0.75** | **0.80** | **0.84** |
| 11 | **0.71** | **0.74** | **0.77** | **0.81** |
| **average** | **0.85** | **0.84** | **0.86** | **0.88** |

**Panel B: 1995 – 2001 window**

| **1995 - 2001** | 2 models | 3 Models | 4 models | 5 models |
|---|---|---|---|---|
| **Horizon (Quarters)** | ***+DSGE*** | ***+ adj. futures*** | ***+BVAR*** | ***+ VAR GDP*** |
| 1 | 1.27 | 1.15 | **0.99** | **0.87** |
| 2 | 1.00 | **0.99** | **0.92** | **0.88** |
| 3 | **0.97** | **0.94** | **0.86** | **0.80** |
| 4 | 1.00 | **0.93** | **0.89** | **0.84** |
| 5 | **0.97** | **0.91** | **0.91** | **0.90** |
| 6 | **0.98** | **0.88** | **0.91** | **0.91** |
| 7 | 1.02 | **0.88** | **0.92** | **0.94** |
| 8 | 1.11 | **0.88** | **0.94** | **0.98** |
| 9 | 1.27 | **0.88** | **0.96** | 1.02 |
| 10 | 1.39 | **0.87** | **0.93** | **0.99** |
| 11 | 1.48 | **0.87** | **0.86** | **0.90** |
| **average** | 1.13 | **0.92** | **0.92** | **0.91** |

**Panel C: 2002 - 2007 window**

| **2002 - 2007** | 2 models | 3 Models | 4 models | 5 models |
|---|---|---|---|---|
| **Horizon (Quarters)** | ***+DSGE*** | ***+ adj. futures*** | ***+BVAR*** | ***+ VAR GDP*** |
| 1 | 1.31 | 1.15 | 1.08 | 1.02 |
| 2 | **0.96** | **0.90** | **0.83** | **0.80** |
| 3 | **0.75** | **0.73** | **0.75** | **0.79** |
| 4 | **0.60** | **0.60** | **0.61** | **0.61** |
| 5 | **0.53** | **0.55** | **0.56** | **0.57** |
| 6 | **0.51** | **0.54** | **0.55** | **0.56** |
| 7 | **0.48** | **0.55** | **0.56** | **0.57** |
| 8 | **0.46** | **0.59** | **0.58** | **0.58** |
| 9 | **0.44** | **0.63** | **0.60** | **0.58** |
| 10 | **0.43** | **0.69** | **0.64** | **0.60** |
| 11 | **0.41** | **0.70** | **0.65** | **0.62** |
| **average** | **0.63** | **0.69** | **0.67** | **0.66** |

**Panel D: 2008 - 2015 window**

| **2008 - 2015** | 2 models | 3 Models | 4 models | 5 models |
|---|---|---|---|---|
| | ***+DSGE*** | ***+ adj. futures*** | ***+BVAR*** | ***+ VAR GDP*** |
| 1 | **0.99** | **0.98** | **0.86** | **0.78** |
| 2 | **0.95** | **0.98** | **0.94** | **0.90** |
| 3 | **0.93** | **0.95** | **0.96** | **0.95** |
| 4 | **0.96** | **0.96** | **1.00** | 1.01 |
| 5 | **0.97** | **0.95** | 1.02 | 1.05 |
| 6 | **0.97** | **0.92** | 1.02 | 1.08 |
| 7 | **0.97** | **0.88** | **0.99** | 1.07 |
| 8 | **0.96** | **0.85** | **0.95** | 1.03 |
| 9 | **0.97** | **0.84** | **0.95** | 1.04 |
| 10 | **0.98** | **0.81** | **0.93** | 1.04 |
| 11 | **1.00** | **0.77** | **0.90** | 1.02 |
| **average** | **0.97** | **0.90** | **0.96** | **1.00** |

Notes: The tables show the MSPE relative to the futures of the different model combinations starting from the futures and each time (per column) adding the model which is indicated in the second row. The 5 model combination, shown in the last two columns, starts from the 4 model combination (including the futures, the DSGE, the risk-adjusted futures and the BVAR) and adds the VAR with GDP. A value lower than one - indicated in bold - means that the model combination outperforms the benchmark on average over the sample period considered. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

**Table 5. Change in forecasting performance of deleting individual models from the forecast combination**

| | PANEL A: full sample | | | | PANEL B: subsamples | | |
|---|---|---|---|---|---|---|---|
| Horizon (quarters) | **- adjusted futures** | **- BVAR** | **- DSGE** | **- futures** | **(1995 - 2001) - futures** | **(2002 - 2007) - futures** | **(2008 - 2015) - futures** |
| 1 | 0.98 | 1.13 | **0.97** | **0.99** | 1.04 | 1.04 | **0.98** |
| 2 | **0.97** | 1.05 | 1.02 | **0.98** | 1.06 | **0.96** | **0.98** |
| 3 | **0.99** | 1.00 | 1.05 | **0.99** | 1.06 | **0.94** | **0.99** |
| 4 | 1.01 | **0.97** | 1.06 | 1.00 | 1.08 | **0.87** | 1.00 |
| 5 | 1.03 | **0.95** | 1.08 | **0.99** | 1.10 | **0.84** | 1.01 |
| 6 | 1.05 | **0.93** | 1.10 | **0.99** | 1.12 | **0.82** | 1.01 |
| 7 | 1.05 | **0.92** | 1.13 | **0.98** | 1.14 | **0.83** | 1.01 |
| 8 | 1.04 | **0.93** | 1.16 | **0.97** | 1.17 | **0.84** | 1.00 |
| 9 | 1.02 | **0.94** | 1.18 | **0.97** | 1.23 | **0.85** | 1.00 |
| 10 | 1.00 | **0.95** | 1.21 | **0.97** | 1.28 | **0.87** | 1.00 |
| 11 | **0.98** | **0.96** | 1.24 | **0.96** | 1.37 | **0.88** | 1.00 |

Notes: The tables show the mean squared prediction errors (MSPE) for different model combinations relative to benchmark four-model combination over the total sample period 1995Q1 – 2015Q2. A '-' sign indicates that the individual projection is excluded from the four-model combination consisting of futures, the DSGE, the risk-adjusted futures and the BVAR. Panel A shows the results over the complete forecast evaluation period for all models included and panel B over different subsamples for deleting the futures. A value lower than one - indicated in **bold** - means that the method outperforms the combination on average over the sample period considered. See Section 2 and 4 for the model explanations.

