TI 2021-103/III
Tinbergen Institute Discussion Paper

Quantifying Time-Varying Forecast Uncertainty and Risk for the Real Price of Oil

Knut Are Aastveit[1,2]
Jamie L. Cross[2]
Herman K. van Dijk[1,3,4]

[1] Norges Bank
[2] BI Norwegian Business School
[3] Erasmus University Rotterdam
[4] Tinbergen Institute

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Quantifying time-varying forecast uncertainty and risk for the real price of oil

Knut Are Aastveit*
Norges Bank & BI Norwegian Business School
Jamie L. Cross
BI Norwegian Business School
Herman K. van Dijk
Erasmus University & Tinbergen Institute & Norges Bank

November 9, 2021

Abstract

We propose a novel and numerically efficient quantification approach to forecast uncertainty of the real price of oil using a combination of probabilistic individual model forecasts. Our combination method extends earlier approaches that have been applied to oil price forecasting, by allowing for sequentially updating of time-varying combination weights, estimation of time-varying forecast biases and facets of miscalibration of individual forecast densities and time-varying inter-dependencies among models. To illustrate the usefulness of the method, we present an extensive set of empirical results about time-varying forecast uncertainty and risk for the real price of oil over the period 1974-2018. We show that the combination approach systematically outperforms commonly used benchmark models and combination approaches, both in terms of point and density forecasts. The dynamic patterns of the estimated individual model weights are highly time-varying, reflecting a large time variation in the relative performance of the various individual models. The combination approach has built-in diagnostic information measures about forecast inaccuracy and/or model set incompleteness, which provide clear signals of model incompleteness during three crisis periods. To highlight that our approach also can be useful for policy analysis, we present a basic analysis of profit-loss and hedging against price risk.

Keywords: Oil price, Forecast density combination, Bayesian forecasting, Instabilities, Model uncertainty

*This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of the Norges Bank. We gratefully acknowledge comments and suggestions from the editor Christian Hansen, the associate editor, two anonymous referees, Lennart Hoogerheide, Rob Luginbuhl, Gary Koop, Kenichiro McAlinn James Mitchell, Francesco Ravazzolo, Mike West, an anonymous referee for our Norges Bank Working Paper and seminar and conference participants at Norges Bank, University of Strathclyde, University of Rome Tor Vergata, CPB Data Science Mini-Symposium, IAAE 2021 Annual Conference, and ISBA 2021 World Meeting. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

1 Introduction

The adverse macroeconomic implications of oil price fluctuations have been known since the 1970's, making the real price of crude oil a key variable in both macroeconomic forecasting and structural analysis (e.g. Hamilton, 1983, 2009; Kilian, 2009; Ravazzolo and Rothman, 2013; Baumeister and Hamilton, 2019). For instance, central banks, private sector forecasters and international organizations view the price of oil as one of the key variables in generating macroeconomic projections and in assessing macroeconomic risks. Oil price forecasts are also crucial for how some sectors of the economy operate their business, e.g. airlines, utilities and automobile manufacturers. Common to all, however, is the notion that the price of oil is difficult to forecast. Hamilton (2009) documents that the statistical regularities of changes in the real price of oil have historically tended to be (1) permanent, (2) hard to predict, and (3) governed by very different regimes over time. He further argues that the price of oil seems to follow a random walk without drift. It is therefore common among professional forecasters to either use the current spot price or the price of oil futures contracts as the forecast of the price of oil. More recently, researchers have explored numerous alternative models and methods in order to forecast the most likely future realization of the oil price (e.g. Alquist and Kilian, 2010; Alquist et al., 2013; Baumeister and Kilian, 2012, 2015; Manescu and Robays, 2016; Bernard et al., 2018; Pak, 2018; Garratt et al., 2019; Baumeister et al., 2020). These papers show that while the random walk is hard to beat in out-of-sample forecasting exercises, careful attention to the economic fundamentals that are driving energy markets can generate improvements in point forecast accuracy.

In this paper, we provide a novel and numerically efficient quantification approach to forecast uncertainty of the real price of crude oil using a combination of probabilistic individual model forecasts. The proposed Forecast Density Combination (FDC) model is based on a probabilistic and econometric interpretation of the Bayesian Predictive Synthesis (BPS) model due to McAlinn and West (2019) and McAlinn et al. (2020). BPS is a coherent Bayesian framework for evaluation, calibration, and data-informed combination of multiple forecast densities, that is based on the earlier literature on Bayesian "agent/expert opinion analysis" (West, 1984; Genest and Schervish, 1985; West, 1988, 1992; West and Crosse, 1992). The proposed combination method extends earlier approaches that have been applied to oil price forecasting models, by allowing for three key features. First, the method features time-varying combination weights, and explicitly factors into the model combination the inherent uncertainty surrounding the estimation of the combination weights. Second, it allows for modelling and estimation of time-varying forecast biases and facets of miscalibration of individual forecast densities and time-varying inter-dependencies among models. Third, it provides diagnostic analysis of model set incompleteness and learning from previous forecast mistakes, which serves to improve model specification.

We make use of the proposed combination model and provide an extensive set of empirical results about time-varying out-of sample forecast performance, forecast uncertainty and risk for the real price of oil. Following Garratt et al. (2019) recent replication of Baumeister and Kilian (2015), we use real-time monthly data, where the full sample covers the period 1973:01-2017:12. Our starting point is to document substantial changes over time in the mean and volatility of the real price of oil, a pattern that also transfers into the shape of the data densities. We then proceed by showing six main results, which extends the current literature on forecasting the price of oil. These conclusions are found to be robust across different oil price series, BPS specifications and model combination sets.

First, our combination approach systematically outperforms all benchmarks we compare it to, both in terms of point and density forecasts. The competing benchmark models, range from the six state-of-the-art individual forecasting models used in Baumeister and Kilian (2015), including the commonly used naive no-change model, to alternative combination approaches such as equal weights for the individual models and Bayesian Model Averaging (BMA). While the gains from the model combination relative to the alternative models are limited at the one month horizon, substantial gains in relative forecast accuracy are obtained at all other horizons. At the six month horizon, the magnitude of reduction in terms of mean squared prediction error (MSPE) and logarithmic score (LS) relative to the no-change model exceeds 10% for MSPE and 12% for LS and are credibly different. For longer horizons, the gains are substantially larger, with reductions in the range of 30% and 40% for MSPE and 50% and 110% for LS, at the 12 and 24 months horizon, respectively.

Second, the favourable forecast performance from the proposed approach is not specific for certain time periods but it holds throughout the evaluation period. Large time variation is found in the relative performance of the various individual models and alternative model combinations. For instance, in line with Baumeister and Kilian (2012), it is seen that the VAR model performs well up until 2010, but then does relatively poorly over the subsequent period, thereby corroborating recent results in Baumeister et al. (2020).

Third, the approach allows for time-varying individual model weights that sequentially adapt according to the recent relative forecast performance of each model within the model combination set. At all forecast horizons, we document a considerable time variation in the weights attached to each model, also reflecting the large time-variation in the individual models forecasting performance. One key feature is that the weights are not restricted to be a convex combination in the unit interval - like most of the combination methods that are currently used within the econometrics literature - but are instead specified as a general linear combination and are thereby permitted to evolve along the real line. This has two advantages. First, allowing for both positive and negative weights means that it's possible to hedge against any potential forecast risk in recent periods. Second, a declining weight on one model does not necessarily imply an increasing weight on another model. Instead, model weights are permitted to change in accordance to the forecasts of the individual models. Such features are clearly desirable to a wide range of practitioners. For instance, just as a financial portfolio manager hedges against risk by assigning a negative weight to an asset, the proposed approach is able to automatically assign a negative weight to mitigate the impact of forecast risk, such as forecast bias. Such natural behaviour is not possible under combination models in which the model weights are restricted to be convex combinations, e.g. equal weights or BMA.

Fourth, BPS has a built-in time-varying intercept that is absent from simpler combination methods such as BMA. It is well known within the econometrics literature on forecast combinations that BMA assumes that the true model is included in the model set. By allowing for an intercept component that can adapt during episodes of low frequency signals from a set of forecasting models, our combination approach is able to better mitigate effects of model set misspecifciaton, i.e., model set incompleteness. We show that this has been particularly important since 2010, as the intercept term for all forecast horizons gradually starts to increase before abruptly dropping during the oil price collapse of 2014.

Fifth, the combination approach has built-in diagnostic information measures about forecast inaccuracy and/or model set incompleteness, which is also absent from simpler combination methods such as BMA. This is measured by the estimated time-varying model combination residual, which provides clear signals of model incompleteness during three crisis periods. This type of diagnostic information gives important signals about specifying models and model set improvements.

Finally, a basic analysis of profit-loss and hedging against price risk is presented in order to highlight the models potential for policy analysis. As a measure of optimal hedge ratio, the Minimum Variance Hedge (MVH) ratio is used which fluctuates between 0.1-0.4 at most horizons. However, notable spikes occur around the turn of the century as well as the two oil price collapses of 2009 and 2014.

Our paper is related to the recent resurgence in interest combination of density forecasts in macroeconomics, econometrics, and statistics. Prominent new developments range from combining predictive densities using weighted linear combinations of prediction models, evaluated using various scoring rules (e.g. Hall and Mitchell, 2007; Amisano and Giacomini, 2007; Jore et al., 2010; Hoogerheide et al., 2010; Kascha and Ravazzolo, 2010; Geweke and Amisano, 2011, 2012; Gneiting and Ranjan, 2013; Aastveit et al., 2014), to more complex combination approaches that allows for time-varying weights with possibly both learning and model set incompleteness (e.g. Terui and Van Dijk, 2002; Hoogerheide et al., 2010; Koop and Korobilis, 2012; Billio et al., 2013; Casarin et al., 2015; Pettenuzzo and Ravazzolo, 2016; Del Negro et al., 2016; Aastveit et al., 2018; McAlinn and West, 2019; McAlinn et al., 2020; Takanashi and McAlinn, 2020; Casarin et al., 2020). See also Aastveit et al. (2019) for a recent survey of these developments. Despite these research activities on several macroeconomic and financial variables, there are currently, to the best of our knowledge, no studies on how to quantify forecast uncertainty associated with the dynamic behaviour of the real price of crude oil.

