Global Research Unit Working Paper #2019-024

Forecasting crude oil prices with DSGE models

Michał Rubaszek, SGH Warsaw School of Economics

Forecasting crude oil prices with DSGE models*

Michał Rubaszek†

SGH Warsaw School of Economics, Collegium of Economic Analysis, Poland

Abstract

We run an oil prices forecasting competition among a set of structural models, including vector autoregressions and dynamic stochastic general equilibrium models. Our results highlights two principles. First, forecasts should exploit the mean reversion of the real oil price over long horizons. Second, models should not replicate the high volatility of oil prices observed in sample. Abiding by these principles, we show that a small scale DSGE model performs much better in real oil price forecasting than the random walk as well as vector autoregressions.

Keywords: Forecasting, oil prices, DSGE models, vector autoregression, Bayesian inference.

JEL classification: C32, Q35, Q43, Q47.

1 Introduction

Large swings in oil prices draw a lot of attention among policymakers, academics and practitioners. The key questions are whether large price changes can be explained by demand or rather supply shocks and whether these changes are predictable. This is why the last decade has witnessed a series of studies aimed at constructing a structural model that would be able to link the dynamics of oil prices and global macroeconomic variables and which could be used for forecasting.

The most recent studies aimed at explaining the dynamics of commodity prices using structural macroeconomic models can be divided into two broad categories, differentiated on the basis of the adopted framework. In the first group, which is focused on the data, researchers use structural vector autoregressions (VAR). In the second one, which is more grounded on the economic theory, authors simulate dynamic stochastic general equilibrium (DSGE) models incorporating the oil sector. It should be mentioned that on top of structural VAR and DSGE based studies, the literature from the 1990s and early 2000s also delivers analyses of the oil market based on traditional macroeconometric models (see Kaufmann, 1995; Dees et al., 2007, and references therein).

The most popular specification of the structural VAR model for the oil market was proposed by Kilian (2009). It explains the dynamics of three endogenous variables (oil production, global

*The project is realized with the financial support from the National Science Centre (grant No. 2017/25/B/HS4/00156)

†michal.rubaszek@sgh.waw.pl

activity and the real price of oil) by three structural shocks (oil supply, global demand and oil demand). A big contribution of Kilian is that he proposes an index of global economic activity based on shipping costs and shows that it is a good indicator of global demand for oil.[1] Next, he presents evidence that real oil prices are predominantly driven by demand shocks, whereas the impact of supply shocks on the dynamics of oil prices is of much lower importance. At that time this claim was rather controversial as, following the study by Hamilton (2003), the consensus was that oil price fluctuations are driven rather by supply than demand factors. The subsequent studies using structural VAR models evolved in few directions. First, the trivariate specification was extended by decomposing global economic activity into developing and developed countries (Lippi and Nobili, 2012; Caldara et al., 2018), taking into account financial speculation (Fattouh et al., 2013) or incorporating the level of oil inventories (Kilian and Murphy, 2014). Second, the literature discussed how to measure global economic activity and pointed to global GDP (Jacks and Stuermer, 2018), global industrial production or the index of industrial metals (Caldara et al., 2018) as well as a common factor extracted from commodity markets (Delle Chiaie et al., 2017; Alquist et al., 2019). Third, some authors investigated a method of shocks identification, which in the benchmark model of Kilian (2009) assumes that oil supply is price inelastic in the short term. Even though the main finding of Kilian (the marginal role of supply shocks) was not overturned by studies applying sign restrictions (Kilian and Murphy, 2012; Baumeister and Peersman, 2013), a more elaborate inference methods applied by Baumeister and Hamilton (2019) or Caldara et al. (2018) indicate that supply shocks constitute an important source of oil price fluctuations.

As regards studies using DSGE models incorporating the oil sector, they can be grouped into three main categories. In the first group, oil prices and production are assumed to be exogenous (e.g. Leduc and Sill, 2004). In the second one, oil prices are endogenous but production remains exogenous (e.g. Bodenstein et al., 2011). In the third group, both oil prices and production are determined within the model, which allows for a two-way feedback between the two variables. In this category there are two distinctive frameworks that were fitted to the data: dominant oil producer and two-stage production setup. The dominant oil producer framework, proposed by Nakov and Pescatori (2010) and modified by Nakov and Nuno (2013), assumes that there are three global regions: oil importing country (US), dominant oil exporter (OPEC) and other oil exporters (non-OPEC). OPEC behaves as a monopolist and chooses the level of oil production (and implicitly prices) that maximizes the present value of profits, subject to the level of oil production by non-OPEC countries, the expected demand for oil in the global economy and the marginal cost of producing oil. In the two-stage production framework, which was proposed by Balke et al. (2010) and Balke and Brown (2018), it is assumed that in stage one competitive drilling firms invest to discover new exploitable oil fields, whereas in stage two oil producing firms rent discovered fields and combine them with labor to extract the oil from the ground. It can be added that the original setup by Balke et al. (2010) was extended by Peersman and Stevens (2013), among others by allowing for the variable utilization rate of discovered oil fields, which increased the responsiveness of oil production to current market conditions, and by adding oil to the utility function to account for the existence of oil inventories.

As regards studies focusing on forecasting oil prices with VAR and DSGE models, their number is low. The most prominent example of this kind of analysis was presented by Baumeister and Kilian (2012, 2015), who assessed the quality of forecasts for real oil prices over years 1991-2010 derived with a set of models, including the VAR proposed by Kilian and Murphy (2014). They

[1] The construction of the index was recently criticized by Hamilton (2019), where the critique was responded to by Kilian (2019).

showed that this model performed well compared with the no-change forecast, especially at the shortest horizons. A less favorable results on the usefulness of VAR models in oil price forecasting was presented by Manescu and Van Robays (2016), who analyzed the out-of-sample performance of various methods over the 1995-2012 horizon. They found that the VAR model by Kilian and Murphy (2014) is not powerful enough to consistently deliver better forecasts than those from the random walk model. The analysis also reveals that the outstanding performance of VAR presented in Baumeister and Kilian (2012) was driven predominantly by the upward trend in oil prices observed in the 2002-2007 period. Its performance was much weaker if forecasts were evaluated over 1995-2001 or 2008-2012 sub-samples. According to our best knowledge, the study by Manescu and Van Robays (2016) is also the only one that analyzes the quality of oil price forecasts produced by a DSGE model. In particular, the authors applied a calibrated model of Nakov and Nuno (2013) to find that its out-of-sample performance is relatively good. It should be added, however, that the DSGE model was given a head start as its calibration was based on the knowledge about the dynamics of observed time series over the entire data span, including the forecast evaluation sample. We are not aware of any study that would use recursively estimated DSGE model to forecast oil prices.

The key aim of this paper is to provide a thorough evaluation of how well VAR and DSGE models perform in real oil price forecasting. In this sense we contribute to both the literature surveyed above as well as the studies that explore the forecasting properties of VAR and DSGE models (see Adolfson et al., 2007; Rubaszek and Skrzypczynski, 2008; Del Negro and Schorfheide, 2012; Ca' Zorzi et al., 2017; Kolasa and Rubaszek, 2018, and references therein). Towards this end, we generate recursive forecasts from the structural VAR model proposed by Kilian (2009) as well as the DSGE model developed by Nakov and Pescatori (2010). These two models were chosen for three reasons. First, they are well recognized in the literature. Second, both articles were available to the public over a decade ago, hence we can design a forecasting "horse race" in which applied methods were available to economists at the start of the competition. Third, both models were checked for their reliability by passing the replicability test of Chang and Li (2015).

