NBER WORKING PAPER SERIES

RISKY OIL: IT'S ALL IN THE TAILS

Christiane Baumeister
Florian Huber
Massimiliano Marcellino

Working Paper 32524
http://www.nber.org/papers/w32524

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

We thank Jim Hamilton for useful suggestions, and participants at the IWEEE 2024 Workshop and the 2024 IIASA Workshop on Agricultural Commodity Prices for helpful comments. Marcellino thanks MUR- Prin 2022- Prot. 20227YZ9JK, financed by the European Union - Next Generation EU, for partial financial support. Huber gratefully acknowledges funding from the Austrian Science Fund: ZK-35. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Risky Oil: It's All in the Tails
Christiane Baumeister, Florian Huber, and Massimiliano Marcellino
NBER Working Paper No. 32524
May 2024
JEL No. C11,C32,C53,Q41,Q47

ABSTRACT

The substantial fluctuations in oil prices in the wake of the COVID-19 pandemic and the Russian invasion of Ukraine have highlighted the importance of tail events in the global market for crude oil which call for careful risk assessment. In this paper we focus on forecasting tail risks in the oil market by setting up a general empirical framework that allows for flexible predictive distributions of oil prices that can depart from normality. This model, based on Bayesian additive regression trees, remains agnostic on the functional form of the conditional mean relations and assumes that the shocks are driven by a stochastic volatility model. We show that our nonparametric approach improves in terms of tail forecasts upon three competing models: quantile regressions commonly used for studying tail events, the Bayesian VAR with stochastic volatility, and the simple random walk. We illustrate the practical relevance of our new approach by tracking the evolution of predictive densities during three recent economic and geopolitical crisis episodes, by developing consumer and producer distress indices that signal the build-up of upside and downside price risk, and by conducting a risk scenario analysis for 2024.

Christiane Baumeister
Department of Economics
University of Notre Dame
3028 Jenkins Nanovic Hall
Notre Dame, IN 46556
and CEPR
and also NBER
cbaumeis@nd.edu

Florian Huber
University of Salzburg
florian.huber@sbg.ac.at

Massimiliano Marcellino
IGIER
Universita Bocconi
Via Roentgen 1
20136 Milan
Italy
Massimiliano.marcellino@unibocconi.it

1 Introduction

The recent large swings in global oil prices associated with the COVID-19 pandemic in 2020, the Russian invasion of Ukraine in 2022, and the Israel-Hamas war in 2023, have emphasized the need to account for tail events when forecasting oil prices. Estimating tail risks is particularly important in the context of global energy markets because of the widespread ramifications that extreme price realizations have for the macroeconomy. In fact, many stakeholders including policymakers, government agencies, business people, industry, and consumers worry about tail outcomes of oil prices since they might adversely affect their business operations, investment and spending decisions, overall economic performance and welfare.

Most of the academic literature on oil price forecasting has focused on forecasting the most likely future value of the real price of oil without paying much attention to the risks surrounding these out-of-sample predictions, which is however key to decision-making under uncertainty.[1] One notable exception is Aastveit, Cross, and van Dijk (2023) who quantify risks in oil price forecasts by combining predictive distributions obtained from a set of individual models using time-varying weights. Another recent contribution that goes beyond point forecasts and uses density forecasts to derive measures of expected oil price risks is Baumeister, Korobilis, and Lee (2022). While these studies take forecast uncertainty more seriously and provide some probabilistic assessment of future oil price risks, they make no attempt to model the tail behavior of oil prices. In fact, many existing studies rely on standard linear models such as vector autoregressions (VARs) that lack the flexibility necessary to capture sudden breaks in mean relations, asymmetries, or heavy tails in predictive densities of oil prices.

Unlike Hamilton (2003) who proposed a flexible approach to model nonlinearities in the predictive relationship of oil price changes on economic activity, our goal is to generate forecasts of the real price of oil itself using nonlinear models to accommodate tail events. This is challenging because fewer observations tend to fall in the tails and these observations are, by definition, extreme. Most nonlinear modeling approaches also require taking a stand on the functional form nonlinearities might take, and this might lead to misspecification if the data-generating process differs from the assumed type of nonlinearity (see Hamilton, 2003). This could translate into a deterioration in forecasting accuracy, giving rise to a situation where linear models, being more parsimonious, perform better than more flexible models.

We consider three model classes that have been shown to produce useful tail forecasts in other settings. One model class that has become very popular recently for assessing and predicting financial and macroeconomic tail risks (see, e.g., Giglio, Kelly, and Pruitt, 2016; Adrian, Boyarchenko, and Giannone, 2019) is quantile regressions (QR).[2] This semiparametric method, originally proposed by Koenker and Bassett (1978), targets the tails explicitly. If interest extends

[1] See, for example, Baumeister and Kilian (2012, 2015), Alquist et al. (2013), Bernard et al. (2018), Garratt et al. (2019), among others.

[2] See also Carriero, Clark, and Marcellino (2022) for the tail nowcasting performance of quantile regressions.

beyond evaluating tail risks to computing functions of the predictive distribution (such as the probability that the oil price falls below a certain threshold in the near future), with QRs one needs to rely on various approximations to recover the full predictive distribution from the quantiles. An alternative approach to modeling tail behavior that does not require approximating the predictive distribution introduces a time-varying volatility process into an otherwise standard linear VAR model. This feature allows the model to adapt, at least in part, to extreme events by attributing them to increases in shock volatility (see, e.g., Carriero, Clark, Marcellino, and Mertens, 2022). Yet, parametric VARs assume that the conditional mean relations are linear and time-invariant. A third model class are nonparametric VARs that remain agnostic about the presence and type of nonlinearity, by assuming that the relationship between the target and predictor variables is unknown and possibly highly nonlinear. Bayesian Additive Regression Trees (BART) are used to approximate this unknown function (see Chipman, George, and McCulloch, 2010), leading to the BART-VAR originally developed in Huber and Rossini (2022). This model comprises both a linear and a nonparametric component in such a way that if there is strong evidence that a linear model is sufficient, this is reflected in estimation; by contrast, if dynamics are far from linear, the nonparametric specification captures this nonlinearity, leading to non-Gaussian properties in the predictive distribution. To further increase flexibility, we assume that the shocks feature time-varying volatility, as in the parametric VAR, which we describe in both cases as a factor stochastic volatility model (see, e.g., Pitt and Shephard, 1999; Aguilar and West, 2000).

We compare how well these three model classes perform when it comes to predicting the tail behavior of oil prices based on a set of traditional and novel predictor variables that capture different sources of risk. Using final revised data, we show that the nonparametric VAR model displays the best tail forecasting performance overall, but the parametric VAR model comes in as a close second. Both models produce better tail forecasts relative to the no-change benchmark, with gains increasing substantially as the forecast horizon lenghtens, especially in the right tail of the predictive distribution. Quantile regressions perform rather poorly and almost never beat the no-change forecast except for a few instances at longer horizons; however, they are clearly inferior to the other two models.

After having established that BART-VAR with stochastic volatility delivers the most accurate tail forecasts on average over a long evaluation period, we track its performance in real time over the recent turbulent period from January 2020 to January 2024 to explore the properties of the predictive distributions. We confirm that this model captures extreme events reasonably well by imparting departures from Gaussianity, in particular long tails. To illustrate how these predictive distributions can be used to inform stakeholders about impending oil price risks, we develop real-time indicators of the probability of extreme oil price realizations tailored to agents on both sides of the market. Specifically, we propose both a producer and a consumer distress index that capture downside and upside price risks. We also look ahead and consider several scenarios related to economic weakness and physical oil market tightness as they might unfold over the course of 2024 and study their implications for future price distributions and risk assessment.

The plan of the paper is as follows. Section 2 presents a unified treatment of three classes of models that are popular for modeling tail behavior of macroeconomic and financial variables with a special focus on the sources of nonlinearities. Section 3 discusses the selection of predictor variables that are economically linked to either the demand or the supply side of the global oil market and incorporate various price risks. We also provide preliminary in-sample evidence on the existence of nonlinear relationships between the real price of oil and its determinants. Section 4 examines the forecasting performance of our three competing models, especially with regard to their tail forecast accuracy. Section 5 conducts a real-time out-of-sample evaluation of tail-risk prediction of oil prices during three recent crisis periods in oil markets. We also develop quantitative tools for characterizing tail risks that either consumers or producers are exposed to and study the potential consequences of a diverse set of risk scenarios for 2024. Section 6 offers some concluding remarks.

2 Alternative Approaches to Modeling Tail Risks in Oil Price Dynamics

We model the joint evolution of the real price of oil and its determinants using multivariate time series models of the generic form

$$\boldsymbol{y}_{t+1} = F(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}_{t+1}, \qquad \boldsymbol{\varepsilon}_{t+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t+1}) \tag{1}$$

where $\boldsymbol{y}_{t+1}$ denotes an M-dimensional vector including the real price of oil and possibly a set of oil market fundamentals to be discussed in the next section. The explanatory variables are stored in a K-dimensional vector $\boldsymbol{x}_t = (\boldsymbol{y}_t', ..., \boldsymbol{y}_{t-p+1}')'$ that includes p lags of $\boldsymbol{y}_{t+1}$ so that $K = pM$.[3] The function $F(\boldsymbol{x}_t) = (f_1(\boldsymbol{x}_t), ..., f_M(\boldsymbol{x}_t))'$ is comprised of a set of equation-specific sub-functions $f_j(\boldsymbol{x}_t)$ that take $\boldsymbol{x}_t \in \mathbb{R}^K$ as input and return the conditional mean of $y_{j,t+1}$, the j^{th} element of $\boldsymbol{y}_{t+1}$ for $j = 1, \ldots, M$. The shocks in $\boldsymbol{\varepsilon}_{t+1}$ are normally distributed and feature a possibly time-varying covariance matrix $\boldsymbol{\Sigma}_{t+1}$. The model in (1) is flexible as it effectively allows for different functional relationships across the elements in $\boldsymbol{y}_{t+1}$ and $\boldsymbol{x}_t$ and nests all model specifications considered in our empirical application as special cases as we will discuss next.