**Table** 6. MSPE of forecast **combination with different combination weights, relative to the equal weighted four-model combination**

| **Horizon (Quarters)** | **OLS** | **Decaying** | **Inverse MSPE ratio** | **Trimming** | **Outperforming** | **OLS with constant** |
|---|---|---|---|---|---|---|
| 1 | 1.01 | 1.02 | 1.02 | **1.00** | **0.96** | 2.16 |
| 2 | 1.04 | 1.03 | 1.04 | 1.04 | 1.04 | **0.62** |
| 3 | 1.05 | 1.05 | 1.06 | 1.06 | 1.08 | 1.47 |
| 4 | 1.03 | 1.07 | 1.08 | 1.02 | 1.07 | 3.18 |
| 5 | 1.03 | 1.09 | 1.09 | 1.03 | 1.07 | 4.77 |
| 6 | 1.00 | 1.11 | 1.10 | 1.02 | 1.04 | 5.06 |
| 7 | **1.00** | 1.13 | 1.11 | 1.08 | 1.03 | 3.73 |
| 8 | 1.04 | 1.17 | 1.13 | 1.13 | 1.01 | 1.97 |
| 9 | 1.19 | 1.21 | 1.16 | 1.14 | **0.97** | 1.46 |
| 10 | 1.20 | 1.24 | 1.21 | 1.17 | **0.96** | 1.25 |
| 11 | 1.22 | 1.28 | 1.28 | 1.25 | 1.11 | 1.41 |

Notes: The table shows the mean squared prediction errors (MSPE) for different weighting schemes for the four-model combination relative to the equal-weighted benchmark combination over the total sample period 1995Q1 – 2015Q2. The different weighting schemes tried are (i) an ordinary least-squares estimate of the weights obtained by regressing the actual oil price realizations on the individual forecasts, with and without intercept term in the regression to account for possible biases in the individual projections (respectively column 'OLS with constant' and column 'OLS'), (ii) performance-based weights generated as the inverse of the MSPE ratio, calculated as the MSPE of the individual projection over the total MSPE in the combination (column 'Inverse MSPE ratio'), (iii) decaying weights which give more importance to recent forecasting performance (column 'Decaying'), (iv) a trimming approach which deletes the worst performing model from the equally-weighted forecast combination (column 'Trimming') and (v) an outperforming weighting scheme which sets the weights equal to the percentage of times that an individual forecast has the minimum MSPE (column 'Outperforming'). A value lower than one - indicated in **bold** - means that using the specific set of combination weights outperforms the equal-weighted benchmark combination over that specific horizon. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

## Table 7. MSPE of real quarterly WTI oil price forecasts

| Horizon (Quarters) | relative to futures | | | | | | | | | relative to futures | | | relative to no-change | relative to no-change | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1995-2015 | | | | | | | | | 1995-2001 | 2002-2007 | 2008-2015 | 1995-2015 | 1995-2001 | 2002-2007 | 2008-2015 |
| | **VAR levels** | **VAR differences** | **Futures** | **Non-oil index** | **VAR GDP** | **DSGE** | **BVAR** | **Adjusted-futures** | **4-model combination** | **4-model combination** | **4-model combination** | **4-model combination** | **4-model combination** | **4-model combination** | **4-model combination** | **4-model combination** |
| 1 | 1.03 | 1.01 | 1.04 | **0.87** | **0.91** | 1.71 | 1.59 | 1.78 | 1.19 | 1.18 | 1.21 | 1.19 | 1.15 | 1.36 | 1.25 | 1.13 |
| 2 | 1.04 | 1.09 | 1.04 | 1.10 | **0.94** | 1.05 | 1.14 | 1.24 | 1.00 | **0.90** | **0.91** | 1.02 | **0.96** | 1.21 | **0.96** | **0.95** |
| 3 | 1.10 | 1.22 | 1.10 | 1.27 | 1.07 | **0.93** | 1.21 | 1.20 | **0.99** | **0.84** | **0.81** | 1.03 | **0.91** | 1.11 | **0.96** | **0.89** |
| 4 | 1.22 | 1.35 | 1.16 | 1.31 | 1.17 | **0.96** | 1.29 | 1.18 | **1.00** | **0.87** | **0.69** | 1.09 | **0.86** | 1.01 | 1.01 | **0.83** |
| 5 | 1.37 | 1.49 | 1.21 | 1.43 | 1.27 | **0.94** | 1.36 | 1.18 | **0.98** | **0.89** | **0.62** | 1.13 | **0.81** | **0.87** | 1.02 | **0.77** |
| 6 | 1.44 | 1.52 | 1.19 | 1.57 | 1.37 | **0.90** | 1.38 | 1.16 | **0.94** | **0.89** | **0.59** | 1.15 | **0.78** | **0.79** | 1.08 | **0.72** |
| 7 | 1.34 | 1.36 | 1.12 | 1.72 | 1.31 | **0.86** | 1.31 | 1.20 | **0.89** | **0.90** | **0.60** | 1.12 | **0.79** | **0.75** | 1.10 | **0.71** |
| 8 | 1.26 | 1.23 | 1.01 | 1.77 | 1.21 | **0.80** | 1.21 | 1.29 | **0.85** | **0.91** | **0.63** | 1.05 | **0.84** | **0.68** | 1.19 | **0.74** |
| 9 | 1.19 | 1.14 | **0.94** | 1.77 | 1.12 | **0.78** | 1.16 | 1.41 | **0.83** | **0.89** | **0.64** | 1.04 | **0.88** | **0.60** | 1.31 | **0.75** |
| 10 | 1.14 | 1.07 | **0.90** | 1.66 | 1.08 | **0.78** | 1.10 | 1.58 | **0.83** | **0.84** | **0.68** | 1.03 | **0.92** | **0.57** | 1.38 | **0.74** |
| 11 | 1.06 | **0.99** | **0.84** | 1.57 | 1.01 | **0.76** | 1.02 | 1.64 | **0.81** | **0.75** | **0.69** | **0.99** | **0.96** | **0.57** | 1.39 | **0.74** |

Notes: The table shows the mean squared prediction errors (MSPE) relative to the futures forecast for various individual WTI forecasting methods and the forecast combination (1) over the total sample period 1995Q1 – 2015Q2 and (2) over different subsamples for the combination, and (3) relative to the no-change for the full sample . A value lower than one means that the method outperforms the benchmark on average over the sample period.The numbers in bold indicate an improvement relative to the respective benchmark. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 for the model explanations.