Apart from the two structural models, to understand better the dynamics of the oil market, we also bring into the forecasting race other competing models. The first one is the "twin" DSGE model, which is identical from a theoretical perspective, but allows for a linear trend in the real price of oil in the measurement equation to improve the in-sample fit. We include this specification in our forecasting race to check if adding a trend is helpful or rather counterproductive in forecasting oil prices, as it was showed for real exchange rates by Ca' Zorzi et al. (2017). The next competitor is the VAR model by Kilian estimated with Bayesian methods. We also add three Bayesian VAR models estimated with data present in the DSGE model. Two of them are standard, with the difference that one is estimated on variables in levels, whereas the second one after their differentiation. The third one exploits the methodology of Villani (2009) to elicit the prior that the real price of oil reverts to its recursive mean, as would suggest the discussion by Dvir and Rogoff (2009) and Kruse and Wegener (2019). The next two models are random walks, with and without drift, which are the most popular benchmarks in real oil price forecasting competitions. The last two models are simple autoregressive processes, which were estimated or calibrated. In the latter case we assume that the half-life of real price of oil is five years, as it was done in the study of Ca' Zorzi et al. (2017) for real exchange rates.

Our key contribution is that we present evidence that, in order to deliver real oil price forecasts of high quality, models must abide by two principles. As a first priority, they must produce "conservative" forecasts, in the sense that they should not attempt to explain a large fraction of oil price volatility out of sample. The second principle to which models should conform is that they ought to exploit mean reverting tendency of the real price of oil. All our core results become entirely intuitive if we keep these principles in the back of our mind. First, the DSGE model perform better than the random walk for all horizons. However, adding the trend to the DSGE model decreases slightly the accuracy of forecasts, as it unnecessarily extrapolates trends observed in the past. The second result is that all VAR models performed relatively poor for short-term horizons, also the framework proposed by Kilian (2009). Even though for longer horizons their performance improved somewhat, it remained inferior to that of the DSGE model. Finally, we find that a calibrated autoregressive mean-reverting model performs relatively well.

The remainder of the paper is structured as follows. Section 2 discusses the models participating in the forecasting race, whereas Section 3 describes the data. Section 4 presents and explains the main results, emphasizing the main takeaways of our analysis. Section 5 concludes.

2 Round-up of forecasting methodologies

We consider the following competitors in the forecasting horse race.

DSGE models

Our key theoretical reference is the DSGE model developed by Nakov and Pescatori (2010). In this model the oil industry is represented by a dominant producer (OPEC) a fringe of competitive oil suppliers (non-OPEC), who are small individually but collectively can restrain the market power of the cartel, and oil importer (US). OPEC maximizes welfare from oil profits, internalizing the effect of its pricing decision on global output and oil demand. Oil is subsequently used as an input in the production of final goods by US producers, which operate in a monopolistically competitive environment and face nominal rigidities á la Calvo. Households maximize their lifetime utility, which depends on consumption and habits. Monetary policy is conducted according to the Taylor-type rule with smoothing.

The model is fitted to the data with exactly the same method as it was done in the original paper, which means that all the model's assumptions and derivations, as well as prior distributions can be found in Nakov and Pescatori (2010). We proceed as follows. First, we calibrate the steady-state and fit the model to four macroeconomic times series: the log change in real GDP ($\Delta\widetilde{y}$), inflation of GDP deflator ($\Delta\widetilde{p}$), the nominal Fed Funds rate $(\widetilde{i})$ and the growth rate of the real WTI price ($\Delta\widetilde{q}$). For that purpose we use Bayesian methods and make the same prior assumptions for the estimated parameters as in the source article. The only difference is that in the baseline DSGE model we fix the constant in the measurement equation to zero, whereas in the version of the model with trend (DSGEt) the constant is estimated. It can be added that since our estimation assumptions are based on the results and model specification that were available at the beginning of forecast evaluation sample, it is unlikely that the DSGE model received an unfair advantage in the forecasting race due to a choice of priors that aims to improve the model fit to the forecast evaluation sample, a concern raised by Gurkaynak et al. (2013).

BVAR counterparts of the DSGE model

It is well known that DSGE models have a restricted infinite-order VAR representation (Fernandez-Villaverde et al., 2007), which explains why VARs have been widely used in the forecasting literature evaluating DSGE models. However, because of the large number of parameters and short time series, classical estimates of unrestricted VAR coefficients are often imprecise and forecasts are of low quality due to large estimation error. A common method to tackle this problem is to apply Bayesian VAR techniques. We follow this route by considering three BVAR models that are estimated using the same times series as in the case of the DSGE model. These three specifications differ in the choice of whether the real oil price and other regressors are differenced prior to estimation, and on whether we impose the prior that the real price of oil is mean reverting. In particular, we consider a BVAR in "levels" (LBVAR, for $\widetilde{y}$, $\widetilde{p}$, $\widetilde{i}$ and $\widetilde{q}$), one expressed as a mixed model of variables expressed in "levels" and "differences" (DBVAR, for $\Delta\widetilde{y}$, $\Delta\widetilde{p}$, $\widetilde{i}$ and $\Delta\widetilde{q}$) and one where we exploit the methodology of Villani (MBVAR, for $\Delta\widetilde{y}$, $\Delta\widetilde{p}$, $\widetilde{i}$ and $\widetilde{q}$) to elicit the prior that the real price of oil is mean reverting, as would suggest the work by Dvir and Rogoff (2009); Kruse and Wegener (2019). In all cases we use the specification with four lags as the models are fitted to the data of quarterly frequency.

As regards the details of the estimation process, we set the standard Normal-Wishart prior proposed by Kadiyala and Karlsson (1997) for LBVAR and DBVAR models, and assume a normal-diffuse prior for the MBVAR as in Villani (2009). For LBVAR we center the prior at random walk parameters, whereas for the mixed models (MBVAR and DBVAR), we follow Adolfson et al. (2007) and Villani (2009), centering the prior for the first own lag at zero for the differenced variables and at 0.9 for the variables in levels. All other VAR coefficients are centered at zero. As regards the dispersion of the prior distributions, we assume that they are tighter for higher lags (*decay* hyperparameter is set to 1) and choose the conventional value of 0.2 for the *overall tightness* hyperparameter. In the case of the MBVAR model, we additionally set the prior variance for cross-variable coefficients to lower values than for their own lags (*weight* hyperparameter equal to 0.5). The steady-state prior for the real oil prices is centered at its recursive mean, with tightness such that the 95% interval coincides with the $\pm 5\%$ range around this mean. As regards the remaining economic variables, we take standard values suggested by the literature. The 95% interval is defined as 0.5% $\pm 0.25\%$ for steady-state (quarterly) inflation and output growth and 1.0% $\pm 0.25\%$ for the (quarterly) interest rate.

Structural VAR

The next competitors in the forecasting horse race are based on the structural VAR model proposed by Kilian (2009), which are estimated using three variables: log-changes in oil production ($\Delta\widetilde{prod}$), global economic activity index ($\widetilde{gea}$) and the real price of oil ($\widetilde{q}$). In the first version the model is estimated with least-squares estimator (Kilian VAR, KVAR), whereas in the second using Bayesian methods with settings described above (Kilian BVAR, KBVAR). As in the previous cases, we use the specification with four lags.

Atheoretical benchmarks

We also let three atheoretical models into the race. The first one is the most widely used benchmark in the forecasting literature, i.e. the naïve random walk model, both with and without trend. From the perspective of the forecasting practitioner, there is nothing more conservative than assuming that no changes occur over the forecast horizon. The second one is an autoregression for the real price of oil ($\widetilde{q}$) estimated with Bayesian methods (BAR). As for BVAR models, we use the specification with four lags and the prior proposed by Kadiyala and Karlsson (1997).

The third atheoretical model is a method that assumes that the real price of oil gradually returns to its historical average (AR-fixed). This method was recently shown to be very competitive relative to several other methods for inflation (Faust and Wright, 2013) as well as real exchange rate forecasting (Ca' Zorzi et al., 2017). We set the autoregressive parameter of the "AR fixed" to be consistent with the half-life adjustment of three years.