The contents of this paper is structured as follows. In Section 2 we present our Forecast Density Combination model using Bayesian inference. In Section 3 a summary of the models used is given. Forecasting results and the risk analysis are presented in Sections 4 and 5, respectively. Section 6 concludes and suggests directions for future research.

2 Forecast Density Combination approach

A basic probabilistic approach to combine forecast information from different sources proceeds as follows. Let y_t be the economic variable of interest; let $\tilde{\boldsymbol{y}}_{\boldsymbol{t}}' = (\tilde{y}_{1,t}, \ldots, \tilde{y}_{n,t})$ be forecasts for this variable from $i = 1, ..., n$ models. In a simulation context $\tilde{y}_{i,t}$ is a draw from the forecast distribution with conditional density $p(\tilde{y}_{i,t}|I_{i,t-1}, M_i)$ given information set $I_{i,t-1}$ and model M_i. Let $\boldsymbol{v}_{\boldsymbol{t}}' = (v_{0,t}, \ldots, v_{n,t})$ be latent continuous random variable parameters where $v_{1,t}, \ldots, v_{n,t}$ will be used to combine the model forecasts and the role of $v_{0,t}$ is discussed below. The decomposition of the joint density of $y_t, \boldsymbol{v}_t, \tilde{\boldsymbol{y}}_t$ for the case of continuous random variables is given as:

$$p(y_t|I_{t-1}, M) = \int\int p(y_t|\boldsymbol{v}_t, \tilde{\boldsymbol{y}}_{\boldsymbol{t}})p(\boldsymbol{v}_t|\tilde{\boldsymbol{y}}_{\boldsymbol{t}})p(\tilde{\boldsymbol{y}}_{\boldsymbol{t}}|I_{t-1}, M)d\boldsymbol{v}_t d\tilde{\boldsymbol{y}}_{\boldsymbol{t}}, \tag{1}$$

where I_{t-1} is the joint information set of all models and M the union of all models. We label $p(y_t|\boldsymbol{v}_t, \tilde{\boldsymbol{y}}_{\boldsymbol{t}})$ as the combination density; $p(\boldsymbol{v}_t|\tilde{\boldsymbol{y}}_{\boldsymbol{t}})$ as the variable parameter density and $p(\tilde{\boldsymbol{y}}_{\boldsymbol{t}}|I_{t-1}, M)$ as the joint forecast density of the different models. Note that the integrals are of dimension $n+1$ and n.

A key step is to give specific content to the different densities. For the case of BPS it follows that:

$$\begin{aligned} p(y_t|\boldsymbol{v}_t, \tilde{\boldsymbol{y}}_{\boldsymbol{t}}) &= n(y_t|v_{0,t} + \sum_{i=1}^{n} v_{i,t}\tilde{y}_{i,t}, \sigma_t^2), & (2)\\ p(\boldsymbol{v}_t|\boldsymbol{v}_{t-1}, \boldsymbol{\Sigma}_{\boldsymbol{t}}) &= n(\boldsymbol{v}_t|\boldsymbol{v}_{t-1}, \boldsymbol{\Sigma}_{\boldsymbol{t}}), & (3)\\ p(\tilde{\boldsymbol{y}}_{\boldsymbol{t}}|I_{t-1}, M) &= \prod_{i=1}^{n} p(\tilde{y}_{i,t}|I_{i,t-1}, M_i). & (4) \end{aligned}$$

We emphasize that the combination density is a multivariate normal one with a time-varying constant $v_{0,t}$ in the conditional mean. This specification adds flexibility to the model combination and allows for forecast adjustments to shocks and regime changes in the data series while σ_t^2 allows for time-varying volatility. The parameter $\boldsymbol{\Sigma}_{\boldsymbol{t}} = \sigma_t^2 \boldsymbol{W}_{\boldsymbol{t}}$ and $\boldsymbol{W}_{\boldsymbol{t}}$ is a diagonal matrix with elements w_{it} given below.

A first feature of this approach, compared to standard combination models like BMA, is an analysis of the dynamic behaviour of the error ε_t implied by the combination density. It is given as:

$$\varepsilon_t = y_t - (v_{0,t} + \sum_{i=1}^{n} v_{i,t}\tilde{y}_{i,t}). \tag{5}$$

The forecast error of the ith model is usually defined as $y_t - \tilde{y}_{i,t}$ due to, for instance, sudden shocks in the series and model misspecification. Given the conditional mean of the combination density in (2), it is seen that the BPS error in (5) can be viewed as weighted combination of model forecast errors from each of the individual models.

We also investigate the dynamic behaviour of the error $\varepsilon_{i,t}$ using only model M_i, using:

$$\varepsilon_{i,t} = y_t - (v_{0,t} + v_{i,t}\tilde{y}_{i,t}). \tag{6}$$

A second feature of the approach is the possibility to learn about the contribution of the different individual model forecasts in the combination. Learning is specified as a random walk process of the continuous latent variable parameters $\boldsymbol{v}_t = (v_{0,t}, ...v_{n,t})'$, see equation (3). We note that the weights may become negative which in some cases helps in dynamic averaging. That is, the proposed approach is able to automatically assign a negative weight which may mitigate the impact of forecast risk, such as forecast bias. Such natural behaviour is not possible under combination models in which the model weights are restricted to be convex combinations, e.g. BMA or equal weighting methods.

Given the specified probability model, there exists a system of equations that has been labeled a latent dynamic factor model by McAlinn and West (2019) and McAlinn et al. (2020). However, we interpret this system as a multivariate regression model with generated regressors $\tilde{y}_t$ and latent time-varying parameters $v_{i,t}$. By construction, this equation system can be represented in the form of a generalized linear State Space Model where the explanatory variables $\tilde{y}_{i,t}$ are not given data but generated draws from the forecast distributions of the n models:

$$\begin{aligned} y_t &= v_{0,t} + \sum_{i=1}^{n} v_{i,t}\tilde{y}_{i,t} + \varepsilon_t, \varepsilon_t \sim NID(0, \sigma_t^2), & (7) \\ v_{i,t} &= v_{i,t-1} + \varepsilon_{v,t}, \varepsilon_{v,t} \sim NID(0, \sigma_{v,t}^2 = \sigma_t^2 w_t), i = 0, \dots, n. & (8) \end{aligned}$$

The time-varying volatility parameters σ_t^2 and $\sigma_{v,t}^2$ play important roles in this model as smoothness parameters: $\sigma_t^2 = \frac{\delta\sigma_{t-1}^2}{\gamma_t}$ is a betagamma volatility model in which $\delta \in (0, 1]$ is a volatility discount factor and $\gamma_t \sim \text{Beta}(\frac{\delta h_{t-1}}{2}, \frac{(1-\delta)h_{t-1}}{2})$ is an independent Beta innovation such that $h_t = \delta h_{t-1} + 1$ and $\mathbb{E}[\gamma_t | h_{t-1}] = \delta$ at all dates $t = 1, \dots, T$. The weight $w_t = \frac{1-\beta}{\beta} w_{t-1}$ is a component discount term in which $\beta \in (0, 1]$ is a state discount factor. For details, see West and Harrison (2006), Sect. 6.3.2 and Sect. 10.8.

In Figure 1 we show in a roadmap the connections between the components of the model. We distinguish between two figure shapes: rectangles which contain data and forecasts from different models and their combination; circles which contain latent time-varying regression parameters and the unobserved random parameters from the stochastic volatility process which have to be filtered/integrated out.

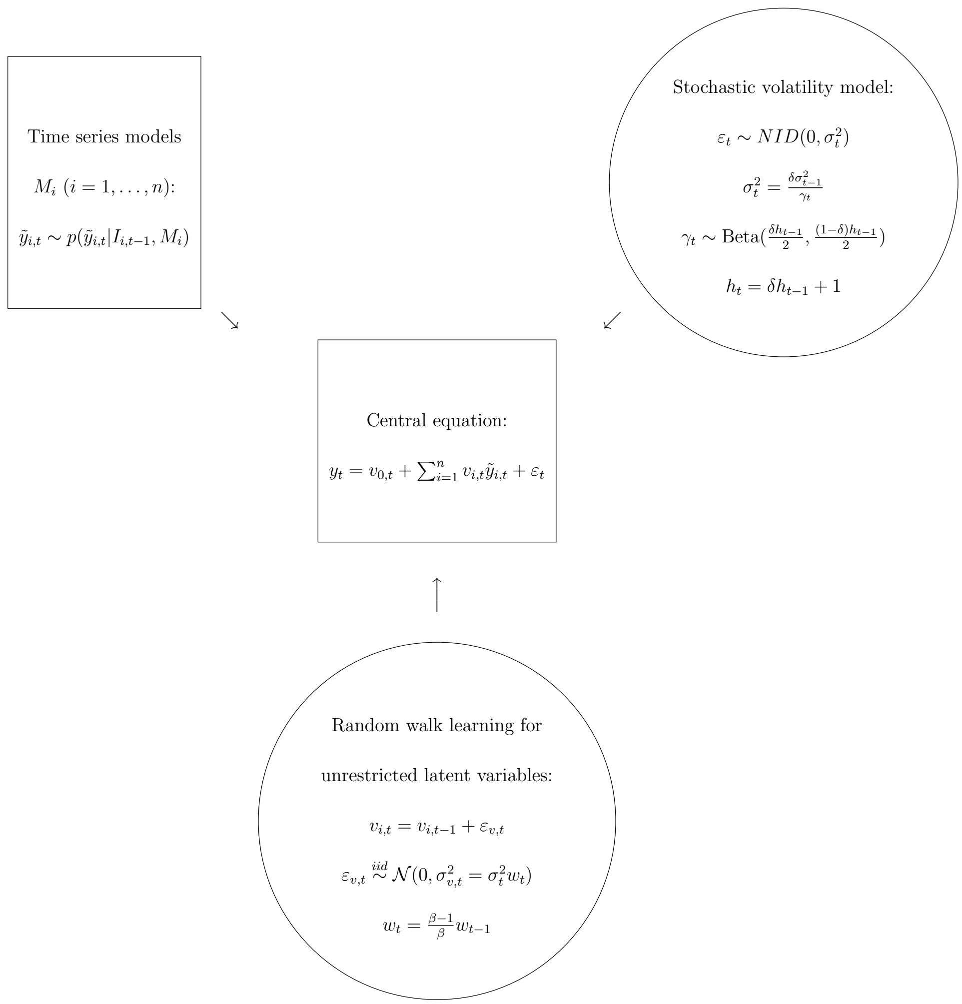

Figure 1: FDC model in generalized linear state space form. Given data, rectangles indicate model forecasts and combined forecasts. Circles refer to latent time-varying regression parameters and the random parameters from the stochastic volatility process where filtering/integration is used.