**Table 8. Average CRPS and average Logarithmic score of real quarterly Brent density forecasts produced by the individual models and the four-model combination**

**PANEL A: Average CRPS**

| Horizon (Quarters) | No-change | VAR levels | VAR differences | Futures | Non-oil Index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Time-varying weights without learning | Time-varying weights with learning | Constant weights |
| 1 | 1,71 | 1,90 | 1,71 | 1,93 | 1,80 | 1,85 | 2,18 | 1,71 | 1,96 | **1,35** | 1,56 | 1,94 |
| 2 | 3,11 | 3,54 | 3,09 | 3,28 | 3,49 | 3,44 | 3,44 | 3,10 | 3,33 | **2,42** | 2,80 | 3,33 |
| 3 | 4,28 | 4,70 | 4,17 | 4,37 | 4,89 | 4,45 | 4,32 | 4,15 | 4,35 | **3,16** | 3,67 | 4,33 |
| 4 | 4,93 | 5,51 | 4,65 | 5,04 | 5,38 | 5,07 | 4,95 | 4,68 | 4,90 | **3,42** | 3,43 | 4,81 |
| 5 | 5,48 | 6,10 | 5,31 | 5,61 | 5,97 | 5,77 | 5,34 | 5,18 | 5,48 | **3,50** | 3,51 | 5,23 |
| 6 | 5,87 | NaN | 5,68 | 5,84 | 6,83 | NaN | 5,51 | 5,51 | 5,68 | **3,28** | 3,28 | 5,33 |
| 7 | 6,09 | NaN | 5,77 | 6,01 | 8,00 | NaN | 5,82 | 5,59 | 6,08 | **3,07** | 4,16 | 5,50 |
| 8 | 6,26 | NaN | NaN | 6,45 | 8,94 | NaN | 7,46 | 5,91 | 6,77 | NaN | NaN | NaN |
| 9 | 6,51 | NaN | NaN | 6,78 | 10,05 | NaN | NaN | 6,13 | 7,36 | NaN | NaN | NaN |
| 10 | 6,67 | NaN | NaN | 6,97 | 10,71 | NaN | NaN | 6,33 | 7,99 | NaN | NaN | NaN |
| 11 | 6,67 | NaN | NaN | 7,06 | 10,86 | NaN | NaN | 6,37 | 8,36 | NaN | NaN | NaN |

**PANEL B: Average Logarithmic Score**

| Horizon (Quarters) | No-change | VAR levels | VAR differences | Futures | Non-oil Index | VAR GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | Time-varying weights without learning | Time-varying weights with learning | Constant weights |
| 1 | -3,85 | -4,47 | -2,78 | -6,02 | -4,24 | -4,66 | -4,93 | -2,47 | -5,40 | **-2,08** | -2,29 | -8,98 |
| 2 | -3,64 | -5,08 | -3,46 | -6,93 | -9,01 | -8,20 | -4,08 | **-3,22** | -6,48 | -3,44 | -3,84 | -10,84 |
| 3 | -4,11 | -9,02 | -4,78 | -10,43 | -11,21 | -9,27 | -4,23 | -3,41 | -10,59 | **-3,16** | -3,94 | -14,62 |
| 4 | -3,90 | -8,77 | -4,83 | -17,56 | -11,68 | -10,33 | -4,21 | -3,62 | -17,95 | **-3,07** | -3,08 | -15,96 |
| 5 | -3,89 | -8,57 | -5,02 | -22,66 | -13,03 | -10,30 | -4,02 | -4,02 | -22,84 | **-2,64** | -2,69 | -12,64 |
| 6 | -3,98 | NaN | -3,73 | -21,45 | -14,24 | NaN | -4,48 | -3,71 | -23,80 | -2,58 | **-2,48** | -11,90 |
| 7 | -4,00 | NaN | -4,71 | -22,59 | -17,26 | NaN | -4,47 | -3,74 | -30,60 | **-2,43** | -3,25 | -10,92 |
| 8 | -4,03 | NaN | NaN | -27,09 | -15,99 | NaN | NaN | -3,77 | -36,41 | NaN | NaN | NaN |
| 9 | -3,92 | NaN | NaN | -25,35 | -16,36 | NaN | NaN | -3,81 | -35,13 | NaN | NaN | NaN |
| 10 | -3,92 | NaN | NaN | -25,76 | -15,35 | NaN | NaN | -3,84 | -37,60 | NaN | NaN | NaN |
| 11 | -3,89 | NaN | NaN | -31,17 | -14,51 | NaN | NaN | -3,80 | -46,24 | NaN | NaN | NaN |

Notes: The average CRPS is computed as $\sum_{t=1}^{T}(\int_{-\infty}^{\infty}(F(z_t)-1\{y_t \leq z_t\})^2 dz)/T = \sum_{t=1}^{T}(E_p|Y_t - y_t| - 0.5E_f|Y_t - Y_t'|)/T$, where *F()* is the cdf of $f(y_y|I_{t-1})$ and $Y_t$ and $Y_t'$ are random draws from $f(y_y|I_{t-1})$. The average logarithmic score is computed as $\sum_{t=1}^{T}\ln(f(y_t|\, I_{t-1}))/T$ where $f(y_y|I_{t-1})$ is a predictive density for $y_t$constructed using information up to time $t-1$. The lower the CRPS and the higher the logarithmic score, the higher the accuracy of the predictive density. The values **in bold** indicate the model with highest accuracy. A value of NaN indicates that the predictive density could not be computed for that model at that horizon. The evaluation statistics are shown over the total sample period 1995Q1 – 2015Q2. For a description of how predictive densities are computed for each model we refer to section 6.