2.1 Bayesian Vector Autoregressions

If we set $f_j(\boldsymbol{x}_t) = \boldsymbol{a}_j'\boldsymbol{x}_t$ where $\boldsymbol{a}_j$ is a K-dimensional vector of regression coefficients, that is if we assume linearity for all j, we can stack the equation-specific coefficients to obtain an $M \times K$ matrix $\boldsymbol{A}$ whose j^{th} row contains $\boldsymbol{a}_j'$. Under the assumption that $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}$, this yields a standard constant-coefficient vector autoregressive (VAR) model that has been shown to work well for forecasting oil prices at short horizons (see, e.g., Baumeister and Kilian, 2012, 2014a; Alquist et al., 2013; Garrett et al., 2019; Baumeister et al., 2022). Instead, the forecasting performance of small-scale

[3]In line with Baumeister et al. (2022), we set $p = 12$ for the estimation of our monthly oil market models.

oil market VAR models that allow for structural changes in the conditional mean relations by setting $f_j(\boldsymbol{x}_t) = \boldsymbol{a}'_{jt}\boldsymbol{x}_t$ with $\boldsymbol{a}_{jt} \sim \mathcal{N}(\boldsymbol{a}_{jt-1}, \boldsymbol{V}_j)$, where $\boldsymbol{V}_j$ is a state innovation variance-covariance matrix that controls the intensity of parameter changes, deteriorates considerably (see, e.g., Baumeister and Kilian, 2014a).

To the extent that nonlinearities in the dynamic relationships across variables are associated with low-frequency changes in volatility, introducing time variation in error variances not only allows for possible heteroskedasticity of the shocks but also produces interaction effects with conditional means via a better characterization of the predictive densities (see, e.g., Primiceri, 2005; Clark and Ravazzolo, 2015; Carriero, Clark, and Marcellino, 2019). Baumeister, Korobilis, and Lee (2022) show that allowing for stochastic volatility leads to substantial improvements in the accuracy of point and density forecasts of the real price of oil, particularly at longer horizons.

Given this evidence, the oil market model we consider for the parametric class features $F(\boldsymbol{x}_t) = \boldsymbol{A}\boldsymbol{x}_t$ where $\boldsymbol{x}_t$ includes a constant term and $\boldsymbol{\Sigma}_{t+1}$ which accounts for volatility shifts in energy markets and is estimated with Bayesian methods. For the VAR coefficients contained in $\boldsymbol{A}$, we use a Minnesota-type prior as in Baumeister et al. (2022); the specification for stochastic volatility is described in Section 2.4. This setup is in line with a recent contribution by Carriero et al. (forthcoming) who conclude that Bayesian VARs with stochastic volatility are flexible enough to generate asymmetries in forecast distributions and to capture tail risks in macroeconomic outcomes.

2.2 Bayesian Quantile Regressions

A more direct approach that has been widely used to forecast upside and downside risks in macroeconomic and financial variables relies on quantile regressions. We set $M = 1$ which means that equation (1) simplifies to a univariate model as is standard in quantile regressions (see, e.g., Adrian et al., 2019) and $F(\boldsymbol{x}_t) = \boldsymbol{\beta}_\tau^{(h)\prime}\boldsymbol{x}_t$ where $\boldsymbol{\beta}_\tau^{(h)}$ is a horizon- and quantile-specific K-dimensional vector of regression coefficients with $h = 1, ..., H$ being the forecast horizon and $\tau \in \{0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95\}$ the quantile under consideration. Since we estimate the regression model using a direct multi-step form, the right-hand side variable is y_{t+h}. The vector of predictor variables $\boldsymbol{x}_t$ contains a constant and the parameter vector $\boldsymbol{\beta}_\tau^{(h)}$ is obtained with quantile regressions:

$$y_{t+h} = \boldsymbol{\beta}_\tau^{(h)\prime}\boldsymbol{x}_t + \nu_{\tau,t+h}$$

where $\nu_{\tau,t+h}$ denotes an error term which follows a distribution that is restricted so that the τ^{th} quantile equals zero. Kozumi and Kobayashi (2011) use an asymmetric Laplace distribution (ALD) centered on zero with scale parameter $\sigma_\tau > 0$ for quantile τ which implies that the variance is allowed to change across quantiles. They approximate the ALD using a location-scale mixture of Gaussians of the form:

$$\nu_{\tau,t+h} = \theta_1 z_{\tau,t+h} + \theta_2\sqrt{\sigma_{\tau,h} z_{\tau,t+h}}\, u_{\tau,t+h}$$

where θ_1 and θ_2 are deterministic functions of the quantile τ, $z_{\tau,t+h} \sim \text{Exp}(1)$ and $u_{\tau,t+h} \sim \mathcal{N}(0,1)$.[4] Conditional on this approximation, the quantile regression is a standard linear regression model which we estimate using Bayesian techniques. To the best of our knowledge, this type of semiparametric model has not yet been used for forecasting oil prices.

2.3 Bayesian Additive Regression Trees

To gain additional flexibility in modeling the tail behavior of oil prices, we augment the linear VAR model with a nonparametric component. Specifically, we postulate $F(\boldsymbol{x}_t) = \boldsymbol{A}\boldsymbol{x}_t + G(\boldsymbol{x}_t)$ where $G(\boldsymbol{x}_t) = (g_1(\boldsymbol{x}_t), ..., g_M(\boldsymbol{x}_t))'$ is an unknown function with M equation-specific functions that are possibly nonlinear. Instead of imposing a specific functional form on G, which would require us to take a stand on the type of nonlinearity, our aim is to learn about G in a data-driven manner. This is achieved through Bayesian additive regression trees (BART), a technique commmonly applied in machine learning, which uses a sum of tree-based models to approximate arbitrary conditional mean functions, as originally proposed by Chipman, George, and McCulloch (2010). Formally, we use BART to approximate the equation-specific functions in G as follows:

$$g_j(\boldsymbol{x}_t) \approx \sum_{s=1}^{S} \mathfrak{t}(\boldsymbol{x}_t | \mathcal{T}_s, \boldsymbol{\mu}_s) \tag{2}$$

where S is an integer that denotes the number of trees.[5] The regression tree function $\mathfrak{t}$ takes two inputs: a tree structure $\mathcal{T}_s$ and a terminal node parameter $\boldsymbol{\mu}_s$ of dimension N_s. The tree $\mathcal{T}_s$ is composed of branches of binary decision rules of the form $\{x_{j,t} > c\}$ or $\{x_{j,t} \leq c\}$ where c is a threshold taken from the range of observed values $\{x_{j,t}\}_{t=2}^{T}$ and $x_{j,t}$ is the so-called splitting variable. A sequence of decision rules decomposes the input space into N_s disjoint sets, each of which is associated with a particular value of the terminal node parameters μ_{is} for $i = 1, ..., R_s$ with R_s indicating the number of regimes.[6]

The evolution of time series tends not to be well captured by simple trees (which essentially resemble step functions) given their limited representation flexibility. The key idea of BART is to sum over many of these simple trees, each of which explains only a fraction of the variation in $\boldsymbol{y}_{t+1}$ and thus acts as a "weak learner" in machine-learning terminology, to attain a high level of flexibility so as to characterize rich dynamics in $\boldsymbol{y}_{t+1}$. While one could grow more complex trees by encouraging the models to put more weight on model fit, this strategy would lead to overfitting and poor out-of-sample forecasting performance (see, e.g., Chipman et al., 2010).

The advantage of this class of nonparametric models is that BART automatically adjusts to

[4] Specifically, $\theta_1 = \frac{1-2\tau}{\tau(1-\tau)}$ and $\theta_2 = \frac{2}{\tau(1-\tau)}$.

[5] Based on different datasets of macroeconomic variables, Chipman et al. (2010) and Huber et al. (2023) show that the forecast accuracy of BART improves with the number of trees until $S \approx 50$. We follow the standard choice of Chipman et al. (2010) and set $S = 250$ in our empirical model.

[6] Appendix A and Figure A1 provide a numerical and graphical illustration of what such an estimated tree could look like for our oil market model.

the unique features of a given dataset relying only on a minimal set of assumptions. The fact that the number of parameters is not imposed a priori allows the parameter space to grow with the number of observations, affording the model additional flexibility.

2.4 Capturing Heteroskedasticity

Another source of flexibility that has the potential to enhance the forecasting performance is modeling time variation in the volatilities of shocks to $\boldsymbol{y}_{t+1}$ (see, e.g., Clark, 2011; Clark and Ravazzolo, 2015; Carriero, Clark, and Marcellino, 2016). We capture heteroskedasticity through the following specification for the reduced-form shocks $\boldsymbol{\varepsilon}_{t+1}$:

$$\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\Lambda} \boldsymbol{q}_{t+1} + \boldsymbol{\eta}_{t+1}, \quad \boldsymbol{q}_{t+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}_{t+1}), \quad \boldsymbol{\eta}_{t+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{t+1})$$

with $\boldsymbol{\Lambda}$ denoting an $M \times Q$ matrix of factor loadings where Q is a small number of factors in $\boldsymbol{q}_{t+1}$. These factors are conditionally heteroskedastic with a time-varying diagonal covariance matrix $\boldsymbol{H}_{t+1} = \text{diag}(e^{h_{1,t+1}}, ..., e^{h_{q,t+1}})$ and $\boldsymbol{\eta}_{t+1}$ are measurement errors with diagonal covariance matrix $\boldsymbol{\Omega}_{t+1} = \text{diag}(e^{h_{q+1,t+1}}, ..., e^{h_{q+M,t+1}})$. The latent logarithmic volatilities h_{it+1} evolve according to $Q + M$ independent $AR(1)$ processes:

$$h_{it+1} = \mu_{hi} + \rho_{hi}(h_{it} - \mu_{hi}) + \sigma_{hi} u_{it+1}, \quad u_{it+1} \sim \mathcal{N}(0,1) \quad \text{for } i = 1, ..., Q + M,$$

with μ_{hi} the unconditional mean of the log volatilities, ρ_{hi} their persistence parameter, and σ_{hi}^2 the variance. Thus, the variance of $\boldsymbol{\varepsilon}_{t+1}$ can be written as $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Lambda} \boldsymbol{H}_{t+1} \boldsymbol{\Lambda} + \boldsymbol{\Omega}_{t+1}$ which yields the factor stochastic volatility (FSV) model proposed in Pitt and Shephard (1999) and Aguilar and West (2000).[7] The key feature of this model is that it allows for common co-volatility between the elements in $\boldsymbol{y}_{t+1}$ through the factor and the factor variances. Hence, if there are periods which imply an increase in the volatility of all series, the common factor structure captures this. By contrast, idiosyncratic movements in volatility are reflected in increases in the measurement error variances. Overall, this model allows for rich volatility dynamics.

We use the same FSV specification in both the BVAR and BART models. This not only ensures that the two model classes are treated symmetrically, but this modeling choice is also preferable to the volatility specification used in Baumeister et al. (2022) for at least two reasons. First, the FSV model is order-invariant, and second, the FSV model is more parsimonious given that the factor structure reduces the number of free parameters substantially if $Q < M$.[8] As for our third model class, there is no need to augment quantile regressions with stochastic volatility because they already allow for heteroskedasticity by estimating variance parameters that are specific to a

[7] Without further restrictions, $\boldsymbol{\Lambda}$ and $\boldsymbol{q}_{t+1}$ are not separately identifiable; however, given that we are only interested in $\boldsymbol{\Sigma}_{t+1}$ for the purpose of tail forecasting, this is not an issue.