Bayesian estimation procedure using MCMC. The analytic solution of the integrals specified in the probability model (1)-(4) is often not known. We make use of simulation methods in order to deal with this problem. We also make use of Bayesian inference specifying prior information on the parameters of the stochastic volatility processes and the time-varying equation parameters which can be interpreted as unobserved states. Apart from the fundamental choice for Bayesian inference, there exists a practical reason in our case. Using simulation-based Bayesian inference the generated forecast draws from the different models are computationally directly carried forward to the estimation of the combination density. Thus, the uncertainty in the forecasts of the different models carries directly into the uncertainty of the combination forecasts. In contrast, frequentist methods like method of moments or maximum likelihood proceed in a two-step fashion by substituting the point forecasts of the different models in the combination equation and as such the second stage results suffer from the generated regressor problem; see, e.g. Pagan (1984).

The specification of the model discussed so far leads to the formulation of the likelihood of a generalized linear State Space model. Our Bayesian inferential procedure requires to choose prior values for the discount factors $\delta \in (0, 1]$, $\beta \in (0, 1]$ and priors for the initial values of the time-varying parameters $(v_{i,t}, \sigma_t^2)$ at $t = 0$. The role of discount factor $\beta \in (0, 1]$ is to operate on the parameter evolution via $w_t = \frac{1-\beta}{\beta} w_{t-1}$. Setting $\beta = 1$ implies a constant coefficients model, i.e. $w_t = 0$, while $\beta \in (0, 1]$ is consistent with time-varying coefficients. The parameter $\delta \in (0, 1]$ operates on the volatility evolution via $\sigma_t^2 = \frac{\delta \sigma_{t-1}^2}{\gamma_t}$ in which γ_t are Beta distributed innovations. Relevant choices of the discount factors β and δ are, of course, always context dependent. In our application, we opted for the value 0.9. As a simple robustness check, we computed LS and RMSFE values for the average of the triple (0.85, 0.9, 0.95). The results were consistent with the single value chosen.

Following McAlinn and West (2019), the prior on the latent time-varying parameters is conditional normal $v_{i,0}|\sigma_0^2 \sim N(m_0, C_0 \frac{\sigma_0^2}{s_0})$ where the hyperparameter: m_0 controls the mean, C_0 controls variance and h_0 and s_0 jointly control mean and variance of the measurement volatility, s_0 indirectly effects the variance of parameters. The initial prior on the measurement variance is marginal inverted gamma $\sigma_0^2 \sim IG(h_0/2, h_0 s_0/2)$. Specific choices for hyper parameters are $m_0 = (0, 1/n, \ldots, 1/n)'$, $C_0 = 10^{-4}\mathbf{I}_p$, $s_0 = 0.002$ and $h_0 = 10$.

Algorithmic outline using Kalman Filter and MCMC We emphasize that the proposed method makes use of a three-step Monte Carlo procedure instead of the usual two-step method. The extra step is due to the generation of random draws from the forecast distributions of the different individual models.

- **Forecast from n models.** Generate draws from the forecast distributions from the n different models which gives $\tilde{y}_{i,t}, i = 1, \ldots, n$

- **Latent variable parameters**. Use the Kalman update with initial value $v_{i,0}, i = 1, \ldots, n$, and generate variable parameters $v_{i,t}, i = 1, \ldots, n$ from the random walk process.

- **SV parameters.** Given draws $\tilde{y}_{i,t}, i = 1, \ldots, n$, $v_{i,t}, i = 0, \ldots, n$, generate a draw of the SV parameters from inverted Gamma distribution.

Forecasting proceeds as follows: Given generated $v_{i,t}, i = 0, \ldots, n$, generated SV values and generated $\tilde{y}_{i,t}, i = 1, \ldots, n$, use (7) to generate a one step forecast value y_{t+1}. Repeating this process gives a synthetic sample of future values and a forecast density at time $t + 1$.

3 Individual models and alternative combinations

Let S_t denote the spot price of crude oil at date t. Forecasts are obtained using a general stochastic volatility model with Student's t-distributed errors given as

$$S_{t+h|t} - \hat{S}_{t+h|t} = \epsilon_{t+h|t}, \quad \epsilon_{t+h|t} \sim T(\mu, e^{h_{t+h|t}}, \nu), \tag{9}$$

$$h_{t+h|t} = \mu + \phi(h_{t+h-1|t} - \mu) + \zeta_{t+h|t}, \quad \zeta_{t+h|t} \sim NID(0, \omega^2), \tag{10}$$

in which $|\phi| < 1$ and $\hat{S}_{t+h|t}$ denotes a point forecast of the real oil price, which is set equal to the conditional mean of the posterior predictive density. The model is estimated using the Metropolis-within-Gibbs sampling algorithm described in Chan et al. (2014), using 10000 draws from the posterior distribution after discarding the first 5000 draws as a burn-in.

Individual models Following Baumeister and Kilian (2015), the point forecasts of the real price of oil, $\hat{S}_{t+h|t}$, are obtained using six state-of-the-art oil price forecasting models. The names and acronyms are listed in Table 1. Next, we summarize the specifications.

Table 1: List of individual forecasting models and various forecast density combination approaches and their acronyms.

Model	Description
NC	No-change model
CRB	Changes in the price index of non-oil industrial raw materials
Futures	West Texas Intermediate (WTI) oil futures prices
Spread	Spread between the spot prices of gasoline and crude oil
TVspread	Time-varying parameter model of the gasoline and heating oil spreads
VAR	Vector autoregression
Equal	Equal weighted linear combinations of forecast densities
BMA	Bayesian Model Averaging
BMA2	Bayesian Model Averaging with a two-year rolling window
BPS	Bayesian Predictive Synthesis

The set of models starts with a no-change forecast, with acronym NC:

$$\hat{S}_{t+h|t} = S_t. \tag{11}$$

The second model includes the changes in the price index of non-oil industrial raw materials and is denoted by CRB:

$$\hat{S}_{t+h|t} = S_{t|t}(1 + \pi_t^{h,rm} - \mathbb{E}_t[\pi_{t+h}^{(h)}]). \tag{12}$$

in which $\pi_t^{h,rm}$ denotes the percent change of an index of the spot price of industrial raw materials (other than oil) over the preceding h months and is obtained from the Commodity Research Bureau (CRB), and $\pi_{t+h}^{(h)}$ denotes the expected rate of inflation over the next h-periods which is proxied by recursively constructed averages of past U.S. consumer price inflation data. This model is based on the intuition that there are broad-based predictable shifts in the demand for globally traded commodities.

The third model includes West Texas Intermediate (WTI) oil futures prices and is denoted by Futures:

$$\hat{S}_{t+h|t} = S_{t|t}(1 + f_t^{WTI,h} - s_t^{\mathrm{WTI}} - \mathbb{E}_t[\pi_{t+h}^{(h)}]), \tag{13}$$

in which $f_t^{WTI,h}$ is the log of the current WTI oil futures price for maturity h and s_t^{WTI} is the log the WTI spot price. This model reflects idea that many practitioners and policy institutions rely on the price of oil future contracts in generating forecasts of the oil price.

The fourth model includes the spread between spot prices of gasoline and crude oil and is denoted by Gasoline:

$$\hat{S}_{t+h|t} = S_{t|t} \exp(\hat{\beta}[s_t^{\text{gas}} - s_t^{\text{WTI}}] - \mathbb{E}_t[\pi_{t+h}^{(h)}]), \tag{14}$$

in which s_t^{WTI} is the log of the nominal U.S. spot price of gasoline. This model reflects the idea of many market practitioners believing that a rising spread between the price of gasoline and the price of oil signals upward pressure on the price of oil.

The fifth model is a time-varying parameter model of the gasoline and heating oil spreads and is denoted by TVSpread:

$$\hat{S}_{t+h|t} = S_{t|t} \exp(\hat{\beta}_{1,t}[s_t^{\text{gas}} - s_t^{\text{WTI}}] + \hat{\beta}_{2,t}[s_t^{\text{heat}} - s_t^{\text{WTI}}] - \mathbb{E}_t[\pi_{t+h}^{(h)}]), \tag{15}$$

in which s_t^{heat} is the log of the nominal U.S. spot price of heating oil which is obtained from the EIA and the time-varying parameters evolve according to a random walk with independent Gaussian white noise errors.

The sixth model is an oil market Vector Autoregressive Model and is denoted by VAR:

$$y_t = b + \sum_{i=1}^{p} B_i y_{t-i} + e_t, \tag{16}$$

where y_t is a 4×1 vector of variables including: the percent change in global crude oil production; global real economic activity index of Kilian (2009); the log of the real price of oil - $\hat{S}_{t+h|t} = \exp(\hat{y}_{3,t+h|t})$ and global above-ground crude oil inventories. This model can be viewed as the reduced-form representation of the global oil market structural VAR model developed by Kilian and Murphy (2014). Recently, Hamilton (2021) argued that an alternative measure, derived from world industrial production, is a better indicator of global real economic activity. Comparing various measures of global real economic activity, Baumeister et al. (2020) find that models based world industrial production or a common factor extracted from a panel of real commodity prices provide the best forecasts of the real price of oil. However, due to the lack of available real-time data vintages, we refrain from using these alternative measures of global economic conditions.

Combination models We use five of the individual models listed in Table 1 to compose model combinations: CRB, Futures, Spread, TVspread and VAR. The NC model is then used as a benchmark model to compare relative forecast performance of the various models.