**Forecasting the Brent Oil Price: Addressing time-variation in forecast performance**

**Appendix**

Cristiana Belu Manescu[27] and Ine Van Robays[28]

## Annex 1 – Data sources

The forecasting exercise is conducted with data as available in real-time. The dataset builds on the real-time dataset constructed by Baumeister and Kilian (2012, 2014), extended to 2015:06 and complemented with the additional series as necessary. It consists of monthly vintages of data for the 1991:01-2015:06 period, each of which dates back to 1973:01, as available or necessary.

Specifically, we use the following data series from the proprietary Baumeister and Kilian (2012, 2014) dataset: spot Brent and WTI oil prices, an index of bulk dry cargo ocean shipping freight rates, world crude oil production, US crude oil inventories, US petroleum inventories, OECD petroleum inventories, the US consumer price index for all urban consumers. The Brent oil price is backcasted prior to mid-1987 on the basis of the US refiner's acquisition cost for crude oil imports, similar to Baumeister and Kilian (2014). Except for Brent and WTI prices, which are available in real time and not subject to revisions, all the other series are available with delays and are subject to revisions. Nowcasting techniques are used to extrapolate missing date due to release delays, as documented in Baumeister and Kilian (2012).

In addition, we also use the following series, for which, unless otherwise specified, monthly vintages of data are constructed for the period 1991:01-2015:06, each series dating back to 1973:01:

1. Monthly **Brent and WTI futures prices** with maturities 1 to 33 months-ahead from Bloomberg. Monthly prices are constructed as monthly averages of business day prices. For certain maturities for which Brent futures prices is available later than 1990:M1, Brent futures prices are backcasted based on WTI futures prices, as available. Moreover, owing to market illiquidity, futures prices for maturities longer than 13 months (Brent) and 18 months (WTI) are missing in certain months. These missing data is interpolated as the average of the previous and next available maturity, which is the approach used at the ECB to generate long-term oil projections. All in all, Brent futures prices data are available for the 1990:01-2015:06 period at all maturities, except those longer than 18-months ahead, for which data starts between November 1990 and April 1991.

2. Monthly **US capacity utilization in manufacturing industries** from the Federal Reserve Board, which is available with one month delay and subject to revisions. A real-time dataset is constructed with monthly vintages for the 1991:01-2015:06 period, each series going back to 1989:01. No nowcasting is required as the series enters with a 1-month lag in the regression with futures prices.

3. Monthly averages of the **industrial raw material price index** from the Commodity research Bureau, available without delays and not subject to revisions. The necessary data covers the period 1980:08-2015:06.

[27] European Commission, 170 Rue de la Loi, 1040, Bruxelles. E-mail: cristiana.manescu@ec.europa.eu

[28] European Central Bank, Kaiserstraße 29, D-60311 Frankfurt am Main. E-mail corresponding author: ine.van-robays@ecb.int.

4. Monthly **OECD measure of a global GDP** comprising the OECD member countries and the six largest non-OECD member countries (Brazil, China, India, Indonesia, Russian Federation and South-Africa). Since April 2012, this is the new reference series for the OECD Composite Leading Indicator (CLI) system, replacing the index of industrial production which has been discontinued since. Monthly estimates of GDP are generated based on the official quarterly estimates. The series is released with three months delay and is not subject to revisions. To nowcast missing data, the monthly GDP series is extrapolated using the average growth rate of monthly GDP up to that point in time. We have explored more complex ways of extrapolation, such as nowcasting GDP via its relationship with e.g. US and global PMI indicators, but the simple method based on past growth rates works equally well and offers the advantage of being easier to implement.

5. **Saudi Arabian oil production** is obtained from the internet version of the Monthly Energy Review of the Energy Information Administration (consistent with the other oil data from Baumeister and Kilian, 2012). Missing data due to release delays is nowcast based on the average rate of change in Saudi oil production up to that point in time.

## Annex 2 – Forecast error density plots and descriptive statistics

Of all possible forecasting statistics, the mean squared prediction error (MSPE) is only one measure of forecasting performance. To look at the forecast properties of the different individual methods from a different perspective, Figure A1 shows the forecast error density plots over different horizons.[29] The main descriptive statistics of these distributions (the bias, test for a zero bias, median, standard deviation, skewness and kurtosis) are also listed in Table A1.

The forecast error density plots reveal several interesting model properties. First, the mean of the forecast error distribution, i.e. the forecast bias, indicates that all models tend to underpredict actual oil prices whereas the DSGE tends to overpredict oil prices (see Figure A 1 and Table A1). However, the bias typically only becomes significant (i.e. has a value higher than 1.96) in the long run. To see this, we include the Z-test for a zero sample mean in the second row of Table A1.[30] Second, the median of the forecast error distribution is negative for most models and horizons, indicating that the probability of underestimating actual oil prices is higher than overestimating them. Again, the exception is the DSGE model whose median is positive at the longer term prediction horizons (see also Panel B of Figure A1). Third, concerning the probability to large forecast errors, as measured by the kurtosis of the density function, the futures and the risk-adjusted futures are sensitive to making infrequent but more extreme forecast errors at short forecast horizons, even though this characteristic weakens in the longer term (see Table A1). The VAR in levels and the VAR with GDP generate more frequent but smaller errors in general over the projection horizon, which is also true for the DSGE in the long term. Fourth, the forecast error distribution of most models becomes more normally distributed with the forecast horizon. At short prediction horizons, the forecast error distributions display a long tail to the right, or positive skewness, reflecting mainly the overestimation of oil prices during the plunge in oil demand in 2008-09 and the mid-2014 oil price. Finally, the differences in the forecast error distributions become more pronounced at longer horizons (see Panel B in Figure A1). In particular

[29] Due to similar forecasting properties of the different VAR models, we only show the VAR with GDP models to make the figure more readable. The density plots for the other VAR models can be obtained from the authors upon request.

[30] A value higher than 1.96 indicates that the models' bias is statistically different from zero at a significance level of 5%. This Z-test assumes normality of the underlying distribution which may be questionable for the forecast error distributions at short forecasting horizons. However, the results remain the same for the distributions of the forecast errors up to 2007Q4, which approximate the normal distribution better as the positive skew disappears due to the exclusion of the financial crisis.

the DSGE model and the futures display different forecasting properties compared to the other models on average.

Concerning the four-model combination, the negative bias inherent in the forecast combination projections is smaller than that of the futures and the no-change at every forecast horizon considered and only becomes significant in the medium term (see Table A1 and Figure A1).