3 Data

We use quarterly data over the period 1984:1 to 2018:3 for six time series that are present in the Nakov and Pescatori (2010) and Kilian (2009) models. The models as well as the estimation and forecast evaluation samples were chosen in a way that ensures that we don't use information from the evaluation sample. In particular, the working paper versions of both models, including the values of the calibrated parameters in the DSGE model, were available in 2008. In both papers the estimation sample ends in 2007, whereas the steady state of the DSGE model was calibrated using a sample starting in the first quarter of 1984. The only advantage is given to the KVAR and KBVAR models, given that we use the revised time series for the global economic activity index, which incorporates part of the critique of Hamilton (2019) and is presented in Kilian (2019). In particular, this gives some time advantage to the KVAR and KBVAR models in the forecasting competition as the revision became available to economists in 2018 and not at the start of the forecasting competition. Bearing in mind the last comment, we can claim that our choices are not contaminated by knowledge from the forecast evaluation sample, such as the development of new theories, availability of new data or estimation methods. In other words, we use only information that was available to an economist living at the beginning of the forecast evaluation sample and has gradually became available with the release of new data.

As regards the exact values of the time series, they were retrieved from Federal Reserve Bank of St. Louis FRED database, International Energy Agency through Quandl and Lutz Kilian webpage. All data sources are provided in Table 1.

4 Results

We assess the out-of-sample forecast performance of all models for horizons ranging from one quarter to five years. The models were estimated using recursive samples, where the first set of forecasts was elaborated with models estimated over the sample 1984:1-2007:4 for the period 2008:1-2012:4. This procedure was repeated with samples ending in each quarter from the period 2008:1-2018:2. As a result the one-quarter-ahead forecasts are evaluated on the basis of 43 observations, two-quarter-ahead forecasts on the basis of 42 observations, and so forth with the 20-quarter-ahead forecasts comprising 24 observations.

For the DSGE model the posterior distribution of parameters is approximated with 200,000 draws obtained with the Metropolis-Hastings (MH) algorithm, after discarding the initial 50,000 draws. This number of draws was sufficient to achieve convergence according to standard diagnostics. Next, for every twentieth realization of the MH chain, we take five sequences of random draws of structural shocks over the forecast horizon. Consequently, at each forecast date we have in total 50,000 draws from the predictive density that can be used to calculate both point (mean across the draws) and density forecasts. For the remaining models we also base our forecasts using 50,000 draws from the predictive density.

4.1 Point forecast

We begin the contest by investigating the mean forecast error (MFE) for the real oil prices (Table 2). The values of MFE for the random walk benchmark (without drift) indicate that in the evaluation sample there was a decline in the real prices of oil. In fact, the price of WTI barrel in 2007:4 stood at 90.8 USD, whereas its price in 2018:3 amounted to 69.7 USD. Taking into account that over this period the level of implicit GDP deflator increased by 18.8%, the real price of oil declined by 35.3%. The values of MFE for equally weighted pooled forecast indicate that on average the investigated methods managed well to predict this decline, especially for shorter horizons. Looking at the performance of individual methods, a general finding is that they deliver unbiased forecasts, apart from the RW with drift and DVAR models, which sizeably and significantly overpredict future movements in oil prices over a longer horizons. Finally, the table also shows that methods assuming mean reversion of real oil prices, in particular AR-fixed and DSGE without trend, delivered lower forecasts than methods extrapolating past trends (RW with drift, DVAR and DSGE with trend).

A good way forward to get a visual impression of the performance of the competing methods is to plot their forecasts at different points in time (Figure 1).[2] A first glance at the figure reveals the upward bias of forecasts from the RWdrift and DVAR models, i.e. two models that are estimated using differentiated variables. These models dully extrapolate past trends in the real price of oil observed in the past, which does not pay out in terms of forecast accuracy. Next, the figure illustrates the difference between the two versions of the DSGE model, which differ only on whether we allow for a constant in the measurement equation for the real price of oil. Even though in both cases the analyzed variable is reverting to a well defined steady state, in the case of DSGEt it is given by a deterministic trend, whereas for DSGE it is a constant. It can be seen that the forecasts from DSGEt model are shifted upward in comparison to the baseline DSGE. The figure also shows that forecasts from the KVAR model are relatively unstable and are strongly influenced by just few observations. This is especially visible at the beginning of the forecast evaluation sample, in which the model delivers very high forecast for two initial observations to dramatically change predictions after the Great Financial Crisis. Finally, the figure reveals that the DSGE (without trend) and AR fixed models are characterized by very similar forecasts, which are mean-reverting and conservative, in a sense that they do not attempt to explain a large fraction of the data variation or to anticipate turning points.

We continue by comparing the second moments of the forecast errors with the most popular measure of ex-post forecast quality: root mean squared forecast errors (RMSFE). Given that the benchmark model in our competition is the naïve random walk, in Table 3 we report the levels of the RMSFEs for this benchmark, whereas the remaining numbers are expressed as the ratios so that the values below unity indicate that a given method outperforms the naïve forecast. Moreover, we test whether a given method dominates the benchmark using the Clark and West (2006) test. A number of key features of the results are immediately evident. First, the baseline DSGE model performs best among all investigated methods for all forecast horizons. Moreover, it delivers forecasts that are significantly better than those from the random walk benchmark. For one-quarter ahead horizon the gain in the forecast accuracy amounts to 2% and it gradually increases to 28% for five-year ahead horizon. Second, baseline DSGE is only slightly better than the AR-fixed model, which would suggest that the outstanding performance of the DSGE model

[2] It should be noted that the figure presents the level for the real price of oil, whereas Table 2 is based on error for logarithms, hence the values are not directly comparable.

is predominantly driven by its two already-mentioned features: it delivers mean-reverting and conservative forecasts. Third, the comparison of both DSGE models shows that allowing for a deterministic trend in the steady state for real oil prices is deteriorating the quality of forecasts. However, forecasts from the DSGE model with trend are still characterized by higher accuracy than those from the random walk. Fourth, the results for VAR models are rather discouraging, which is in line with the findings by Manescu and Van Robays (2016) and less optimistic than the results of Baumeister and Kilian (2012). In most cases the RMSFE ratios are above unity, especially for short term horizons. In this case MBVAR is the best performing model, which would emphasize the role of mean-reversion is real oil prices forecasting. On he contrary, DVAR is the worst performing method, which would suggest that differencing variables before estimation is not successful strategy in forecasting the real price of oil.

We proceed by comparing the realized $(\widetilde{q}_{t+h} - \widetilde{q}_t)$ and forecast $(\widetilde{q}^f_{t,h} - \widetilde{q}_t)$ changes in the real price of oil. In Table 4 we report the correlation between both variables for different horizons h. It can be noted that the values in the table can be interpreted as the fit of the Mincer-Zarnowitz type of regression of forecast efficiency:

$$(\widetilde{q}_{t+h} - \widetilde{q}_t) = \alpha + \beta(\widetilde{q}^f_{t,h} - \widetilde{q}_t) + \eta_{t+h}$$

so that negative or low values indicate that a given method do not pass the efficiency test. The table shows that for the baseline DSGE model the correlation coefficients are positive for all horizons, ranging between 0.22 for the shortest one and 0.56 for six-quarter ahead forecasts. Only two other methods deliver comparably good (AR-fixed) or slightly worse (DSGEt) results. On the contrary, forecasts from two models estimated on first differences (RW with drift and DVAR) are negatively correlated with realizations, which confirms their low quality. For the remaining VAR models the results are mixed, as for some horizons the correlation coefficients are negative, whereas for other positive. It is unambiguous, however, that they deliver less effective forecasts than DSGE and AR-fixed models.

To shed more light on the above result, on Figure 2 we present scatter plots of realizations (y-axis) versus model-based forecasts (x-axis) for selected models and horizons. Points along the 45 degree line correspond to perfect predictions. Observations that fall in the top-right and bottom-left quadrants are forecasts that anticipate correctly the directional change of the real price of oil. Based on their position relative to the 45 degree line, the predictions that have the correct sign can be further split between those where the forecasted absolute change in the real price of oil is larger (overprediction) or smaller (underprediction) than realization. The figure shows that in most cases the DSGE model delivers forecasts of correct sign, which underpredict the scale of subsequent oil price adjustment. Adding a trend to the model (DSGEt) is just shifting the "dots" to the right, which is detrimental to forecast quality. As regards the DVAR model, the forecasts are very tightly concentrated around the value that is consistent with the deterministic trend observed in the past. They are therefore unable to predict the wide dispersion of realizations. In turn, the KVAR model is the only one that delivers forecasts that are comparably dispersed as realizations. The problem of this method is that, apart from longer horizons, the correlation between both is negative or low. Finally, the figure shows that the pooled forecast is very tightly dispersed and actually not very distant from forecasts derived from the benchmark random walk model.