In addition to the BPS model discussed in Section 2, we also consider Bayesian model averaging (BMA). BMA is a popular ensemble learning method that has been widely used within the econometrics literature (see, e.g. Aastveit et al. (2019) and references therein). When using BMA the individual forecast densities, $p(\tilde{y}_{t,h}|M_i, I_t)$, from model M_i are pooled into a combined posterior/forecast density, $p(\tilde{y}_{t,h}|I_t)$, given as

$$p(\tilde{y}_{t,h}|I_t) = \sum_{i=1}^{N} w_{i,t,h} p(\tilde{y}_{t,h}|M_i, I_t), \tag{17}$$

where the weights, $w_{i,t,h}$, are specified in one of two ways. In the first instance, following, among others, Amisano and Giacomini (2007); Hall and Mitchell (2007) and Jore et al. (2010), we use recursive weights based on the logarithmic score which take the form

$$w_{i,t,h} = \frac{\exp(\sum_{t=T_0}^{T-1-h} \ln p(\tilde{y}_{t,h}|M_i, I_t))}{\sum_{i=1}^{N} \exp(\sum_{t=T_0}^{T-1-h} \ln p(\tilde{y}_{t,h}|M_i, I_t))}, \tag{18}$$

in which T_0 denotes the start date of the forecast evaluation period and T denotes the end date of the period. In addition to this, we consider a version of BMA which uses a two-year rolling window when updating the weights.

Finally, we also consider equally weighted forecasts (equal) where we set the weight attached to each model to $w_{i,t,h} = 1/N$ in equation (17). In fact, such a simple combination of forecasts is commonly used, see e.g. Timmermann (2006), Stock and Watson (2006), Clark and McCracken (2010) and Baumeister and Kilian (2015), and is often found to outperform more sophisticated adaptive forecast combination methods.

4 Forecasting results

In this section we present results from a real-time, out-of-sample forecast study, in which we generate both point and density forecasts of the real price of oil in the global market for crude oil. Following Garratt et al. (2019) recent replication of Baumeister and Kilian (2015), we use real-time monthly data, where the real price of oil in the global market is approximated by deflating the U.S. refiners acquisition cost for crude oil imports (IRAC) by the seasonally adjusted U.S. consumer price index for all urban consumers (CPI). The data set also includes monthly real-time vintages of variables used for estimating the various individual models, such as, e.g., world oil production, oil inventories and the global real economic activity index. The full data cover the period 1973:01-2017:12. The initial forecasts discussed in Section 3 are estimated on data from 1973:01-1991:12 and forecasts are then made over the remaining data for the period 1992:012017:12 using real-time data vintages. When constructing the combinations, BPS requires an initial training data period which we set to 50 months. Since the first 24 months worth of forecasts account for differences in the forecast horizons, all forecasting models are evaluated on the same period of 1998:03-2017:12. Our objective is to forecast the final release of the real oil price data.

4.1 Typical data patterns of the real price of oil

We begin the analysis by examining the real price of oil in various transformations as shown in Figure 2. The shaded regions highlight various episodes of historical significance for the global market for crude oil: The 1979 oil crisis, the commencement of the IranIraq War in 1980, the disbandment of OPEC in 1985, the 1990/91 Persian Gulf War, the Asian Financial Crisis 1997/98, the oil price surge of mid 2003-08, the collapse of the oil price during the Great Recession and the oil price decline of mid-2014 to early 2015. The various transformations collectively highlight three typical data features in the period 1973-2018. First, the log-level series show substantial changes in the mean of the series which suggests that a time-varying autoregressive mean process may be beneficial. Second, the returns and squared returns series show volatility clustering suggesting that stochastic volatility is an important data feature to model. Third, using sub-period analysis, a changing mean and volatility pattern in the log-level series indicate that a time-invariant autoregressive mean model with SV may provide reasonably accurate forecasts over the initial data period, 1974-2002, as well as the periods 2010-2014 and 2015-2018. This confirms that in sub-periods stable patterns are present but periodic shocks at the mean level have occurred as indicated above. Further evidence of time-varying elements of the oil price distribution can be seen in Figure 3, which shows data distributions over the full data period and notable

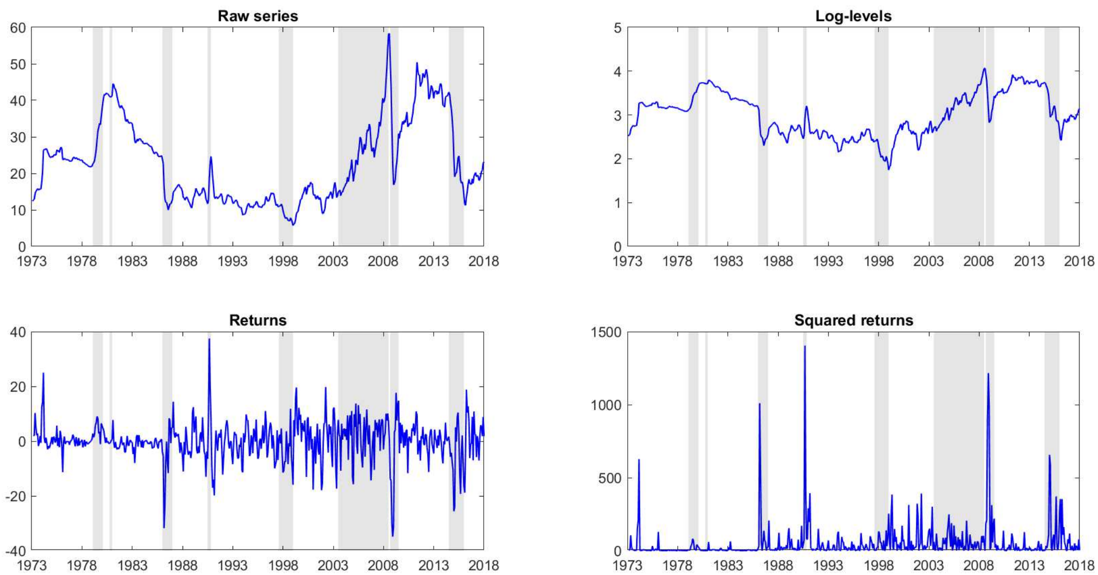

Figure 2: Real IRAC price of oil at monthly frequency over the period: 1973:01-2017:12.

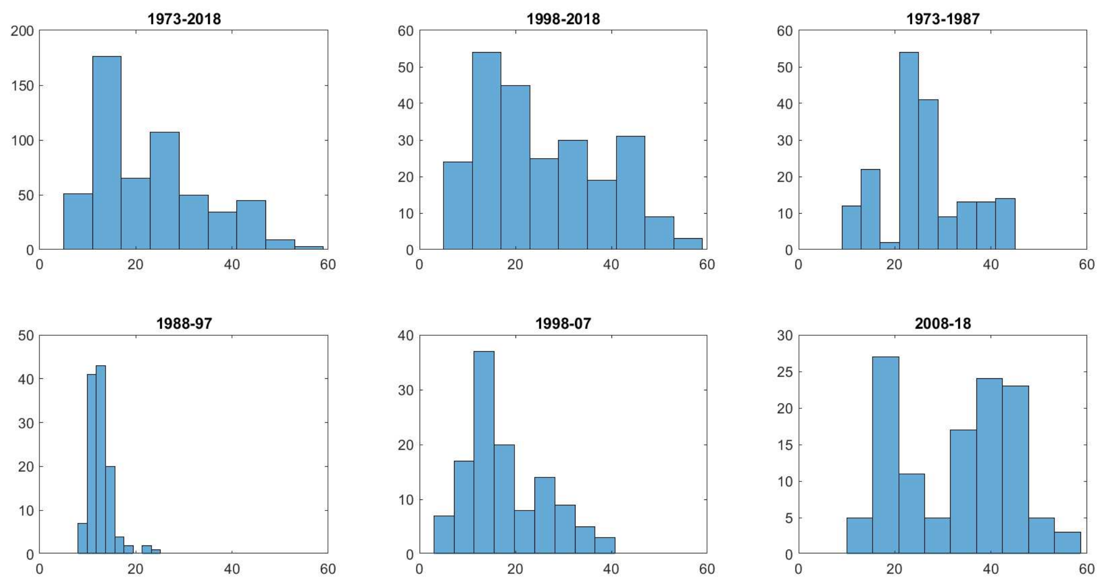

Figure 3: Distributions of the real IRAC price of oil in levels at monthly frequency over the period: 1973:01-2017:12, the forecast evaluation period: 1998:03-2017:12, and various sub-sample periods. The horizontal axis represents the real price of oil in USD and the vertical axis represents the pooled counts of observations in the respective bins.

sub-periods: The forecast evaluation period (1998-2018), a period of turmoil (1973-1987), a period of tranquility (1988-97), the oil price surge (1998-07) and the most recent decade (2008-18). In each case, the horizontal axis represents the level of the real price of oil in USD, as also shown in the top left panel in Figure 2, pooled into nine bins. The vertical axis represents the pooled count of observations over the respective (sub-)periods. Substantial time-variation in the shape of the data distributions is seen. It is noteworthy that asymmetry, fat tails and bi-modality are important features of the data. This suggests that models which allow for non-linearities in both mean and variance, as well as fat-tails, such as in our proposed BPS framework, may provide forecast improvements over simpler linear models, such as the suite of individual forecasting models, and the commonly used equal weight and BMA combination schemes discussed in Section 3. In the next section we discuss the accuracy of the estimated densities compared to the different data distributions.

4.2 Forecast accuracy of individual models and combinations

Forecast results across the evaluation period are provided in Table 2. The upper panel provides density forecast results evaluated by the log score relative to a no-change model benchmark. The lower panel provides point forecast results evaluated by the RMSFE relative to a no-change model benchmark. For interpretation purposes, log score values that are greater than zero indicate that the model outperforms the benchmark and vice versa. In contrast, RMSFE values that are less than one indicate that the model outperforms the benchmark and vice versa. To determine whether the forecast improvements or deteriorations are credibly different from zero, we report results from the Diebold-Mariano test with both 95% and 99% credible intervals. To determine the best model with a given level of credibility, we also report tail probabilities (p-values) for the model credible set (MCS)—a Bayesian interpretation of the model confidence set of Hansen et al. (2011)—in Table 3.