## Annex 3 – Forecast bias and success ratio of the real quarterly Brent oil price forecasts: different subsamples

Table A2 shows the forecast bias and success ratio of the real quarterly Brent oil price forecasts over different subsamples. As is clear from these results, there is also considerable instability over time in the forecast bias and the success ratio for the individual forecasting methods. For example, although the DSGE model has the smallest bias at almost every forecast horizon over the complete sample (see Table 2 in the paper), it is clear that this no longer holds when looking at smaller subsamples. Panel A in Table A2 shows that in the second half of the 1990s, the DSGE model generated the largest absolute bias of all models which was positive, whereas in the most recent period, the bias was negative across all forecast horizons except for the 1-quarter ahead. Evidently, this averages out when looking at the total sample. This time-variation can also be found for the other forecast approaches and for the directional accuracy.

## Annex 4 – Alternative prior specification for the BVAR

The prior specification used in this paper was found to have the highest forecast accuracy when compared to a number of alternative prior specifications and data-transforms. Table A3 illustrates the MSPE ratio relative to the VAR in levels specification of four different prior specifications estimated with the Giannone et al. (2015) method, estimated during the period 1992Q1-2012Q2. Two prior specifications only impose a Minesota prior on the VAR coefficients (Columns 1 and 2) while the other two also include cointegration priors. The Minesota-only priors differ in the assumption made on the VAR coefficient for the log-level of the price of oil, either shrunk towards white noise in Column 1 (i.e. the Buameister and Kilian 2012 specification), or towards a random walk in Column 2. The prior on the other variables – all transformed to stationarity – is shrunk towards white noise in both cases. In contrast, the priors that allow for cointegration differ only in terms of data transforms: either all data in levels in Column 3, or all data but the oil price transformed to stationarity in Column 4 (i.e. the prior specification used in the paper).

Results indicate that including the two cointegration priors on the VAR in level data transforms, in addition to the Minesota prior, significantly improves the MSPE accuracy gains especially at the longer-horizons (e.g. 17% at the 11-quarters). For this reason, this particular prior specification was chosen in the paper. Interestingly, the data transforms seem to matter in this case, as the cointegration priors applied to all data in levels actually worsen performance relative to the VAR in levels. By contrast, imposing only a Minesota prior, irrespective of the assumption on the oil price, does not improve relative to the classical VAR specification.

**References**

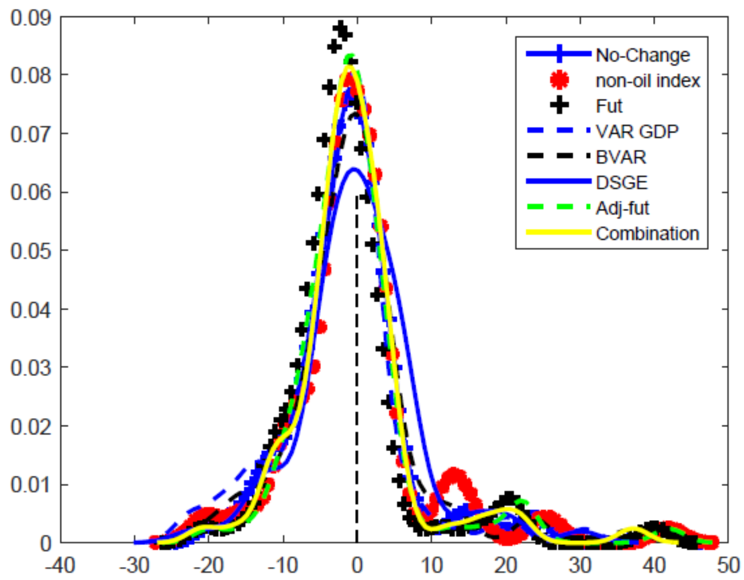
Baumeister, C. and L. Kilian (2012), “Real-Time Forecasts of the Real Price of Oil,” Journal of Business and Economic Statistics, 30, 326-336.

Baumeister, C. and L. Kilian (2014), “What Central Bankers Need to Know about Forecasting Oil Prices,” International Economic Review, 55, 869-889, August 2014.

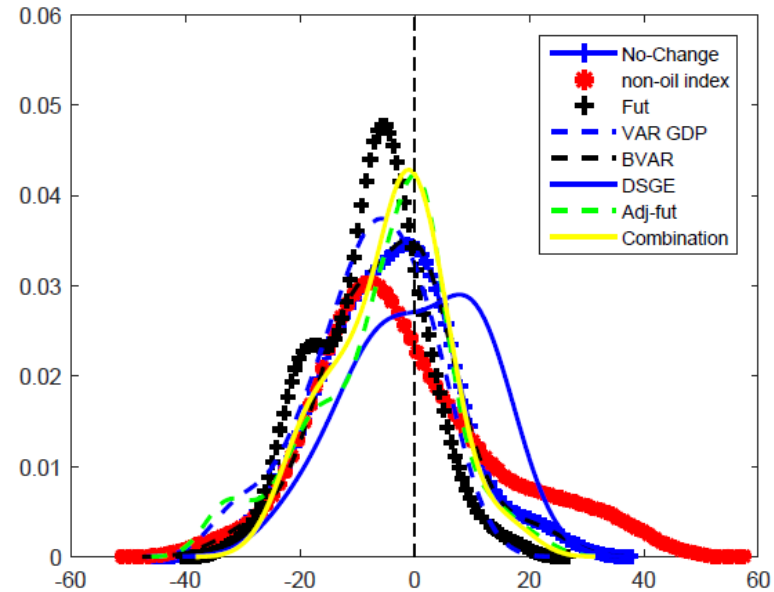
Giannone, D., M. Lenza, and G. Primiceri (2015), "Prior Selection for Vector Autoregressions," Review of Economics and Statistics, 97(2), 436-451.

**Figure A1. Forecast error density plots for the individual forecast methods over different horizons**

**PANEL A: 3 quarter-ahead forecast horizon**




**PANEL B: 11 quarter-ahead forecast horizon**




Notes: The figures show the forecast error densities of the different forecasting methods including the combination. The sample period for the individual projections is 1995Q1 – 2015Q2. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 of the paper for the model explanations.