[8] In our empirical model $M = 7$ and we set $Q = 3$ which is motivated by the Ledermann bound. The Ledermann bound is the largest number of factors that allows for a unique decomposition of the covariance matrix $\boldsymbol{\Sigma}_{t+1}$ and is obtained as the solution Q^* for $(M - Q)^2 \geq M + Q$.

particular quantile τ, as discussed above. Moreover, the location-scale mixture representation to approximate the ALD implies that the resulting Gaussian shocks are heteroskedastic. Adding SV would, in principle, only introduce persistence in the shock volatility. To avoid identifiability issues, we stick to the standard Bayesian quantile regression.

For details on the choices of priors and hyperparameters for all three models and Bayesian estimation and inference, see Appendix B.

3 The Role of Nonlinearities in Oil Markets

The selection of our predictor variables is guided by their relationship with the real price of oil based on economic fundamentals and potential sources of risk that extreme realizations of these variables might entail for oil price fluctuations. In Section 3.1, we discuss the choice of variables motivated by economic and risk considerations, and in Section 3.2, we provide some preliminary evidence on the nonlinear nature of the interplay between the real oil price and its fundamental drivers.

3.1 A Comprehensive Set of Economically-Motivated Predictors

Our variable of interest is the real price of Brent crude oil which is the global reference for oil and petroleum product pricing and is widely followed by policymakers and the press. We model the price of Brent, deflated by the US consumer price index (CPI), as being dynamically determined by a set of six predictor variables that are economically linked to either the demand or the supply side of the global oil market.

A first relevant predictor that measures the quantity of petroleum products demanded is global fuel consumption, which in turn determines the demand for crude oil by refineries and thus has a direct impact on the price formation in the oil market. This variable has been shown to yield more accurate oil price forecasts than world oil production, which is the corresponding quantity measure on the supply side (see Baumeister, Korobilis, and Lee, 2022). Another predictor variable that encapsulates the idea of derived demand is the price differential between gasoline and Brent, often referred to as crack spread, since it approximates the profit margin an oil refinery can expect to earn from transforming crude oil into petroleum products. A widening spread between the price of gasoline and the price of Brent incentivizes refineries to produce more and thus signals upward pressure on the price of oil. Baumeister, Kilian, and Zhou (2018) provide evidence that the gasoline-Brent spread has predictive power for the future path of the real price of oil. In the case of both predictor variables, changes in final product demand may be induced by business cycle fluctuations, speculation about peak oil, or shifts in investor beliefs or consumer preferences. Oil market tightness arising from exceedingly strong consumption growth or refinery shutdowns might affect oil prices nonlinearly.

We consider petroleum inventories in OECD countries as another key predictor of oil prices given that they capture the storability of crude oil and refined products which serves as a buffer on both sides of the market. Alquist et al. (2013) and Baumeister, Guérin, and Kilian (2015) report some forecasting success for oil inventories. Since fluctuations in the level of inventories are limited by available storage capacity, large deviations from normal inventory levels might induce stronger price effects as risks of stockouts or capacity constraints mount.

Another important determinant of the real price of oil is the state of the world economy, which is summarized by the global economic conditions (GECON) indicator recently developed by Baumeister et al. (2022) that measures cyclical fluctuations based on developments in real economic activity, financial markets, geopolitical uncertainty, weather phenomena, transportation demand, and changes in expectations. This comprehensive indicator not only outperforms other measures of global activity in forecasting oil prices (see Baumeister et al., 2022), but also covers exposures to a broad range of risks. To capture trends more specifically related to commodity markets, we also include the Commodity Research Bureau (CRB) raw industrial materials index, deflated by US CPI, which tracks price movements of a basket of globally traded commodities excluding energy and foodstuff used for industrial purposes. As such, it serves as an early indicator of impending changes in business activity. Its promise as a predictor of oil prices has been documented by Baumeister and Kilian (2012), Alquist et al. (2013), and Aastveit et al. (2023). Tail risks in oil prices might arise from both indicators due to weak economic growth and recessionary episodes, geopolitical turmoil, and maybe even the energy transition.

Worldwide rig counts, which are an important business barometer for the drilling industry and its suppliers, complete our set of predictor variables. The number of active drilling rigs depends on the expected future price of oil and is an early indicator of future production. In particular, a decrease in rig counts exerts upward pressure on oil prices. To the best of our knowledge, the predictive content of this variable has not been evaluated for oil price forecasting to date despite the tight link between oil price fluctuations and drilling decisions (see, e.g., Dossani and Elder, 2023). Tail behavior of real oil prices might be driven by a slowdown in the pace of rig additions which leads to a decrease in the inventory of drilled oil wells raising price risks by eroding available weeks of supply (see Baumeister, 2023).

All variables are monthly and are transformed to growth rates by taking the first difference of the natural logarithm except for the GECON indicator and the gasoline-Brent spread which are stationary by construction. The sample spans the period from 1975M1 to 2019M12 based on the January 2024 vintage of data. More details on the data sources can be found in Appendix C.

3.2 An In-Sample Primer on Nonlinear Relationships

Before we move on to a formal comparison of the out-of-sample forecasting performance of our three classes of models, it is instructive to explore potential nonlinearities between the real oil price and its economic drivers with a simple exercise. Specifically, we estimate a simplified version of our nonparametric model:

$$\Delta p_t^{oil} = f(\boldsymbol{z}_t) + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma_u^2), \tag{3}$$

where p_t^{oil} is the log of the real Brent price and $\boldsymbol{z}_t$ is a j-dimensional vector that includes all other predictor variables $j = 1, ..., 6$. The form of the function $f(\cdot)$ is unknown and estimated using BART, as described above. Stacking the variables yields full-sample data matrices $\boldsymbol{\Delta p}^{oil} = (\Delta p_1^{oil}, ..., \Delta p_T^{oil})'$ and $\boldsymbol{Z} = (\boldsymbol{z}_1', ..., \boldsymbol{z}_T')'$.

The nonlinear and nonparametric nature of the model renders the interpretation of marginal effects difficult. As a solution, Friedman (2001) suggests using partial dependence plots (PDPs) to visualize possible nonlinearities between oil price growth and its fundamental determinants that the model learnt. Let $\boldsymbol{z}_j$ denote the variable of interest and $\boldsymbol{Z}_{-j}$ the matrix with the j^{th} column excluded. To get a sense of the (nonlinear) effect of $\boldsymbol{z}_j$ on $\boldsymbol{\Delta p^{oil}}$, we need to integrate out the effects the other variables might have. Doing so for different values of z_{jt} gives rise to the PDPs, which are computed as follows:

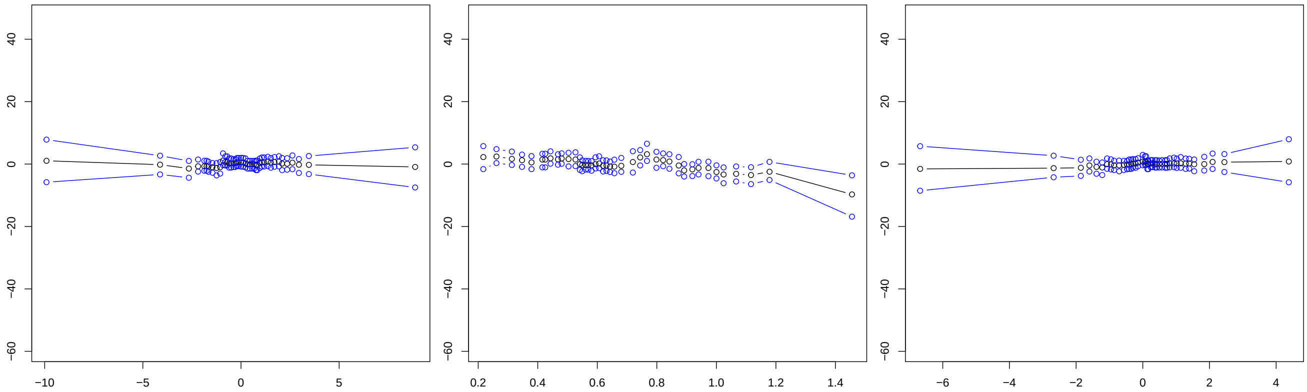
$$\widehat{f}_j(\boldsymbol{z}_j) = \int \widehat{f}(\boldsymbol{z}_j, \boldsymbol{Z}_{-j}) d\mathbb{P}(\boldsymbol{Z}_{-j}).$$

It is worth stressing that one of the assumptions behind PDPs is that the elements in $\boldsymbol{z}_t$ are uncorrelated. In our case, the elements in $\boldsymbol{z}_t$ are obviously correlated, as already apparent from the narrative above. This does not invalidate the PDPs but could lead to combinations of covariate values that do not appear in our dataset. We leave this issue aside since our modest goal here is to provide some preliminary evidence on the possible nonlinear relationship between the growth rate of the real price of oil and the variables contained in different columns of $\boldsymbol{Z}$.

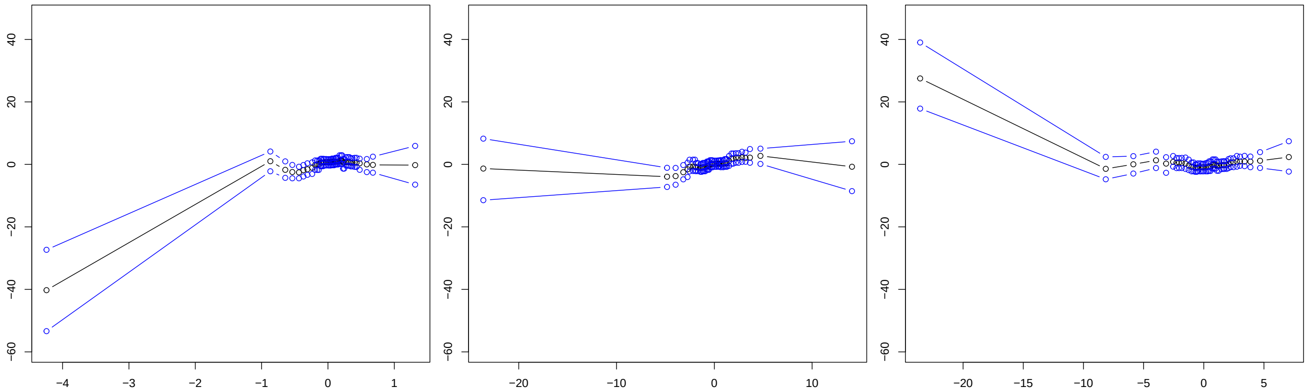
Figure 1 displays the median PDPs together with the 90% probability regions for six bivariate relationships where the x-axis spans the range of values a particular element in $\boldsymbol{z}_j$ can take and the y-axis indicates the corresponding prediction for the percent change of the real Brent price. In the absence of nonlinearities, we would expect the partial dependence profile to be flat. This is largely what we find for global fuel consumption, petroleum inventories, and the real CRB raw industrials index, whereas the other three predictor variables exhibit important nonlinear features. When the GECON indicator reaches its minimum of -4.6%, real oil prices are predicted to drop by 40%, while the effect becomes much weaker for less severe contractions and turns insignificant for values between -1% and 1%. Similarly, the effect of worldwide rig counts on the price of oil is significantly stronger when rig counts are declining steeply than when changes are close to zero or positive, highlighting that sizeable reductions in future oil production growth push oil prices up disproportionately. A substantial rise in the gasoline-Brent spread also leads to a more pronounced response of oil prices. Overall, this analysis shows that moderate variations of predictor variables tend to have limited effects on oil prices, whereas more extreme realizations can impart nonlinear behavior with notable asymmetries of downside and upside risks.