First focusing on density forecasts, we observe that the proposed BPS approach provides the best forecasts when forecasting the real price of oil beyond the immediate one-month-ahead forecast horizon, and that these improvements are credibly different from zero. The no-change model is difficult to beat when producing one month ahead density forecasts, however the performance difference between BPS and the no-change model is not credibly

Table 2: Density (Log Score) and point (RMFSE) forecast results relative to a no-change benchmark. Bold numbers indicate the best forecast performance at each horizon. One or two asterisks indicate that differences are, respectively, credibly different from zero according to the Diebold-Mariano test using 95% and 99% credible intervals.

Log Score									
Horizon	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0,46	**1,69***	-0,55	-3,62	-14,26	-298,95**	-297,04**	-91,44**	-13,23
6	-0,21	3,11	-5,10*	-8,14	-29,57**	-161,18**	-158,22**	-2,55*	**12,59****
12	-12,50	10,02*	-16,56**	-19,14**	-28,74**	-141,46**	-139,63**	25,85*	**47,73****
24	-32,48**	26,10**	-16,91**	-35,25**	2,34	-152,15**	-142,21**	59,14**	**110,96****

RMSFE									
Horizon	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0,95	0,99*	1,00	1,01	0,99	0,96*	0,96*	**0,90***	0,97
6	1,06*	0,97*	1,01	1,04	1,05*	0,99	0,99	0,96**	**0,89****
12	1,05	0,91**	1,01	1,02	1,04	0,96**	0,96**	0,88**	**0,71****
24	1,13**	0,89**	1,07	1,21**	1,01**	0,98	0,97**	0,78**	**0,57****

different from zero at this horizon. Interestingly, we find that the Futures model improves upon the no-change model at each forecast horizon, while the other individual models generally fail to outperform the benchmark; exceptions include the CRB model at the one-step-ahead horizon and the VAR at horizon 24. Due to aggregated forecast uncertainty, the equal weight and BMA combination methods do substantially worse than the no-change model at all horizons. By allowing for a more flexible weighting procedure the two-year rolling window BMA2 outperforms both equal weighted and expanding window BMA approaches, and also improves upon the no-change benchmark at longer horizons.

Shifting focus to point forecasts, we observe that the proposed BPS model provides substantial improvements over the no-change benchmark at all forecast horizons, and that these improvements are credibly different from zero beyond the one month horizon. Consistent with Baumeister and Kilian (2012), we observe that the commodity price-based model improves upon the no-change model at the one-month horizon but does worse at longer horizons. Also in line with Baumeister and Kilian (2015), we find that the equal weights combination model outperforms the no-change benchmark at all forecast horizons, and that the importance of futures-based information improves with the forecast horizon. In contrast to the density forecast results, all combination methods are found to improve upon the point forecast accuracy of the benchmark, however it is clear that specifying time-varying weights as in the BPS and BMA2 produces the largest gains. In line with the density forecast results, the point forecast accuracy of BMA is found to be similar to the equal weights model—a result that is commonly referred to as the "forecast combination puzzle" within the broader literature on model combinations of competing point forecasts. That being said, we find that by specifying more dynamic weighting schemes in BMA2 and learning weights in BPS, we are able to generate greater forecast accuracy at all horizons. Moreover, the proposed BPS approach provides substantial improvements beyond all individual and combination models. The size of these improvements is also increasing with the forecast horizon, in which oil prices are generally assumed to exhibit near random walk behavior. This suggests that the proposed BPS model may be particularly useful for practitioners who hedge against oil price risk. We further explore this in Section 5.

The usefulness of BPS for forecasting the real oil price is further supported by results in Table 3 which show that BPS is generally within the set of superior models, with one exception being the one-step-ahead density forecasts. The observation that combinations with equal and BMA weights are excluded from the set of superior models at all horizons for the density forecasts, while BMA2 and BPS are included, is particularly noteworthy. This emphasizes the importance of allowing for time-varying weights in combination models.

As a next step, we determine whether differences in forecast accuracy between models and model combinations hold throughout the time series periods for the different forecast horizons. To this end, we show the time patterns of cumulative Log Scores and RMSFEs relative to a no-change model benchmark in Figures 4 and 5 respectively. For interpretation purposes, values in Figure 4 that are greater than zero indicate that the model outperforms the benchmark and vice versa, while values in Figure 5 that are less than one indicate that the model outperforms the benchmark and vice versa.

The results in Figure 4 reveal considerable time variation in the relative performance

Table 3: Model credible set (MCS) tail probabilities (p-values) for density (Log Score) and point (RMSFE) forecasts. The MCS tail probabilities are computed with 100,000 block bootstrap replications using a block size of 10. Bold numbers indicate the highest ranked model at each horizon. One or two asterisks indicate that differences are in the 95% and 99% MCS, respectively.

Log Score										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.95**	**1.00****	0.95**	0.95**	0.21**	0.14**	0.00	0.00	0.01	0.01
6	0.36**	0.76**	0.76**	0.36**	0.36**	0.01	0.00	0.00	0.76**	**1.00****
12	0.00	0.00	0.09**	0.00	0.00	0.00	0.00	0.00	0.28**	**1.00****
24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	**1.00****

RMSFE										
Hor	NC	CRB	Futures	Spread	TVspread	VAR	Equal	BMA	BMA2	BPS
1	0.18**	0.19**	0.18**	0.18**	0.18**	0.18**	0.19**	0.19**	**1.00****	0.18**
6	0.10**	0.09**	0.12**	0.10**	0.09**	0.10**	0.12**	0.12**	0.12**	**1.00****
12	0.02*	0.03*	0.03*	0.02**	0.03*	0.03*	0.03*	0.03*	0.03*	**1.00****
24	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	**1.00****

of both individual and combination model specifications. In line with the full data period results, the proposed BPS model is competitive at the one-step horizon and outperforms the no-change benchmark, all individual models, and all combination models, at longer forecast horizons. In contrast, both equal weight and BMA models become progressively worse relative to the no-change benchmark, while the individual model specifications tend to cluster around similar values. Finally, while the recursive window BMA model (BMA2) does quite poorly when forecasting one-month-ahead, the forecast accuracy is competitive with the best individual model forecasts when forecasting six-months-ahead, and is second only to the BPS model at the longer 12 and 24 month horizons.

Turning our attention to the point forecast results, reported in Figure 5, we again find considerable time variation in the relative performance of each model specification. In line

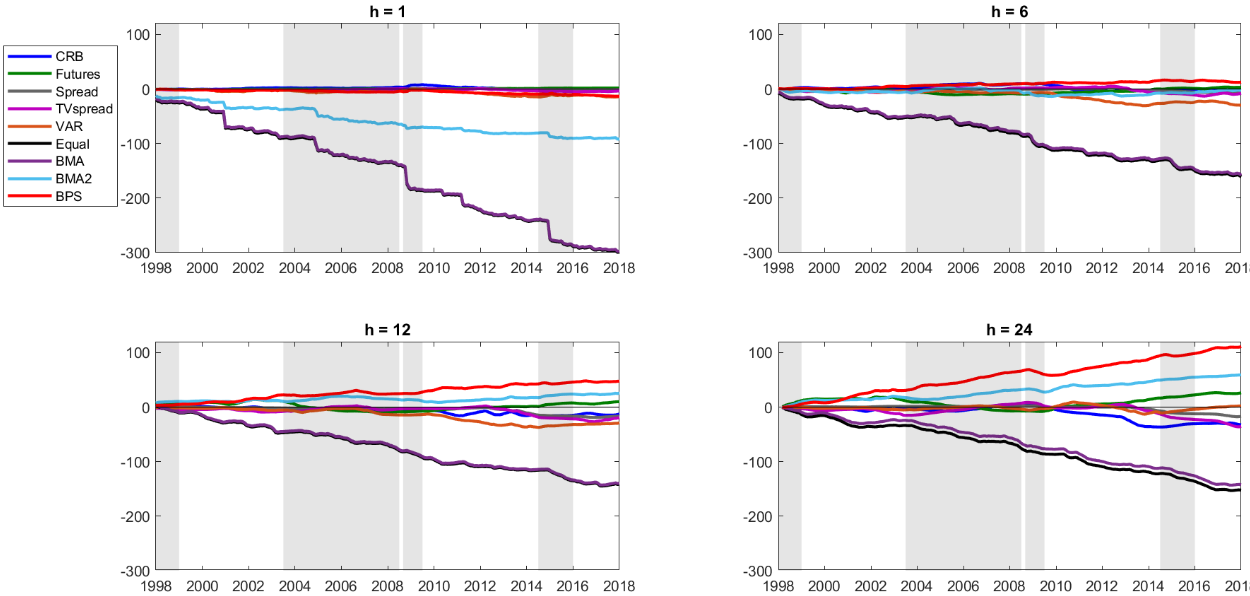

Figure 4: Time patterns of cumulative log scores relative to a no-change model benchmark over the forecast evaluation period: 1998:03-2017:12. Forecast horizons: 1, 6, 12 and 24 months ahead.