## Table A1. Descriptive statistics of forecast error distributions

| Horizon 1 | Futures | VAR levels | VAR differences | VAR GDP | No-change | Non-oil index | DSGE | Adjusted-futures | BVAR | 4-model combination |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | -0.07 | -0.66 | -0.02 | -0.71 | -0.22 | -0.16 | 0.07 | 0.10 | -0.25 | -0.04 |
| Z-test mean | -0.15 | -1.67 | -0.05 | -1.91 | -0.56 | -0.45 | 0.14 | 0.22 | -0.64 | -0.09 |
| Median | -0.59 | -0.76 | -0.38 | -0.67 | -0.56 | -0.38 | -0.17 | -0.45 | -0.37 | -0.43 |
| Stand. Dev | 3.95 | 3.59 | 3.61 | 3.38 | 3.53 | 3.28 | 4.50 | 3.99 | 3.48 | 3.72 |
| Skewness | 2.37 | 0.77 | 1.58 | 0.17 | 2.12 | 1.12 | 1.88 | 2.36 | 0.99 | 1.97 |
| Kurtosis | 12.52 | 6.62 | 8.30 | 4.85 | 12.44 | 7.31 | 13.31 | 12.44 | 6.27 | 12.49 |

| Horizon 3 | Futures | VAR levels | VAR differences | VAR GDP | No-change | Non-oil index | DSGE | Adjusted-futures | BVAR | 4-model combination |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | -1.32 | -2.18 | 0.01 | -2.19 | -0.56 | -0.03 | 0.35 | -0.30 | -0.56 | -0.46 |
| Z-test mean | -1.42 | -2.26 | 0.01 | -2.34 | -0.58 | -0.03 | 0.39 | -0.32 | -0.59 | -0.50 |
| Median | -2.31 | -1.82 | -1.15 | -1.54 | -1.16 | -1.03 | -0.07 | -0.96 | -0.74 | -1.19 |
| Stand. Dev | 8.32 | 8.61 | 9.28 | 8.37 | 8.58 | 9.15 | 8.04 | 8.45 | 8.48 | 8.12 |
| Skewness | 2.02 | 1.23 | 1.96 | 0.57 | 1.74 | 1.49 | 0.59 | 2.11 | 1.25 | 1.57 |
| Kurtosis | 10.58 | 7.75 | 10.03 | 6.15 | 9.43 | 8.54 | 5.53 | 10.94 | 7.77 | 8.87 |

| Horizon 7 | Futures | VAR levels | VAR differences | VAR GDP | No-change | Non-oil index | DSGE | Adjusted-futures | BVAR | 4-model combination |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | -4.62 | -5.37 | -1.82 | -5.73 | -2.36 | -0.81 | 0.46 | -2.55 | -2.41 | -2.28 |
| Z-test mean | -4.72 | -4.55 | -1.38 | -5.07 | -2.03 | -0.58 | 0.43 | -2.52 | -2.09 | -2.32 |
| Median | -4.78 | -4.17 | -2.16 | -4.02 | -1.95 | -2.38 | 1.56 | -2.96 | -1.77 | -2.03 |
| Stand. Dev | 8.53 | 10.28 | 11.50 | 9.86 | 10.14 | 12.33 | 9.32 | 8.85 | 10.05 | 8.58 |
| Skewness | 0.46 | -0.14 | 0.76 | -0.55 | 0.28 | 0.54 | -0.40 | 0.78 | 0.22 | 0.26 |
| Kurtosis | 3.72 | 2.72 | 4.14 | 3.17 | 3.24 | 3.52 | 2.45 | 4.22 | 3.22 | 3.27 |

| Horizon 11 | Futures | VAR levels | VAR differences | VAR GDP | No-change | Non-oil index | DSGE | Adjusted-futures | BVAR | 4-model combination |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | -7.59 | -8.30 | -4.21 | -8.14 | -4.33 | -1.30 | 0.70 | -5.54 | -4.41 | -4.21 |
| Z-test mean | -7.23 | -6.98 | -3.04 | -6.84 | -3.40 | -0.71 | 0.54 | -4.00 | -3.46 | -3.80 |
| Median | -6.31 | -6.25 | -3.46 | -7.10 | -3.71 | -5.17 | 0.07 | -3.12 | -3.61 | -2.78 |
| Stand. Dev | 8.91 | 10.08 | 11.76 | 10.10 | 10.83 | 15.51 | 10.93 | 11.76 | 10.82 | 9.40 |
| Skewness | -0.08 | -0.69 | 0.15 | -0.54 | 0.07 | 0.70 | -0.30 | -0.52 | -0.01 | -0.14 |
| Kurtosis | 2.83 | 2.76 | 3.05 | 2.84 | 3.18 | 3.20 | 2.17 | 3.14 | 3.34 | 2.73 |

Notes: This table shows the descriptive statistics underlying the forecast error density plots (see also Figure A1) for the different individual forecast methods and the forecast combination (in the grey column) over the period 1995Q1 – 2015Q2. The 'Z-test mean' row contains the test statistics for the null-hypothesis of a zero mean (or alternatively, a zero bias); a value higher than 1.96 indicates that the models' bias is statistically different from zero at a significance level of 5%. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 of the paper for the model explanations.

## Table A2. Bias and success ratio of real quarterly Brent oil price forecasts: different subsamples

### PANEL A: Forecast bias

| Horizon (Quarters) | 1995-2001 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | -0.01 | 0.02 | -0.03 | -0.15 | -0.15 | -0.11 | 0.88 | -0.06 | 0.07 | 0.18 |
| 2 | -0.13 | 0.02 | -0.17 | -0.76 | -0.40 | -0.21 | 1.90 | -0.15 | -0.17 | 0.20 |
| 3 | -0.28 | -0.08 | -0.30 | -1.27 | -0.64 | -0.38 | 2.89 | -0.28 | -0.91 | 0.11 |
| 4 | -0.42 | -0.11 | -0.47 | -1.77 | -0.84 | -0.46 | 3.81 | -0.44 | -1.77 | -0.04 |
| 5 | -0.59 | -0.15 | -0.69 | -2.19 | -1.04 | -0.53 | 4.59 | -0.61 | -2.35 | -0.14 |
| 6 | -0.92 | -0.34 | -1.08 | -2.59 | -1.49 | -0.71 | 5.46 | -0.91 | -2.77 | -0.20 |
| 7 | -1.18 | -0.58 | -1.47 | -2.83 | -1.80 | -0.92 | 6.37 | -1.20 | -3.21 | -0.22 |
| 8 | -1.37 | -0.65 | -1.78 | -2.89 | -1.87 | -0.91 | 7.27 | -1.37 | -3.25 | -0.06 |
| 9 | -1.45 | -0.59 | -1.94 | -2.82 | -1.80 | -0.83 | 8.29 | -1.43 | -2.98 | 0.27 |
| 10 | -1.54 | -0.62 | -2.11 | -2.78 | -1.64 | -0.87 | 9.35 | -1.51 | -2.64 | 0.61 |
| 11 | -1.52 | -0.59 | -2.17 | -2.82 | -1.38 | -0.89 | 10.37 | -1.50 | -2.09 | 0.99 |