The figure also calls attention to heterogeneities of partial dependence across variables. While

Figure 1: Partial dependence plots of the real Brent price with individual oil market fundamentals



a) Global fuel consumption **b)** Gasoline-Brent Spread **c)** OECD petroleum inventories



d) GECON indicator **e)** Real CRB raw industrials index **f)** Worldwide rig counts

Notes: The month-on-month growth rate of the real Brent price is on the y-axis and different values of the oil market fundamentals are on the x-axis. The black lines are the median of PDPs and the blue lines refer to the 5^{th} and 95^{th} credible intervals.

suggestive of the key drivers of nonlinearities, there could also be hidden dependencies resulting from the interaction of various oil market fundamentals which are not taken into account in this univariate setting. We also focus only on contemporaneous relationships, while nonlinearities might arise dynamically. Overall, this analysis provides a first useful indication of the underlying dependence structure.

4 (Tail) Forecasting the Real Price of Oil

We now turn to evaluating the accuracy of point and tail forecasts of the monthly price of oil. In Section 4.1, we describe the forecasting environment and evaluation criteria, and in Section 4.2, we compare the forecasting ability of our three competing models and examine the stability of our results across horizons and over time.

4.1 Forecasting Environment

We generate out-of-sample density forecasts for the real price of oil for horizons $h = 1, 3, 6, 9$, and 12 months ahead. Our forecasting design is recursive. We use an initial estimation period that runs from 1976M1 to 1991M12. After producing h-step-ahead forecast distributions, we extend the sample by adding observations for 1992M1, reestimate the models, and obtain predictive distributions for all forecast horizons. We repeat this procedure until we reach the final point in our sample for a hold-out period that spans 1992M1 to 2019M12.[9]

Since our focus is on tail forecasts of the real price of oil, we use quantile scores (QS) to evaluate the forecasting performance of our models. The QS is typically associated with the check loss function (Giacomini and Komunjer, 2005) which for the τ^{th} quantile is given by:

$$\text{QS}_{\tau i,t} = (p_t - Q_{\tau i,t})(\tau - \mathbb{I}\{p_t \leq Q_{\tau i,t}\})$$

where $Q_{\tau i,t}$ is the τ^{th} forecast quantile of the real price of oil for $\tau \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95\}$ and $\mathbb{I}\{p_t \leq Q_{\tau i,t}\}$ denotes an indicator function that takes the value of unity if the outcome is at or below the forecast quantile and zero otherwise. Note that the quantile score for $\tau = 0.5$ corresponds to the median absolute error and thus serves as a metric for point forecasts.

Given that quantile scores are specific to selected quantiles, they do not provide information on particular regions of the predictive densities. For example, it might be of separate interest to assess how well a model predicts downside risks which requires a summary measure that puts more emphasis on the area in the left tail. This is achieved by the quantile-weighted CRPS criterion (qw-CRPS) proposed in Gneiting and Ranjan (2011). The qw-CRPS is a proper scoring rule that takes into account information across the entire predictive density but allows to place more weight on regions of interest to the analyst. We compute the qw-CRPS by weighting the QS for $J - 1$ different quantiles:

$$qw\text{-CRPS}_t = \frac{2}{J-1} \sum_{j=1}^{J-1} \omega(\tau_j) \text{QS}_{\tau i,t,j}$$

with $\tau_j = j/J$ and $\omega(\tau_j)$ denoting a deterministic weighting function. We consider four specifications. The first specification gives more weight to the center of the predictive distribution by setting $\omega(\tau_j) = \tau_j(1 - \tau_j)$. The next two specifications assign greater weight to either the left tail or the right tail by defining $\omega(\tau_j) = (1 - \tau_j)^2$ to emphasize downside risks and $\omega(\tau_j) = \tau_j^2$ to emphasize upside risks. The final specification uses $\omega(\tau_j) = (2\tau_j - 1)^2$ to attach weight to both tails. Each of

[9] We end the evaluation period at the end of 2019 since we conduct a real-time assessment for the subsequent years in Section 5. However, we report results for the entire period up to 2024M1 in Tables A1 and A2, which corroborate the robustness of our findings to including some turbulent episodes and accounting for real-time data constraints.

the qw-CRPS measures is computed based on a relatively fine grid of $J = 99$ quantiles which runs from $\tau = 0.01$ to $\tau = 0.99$ in 0.01 increments.

4.2 Forecast Evaluation

We assess the forecast accuracy of our three models—the Bayesian VAR with stochastic volatility (BVAR-SV), the BART model that mixes linear and nonlinear components and features stochastic volatility (mixBART-SV), and the Bayesian quantile regression (BQR)—against a random walk without drift, which has been identified as a tough benchmark to beat (see, e.g., Hamilton, 2009; Alquist et al., 2013; Bernard et al., 2018; Garratt et al., 2019; Baumeister et al., 2022).

4.2.1 Point Forecasts

Table 1 presents the average ratios of quantile scores for a given model relative to the no-change forecast. A ratio below one indicates a gain in accuracy, whereas a value above one indicates a deterioration in accuracy. Column 4 reports the results for quantile $\tau = 0.5$ which represents the median forecast and thus $\text{QS}_{0.5}$ equals the median absolute error (MAE), a standard measure to evaluate point forecasts. The MAE ranges between 0.98 and 1.02 across all models and forecast horizons, underscoring conventional wisdom that it is hard to outperform the random walk in out-of-sample oil price point forecasting.

4.2.2 Tail Forecasts

Columns 1-3 and 5-7 in Table 1 show the relative QS for quantiles to the left and right of the median. These allow us to assess how well the different models perform in the tails. At the one-month forecast horizon, both BVAR-SV and mixBART-SV improve upon the random walk in the left tail but not in the right tail, with gains as high as 12% in the 5-percent quantile. The picture flips for $h = 3$ where the largest improvements are found in the far-right tail with accuracy gains of about 20%. Despite its forecasting success for standard macroeconomic and financial variables, the BQR model performs abysmally at short horizons. For 6- to 12-month horizons, all three models improve upon the no-change forecast for the 0.25 and 0.75 quantiles but to a differing degree. This changes in the more extreme part of the tails where forecasts generated with the BVAR-SV and mixBART-SV models deliver gains in the range of 11% to 39% in the left tail and 34% to 54% in the right tail depending on the horizon, while BQR records substantial losses. The most striking feature is that tail forecast accuracy of both BVAR-SV and mixBART-SV increases dramatically as the forecast horizon lengthens. These gains are more pronounced in the right tail than in the left tail in the medium term with mixBART-SV slightly dominating BVAR-SV.

These general patterns are corroborated by the quantile-weighted CRPS scores which provide a broader assessment of the predictive distributions. Table 2 presents the average ratios of the qw-CRPS relative to the random walk for the center of the predictive distribution as well as its left

Table 1: Average quantile score ratios relative to no-change forecast of real Brent price
Hold-out period: 1992.1-2019.12

Monthly horizon	Model	Quantiles						
		0.05	0.10	0.25	0.50	0.75	0.90	0.95
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
$h = 1$	BVAR-SV	**0.88**	**0.94**	**0.98**	**0.99**	1.04	1.03	1.02
	mixBART-SV	**0.89**	**0.95**	**0.98**	**0.99**	1.01	1.01	1.00
	BQR	2.39	1.63	1.09	**0.98**	1.29	2.21	3.50
$h = 3$	BVAR-SV	1.01	**0.95**	**0.96**	1.02	**0.95**	**0.82**	**0.80**
	mixBART-SV	1.00	**0.94**	**0.96**	1.02	**0.96**	**0.81**	**0.79**
	BQR	2.51	1.59	1.06	1.02	1.26	1.80	2.77
$h = 6$	BVAR-SV	**0.89**	**0.82**	**0.84**	1.00	**0.79**	**0.65**	**0.66**
	mixBART-SV	**0.87**	**0.81**	**0.83**	1.01	**0.79**	**0.63**	**0.63**
	BQR	2.01	1.32	**0.93**	1.01	1.03	1.37	2.05
$h = 9$	BVAR-SV	**0.75**	**0.70**	**0.74**	**0.98**	**0.68**	**0.56**	**0.60**
	mixBART-SV	**0.73**	**0.68**	**0.73**	**0.99**	**0.67**	**0.53**	**0.54**
	BQR	1.66	1.10	**0.81**	**0.99**	**0.88**	1.13	1.70
$h = 12$	BVAR-SV	**0.66**	**0.62**	**0.68**	**0.99**	**0.60**	**0.50**	**0.54**
	mixBART-SV	**0.64**	**0.61**	**0.67**	**0.99**	**0.59**	**0.46**	**0.49**
	BQR	1.46	**0.96**	**0.73**	1.00	**0.78**	**0.98**	1.48

Notes: Boldface indicates improvements relative to the no-change forecast. Green shaded cells indicate the best-performing model for a particular forecast horizon.

and right tails separately and jointly for each model. While the point forecasts were essentially tied with the no-change benchmark for all horizons, when we consider the region of central tendency, all three models outperform the random walk at horizons 9 and 12 (column 2). BVAR-SV and mixBART-SV offer consistently better tail forecasts of the real oil price than does the random walk from the three-month horizon onward, with gains as high as 28% in the left tail (column 1) and 35% in the right tail (column 3) for one-year-ahead predictions. The tail forecasting performance of BQR is rather disappointing overall. Since our ultimate goal is to use the best-performing model for monitoring risks in the oil market, the most important metric is the joint forecasting performance in the tails. Column 4 shows that mixBART-SV produces the most accurate forecasts of extreme realizations in both tails with gains growing with the horizon, ranging from 2% at $h = 1$, 22% at $h = 6$, and a staggering 40% at $h = 12$.