4.2 Density forecasts

We complement point forecast evaluation with the assessment of density forecasts. The aim is to check to what extent the analyzed forecasts provide a realistic description of actual uncertainty. We start by looking at fan charts from the benchmark and five selected models for two characteristic periods. Figure 3 presents density forecasts formulated in 2008:2, i.e. just before the plummeting of the oil prices during the Great Financial Crisis. It illustrates that the initial depth of oil price decline was deeper than what any method would predict. It also shows that the mean-reverting methods, i.e. DSGE and AR-fixed, provided more reliable density forecasts than DBVAR and KVAR, even though the width of the interval forecast they delivered was lower than from DBVAR. In turn, Figure 4 presents fan charts for forecasts formulated using data up to 2009:1, i.e. just after the oil price collapse. In this case the subsequent rebound is within the 99% confidence interval generated by all presented methods. It can be noted that fan charts can only be used for illustrative purposes, but cannot determine which model is the best.

To analyze if density forecasts from one model outperform another one, we compare their quality by using the log predictive scores (LPS) calculated with the method proposed by Adolfson et al. (2007). Table 5 presents the average LPS differences of a given model in comparison to the random walk so that positive values indicate that the investigated model outperforms the benchmark. We test whether the values are significantly above zero with the one-tailed Amisano and Giacomini (2007) test. The table confirms the results for point forecasts, namely that the baseline DSGE model and AR-fixed perform best. In this case, however, their superiority over the benchmark random walk model is not significant. As regards the VAR models, the average values of LPS are usually below the random walk benchmark, apart from MBVAR.

We end the evaluation of density forecasts with Probability Integral Transform (PIT), a measure developed by Rosenblatt in 1952, imported to the economic literature by Diebold et al. (1998) and applied in evaluating the quality of DSGE based forecasts by Herbst and Schorfheide (2012) and Kolasa et al. (2012). The PIT is defined value of cumulative predictive density at realization. If the density forecast is well calibrated, PITs should drawn from independent uniform distribution on the interval (0, 1). One way to illustrate if this is the case is to divide the unit interval into J subintervals and check if the fraction of PITs in each of them is close $1/J$. Following Herbst and Schorfheide (2012) we set $J = 10$ and present the results in Figure 5. It shows that for the shortest horizon the distribution of PITs does not deviate to much from uniformity. As regards longer horizons, it indicates that density forecasts from the baseline DSGE model are too narrow as too many PITs fall into the corner bins. As regards RW and especially DBVAR, an unproportional number of PITs fall into the lowest bin, reflecting the fact that both models tend to overpredict the level and analyzed variable. The opposite picture emerges for KVAR model, for which a large number of PITs fall into the upper bins, suggesting that the model produces density forecasts that are on average too low. In general, the figure shows that none of the models delivers a well calibrated density forecast.

Overall, a general picture that emerges from the above analysis is that the DSGE model, followed closely by AR-fixed, performs best among all analyzed methods. Its forecasts turned out to be unbiased, strongly correlated with realizations and characterized by the lowest RMSFE and highest LPS. On the downside, density forecasts from this model turned out to be too narrow, especially for longer horizons. Our interpretation of this results is that the strength of this model can be attributed to the fact that it assumes gradual mean-reversion of real oil prices and conservative short term dynamics.

5 Conclusion and Policy Implications

In this paper we have verified the forecasting performance of a set of methods for the quarterly real WTI oil prices. We were particularly interested in the performance of two structural models for the oil market, which were proposed in two articles published a decade ago. The first one is a structural VAR model by Kilian (2009), whereas the second one is a DSGE model developed by Nakov and Pescatori (2010). For the benchmark model we have chosen the naïve random walk, as it is was usually done in the previous studies. We have also added nine other competitors to the forecasting contest. Next, we have collected quarterly data from the period 1984:1-2018:3 and calculated recursive point and density forecasts from all analyzed methods. Finally, we have compared the quality of point and density forecasts using the data from the evaluation period 2008:1-2018:3 and a wide range of statistics.

The main conclusion is that, in order to deliver real oil prices forecasts of good quality, models must fulfill two principles. First, they should be conservative, i.e. they must not attempt to replicate out of sample the high volatility of oil prices observed in-sample. Second, they have to deliver real oil prices that are mean reverting. Given that the baseline DSGE model conforms with both principles, it becomes intuitive why its performance is relatively good over all forecast horizons. The side message of this paper is that the ability of DSGE models to forecast real oil prices should not be overplayed as another model that conforms with the same principles performs equally well. We have shown that there is indeed not an appreciable difference in the forecasting performance of the DSGE and AR-fixed models. The decision on which of the two models is more useful depends on the preferences of the final user. The argument in favor of our baseline DSGE is that it provides a fully consistent story and accounts for the feedback between the oil market and the macroeconomy. The strength of the AR-fixed method is that its is very easy to implement. However, it does not provide any insights into what drives oil prices nor explanations of how the adjustment to unfolds. The final finding from our analysis is that a good performance of VAR models presented in the earlier studies has not been confirmed in our sample. The accuracy of forecasts from this class of models turned out to be visibly worse than those from the DSGE model.

References

Adolfson, M., Linde, J., Villani, M., 2007. Forecasting performance of an open economy DSGE model. Econometric Reviews 26 (2-4), 289–328.

Alquist, R., Bhattarai, S., Coibion, O., 2019. Commodity-price comovement and global economic activity. Journal of Monetary Economics, in Press.
URL `https://doi.org/10.1016/j.jmoneco.2019.02.004`

Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economic Statistics 25, 177–190.

Balke, N. S., Brown, S. P., 2018. Oil supply shocks and the U.S. economy: An estimated DSGE model. Energy Policy 116, 357 – 372.

Balke, N. S., Brown, S. P. A., Yucel, M. K., 2010. Oil price shocks and U.S. economic activity: An international perspective. Working Papers 1003, Federal Reserve Bank of Dallas.

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

Baumeister, C., Kilian, L., 2012. Real-time forecasts of the real price of oil. Journal of Business & Economic Statistics 30 (2), 326–336.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: A forecast combination approach. Journal of Business & Economic Statistics 33 (3), 338–351.

Baumeister, C., Peersman, G., 2013. The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. Journal of Applied Econometrics 28 (7), 1087–1109.

Bodenstein, M., Erceg, C., Guerrieri, L., 2011. Oil shocks and external adjustment. Journal of International Economics 83, 168–184.

Ca' Zorzi, M., Kolasa, M., Rubaszek, M., 2017. Exchange rate forecasting with DSGE models. Journal of International Economics 107 (C), 127–146.

Caldara, D., Cavallo, M., Iacoviello, M., 2018. Oil price elasticities and oil price fluctuations. Journal of Monetary Economics 103, 1–20.

Chang, A. C., Li, P., 2015. Is economics research replicable? Sixty published papers from thirteen journals say "usually not". Finance and Economics Discussion Series 2015-83, Board of Governors of the Federal Reserve System.

Clark, T. E., West, K. D., 2006. Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. Journal of Econometrics 135 (1-2), 155–186.

Dees, S., Karadeloglou, P., Kaufmann, R. K., Sanchez, M., 2007. Modelling the world oil market: Assessment of a quarterly econometric model. Energy Policy 35 (1), 178–191.

Del Negro, M., Schorfheide, F., 2012. DSGE model-based forecasting. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2 of Handbook of Economic Forecasting. Elsevier.

Delle Chiaie, S., Ferrara, L., Giannone, D., 2017. Common factors of commodity prices. Working Paper Series 2112, European Central Bank.

Diebold, F. X., Gunther, T. A., Tay, A. S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39 (4), 863–83.

Dvir, E., Rogoff, K. S., 2009. Three epochs of oil. NBER Working Papers 14927, National Bureau of Economic Research, Inc.