| Horizon (Quarters) | 2002-2007 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | -1.05 | -0.83 | -0.54 | -1.06 | -0.63 | -0.97 | -1.02 | -0.88 | -0.91 | -0.97 |
| 2 | -1.99 | -1.84 | -1.27 | -2.18 | -1.08 | -2.07 | -1.09 | -1.74 | -1.57 | -1.64 |
| 3 | -2.77 | -2.32 | -1.74 | -3.37 | -1.58 | -3.01 | -0.99 | -2.61 | -2.18 | -2.29 |
| 4 | -3.45 | -2.93 | -2.50 | -4.57 | -2.12 | -3.87 | -0.71 | -3.35 | -2.90 | -2.88 |
| 5 | -4.13 | -3.93 | -3.67 | -5.88 | -2.99 | -4.74 | -0.29 | -4.07 | -3.95 | -3.55 |
| 6 | -4.59 | -4.44 | -4.24 | -7.08 | -3.51 | -5.13 | -0.17 | -4.51 | -5.30 | -4.27 |
| 7 | -5.28 | -5.09 | -5.12 | -8.37 | -4.36 | -5.66 | 0.01 | -5.20 | -7.39 | -5.24 |
| 8 | -5.90 | -5.82 | -6.08 | -9.63 | -5.22 | -6.20 | 0.35 | -5.83 | -9.86 | -6.24 |
| 9 | -6.47 | -6.46 | -6.87 | -10.88 | -6.15 | -6.63 | 0.68 | -6.39 | -12.31 | -7.22 |
| 10 | -7.24 | -7.10 | -7.76 | -12.01 | -7.50 | -7.10 | 0.79 | -7.12 | -14.81 | -8.29 |
| 11 | -8.01 | -7.80 | -8.65 | -12.96 | -9.01 | -7.73 | 1.07 | -7.94 | -16.63 | -9.12 |

| Horizon (Quarters) | 2008-2015 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | 0.25 | -1.16 | 0.41 | 0.81 | 0.20 | -1.06 | 0.19 | 0.09 | 0.93 | 0.50 |
| 2 | 0.69 | -2.84 | 1.21 | 0.46 | 0.95 | -2.37 | -0.17 | 0.62 | 1.13 | 0.51 |
| 3 | 0.96 | -3.88 | 1.68 | 0.28 | 1.75 | -3.09 | -0.78 | 0.85 | 1.73 | 0.52 |
| 4 | 0.61 | -5.14 | 1.57 | -0.52 | 1.81 | -4.91 | -1.67 | 0.39 | 1.78 | 0.00 |
| 5 | 0.07 | -6.47 | 1.48 | -1.32 | 2.02 | -6.90 | -2.49 | -0.05 | 1.86 | -0.50 |
| 6 | -0.44 | -7.67 | 1.30 | -2.11 | 2.27 | -8.32 | -3.02 | -0.50 | 1.86 | -0.94 |
| 7 | -0.88 | -9.10 | 0.55 | -2.93 | 2.75 | -9.32 | -3.51 | -1.06 | 1.80 | -1.43 |
| 8 | -1.43 | -10.45 | -0.19 | -3.69 | 3.30 | -10.38 | -4.05 | -1.66 | 1.67 | -1.93 |
| 9 | -1.95 | -11.55 | -0.71 | -4.47 | 4.20 | -11.29 | -4.52 | -2.16 | 1.55 | -2.40 |
| 10 | -2.44 | -12.40 | -1.09 | -5.23 | 4.92 | -12.04 | -4.93 | -2.68 | 1.51 | -2.83 |
| 11 | -3.08 | -13.32 | -1.89 | -6.16 | 4.91 | -12.82 | -5.40 | -3.34 | 1.25 | -3.41 |

### PANEL B: Success ratio

| Horizon (Quarters) | 1995-2001 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | 0.57 | 0.75 | 0.68 | 0.57 | 0.68 | 0.64 | 0.57 | 0.68 | 0.68 | 0.71 |
| 2 | 0.63 | 0.63 | 0.59 | 0.52 | 0.63 | 0.67 | 0.48 | 0.74 | 0.56 | 0.44 |
| 3 | 0.69 | 0.50 | 0.62 | 0.58 | 0.50 | 0.62 | 0.46 | 0.73 | 0.46 | 0.46 |
| 4 | 0.68 | 0.64 | 0.64 | 0.64 | 0.48 | 0.72 | 0.48 | 0.68 | 0.56 | 0.68 |
| 5 | 0.54 | 0.63 | 0.46 | 0.67 | 0.54 | 0.63 | 0.54 | 0.50 | 0.58 | 0.67 |
| 6 | 0.57 | 0.70 | 0.52 | 0.65 | 0.52 | 0.65 | 0.57 | 0.52 | 0.57 | 0.65 |
| 7 | 0.50 | 0.59 | 0.41 | 0.64 | 0.55 | 0.59 | 0.59 | 0.45 | 0.50 | 0.68 |
| 8 | 0.52 | 0.57 | 0.48 | 0.62 | 0.43 | 0.48 | 0.62 | 0.48 | 0.57 | 0.67 |
| 9 | 0.40 | 0.60 | 0.40 | 0.65 | 0.30 | 0.50 | 0.60 | 0.45 | 0.70 | 0.70 |
| 10 | 0.37 | 0.58 | 0.32 | 0.68 | 0.16 | 0.47 | 0.58 | 0.42 | 0.74 | 0.63 |
| 11 | 0.28 | 0.50 | 0.39 | 0.67 | 0.11 | 0.44 | 0.61 | 0.33 | 0.72 | 0.61 |