4.2.3 Tracking the Forecasting Performance over Time

The results in Tables 1- 2 summarize the average tail forecast accuracy of all three models over the hold-out period. Focusing on average QS and qw-CRPS ratios might mask important changes in the models' forecasting performance over time and raises the question of how robust accuracy gains

Table 2: Average quantile-weighted CRPS ratios relative to no-change forecast of real Brent price
Hold-out period: 1992.1-2019.12

Monthly horizon	Model	Quantile weights			
		left tail (1)	center (2)	right tail (3)	both tails (4)
$h = 1$	BVAR-SV	**0.97**	1.00	1.02	**0.99**
	mixBART-SV	**0.97**	**0.99**	1.00	**0.98**
	BQR	1.28	1.13	1.47	1.78
$h = 3$	BVAR-SV	**0.98**	**0.98**	**0.94**	**0.93**
	mixBART-SV	**0.97**	**0.98**	**0.94**	**0.92**
	BQR	1.28	1.14	1.41	1.68
$h = 6$	BVAR-SV	**0.88**	**0.89**	**0.82**	**0.79**
	mixBART-SV	**0.88**	**0.89**	**0.82**	**0.78**
	BQR	1.15	1.04	1.21	1.38
$h = 9$	BVAR-SV	**0.79**	**0.81**	**0.73**	**0.69**
	mixBART-SV	**0.78**	**0.81**	**0.72**	**0.67**
	BQR	1.02	**0.94**	1.06	1.18
$h = 12$	BVAR-SV	**0.73**	**0.76**	**0.67**	**0.62**
	mixBART-SV	**0.72**	**0.75**	**0.65**	**0.60**
	BQR	**0.92**	**0.87**	**0.96**	1.04

Notes: See Table 1.

are to the hold-out period. Instead of tracking the evolution of each model against the random walk, we compare mixBART-SV, which emerged as the overall winner of the forecasting horserace, with the second-best model. To shed light on possible time variation in the forecasting performance of extreme tail events, Figure 2 displays the cumulative quantile scores for selected forecast horizons for mixBART-SV relative to BVAR-SV for the one and five percent quantiles (left column) and for the 95 and 99 percent quantiles (right column).[10]

Panel (a) shows that at the one-month horizon there is substantial heterogeneity in the forecasting performance across left and right tails throughout the hold-out period. Initially, mixBART-SV gains a considerable advantage over BVAR-SV in predicting the far-left tail of the distribution which gradually subsides as more observations accumulate and vanishes in the mid-2010s for both quantiles. Instead, BVAR-SV forecasts for the right tail always dominate mixBART-SV at the most extreme quantile but the difference levels off in the early 2000s with both models performing similarly. The evolution at the 95-percent quantile mimics the left-tail experience but the shift in favor of mixBART-SV occurs later and is less stark, yet the relative QS stays below unity.

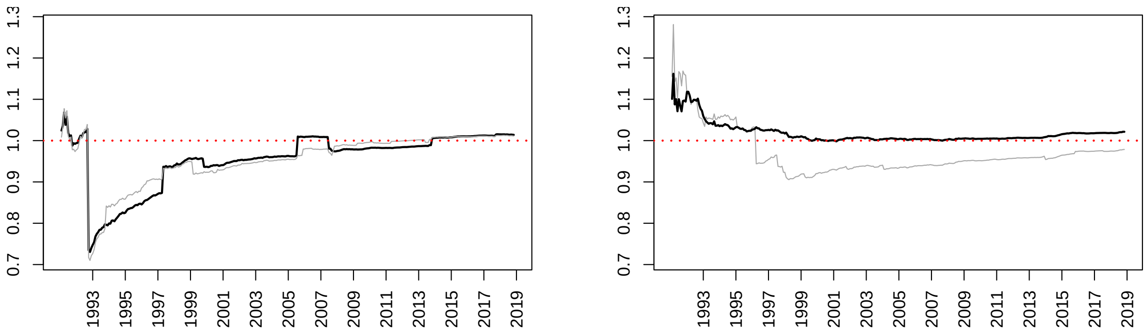
The patterns of time variation are quite different for the other two forecast horizons ($h = 6$ and $h = 12$). Panels (b) and (c) show that in both cases the left-tail and right-tail quantile scores

[10] The results for other quantiles and forecast horizons are qualitatively similar to the ones reported here.

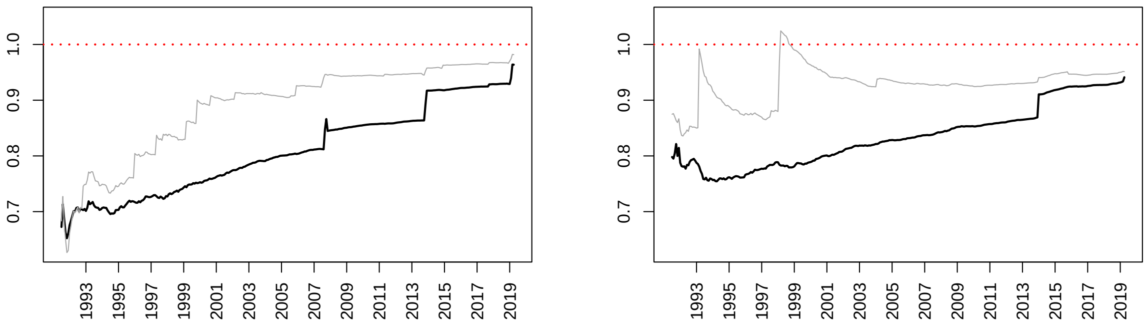
Figure 2: Cumulative quantile scores of mixBART-SV relative to BVAR-SV

Left tail **Right tail**

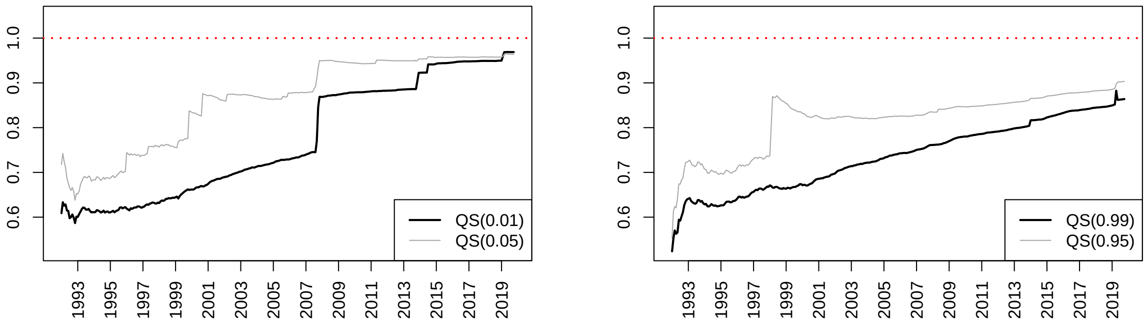
(a) One-month-ahead



(b) Six-months-ahead



(c) 12-months-ahead



evolve in a more similar way. MixBART-SV produces more precise tail forecasts for the entire hold-put period except for one brief episode in 1999 for the 95-percent quantile six-months-ahead. Over time, BVAR-SV gains some mileage and gradually closes some of the gap with mixBART-SV. For example, there is a notable deterioration in the relative performance of mixBART-SV around the time of the global financial crisis for left-tail predictions.

5 Monitoring Oil Price Risks in Real Time

The evidence presented thus far clearly suggests that the mixture BART model with stochastic volatility is the most accurate model in terms of tail forecasts and competitive in terms of point forecasts. Up to this point, we have relied on final-vintage data and thus abstracted from the fact that some of our predictor variables get revised over time and become available with a lag. We now track the evolution of predictive densities over the period 2020 to 2023, which was characterized by three major events that generated substantial turmoil in global oil markets. To assess to what extent the model captures time-varying tail risks in extreme circumstances and to conduct model-based risk assessment in real time, we compile data vintages from December 2019 to January 2024 that reflect only the information as was available at the end of each month.

To fill in gaps at the end of each vintage, we follow Baumeister and Kilian (2012, 2014a) who show that simple nowcasting rules based on the time-series properties of the data perform best. For data on fuel consumption which become available with a delay of two months and for data on OECD petroleum stocks which are lagging three months, we extrapolate the missing observations based on their average growth rates. While the nominal Brent price and the CRB raw industrial materials index are available in real time, the release of U.S. CPI is delayed by one month. To deflate both the Brent price and the CRB index in real time, we nowcast the CPI based on past average inflation. Given that the retail gasoline price lags one month behind, we assume that it follows a random walk in order to obtain the current value for the gasoline-Brent spread. While worldwide rig counts are updated with a one-month delay, rig counts for the US and Canada are released at the end of every week. To nowcast the missing value for the total number of drilling rigs, we add the change in North American oil rigs, computed as the difference between the average weekly counts this month and the previous month, to the most recent observation for worldwide rig counts. The 16 variables underlying the construction of the GECON indicator are subject to varying publication lags (see Baumeister et al., 2022). When extracting GECON as the first principal component from this unbalanced panel, we exploit the EM algorithm to fill in the missing data, except for the period March to June 2020. During these tumultuous months, we use the data vintages assembled in real time by Baumeister and Guérin (2021) who apply crisis-specific nowcasting rules to account for the rapid deterioration in several variables underlying GECON, following Foroni, Marcellino, and Stevanovic (2022) who stress the importance of informative nowcasts in the wake of the COVID-19 pandemic.[11]