Fattouh, B., Kilian, L., Mahadeva, L., 2013. The role of speculation in oil markets: What have we learned so far? The Energy Journal 34 (3), 7–33.

Faust, J., Wright, J. H., 2013. Forecasting inflation. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2. Elsevier, Ch. 1.

Fernandez-Villaverde, J., Rubio-Ramirez, J. F., Sargent, T. J., Watson, M. W., 2007. ABCs (and Ds) of understanding VARs. American Economic Review 97 (3), 1021–1026.

Gurkaynak, R. S., Kisacikoglu, B., Rossi, B., 2013. Do DSGE models forecast more accurately out-of-sample than VAR models? CEPR Discussion Papers 9576, The Center for Economic and Policy Research.

Hamilton, J. D., 2003. What is an oil shock? Journal of Econometrics 113 (2), 363 – 398.

Hamilton, J. D., 2019. Measuring global economic activity. NBER Working Papers 25778, National Bureau of Economic Research, Inc.

Herbst, E., Schorfheide, F., 2012. Evaluating DSGE model forecasts of comovements. Journal of Econometrics 171 (2), 152–166.

Jacks, D. S., Stuermer, M., 2018. What drives commodity price booms and busts? Energy Economics, in Press.
URL `https://doi.org/10.1016/j.eneco.2018.05.023`

Kadiyala, K. R., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR models. Journal of Applied Econometrics 12 (2), 99–132.

Kaufmann, R. K., 1995. A model of the world oil market for project link integrating economics, geology and politics. Economic Modelling 12 (2), 165 – 178.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. The American Economic Review 99 (3), 1053–1069.

Kilian, L., 2019. Measuring global real economic activity: Do recent critiques hold up to scrutiny? Economics Letters 178, 106 – 110.

Kilian, L., Murphy, D. P., 2012. Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models. Journal of the European Economic Association 10 (5), 1166–1188.

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

Kolasa, M., Rubaszek, M., 2018. Does the foreign sector help forecast domestic variables in DSGE models? International Journal of Forecasting 34 (4), 809–821.

Kolasa, M., Rubaszek, M., Skrzypczynski, P., October 2012. Putting the New Keynesian DSGE model to the real-time forecasting test. Journal of Money, Credit and Banking 44 (7), 1301–1324.

Kruse, R., Wegener, C., 2019. Time-varying persistence in real oil prices and its determinant. Energy Economics in Press.

Leduc, S., Sill, K., 2004. A quantitative analysis of oil-price shocks, systematic monetary policy, and economic downturns. Journal of Monetary Economics 51 (4), 781–808.

Lippi, F., Nobili, A., 2012. Oil and the macroeconomy: A quantitative structural analysis. Journal of the European Economic Association 10 (5), 1059–1083.

Manescu, C., Van Robays, I., 2016. Forecasting the brent oil price: Addressing time-variation in forecast performance. CESifo Working Paper Series 6242, CESifo Group Munich.

Nakov, A., Nuno, G., 2013. Saudi Arabia and the oil market. Economic Journal 123 (12), 1333–1362.

Nakov, A., Pescatori, A., 2010. Oil and the great moderation. The Economic Journal 120, 131–155.

Peersman, G., Stevens, A., 2013. Analyzing oil demand and supply shocks in an estimated DSGE model. In: Stevens, A. (Ed.), Applications in dynamic stochastic general equilibrium macroeconomics. PhD Dissertation at Ghent University, Ch. 1.

Rubaszek, M., Skrzypczynski, P., 2008. On the forecasting performance of a small-scale DSGE model. International Journal of Forecasting 24 (3), 498–512.

Villani, M., 2009. Steady-state priors for vector autoregressions. Journal of Applied Econometrics 24 (4), 630–650.

Tables and figures

Table 1: Variable description and data sources

	Variable description	Source
$\widetilde{y}$	log of real GDP	FRED (GDPC1)
$\widetilde{p}$	log of GDP implicit price deflator	FRED (GDPDEF)
$\widetilde{i}$	effective federal funds rate	FRED (FEDFUNDS)
$\widetilde{s}$	log of WTI price	FRED (WTISPLC)
$\widetilde{q}$	log of real WTI price	$\tilde{q} = \tilde{s} - \tilde{p}$
$\widetilde{prod}$	log of global oil production	Quandl (EIA/INTL_53_1_WORL_TBPD_Q)
$\widetilde{gea}$	global economic activity index	https://sites.google.com/site/lkilian2019/

Notes: FRED – Federal Reserve Bank of St. Louis database. The data for oil production prior to 1994 were interpolated from annual values sourced from Quandl, ticker: EIA/INTL_53_1_WORL_TBPD_A.

Table 2: Mean Forecast Error (MFE)

	H=1	H=2	H=4	H=6	H=8	H=12	H=16	H=20
RW	-0.01	-0.02	-0.07	-0.08	-0.10	-0.21	-0.32	-0.37
RW with drift	-0.01	-0.03	-0.08	-0.09	-0.12	-0.25	-0.38*	-0.45$^{\bullet}$
Bayesian AR	0.02	0.03	0.04	0.07	0.09	0.07	0.04	0.03
AR-fixed	0.02	0.04	0.05	0.10	0.14	0.16	0.17	0.19
Difference BVAR	-0.02	-0.06	-0.15	-0.21°	-0.28°	-0.50$^{\bullet}$	-0.71$^{\bullet}$	-0.85$^{\bullet}$
Level BVAR	0.00	-0.01	-0.05	-0.04	-0.05	-0.10	-0.16	-0.20
Mean-reverting BVAR	0.00	0.00	-0.04	-0.03	-0.04	-0.11	-0.19	-0.26°
Kilian VAR	0.05	0.10	0.15	0.21	0.25	0.21	0.11	0.05
Kilian BVAR	0.04	0.09	0.13	0.19	0.23	0.19	0.10	0.04
DSGE with trend	-0.01	-0.02	-0.06	-0.05	-0.06	-0.10	-0.16	-0.21
DSGE without trend	0.02	0.04	0.05	0.09	0.11	0.11	0.09	0.06
Pool	0.01	0.01	0.00	0.01	0.01	-0.05	-0.13	-0.18

Notes: The errors are calculated for log real prices so that the presented numbers represent average percent difference between realizations and forecasts. Positive and negative values stand for under- and overprediction, respectively. Asterisks $^{\bullet}$, * and $^{\circ}$ denote, respectively, the 1%, 5% and 10% significance levels for the unbiasedness null, where the long-run variance is calculated with the Newey-West method.

Table 3: Root Mean Squared Forecast Error (RMSFE)

	H=1	H=2	H=4	H=6	H=8	H=12	H=16	H=20
RW (RMSFE value)	0.18	0.28	0.35	0.39	0.46	0.52	0.58	0.59
RW with drift	1.01	1.01	1.03	1.04	1.05	1.06	1.09	1.11
Bayesian AR	1.00	1.03	1.01	1.01	1.00	0.94°	0.85^{*}	$0.74^{\bullet}$
AR-fixed	0.98	0.96	0.90°	0.91°	0.90^{*}	0.88^{*}	$0.80^{\bullet}$	$0.75^{\bullet}$
Difference BVAR	1.09	1.14	1.17	1.21	1.26	1.39	1.55	1.70
Level BVAR	1.02	1.05	1.06	1.08	1.07	1.06	1.00	0.92°
Mean-reverting BVAR	1.02	1.04	0.99	0.97	0.96	0.94°	0.92°	0.91^{*}
Kilian VAR	1.11	1.21	1.40	1.48	1.45	1.15	0.84^{*}	0.82^{*}
Kilian BVAR	1.09	1.19	1.38	1.45	1.41	1.12	0.85^{*}	0.83^{*}
DSGE with trend	1.00	0.97	0.91^{*}	0.92^{*}	$0.91^{\bullet}$	$0.91^{\bullet}$	$0.86^{\bullet}$	$0.80^{\bullet}$
DSGE without trend	0.98	0.95	0.85°	0.86^{*}	0.87^{*}	$0.85^{\bullet}$	$0.77^{\bullet}$	$0.72^{\bullet}$
Pool	1.00	1.02	1.01	1.00	0.99	0.90°	0.82^{*}	$0.76^{\bullet}$

Notes: The table shows the ratios of the RMSFE from a given model in comparison to the RW benchmark. Values below unity indicate that forecasts from the model are more accurate than from this benchmark. Asterisks $^{\bullet}$, * and $^{\circ}$ denote, respectively, the 1%, 5% and 10% significance levels of the Clark-West test, where the long-run variance is calculated with the Newey-West method.