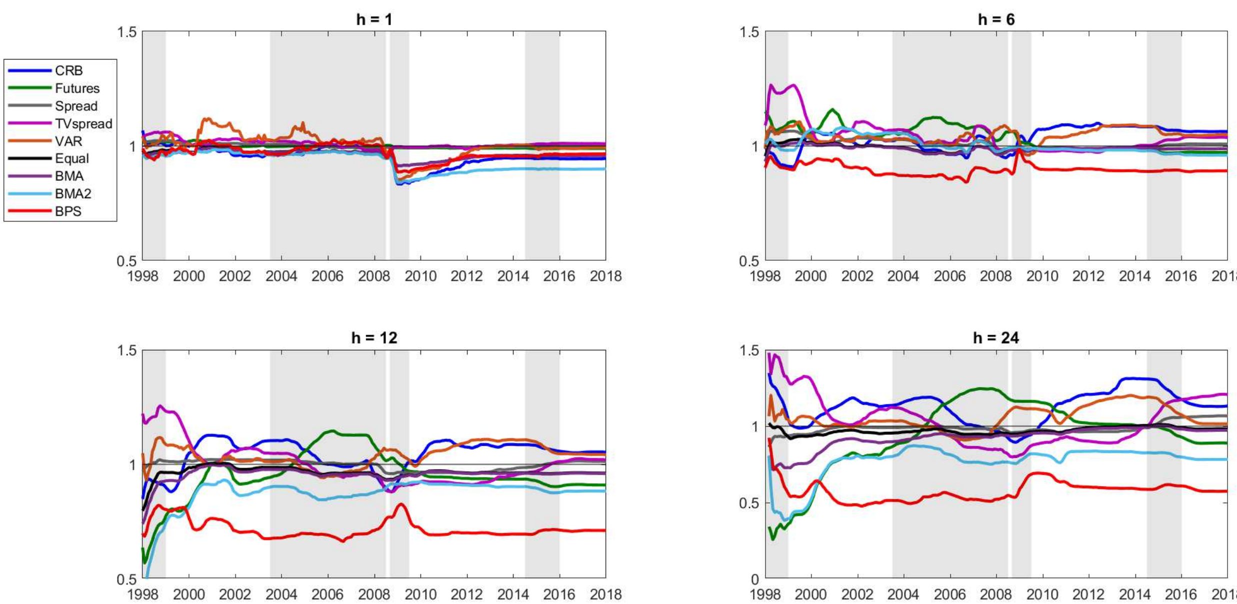

Figure 5: Time patterns of RMSFEs relative to a no-change model benchmark over the forecast evaluation period 1998:03-2017:12. Forecast horizons: 1, 6, 12 and 24 months ahead.

with the density forecast results, we find that the BPS model is competitive at the one-month horizon and provides substantial improvements at longer horizons. The fact that most of the models RMSFEs cluster around one at the one-month horizon suggests that the oil price exhibits near random walk behavior at this horizon. A notable exception occurs around the oil price collapse in 2009, during which the combination models provide notable improvements over the no-change benchmark, however these gains gradually dissipate over the next few years. In contrast to the relatively similar forecast performance at the one-month horizon, the longer horizons exhibit much more dispersion. For instance, at each horizon, the futures price model produces generally superior forecasts relative to the no-change benchmark with the exception of the early to mid 2000's. This is in line with existing results that the real price of oil between mid 2003-08 was driven by unexpectedly high growth mainly in emerging Asia (Aastveit et al., 2015). We also find that the VAR model performs well up until 2010, but then does relatively poorly over the subsequent period, thereby corroborating recent results in Baumeister et al. (2020).

In the Online Appendix, Table A1, we report results for absolute forecast accuracy by testing if the density forecasts are correctly calibrated over the entire forecast evaluation period, using the test in Knüppel (2015). We also report in Table A2 results from a two-sample Kolmogorov-Smirnov test between the empirical cumulative distribution function of the data and the cumulative distribution function of the BPS forecasts at each forecast horizon. The results from these tests suggest that the BPS forecasts are well calibrated and provide a good approximation of the data distribution at each of the forecast horizons.

4.3 Learning about time-varying combination weights and model diagnostics

An important feature of BPS is that it allows for time-varying individual model weights that adapt according to the recent relative forecast performance of each model within the model combination set. The means of the densities of the individual model weights are shown in Figure 6. The general observation across all forecast horizons is that while considerable time variation exists, there are similarities among models. First, focusing on the one-step-ahead results we observe that the mean weights for both the CRB and futures price models tend to follow a similar trajectory, while the Futures, Spread and TV spread models respectively follow a similar path that is distinct from the former trajectory. It is particularly notable that the mean weights for the two series in the former group abruptly increase during the two oil price collapses of 2009 and 2014 respectively, but gradually decline in the subsequent years surrounding these events. In contrast, the mean weights for the three series in the latter group sharply declined during the oil price collapse of 2009, but then gradually increase, with all mean weights sharing roughly similar time patterns by the end of the period. Interestingly, the same weighting clusters are not observed at longer horizons. At the 12-month-ahead horizon, we find that each of the mean weights tends to follow a similar trajectory. One notable exception is the abrupt increase in the mean weight of the Futures model following the 2014 oil price collapse.

It is also worth noting that the BPS mean weights are not restricted to be a convex combination in the unit interval—like most of the combination methods that are currently used within the econometrics literature—but are instead specified as a general linear combination and are thereby permitted to evolve along the real line. While this may have a disadvantage relative to the natural interpretation of mean weights within convex combinations as representing a probability distribution over different possible models, specifying a linear combination offers two practical advantages. In the first instance, allowing for both positive and negative mean weights means that it's possible to hedge against any potential forecast risk in recent periods. Second, a declining mean weight on one model does not necessarily imply an increasing mean weight on another model. Instead, mean weights of models are permitted to change in accordance to the forecasts of the individual models. Such features are clearly desirable to a wide range of practitioners. For instance, just as a financial portfolio manager hedges against risk by assigning a negative mean weight to an asset, the BPS approach is able to automatically assign a negative mean weight to mitigate the impact of forecast risk, such as forecast bias. Such natural behaviour is not possible under combination models in which the mean weights are restricted to be convex combinations, e.g. BMA or equal weighting methods.

Another important feature of BPS is that it has a built-in time-varying intercept that is absent from simpler combination methods such as BMA. It is well known within the

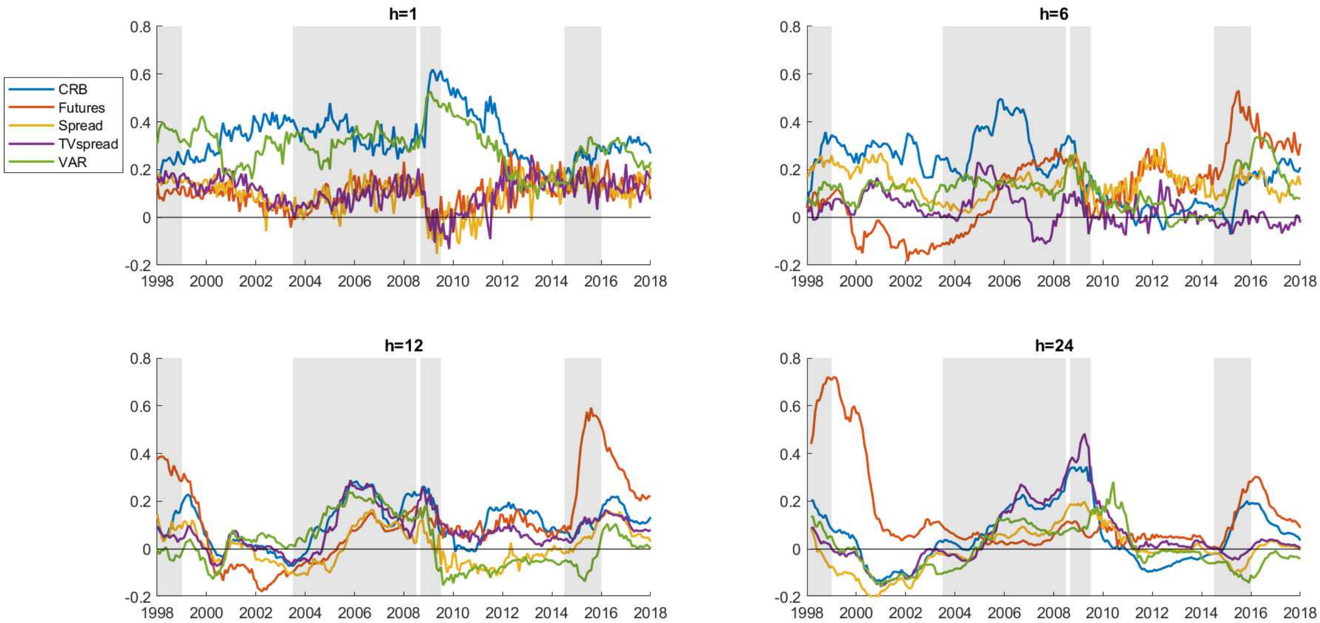

Figure 6: Time-varying posterior predictive mean of the individual model weights ($v_{i,t}$) in the BPS model, sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12.

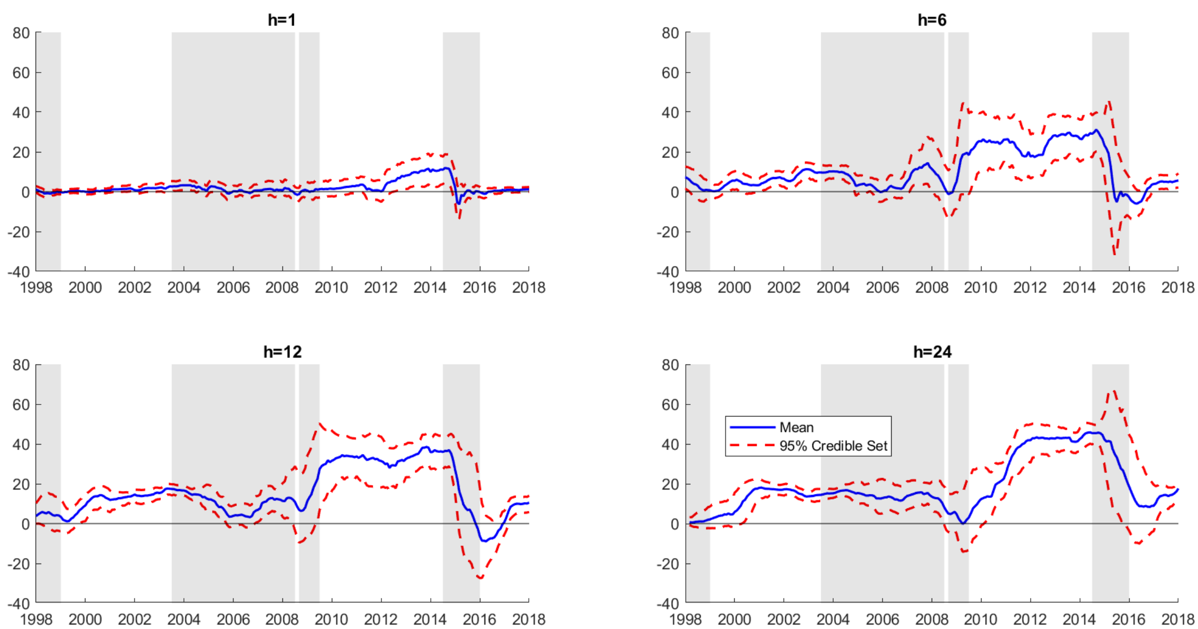

Figure 7: Time-varying posterior predictive mean of the intercept coefficient weight ($v_{0,t}$) in the BPS model, sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12.

econometrics literature on forecast combinations that BMA assumes that the true model is included in the model set. However, the model set could be misspecified due to incompleteness. By allowing for an intercept component that can adapt during episodes of low frequency signals from a set of forecasting models, BPS is able to better mitigate the effects of this problem. For instance, from the previous section, we know that there exists a model within the combination set–e.g. the VAR model–that provided superior one-step-ahead forecasts of the real price of oil up to and during the oil price drop of Great Recession relative to the no-change benchmark. After 2010, however, none of the models forecasted the oil price collapse of 2014. This suggests that there exists a degree of model set incompleteness since 2010, and we expect that this is reflected in the estimated BPS intercept for the one-step-ahead forecast. This is exactly what we observe in Figure 7 which shows the mean intercept terms at each forecast horizon. The one-step-ahead mean intercept is around zero up until 2010, when it starts to gradually increase before abruptly dropping and becoming negative during oil price collapse of 2014. Moreover, as shown previously in Figure 5, this feature allows BPS to improve upon the no-change benchmark despite the relatively weak signals stemming from the models during this period. Similar scenarios can be observed in the remaining estimates, which each exhibit a hump shaped response for the respective estimated mean intercepts over the period 2008 to 2014.