5.1 Real-Time Assessment of Tail Risks

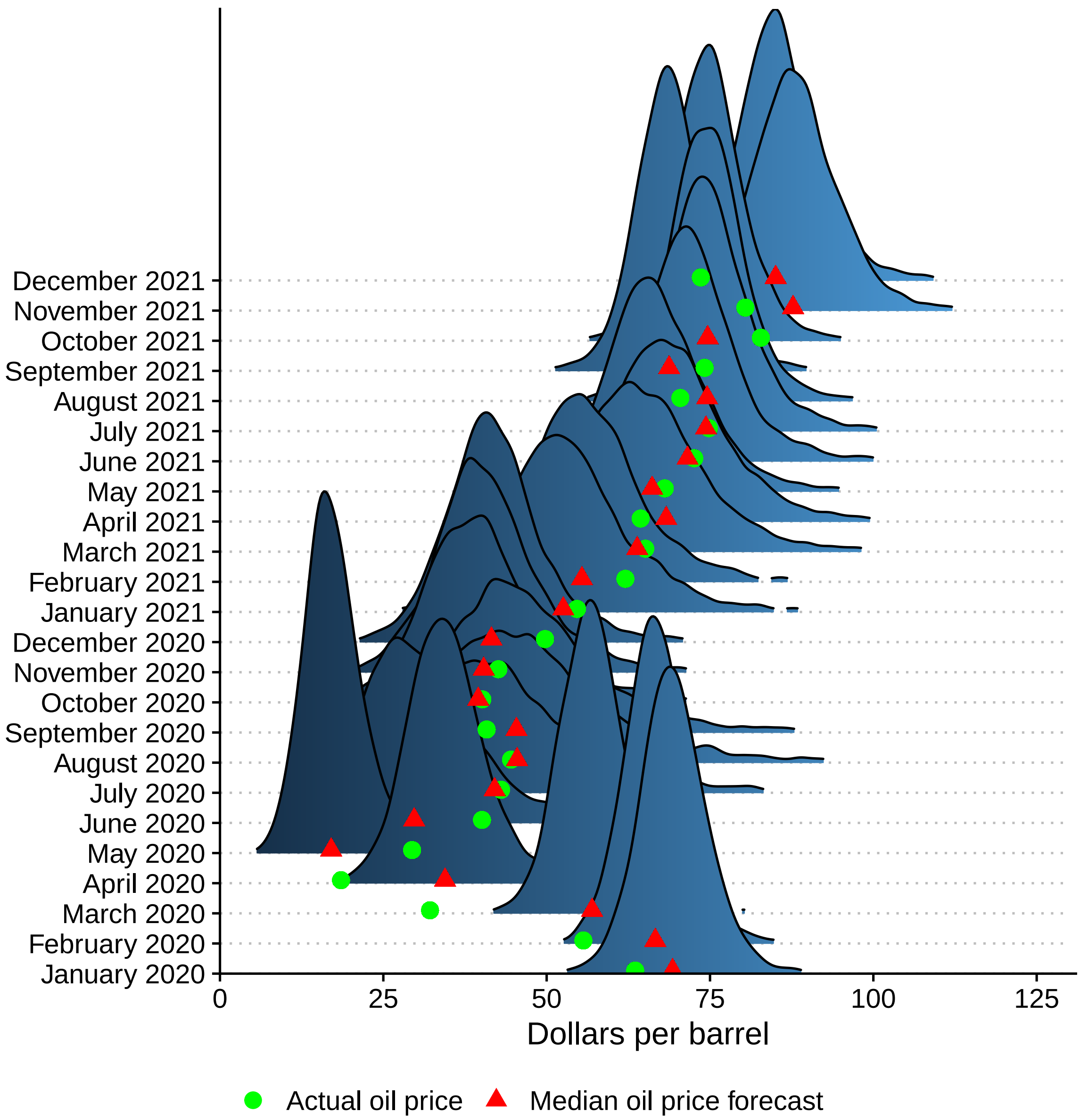
Based on these real-time data vintages, we generate predictive densities to evaluate the role of time-varying tail risks in a true out-of-sample context. We focus on the COVID-19 pandemic, the Russian invasion of Ukraine, and the Israel-Hamas war. These three case studies cover a diverse

[11] For a detailed description of these crisis-specific adjustments, the reader is referred to Baumeister and Guérin (2021).

set of dramatic events that caused large swings in oil prices and thus are well-suited to analyze the properties of the predictive distributions and track the overall performance of our preferred model. In fast-evolving situations like those experienced over the past four years, price forecasts for the next month are particularly valuable given the high level of uncertainty that these events entail. Figure 3 presents a sequence of real-time one-month-ahead predictive distributions from the mixBART-SV model, where the red triangles indicate the median oil price forecast, while the green dots denote the actual outcome.[12]

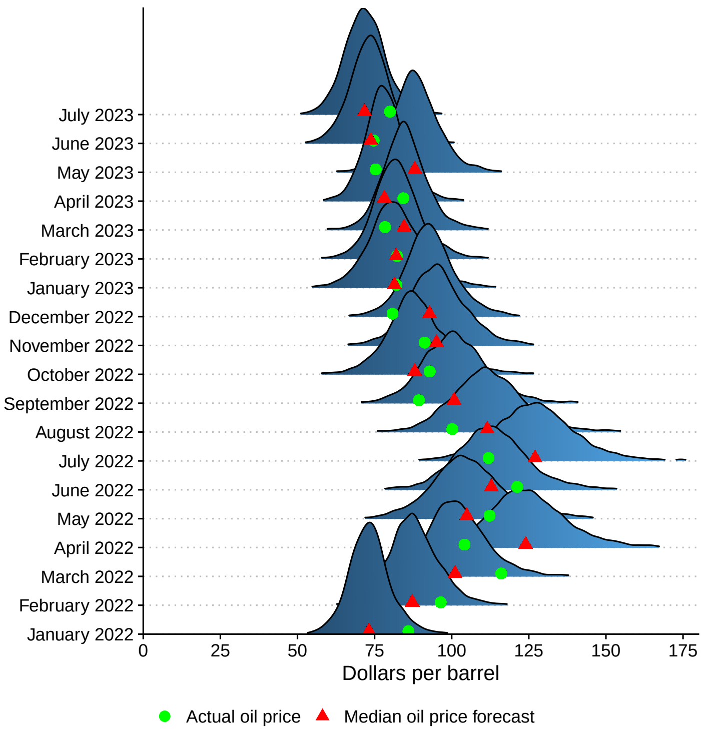
Figure 3: Real-time one-step-ahead predictive distributions of the mixBART-SV model

(a) The COVID-19 Pandemic and the Russia-Saudi Oil Price War

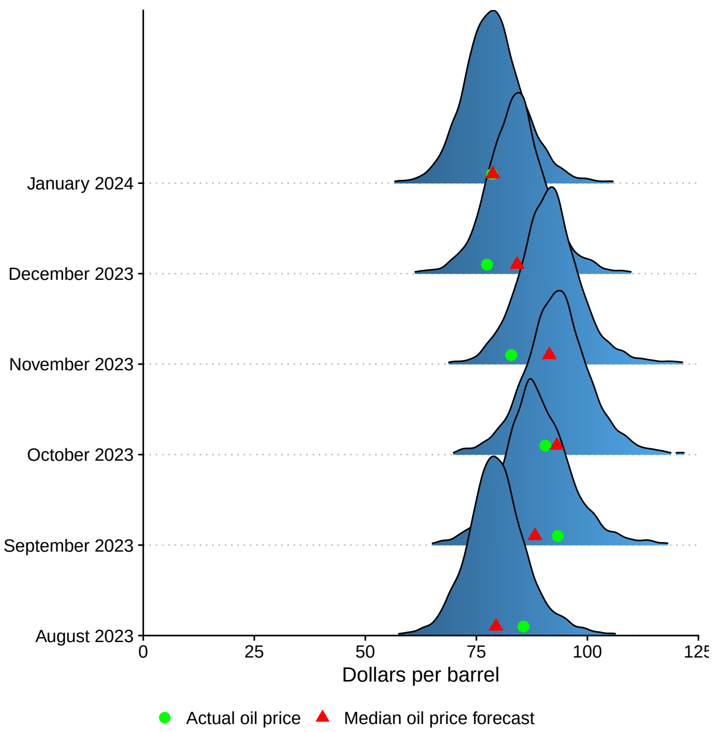


[12]Note that we have mapped the real Brent price back into nominal terms by factoring in inflation.

(b) The Russian Invasion of Ukraine and the Energy Crisis



(c) The Israel-Hamas War and the Red Sea Crisis



5.1.1 The COVID-19 Pandemic and the Russia-Saudi Oil Price War

The outbreak of the COVID-19 pandemic in early 2020 led to significant disruptions in global energy markets. Worldwide restrictions on mobility and other containment measures triggered an unprecedented reduction of fuel consumption and an abrupt deterioration of global economic conditions.

Panel (a) shows that while our model's point forecast is on target at the beginning of 2020, it cannot keep up with the rapid decline in oil prices as the spread of the coronavirus accelerated in February and March of 2020. In addition to the unfolding economic crisis, the collapse of oil prices was exacerbated by a price war raging between Russia and Saudi Arabia over oil production cuts proposed by OPEC to stem the freefall of demand (see Baumeister, 2023). In response to Russia's reluctance to comply with production cutbacks, Saudi Arabia retaliated by flooding the market with oil thereby adding to the global supply glut and putting additional pressure on limited storage capacity. Initially, the model is not able to pick up all these developments, but then it quickly incorporates this new information and from the end of June 2020 onward delivers point forecasts that are closely aligned with subsequent realizations except for September 2020 when it slightly misses the reversal of oil prices in connection with the start of the second wave of the pandemic as new variants of coronavirus emerged. Some larger discrepancies between median forecast and actual outcome are observed in the last quarter of 2021 which coincides with another surge in covid cases and travel restraints in various parts of the world. In general, however, our model produces precise point forecasts in real time during this turbulent period.

When we consider the shape of the predictive distributions throughout the pandemic, some interesting patterns emerge. In the early stages of the pandemic, the predictive distributions shift to the left but remain rather symmetric until May 2020. From June onward, the right tail extends considerably as the model attributes a higher probability to rising oil prices. This is in line with the rebound in real activity that gained speed in the third quarter of 2020. Uncertainty increases further at the beginning of 2021 with predictive densities becoming flatter and more spread out but they remain skewed to the upside until December 2021.

5.1.2 The Russian Invasion of Ukraine and the Energy Crisis

In early 2022, when tensions between Russia and Ukraine were mounting, the predictive distributions start to gravitate to the right of the density that was forecast for January 2022 at the end of 2021, as can be seen in panel (b). What is noteworthy is that after Russia's full-scale invasion of Ukraine at the end of February 2022, the distributions become increasingly widespread, reflecting the rise in uncertainty as the war unfolded. One source of concern was the disruption of established oil trading relationships, in particular between Russia and the European Union (EU), which triggered a reorganization of global oil flows (see Baumeister, 2023).

The heightened volatility during the first phase of the military conflict is also visible in the misses of the point forecasts. While the model is quick to recognize the higher upward risks to oil prices, it underpredicts the price jump in March 2022 by a wide margin and it predicts a price spike in April 2022 when oil prices temporarily receded. Albeit less stark, this pattern persists until September, after which our mixBART-SV model produces reliable point forecasts. For the first few months of the war, the right tail of the predictive distributions is particularly long indicating that tail outcomes remain a distinct possibility. In fact, throughout this period, there was a lot of speculation about oil prices climbing to new record levels as a result of the EU oil embargo against Russia and the ensuing energy crisis (see Baumeister, 2022, 2023), a risk that is captured by our model.

Upside risks gradually subside in the last quarter 2022, as it became clear that war-related disruptions in oil markets would be limited. In the first half of 2023, increasing recessionary fears ignited by the rapid tightening cycle of the world's major central banks were counteracted by fears of market tightness induced by OPEC's plan to reduce oil production, making the predictive distributions more symmetric again, which indicates that upside and downside risks are balanced.