Table 4: Correlation of forecasts and realizations

	H=1	H=2	H=4	H=6	H=8	H=12	H=16	H=20
RW	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RW with drift	-0.25	-0.44	-0.61	-0.58	-0.59	-0.56	-0.56	-0.53
Bayesian AR	0.12	-0.06	0.01	0.07	0.12	0.02	-0.09	0.36
AR-fixed	0.24	0.37	0.51	0.51	0.54	0.48	0.51	0.53
Difference BVAR	-0.05	-0.34	-0.31	-0.40	-0.46	-0.38	-0.40	-0.46
Level BVAR	0.08	-0.06	-0.01	0.02	0.03	-0.15	-0.60	-0.35
Mean-reverting BVAR	-0.04	-0.17	0.11	0.18	0.20	0.13	0.01	0.14
Kilian VAR	0.08	-0.15	-0.36	-0.33	-0.28	0.06	0.41	0.46
Kilian BVAR	0.02	-0.20	-0.39	-0.33	-0.27	0.06	0.38	0.43
DSGE with trend	0.13	0.23	0.40	0.39	0.40	0.27	0.22	0.38
DSGE without trend	0.22	0.34	0.55	0.56	0.54	0.45	0.42	0.40
Pool	0.10	-0.07	-0.03	0.02	0.07	0.19	0.45	0.61

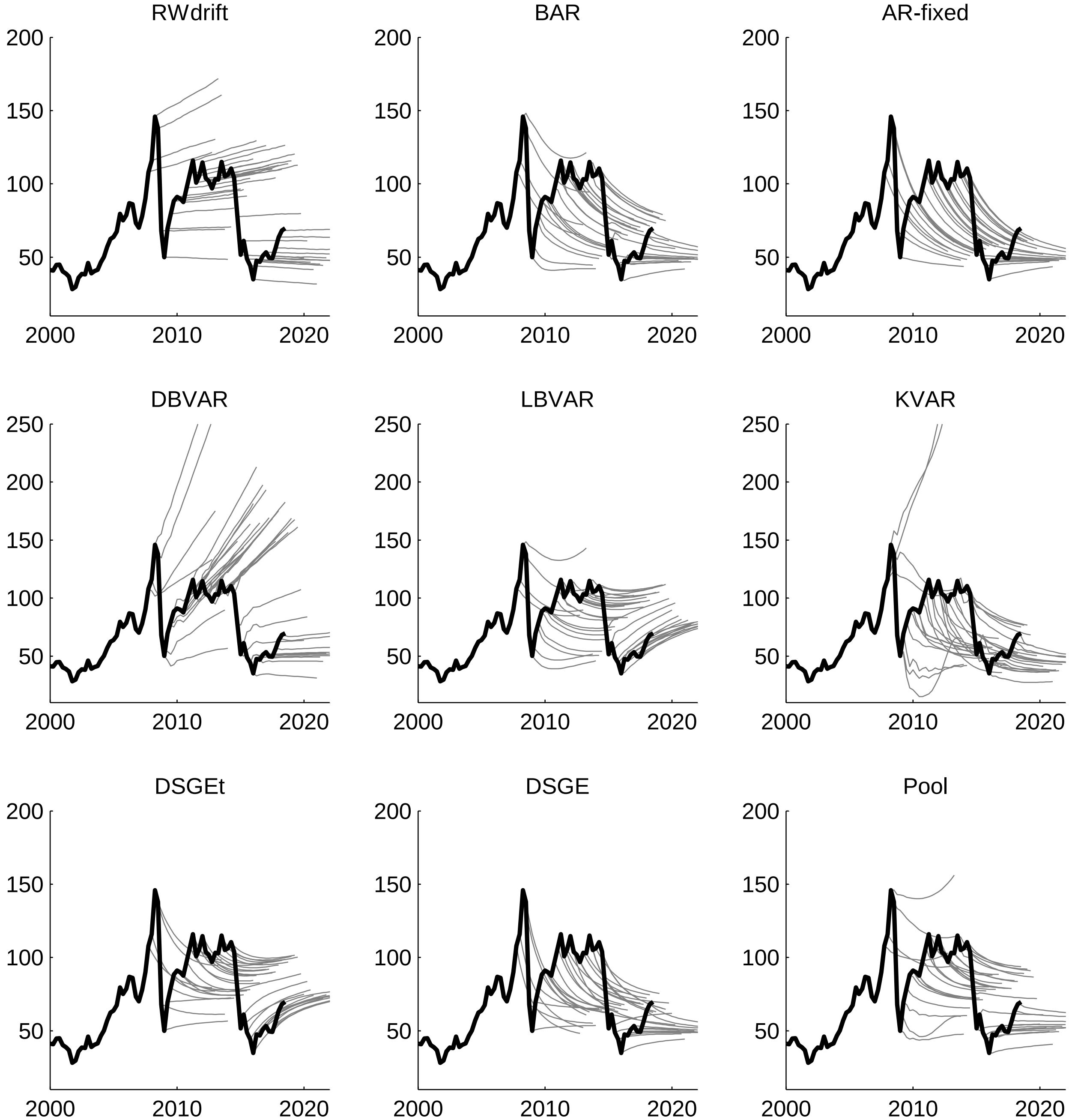
Notes: The table shows the correlation of realized ($\widetilde{q}_{t+h} - \widetilde{q}_t$) and forecasted ($\widetilde{q}^f_{t,h} - \widetilde{q}_t$) changes for log real oil prices.

Table 5: Log Predictive Scores (LPS)

	H=1	H=2	H=4	H=6	H=8	H=12	H=16	H=20
RW with drift	-0.01	-0.03	-0.05	-0.05	-0.06	-0.07	-0.09	-0.10
Bayesian AR	0.05	0.12	0.11	0.01	0.00	0.00	0.05	0.08
AR-fixed	0.04	0.08	0.16	0.12	0.11	0.11	0.19	0.26
Difference BVAR	-0.09	-0.09	-0.14	-0.28	-0.34	-0.50	-0.64	-0.69
Level BVAR	0.03	0.06	0.01	-0.11	-0.13	-0.15	-0.08	0.01
Mean-reverting BVAR	0.06	0.11	0.14	0.05	0.04	0.02	0.03	0.00
Kilian VAR	-0.07	-0.11	-0.29	-0.33	-0.30	-0.15	-0.03	-0.05
Kilian BVAR	-0.05	-0.12	-0.27	-0.31	-0.29	-0.13	-0.01	-0.03
DSGE with trend	0.01	0.05	0.12	0.03	0.02	0.00	0.10^{*}	$0.23^{\bullet}$
DSGE without trend	0.05	0.11	0.23	0.17	0.12	0.11	0.19	0.26