| Horizon (Quarters) | 2002-2007 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | 0.63 | 0.71 | 0.67 | 0.54 | 0.67 | 0.75 | 0.46 | 0.63 | 0.63 | 0.67 |
| 2 | 0.63 | 0.71 | 0.71 | 0.50 | 0.71 | 0.71 | 0.63 | 0.67 | 0.71 | 0.79 |
| 3 | 0.58 | 0.67 | 0.67 | 0.50 | 0.71 | 0.71 | 0.75 | 0.58 | 0.67 | 0.71 |
| 4 | 0.63 | 0.75 | 0.58 | 0.46 | 0.71 | 0.67 | 0.83 | 0.54 | 0.54 | 0.71 |
| 5 | 0.58 | 0.63 | 0.54 | 0.33 | 0.67 | 0.58 | 0.75 | 0.54 | 0.50 | 0.63 |
| 6 | 0.50 | 0.58 | 0.54 | 0.33 | 0.71 | 0.54 | 0.83 | 0.46 | 0.42 | 0.75 |
| 7 | 0.58 | 0.63 | 0.54 | 0.25 | 0.54 | 0.54 | 0.88 | 0.50 | 0.33 | 0.75 |
| 8 | 0.50 | 0.58 | 0.46 | 0.33 | 0.67 | 0.54 | 0.79 | 0.42 | 0.38 | 0.63 |
| 9 | 0.54 | 0.58 | 0.46 | 0.25 | 0.58 | 0.63 | 0.79 | 0.46 | 0.33 | 0.46 |
| 10 | 0.50 | 0.58 | 0.50 | 0.29 | 0.58 | 0.67 | 0.75 | 0.46 | 0.33 | 0.46 |
| 11 | 0.63 | 0.75 | 0.50 | 0.17 | 0.42 | 0.71 | 0.88 | 0.58 | 0.17 | 0.29 |

| Horizon (Quarters) | 2008-2015 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No-Change | VAR levels | VAR differences | Futures | Index | GDP | DSGE | BVAR | Adjusted-futures | 4-Model Combination |
| 1 | 0.60 | 0.63 | 0.73 | 0.67 | 0.63 | 0.60 | 0.57 | 0.67 | 0.67 | 0.70 |
| 2 | 0.63 | 0.50 | 0.57 | 0.70 | 0.53 | 0.47 | 0.63 | 0.50 | 0.80 | 0.63 |
| 3 | 0.60 | 0.47 | 0.53 | 0.77 | 0.77 | 0.57 | 0.70 | 0.53 | 0.77 | 0.73 |
| 4 | 0.60 | 0.53 | 0.57 | 0.90 | 0.73 | 0.53 | 0.73 | 0.50 | 0.93 | 0.80 |
| 5 | 0.53 | 0.50 | 0.57 | 0.87 | 0.57 | 0.57 | 0.77 | 0.53 | 0.87 | 0.93 |
| 6 | 0.50 | 0.53 | 0.43 | 0.90 | 0.50 | 0.57 | 0.83 | 0.50 | 0.83 | 0.90 |
| 7 | 0.47 | 0.57 | 0.43 | 0.83 | 0.50 | 0.57 | 0.80 | 0.53 | 0.77 | 0.83 |
| 8 | 0.47 | 0.53 | 0.43 | 0.80 | 0.40 | 0.57 | 0.80 | 0.53 | 0.73 | 0.83 |
| 9 | 0.47 | 0.63 | 0.47 | 0.83 | 0.27 | 0.63 | 0.80 | 0.50 | 0.77 | 0.90 |
| 10 | 0.57 | 0.53 | 0.47 | 0.83 | 0.30 | 0.53 | 0.77 | 0.57 | 0.77 | 0.90 |
| 11 | 0.60 | 0.47 | 0.50 | 0.83 | 0.40 | 0.47 | 0.73 | 0.60 | 0.70 | 0.87 |

Notes: The tables show the forecast bias (Panel A) and the success ratio (Panel B) over different sample periods for various forecasting methods including the forecast combination (in the grey column). A negative number for the bias indicates and underprediction of actual oil prices on average and vice versa. The success ratio gives the probability of predicting the correct change in the oil price. A value higher than 0.5 - shown in **bold** - denotes a higher directional accuracy than at random. The forecast combination consists of the futures, the risk-adjusted futures, the BVAR and the DSGE model. See Section 2 and 4 of the paper for the model explanations.

**Table A3. Recursive MSPE <u>relative to the VAR in levels specification</u>, over the period 1992Q1 – 2012Q4**

| Horizon (Quarters) | No cointegration priors | | With cointegration priors | |
|---|---|---|---|---|
| | All VAR coeffs are shrunk towards white noise | The price of oil VAR coefficient is shrunk towards random walk, all others towards white noise | All VAR coeffs are shrunk towards random walk | The price of oil VAR coefficient is shrunk towards random walk, all others towards white noise |
| | **Oil prices in levels, other variables stationary** | | **All data in levels** | **Oil prices in levels, other variables stationary** |
| 1 | **0,99** | 0,99 | 1,00 | 1,02 |
| 2 | **1,00** | 1,00 | 1,03 | **0,99** |
| 3 | 1,01 | 1,01 | 1,06 | **0,96** |
| 4 | 1,02 | 1,02 | 1,07 | **0,91** |
| 5 | 1,02 | 1,01 | 1,07 | **0,87** |
| 6 | 1,01 | 1,00 | 1,07 | **0,84** |
| 7 | 1,02 | 1,01 | 1,11 | **0,84** |
| 8 | 1,02 | 1,02 | 1,13 | **0,85** |
| 9 | 1,03 | 1,03 | 1,15 | **0,87** |
| 10 | 1,03 | 1,02 | 1,15 | **0,86** |
| 11 | 1,02 | 1,01 | 1,16 | **0,83** |

Notes: A value lower than one means that, on average, the method outperforms the original VAR over the sample period.The numbers in **bold** indicate an improvement relative to the VAR in levels. The highlighted column indicates the prior selected in the paper.