5.1.3 The Israel-Hamas War and the Red Sea Crisis

In the last quarter of 2023, the Hamas-led attack on Israel in early October triggered the Israel-Hamas war. Panel (c) shows that, in the run-up to this event, a market imbalance tilted toward supply concerns and thus slightly higher upside risks seemed to develop with oil price realizations landing to the right of the median forecasts made in July and August 2023. Shortly after the outbreak of the Israel-Hamas war, unrest in the region was compounded by militant activities in the Red Sea, including repeated attacks on commercial vessels.

Based on past experience, this geopolitical turmoil should move the predictive distribution upward and induce both a higher variance and a longer right tail. What stands out over the period October 2023 to January 2024 is not only that the one-month-ahead point forecasts are particularly accurate, but also that the model does not indicate a surge in right-tail risks. This is actually consistent with narrative accounts of oil markets remaining relatively calm with limited reaction by market participants to this crisis. What most likely has kept risks in check is the perception that the market is well supplied and that enough petroleum products are held in storage, while demand is rather subdued.

5.2 Tail-Risk Indicators for Producers and Consumers

The changing pattern of asymmetric shapes in the predictive distributions analyzed above can be mapped into real-time indicators that carry information about tail risks various stakeholders in energy markets might be exposed to. For example, what matters to oil producers and governments whose budget depends on revenues from oil extraction is how likely it is that the oil price will end up in the left tail next month as an early signal for spending cuts and investment deferral. Instead, firms and households who are buyers of oil and petroleum products have a keen interest in knowing the probability of Brent exceeding a certain threshold or falling within a specific range for planning purposes. Thus, having a timely indicator of tail probabilities in oil prices is key to decision-making under uncertainty. Such indicators can be computed numerically as functions of the predictive density in the following way:

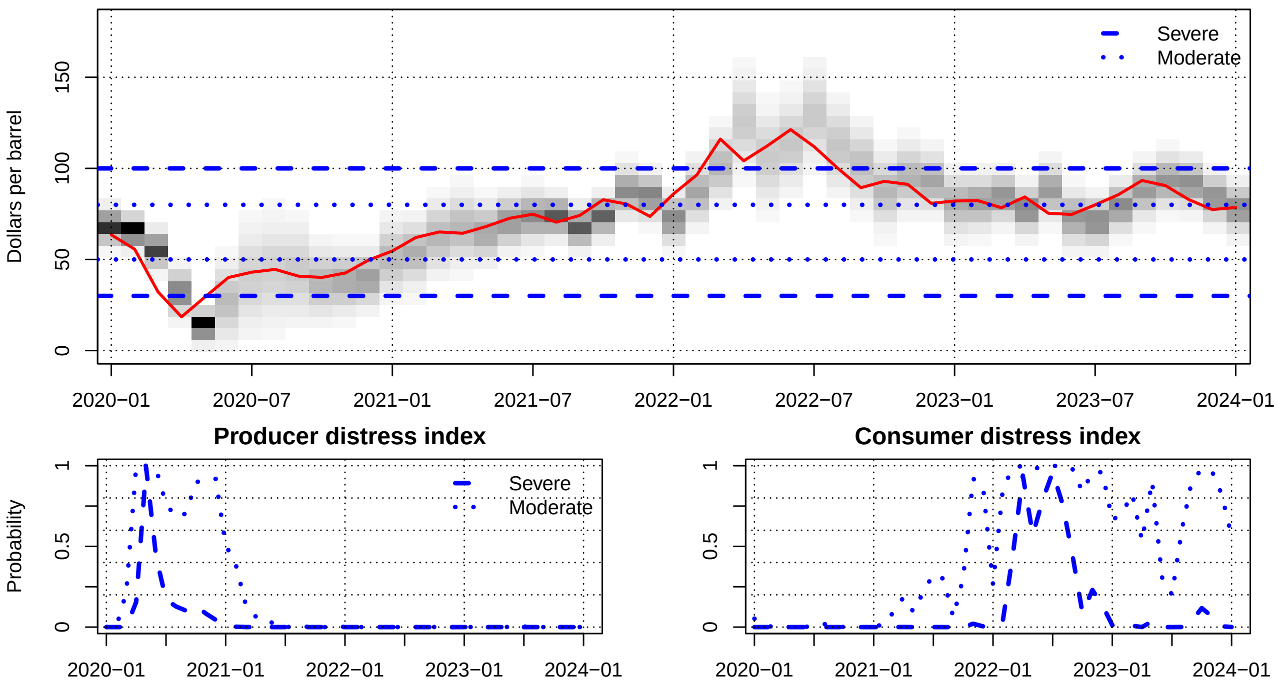
$$
\begin{aligned}
\text{Prob}(p^{oil}_{T+h} &\geq c^U) = \int_{c^U}^{\infty} p(p^{oil}_{T+h}|\mathcal{I}_{1:T}) \approx \frac{\sum_{n=1}^{N} \mathbb{I}(p^{oil,(n)}_{T+h} \geq c^U)}{N} \\
\text{Prob}(p^{oil}_{T+h} &\leq c^L) = \int_{-\infty}^{c^L} p(p^{oil}_{T+h}|\mathcal{I}_{1:T}) \approx \frac{\sum_{n=1}^{N} \mathbb{I}(p^{oil,(n)}_{T+h} \leq c^L)}{N}
\end{aligned}
$$

where c^U and c^L are respectively an upper and lower threshold, $p(p^{oil}_{T+h}|\mathcal{I}_{1:T})$ is the predictive density for the price of oil at horizon h that depends on all available information up to time T, $\mathcal{I}_{1:T}$, and $p^{oil,(n)}_{T+h}$ denotes the n^{th} draw from the predictive distribution with N the total number of draws stored and $\mathbb{I}$ an indicator function that equals one if its argument is true.

The threshold is typically tailored to the economic agent under consideration and can vary over time depending on sector-specific circumstances. Likewise, the relevant horizon for tail-risk indicators might differ across stakeholders. For illustration, we consider four realistic cut-offs, two on the upside and two on the downside. In the five years prior to the spread of the COVID-19 pandemic, Brent oil prices have been fluctuating between \$30 and \$80, which could be conceived as a price range during relatively stable times. Thus, we treat \$80 as a plausible threshold above which consumers would get alarmed; we also assess the upside probability of crossing the psychological barrier of \$100. On the producer side, \$30 should be considered a threshold that could jeopardize the future growth prospects of oil companies; we set the second threshold at the level of \$50, below which concerns about profitability arise since it corresponds to the average breakeven price according to the Dallas Fed Energy Survey (see Plante and Patel, 2019). We refer to the tail-risk indicators implied by these four thresholds as the moderate and severe consumer distress indices and the moderate and severe producer distress indices.

The top panel of Figure 4 presents a heatmap that shows the one-month-ahead predictive densities from January 2020 to January 2024 evaluated over a grid of values for the Brent oil price where each cell in the heatmap corresponds to the (predictive) likelihood of a given value occurring. Dark-shaded cells indicate that the probability of observing a particular value for the oil price is high, whereas grayish cells indicate that the probability of observing that value is small with lighter and lighter shades approaching zero. The red solid line is the actual outcome and the blue dotted and dashed lines refer respectively to the moderate and severe thresholds for the tail-risk indicators. The heatmap reveals that mixBART-SV attributes more mass to regions which include the realized oil price. In fact, more often than not, the actual price observations land in darker-shaded cells which are high-density areas. There are, however, several periods during which values far away from the actual outcome received considerable probability mass. These periods correspond to turbulent times when there was an elevated risk of extreme realizations.

Figure 4: Probability heatmap and tail-risk indicators for producers and consumers 2020.1-2024.1



Notes: The probability heatmap shows the one-month-ahead predictive density of nominal oil prices against the actual outcome (in red). The darker the shades, the higher the probability of a particular value on the y-axis. The blue dotted lines indicate the range from \$50 to \$80 and the blue dashed lines the range from \$30 to \$100.

The probabilities of such tail events conditional on crossing a particular upper or lower threshold are shown in the lower two panels of Figure 4. The left panel highlights that oil producers came under severe distress in the early stages of the coronavirus pandemic when the probability of oil prices dropping below \$30 rose rapidly, adequately capturing the collapse in oil prices later observed in March and April of 2020. However, this episode was relatively short-lived and tail probabilities diminished over subsequent months. Considering the more moderate threshold of falling below the breakeven price, tail risks were mounting even earlier and stayed above a probability of 50 percent until the beginning of 2021. As risks for producers subsided, the right panel shows that moderate consumer distress started to build up gradually during 2021, reaching a first peak toward the end of that year. Rising upside probabilities of exceeding \$80 are consistent with actual oil prices climbing back to previous levels throughout 2021, as can be seen in the top panel. Tail probabilities were highest in 2022 and remained elevated for most of 2023. Signs of the possibility of breaching the \$100 threshold started flashing at the beginning of 2022 with the severe consumer distress index shooting up in February, attaching a probability of almost 100 percent to this event. Upside risks for oil consumers persisted for the next six months with more than a 50 percent chance for the Brent price to exceed \$100, which is indeed what happened.

5.3 Risk Scenarios for 2024

Next we conduct a scenario analysis of possible risks that might impinge on future price developments conditioning on information available as of the end of January 2024. Figure 5 compares the multi-step-ahead unconditional predictive densities with their conditional counterpart where we fix one predictor variable ω at the time to take on historical values corresponding to various quantiles. Specifically, we construct the implied predictive distributions for our scenarios as follows:

$$p(p^{oil}_{T+h}|\mathcal{I}_{1:T}, \omega_{T+h} = \{Q_m(\boldsymbol{\omega}_R), \boldsymbol{\omega}_U\})$$

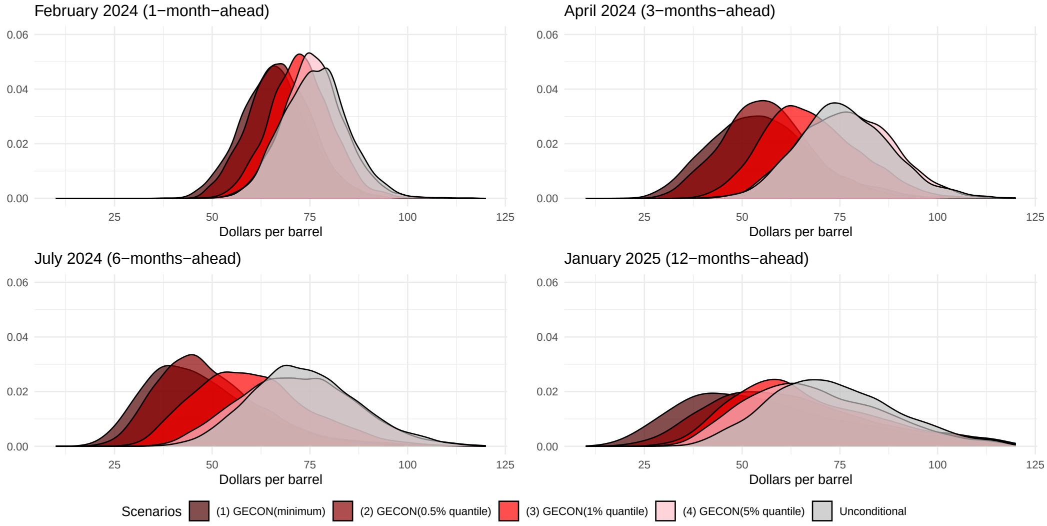
where $Q_m(\boldsymbol{\omega}_R)$ denotes the m^{th} quantile of $\boldsymbol{\omega}_R = (\omega_{T+1}, ..., \omega_{T+r})$ up to $r = 6$ and $\boldsymbol{\omega}_U = (\omega_{T+r+1}, ..., \omega_H)$ evolves according to the unrestricted model dynamics up to $H = 12$. This allows us to assess by how much an alternative path for that one variable would change the distribution of possible outcomes and thus lead us to revise our forecast and risk assessment.