The final important feature of BPS is that it has built-in diagnostic information measures about forecast inaccuracy and/or model set incompleteness which is also absent from simpler combination methods such as BMA. We first present this diagnostic measure for σ_t^2 for the model set in Figure 8. It shows clearly that during three crisis periods this measure increases. In Figure A2 in the Online Appendix we also provide estimates of this diagnostic measure for each individual model, $\sigma_{i,t}^2$, within the BPS framework. It is seen from the longer term forecasts that none of the individual models is capable to accurately forecast crises, in particular the VAR does poorly in crisis periods estimated over short as well as long horizons. This type of diagnostic information gives important signals about specifying model and model set improvements. We leave this as a topic for further research.

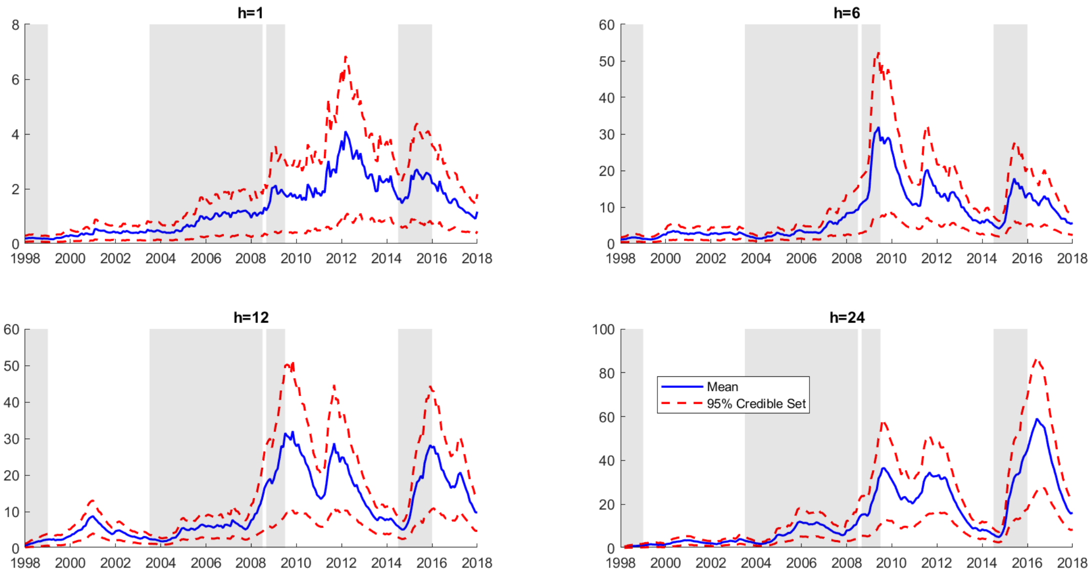

Figure 8: Time-varying variance, measured as the posterior predictive mean of the measurement variance, in the BPS model (σ_t^2), sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12. The red dotted line show the 95% credible bands.

4.4 Robustness Checks

In this Section we consider various robustness checks to our main forecasting exercise.

4.4.1 Alternative oil price series

We have focused on forecasting the IRAC price of crude oil, which is commonly viewed as a proxy for the global price of oil. Two alternative series that are frequently cited in the press are the Brent and West Texas Intermediate (WTI) prices of crude oil. We therefore repeated the main forecasting exercise using both of these series. Results in Tables A3 and A4 in the Online Appendix show that while some quantitative differences emerge, our qualitative conclusion that BPS provides the best forecast results at all but the one-step-ahead horizon remains robust to the choice of oil price series. Associated MCS and PITs tests Tables A5–A7 are also broadly consistent with those from the IRAC.

4.4.2 Alternative BPS specification

The BPS coefficients are able to simultaneously change over time and learn from previous performance. It is therefore natural to explore the practical significance of the random walk state equation in the main BPS specification. To this end, we redo the main forecasting exercise using an alternative specification in which we maintain the same stochastic volatility structure as in the BPS model, however the random walk component is shut off and the combination weights and intercept are instead estimated with standard linear regression techniques. Given the recursive nature of any forecasting exercise, this enables the combination weights to update over the forecast evaluation period, however there will be substantially less flexibility in the learning process. This alternative specification can be estimated with straight forward modifications of the equation system in equations (7)-(8) and textbook algorithms in West and Harrison (2006), Sect. 6.3.2 and Sect. 10.8.

In Table A8 in the Online Appendix we show the density and point forecast results using the alternative specification in which the combination weights and intercept are estimated with linear regression techniques. The main insight is that the alternative specification provides comparable results to the main BPS specification at the shorter horizons, however the main specification with combination weights and intercept with random walk learning provide superior results at longer horizons. This result highlights the importance of allowing for both flexible combination weights and intercept at longer forecast horizons.

4.4.3 Alternative Models

In our main analysis we have expanded on the empirical results in Baumeister and Kilian (2012, 2015) by investigating whether a combination forecast using BPS can outperform their six individual models and conventional combination methods. Extensive analysis in Alquist et al. (2013) also suggests that these models tend to produce better point forecasts than simple univariate time series models such as AR and ARMA models. That being said, forecasters may not necessarily use such six models in practice, and may opt for simpler regressions with alternative predictors such as exchange rates or interest rates. To conserve space we here present a concise discussion of the results and defer all tables with results to the Online Appendix.

With this in mind, we estimate an additional seven predictive regressions for which the selection of variables is motivated by the tests of Granger causality in Table 8.1 of Alquist et al. (2013), and include measures of exchange rates, interest rates, money and inflation. Each of these specifications take the same form as in equation (12), where we replace the CRB commodity price index with the alternative predictor. The main insight is that none of the additional models increase the forecast accuracy of the no-change model.

We have further investigated how adding some of these predictive models to the combination set effects the results from the various combination methods considered in our paper. The results, detailed in the Online Appendix, show that including these regression models in the combination set have very little effect on the BPS forecasting performance. In contrast, the alternative combination methods generally yield worse results. This is particularly the case for density forecasts for equal weights and BMA combinations.

Finally, to investigate the relative importance of time-varying combination weights and time-varying model parameters, we have estimated each of the predictive regressions with time-varying parameters via a random walk state equation. Overall, the results for the individual models improve upon the constant parameter regression models, particularly for density forecasts. This suggests that using time-varying parameter models is somewhat useful when forecasting the price of oil. To further explore this insight, we compute forecast combinations where we include TVP regression models in the combination set. We find that the forecasting accuracy from combination models with individual TVP regeression models are very similar to the ones of combination models with constant coefficient individual models. This indicates that it is more important to account for time-varying combination weights than individual time-varying parameters when forecasting the real price of oil.

5 Risk analysis

In this Section we analyse the risk and return properties of investing in the global market for crude oil using the BPS modelling approach as an investment tool. The means of the profit and loss distribution including 95% credibility regions associated with the forecasted spot prices from BPS are shown in Figure 9. For interpretation purposes, we highlight that positive mean values indicate a profit and negative mean values indicate a loss. There are no periods of profits or losses that are credibly different from zero, nor is there any observable serial correlation pattern across the entire data period, thus corroborating the widely held view that, in the short run, oil prices exhibit random walk behavior. That being said, there are periods in which BPS does well, and not so well. For instance, the means signify notable profits could have been made during the two oil price collapses of 2008-09 and 2014-15, however they then tend to gradually revert to zero and the credible set contains negative values.

Figure 9: Means of the profit and loss distribution (profit positive and loss negative) associated with the forecasts from the BPS model, sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12. The red dotted line show the 95% credible bands.

To quantify the measure of risk of loss associated with the profit and loss distributions, we next compute the value-at-risk (VaR). The VaR is widely used by regulators and practitioners in the financial industry to measure the quantity of assets needed to cover possible losses. The implied 1% and 5% VaR for the BPS model profit and loss distribution over the forecast evaluation period are shown in Figure A3 in the Online Appendix. The vertical axis are in percent. For interpretation purposes, this means that, for instance, a one-month 1% VaR of -2 means that there is a 1% chance of a 2% loss during the one-month period.