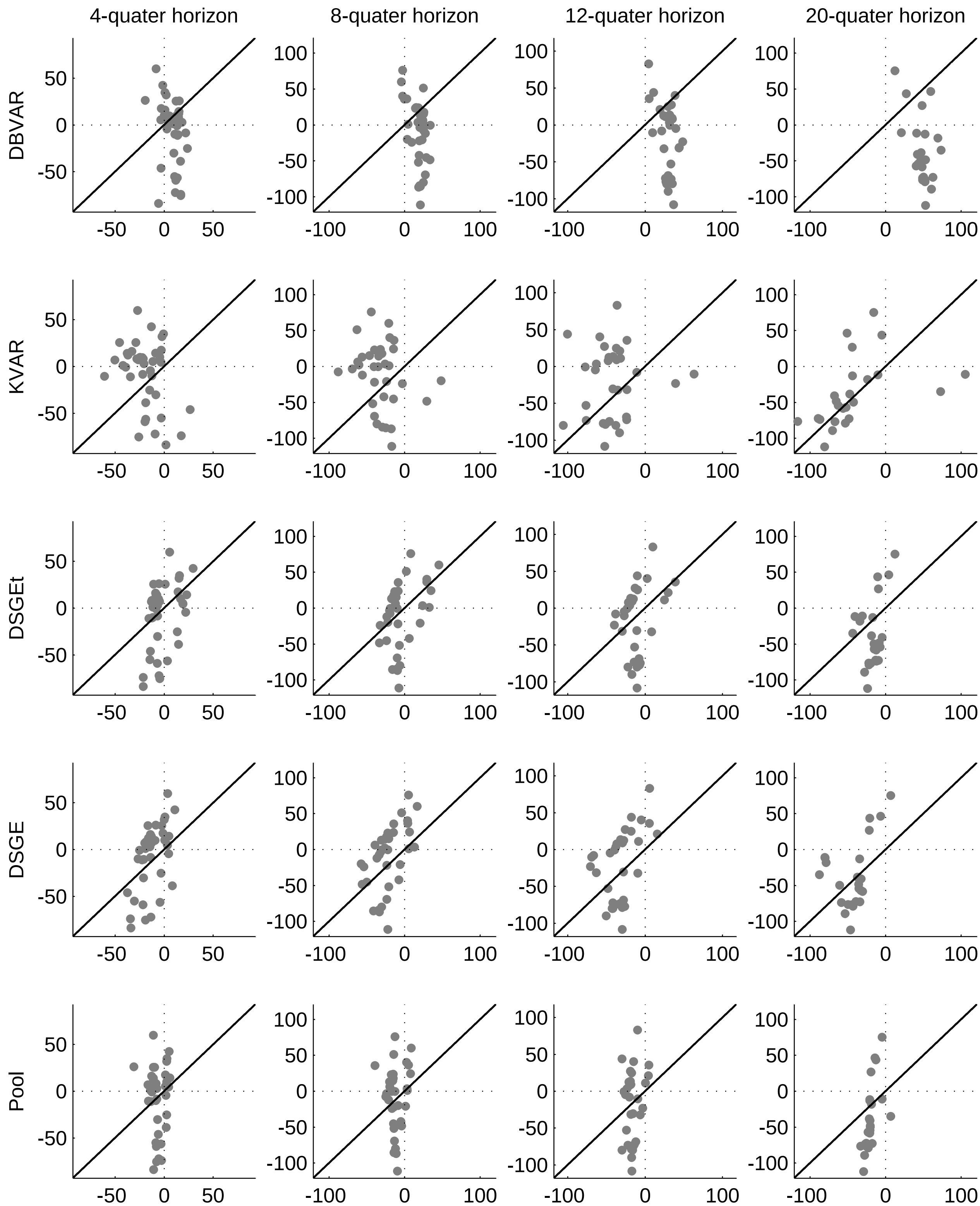
Notes: The figures represent the average differences of the LPS from a given model in comparison to the RW benchmark. Positive values indicate that density forecasts from a given model are better calibrated than those from the benchmark. The values can be interpreted as a percentage gain in the fit of forecasts to realizations in comparison to the benchmark. Asterisks $^{\bullet}$, * and $^{\circ}$ denote, respectively, the 1%, 5% and 10% significance levels of the one-tailed Amisano and Giacomini (2007) test, where the long-run variance is calculated with the Newey-West method.

Figure 1: Sequential forecasts for real WTI prices



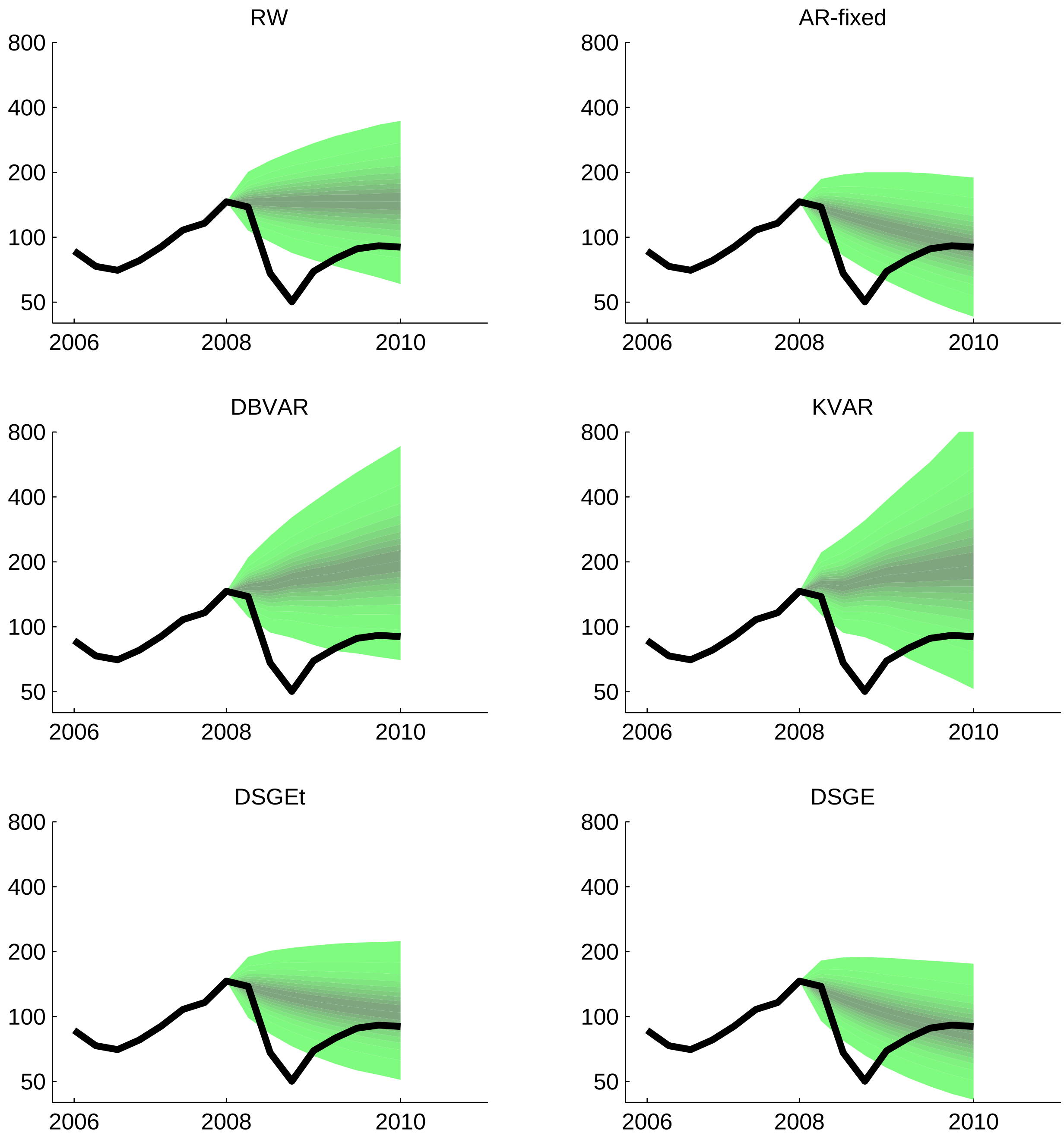
Notes: The values are expressed in real USD per barrel, where the reference date is set to the last observation from the sample (2018q3).