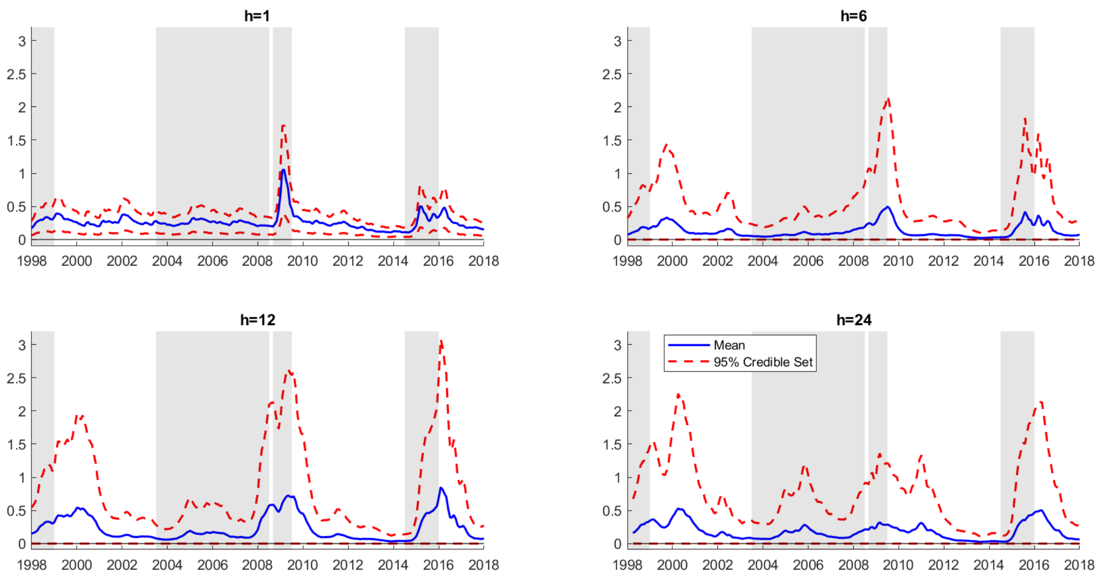

Figure 10: The minimum variance hedge ratio (MVH) from the BPS model, sequentially computed at each point in time over the forecast evaluation period 1998:03-2017:12. The red dotted line show the 95% credible bands.

Faced with such risks when operating in global oil markets, firms and portfolio managers naturally face the decision whether or not to hedge against unanticipated fluctuations in the price of oil. For instance, a petroleum company may wish to hedge its purchase price of crude oil by purchasing a futures contract. A widely used strategy for computing the optimal number of contracts needed to hedge a position is the ratio of the product of the optimal hedge ratio and the units of the position being hedged, to the size of a futures contract. The most common optimal hedge ratio is the Minimum Variance Hedge (MVH) ratio which aims to minimize the variance of the position's value. It is calculated as the product of (1) the correlation coefficient between the changes in the spot and futures prices, $\rho_{S,F}$, and (2) the ratio of the standard deviation of the changes in the spot price, σ_S, to the standard deviation of the futures price, σ_F.

Means of the MVH ratios, including 95% credibility regions, are shown for the forecasted spot price from the BPS model in Figure 10. Since the Brent price of crude oil is often used as a global price benchmark, we have used Brent futures data from 1992:1-2017:12 as provided by Garratt et al. (2019). The results show that the means of the optimal hedge ratio differ based on the forecast horizon. For instance, the mean of the MVH ratio tends to fluctuate between 0.2-0.4 at the one-step-ahead horizon, compared to 0-0.1 at the six-step-ahead horizon. That being said, at each horizon we observe notable spikes occur around the turn of the century as well as the two oil price collapses of 2009 and 2014.

6 Conclusion

Given the typical data pattern of the real price of oil over a substantial time period and some well-known models that describe these data, discussed in Section 3 and Section 4, we have successfully specified a basic probabilistic model structure and corresponding state space equation system based on the Bayesian Predictive Synthesis approach. Compared to more standard approaches like BMA, our BPS approach contains important extensions about diagnostic analysis of model set incompleteness and time-varying learning weights in the combination. This approach also leads to the use of numerically efficient Markov Chain methods in order to evaluate the Forecast Density Combination.

Applying a Bayesian procedure to estimate this model, we have obtained an extensive set of empirical results about time-varying forecast uncertainty and risk for the real price of oil over the period 1974-2018. This yielded substantial gains in forecast accuracy from point and, in particular, density forecasts using model combinations compared to individual models. These forecast gains are confirmed by exploring the estimated forecast time patterns, especially in the long term like 12-24 months, which is relevant information when forecasts are used for policy decisions. Dynamic patterns of the estimated individual model weights showed the relative contribution of individual models in the forecast combination. In addition, time patterns of diagnostic information about model incompleteness were obtained which give information on possible improvements about model specifications. We ended our analysis by showing results of time-varying risk in the oil price forecasts and presented a basic analysis of profit-loss and hedging against price risk.

The research presented can be extended in several directions and for classes of many economic data sets that are of interest for forecasters and policy makers. Exchange rate forecasting and risk analysis using sets of countries is an obvious example. Using micro-data to strengthen the information contained in macroeconomic forecasts and using large sets of finance data for dynamic portfolio analysis are further research topics with potential interesting policy implications.

We end with a remark on the possible connections between typical data patterns of economic variables of interest, the complexity of an FDC model and the class of Monte Carlo simulation algorithms which has to be used for the numerical evaluation of the densities involved. The literature of this field is extensive and still expanding; for a recent survey we refer to Aastveit et al. (2019). Which FDC approach is the most useful to apply depends on data patterns and model specification. This is an interesting topic of future research but beyond the scope of the present paper.

References

Aastveit, K., K. Gerdrup, A. Jore, and L. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics 32*(1), 48–68.

Aastveit, K. A., H. C. Bjørnland, and L. A. Thorsrud (2015). What drives oil prices? emerging versus developed economies. *Journal of Applied Econometrics 30*(7), 1013–1028.

Aastveit, K. A., J. Mitchell, F. Ravazzolo, and H. K. van Dijk (2019). The evolution of forecast density combinations in economics. In *Oxford Research Encyclopedia of Economics and Finance*.

Aastveit, K. A., F. Ravazzolo, and H. K. Van Dijk (2018). Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics 36*(1), 131–145.

Alquist, R. and L. Kilian (2010). What do we learn from the price of crude oil futures? *Journal of Applied econometrics 25*(4), 539–573.

Alquist, R., L. Kilian, and R. J. Vigfusson (2013). Forecasting the price of oil. In *Handbook of economic forecasting*, Volume 2, pp. 427–507. Elsevier.

Amisano, G. and R. Giacomini (2007). Comparing density foreasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

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

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

Baumeister, C. and L. Kilian (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., D. Korobilis, and T. K. Lee (2020). Energy markets and global economic conditions. *The Review of Economics and Statistics*.

Bernard, J.-T., L. Khalaf, M. Kichian, and C. Yelou (2018). Oil Price Forecasts For The Long Term: Expert Outlooks, Models, Or Both? *Macroeconomic Dynamics 22*(3), 581–599.

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

Casarin, R., S. Grassi, F. Ravazzolo, and H. K. van Dijk (2020). A bayesian dynamic compositional model for large density combinations in finance. Technical report.

Casarin, R., S. Grassi, F. Ravazzolo, and H. vanDijk (2015). Parallel sequential monte carlo for efficient density combination: The deco matlab toolbox. *Journal of Statistical Software (Online) 68*.

Chan, J. C., C. Y. Hsiao, et al. (2014). *Estimation of stochastic volatility models with heavy tails and serial dependence*. Wiley Online Library.

Clark, T. E. and M. W. McCracken (2010). Averaging forecasts from vars with uncertain instabilities. *Journal of Applied Econometrics 25*(1), 5–29.

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics 192*(2), 391–405.

Garratt, A., S. P. Vahey, and Y. Zhang (2019). Real-time forecast combinations for the oil price. *Journal of Applied Econometrics 34*(3), 456–462.

Genest, C. and M. J. Schervish (1985). Modelling expert judgements for Bayesian updating. *Annals of Statistics 13*, 1198–1212.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130 – 141.

Geweke, J. and G. G. Amisano (2012). Prediction with misspecified models. *The American Economic Review 102*, 482–486.

Gneiting, T. and R. Ranjan (2013). Combining predictive distributions. *Electron. J. Statist. 7*, 1747–1782.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

Hamilton, J. D. (1983). Oil and the macroeconomy since world war ii. *Journal of Political Economy 91*(2), 228–248.

Hamilton, J. D. (2009). Understanding Crude Oil Prices. *The Energy Journal 30*(2), 179–206.

Hamilton, J. D. (2021). Measuring global economic activity. *Journal of Applied Econometrics 36*(3), 293–303.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Hoogerheide, L., R. Kleijn, F. Ravazzolo, H. K. Van Dijk, and M. Verbeek (2010). Forecast accuracy and economic gains from bayesian model averaging using time-varying weights. *Journal of Forecasting 29*(1-2), 251–269.

Hoogerheide, L., R. Kleijn, R. Ravazzolo, H. K. van Dijk, and M. Verbeek (2010). Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights. *Journal of Forecasting 29*(1-2), 251–269.

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.

Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of Forecasting 29*(1-2), 231–250.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review 99*(3), 1053–69.

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

Knüppel, M. (2015). Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. *Journal of Business & Economic Statistics 33*(2), 270–281.

Koop, G. and D. Korobilis (2012). Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review 53*(3), 867–886.

Manescu, C. B. and I. V. Robays (2016). Forecasting the Brent Oil Price: Addressing Time-Variation in Forecast Performance. Technical report.

McAlinn, K., K. A. Aastveit, J. Nakajima, and M. West (2020). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association 115*(531), 1092–1110.

McAlinn, K. and M. West (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of Econometrics 210*(1), 155–169.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 221–247.

Pak, A. (2018). Predicting crude oil prices: Replication of the empirical results in what do we learn from the price of crude oil?. *Journal of Applied Econometrics 33*(1), 160–163.

Pettenuzzo, D. and F. Ravazzolo (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics 31*, 1312–1332.

Ravazzolo, F. and P. Rothman (2013). Oil and U.S. GDP: A Real-Time Out-of-Sample Examination. *Journal of Money, Credit and Banking 45*(2-3), 449–463.

Stock, J. H. and M. W. Watson (2006). *Forecasting with Many Predictors*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 10, pp. 515–554. Elsevier.

Takanashi, K. and K. McAlinn (2020). Predictive properties and minimaxity of bayesian predictive synthesis.

Terui, N. and H. K. Van Dijk (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting 18*(3), 421–438.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 136–96. Amsterdam: Elsevier.

West, M. (1984). Bayesian aggregation. *Journal of the Royal Statistical Society: Series A (General) 147*(4), 600–607.

West, M. (1988). Modelling expert opinion (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 493–508. Oxford University Press.

West, M. (1992). Modelling agent forecast distributions. *Journal of the Royal Statistical Society (Series B: Methodological) 54*, 553–567.

West, M. and J. Crosse (1992). Modelling of probabilistic agent opinion. *Journal of the Royal Statistical Society (Series B: Methodological) 54*, 285–299.

West, M. and J. Harrison (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.