Figure 2: Scatter-plot of realizations and forecasts



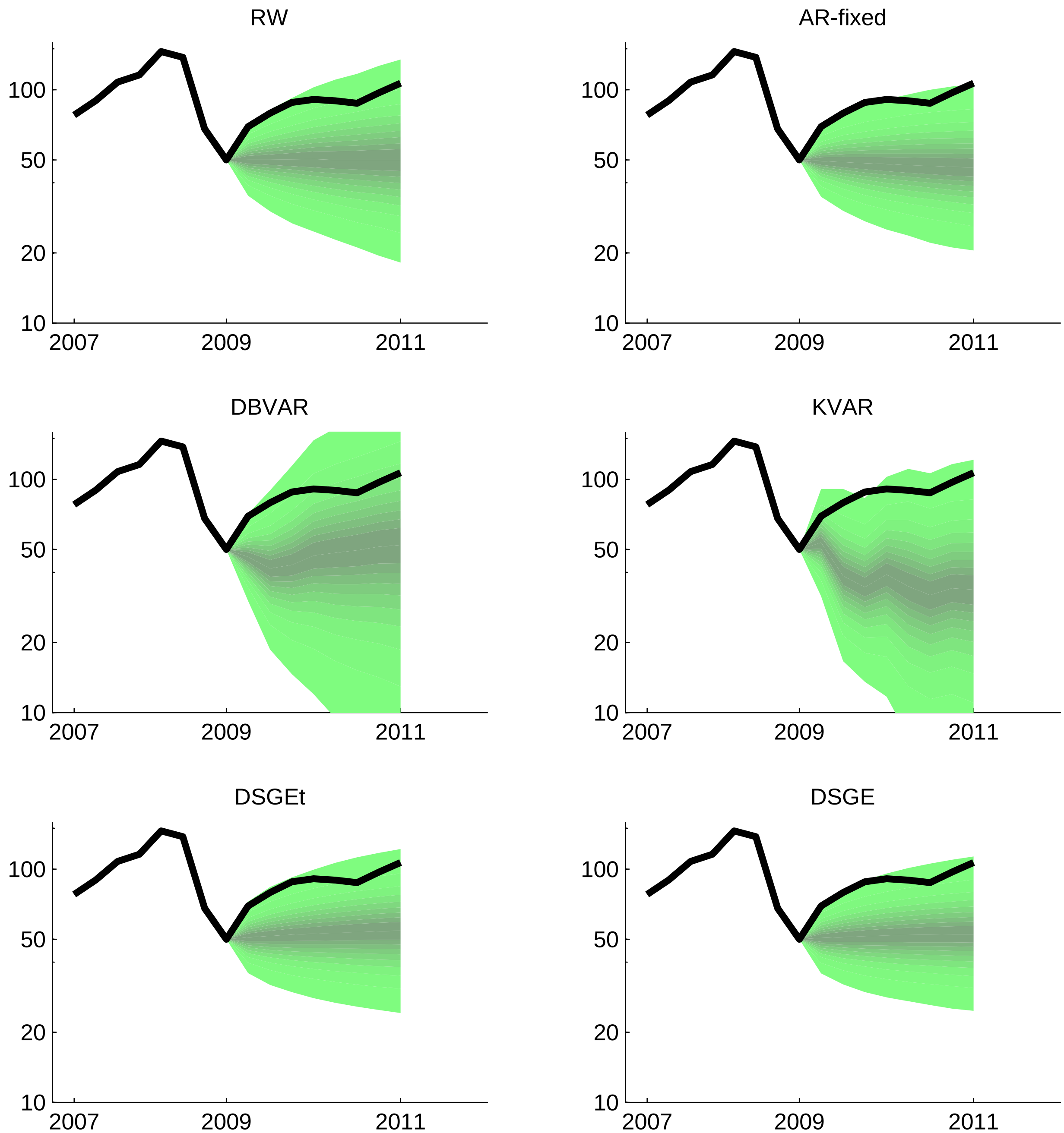
Notes: Forecasted values (expressed as percents) for log-changes ($\widetilde{q}^{f}_{t,h} - \widetilde{q}_t$) are on the horizontal axis, whereas for realizations ($\widetilde{q}_{t+h} - \widetilde{q}_t$) on the vertical one.

Figure 3: Fan chart for 2008q2



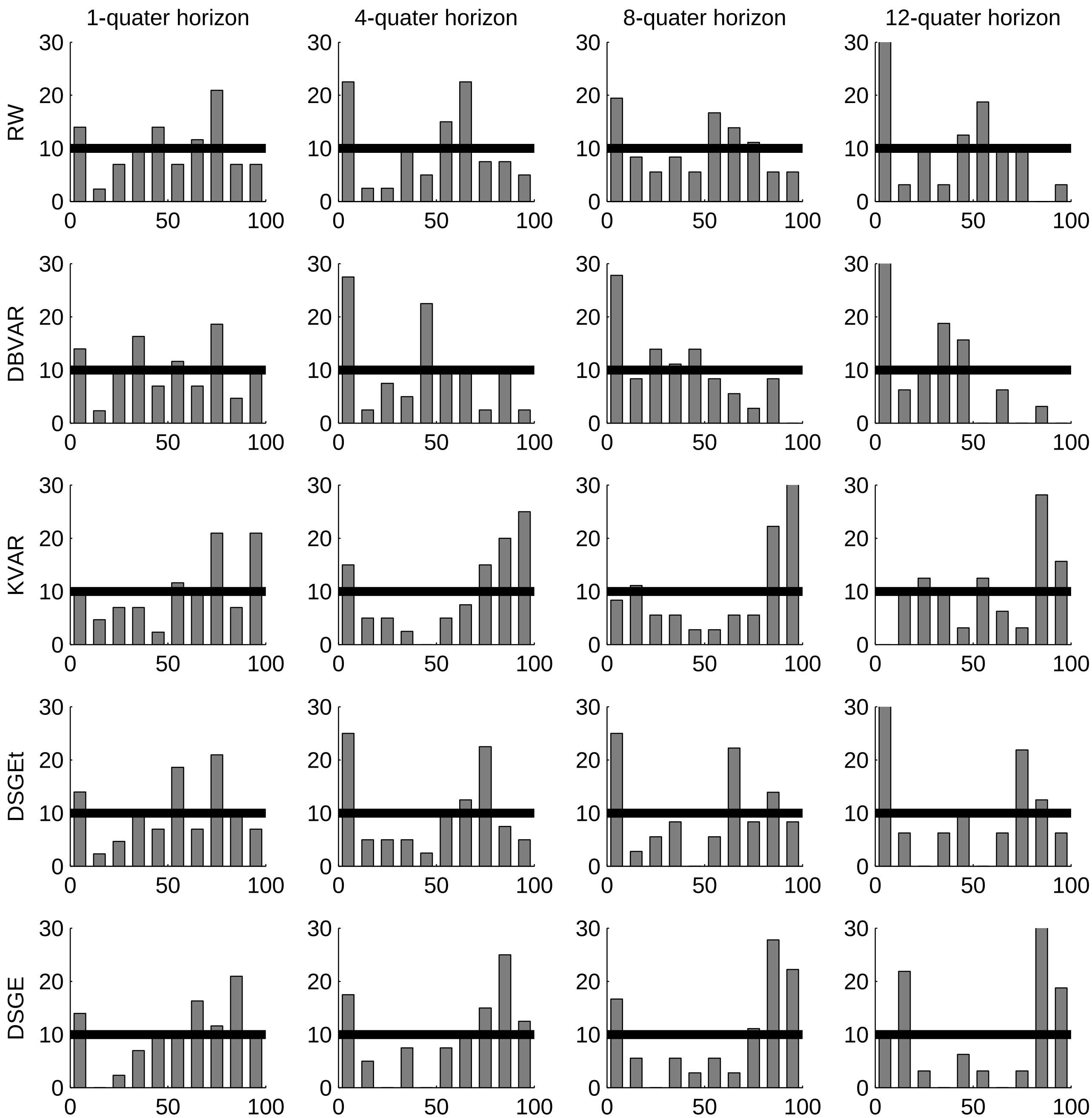
Notes: The values are expressed in real USD per barrel, where the reference date is set to the last observation from the sample (2018q3). The shaded areas cover up to 99% confidence interval.

Figure 4: Fan chart for 2009q1



Notes: The values are expressed in real USD per barrel, where the reference date is set to the last observation from the sample (2018q3). The shaded areas cover up to 99% confidence interval.

Figure 5: Probability Integral Transform (PIT) histograms



Notes: Bars represent the fraction of realized observations falling into deciles of density forecasts. The solid line represents the theoretical value of 10% for a well-calibrated model.