At the beginning of 2024 there has been a lot of discussion about the gloomy economic outlook for the global economy with several advanced economies showing marked signs of anemic growth in a longer-than-expected high interest rate environment. While the US economy is in pretty good shape, China and Europe continue to struggle with below-expectations manufacturing activity and overall economic weakness. To see how these macroeconomic concerns could affect the oil price over the next year, we condition on values in four different lower quantiles of the GECON indicator. Panel (a) shows that a deterioration of global economic conditions would shift the predictive densities for February 2024 to the left relative to the unconditional one with the magnitude of the downward shift depending on the severity of the economic slowdown. While there is hardly any difference in the unconditional and conditional densities for values in the 5% quantile corresponding to a mild deceleration of growth, values in the 1% quantile which are typical of an average recessionary episode imply noticeable downward pressures on oil prices with the highest-density region shifting from around \$75-\$80 unconditionally to \$70-\$75 and with more probability mass on lower oil price forecasts. The strongest adjustment is observed for values in the lowest two quantiles which imply a severe contraction comparable to that of the Great Recession or the COVID-19 recession. What is noteworthy is that adverse changes in GECON have limited effects on the tails of the predictive densities. As the forecast horizon extends beyond one month, not only get the downward shifts more pronounced and the distributions widen, but also the shapes of the conditional densities start to change with clear asymmetries. Interestingly, the 1% quantile density three-months-ahead is skewed to the right indicating that the model attaches predictive weight to oil price increases and six-months-ahead it has an extremely flat mode assigning almost equal probability to outcomes between \$50 and \$60. Since our scenario assumes a relatively short-lived downturn of only two quarters, at the 12-month horizon the differences between the conditional and unconditional densities are less stark, also because forecast uncertainty increases across the board, but the conditional densities still suggest higher downside risks.

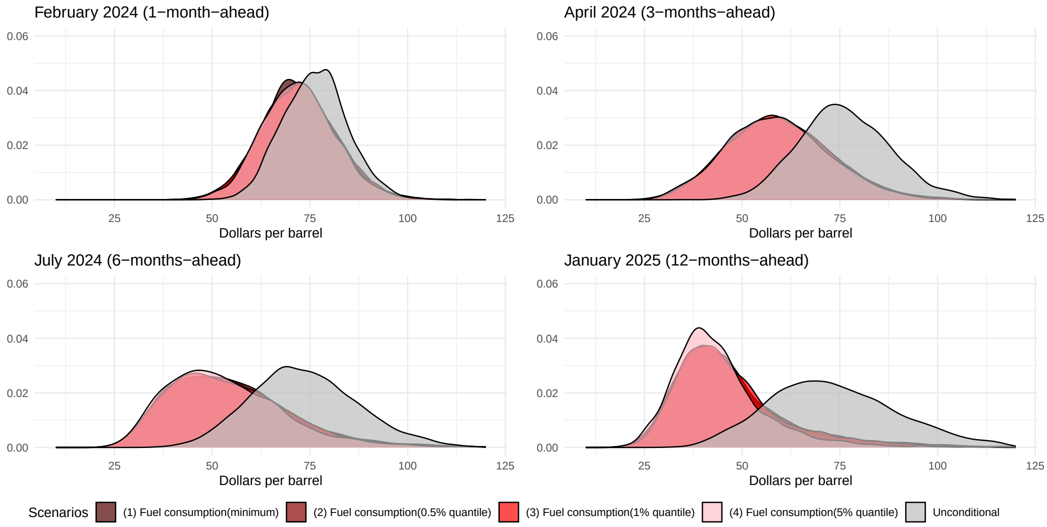
More directly related to the oil market are ongoing doubts about the strength of Chinese oil

Figure 5: Predictive distributions conditional on different paths for selected covariates

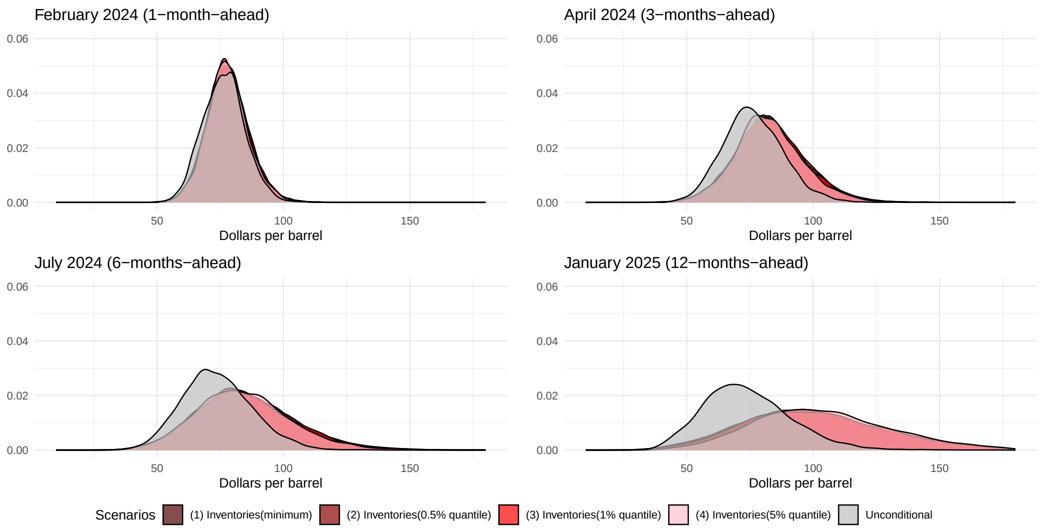
(a) Economic weakness



(b) Dwindling oil demand from China



(c) Physical market tightness



demand. China accounts for a large share of global fuel consumption which is why we now consider a scenario that explores the price impact of different paths of petroleum consumption across lower quantiles. Panel (b) reveals that the modal oil price forecast for February 2024 conditioning on a decrease in fuel consumption is similiar to the one in a typical-recession scenario, but with somewhat greater forecast uncertainty. In contrast to global economic conditions, the degree of the shortfall in fuel consumption does not seem to play much of a role for the lower quantile forecasts of the price of oil across all horizons. However, at longer horizons a reduction in petroleum consumption has more sizeable and more persistent effects on the predictive distributions. For example, six months out, 44% of the probability mass is below \$50 (an important threshold that we considered above) for the 1% consumption quantile forecast, relative to 3% for the unconditional forecast and 25% for the 1% GECON quantile forecast. Despite the fact that fuel consumption returns to its model-implied path after six months, the year-ahead predictive densities are quite concentrated around much lower modal forecasts of around \$38 compared to \$70 unconditionally. What stands out is the lower dispersion of the conditional densities at the one-year horizon and the long right tail, again emphasizing important asymmetric features.

Another relevant concern at the start of the year was the decision of the OPEC+ alliance to extend its voluntary oil production cuts to mid-2024 which might create tightness in the physical market. One interesting scenario against this background is how a draw on oil and petroleum inventories as a substitute for OPEC+ supplies would affect oil price developments in the next 12 months. Panel (c) shows that initially there is no difference between unconditional and conditional predictive densities, no matter how much inventories are released. As the forecast horizon lengthens, the conditional densities keep shifting to the right of the unconditional ones and widen considerably more than the unconditional ones, indicating a tightening of supply conditions, most likely related to concerns about inventory depletion. Implications for the one-year horizon are particularly striking with the predictive densities symmetrically centered on \$100 with a very large variance, suggesting extremely high upside risks for oil consumers.

It is important to stress that these forecast densities are time-specific given that the nonlinearities implied by mixBART-SV evolve continuously as the model produces different combinations of covariates which might induce different tree allocations. This highlights the flexibility and adaptability of mixBART-SV to different circumstances at any given point in time when relevant scenarios are contemplated.

6 Conclusions

Knowing the probability of future extreme oil price realizations is key for making informed policy, investment, and spending decisions, and for generating robust macroeconomic projections. In this paper, we evaluated the usefulness of three popular model classes for modeling and predicting the tail behavior of the real Brent price. We showed that the most flexible model that allows both for nonlinearities of unknown form in the mean relations of a set of fundamental determinants of oil prices and for time variation in the volatilities of shocks delivers the most accurate out-of-sample tail forecasts. We then generated predictive densities from this mixBART-SV model to monitor the build-up of tail risks in real time during three periods of economic and geopolitical turmoil. To convey the probability of severe upside and downside price risks for various stakeholders in the oil market, we devised a consumer and a producer distress index. We also provide a forward-looking risk analysis where we assess the implications of scenario-based tail events as they were looming at the beginning of the year for the oil price throughout 2024. These various real-time risk assessments underscore the importance of allowing for nonlinearities when forecasting oil prices, in particular in the face of turbulent episodes that could hardly be captured by linear models with stable parameters.

Appendix

A Illustration of a Tree-Based Model

We draw on a simple example from our empirical application to show how a tree-based model explains the dependent variable, the growth rate of the real price of oil. To this end, we focus on the case of a single tree ($S = 1$) to approximate f in equation (3). Using (2) implied by this assumption, the conditional mean amounts to a piece-wise constant function of the following form:

$$E(\Delta p_t^{oil}|\boldsymbol{x}_t) = t(\boldsymbol{X}|\mathcal{T}, \boldsymbol{\mu}) \approx \sum_{n=1}^{b} \mu_n \mathbb{I}(\boldsymbol{X} \in \mathcal{S}_n)$$

where b indicates the number of breaks or regimes, and within each regime, the predicted mean is given by μ_n if a specific configuration of $\boldsymbol{X}$ is in the set $\mathcal{S}_n$. The resulting tree is displayed in Figure A1.

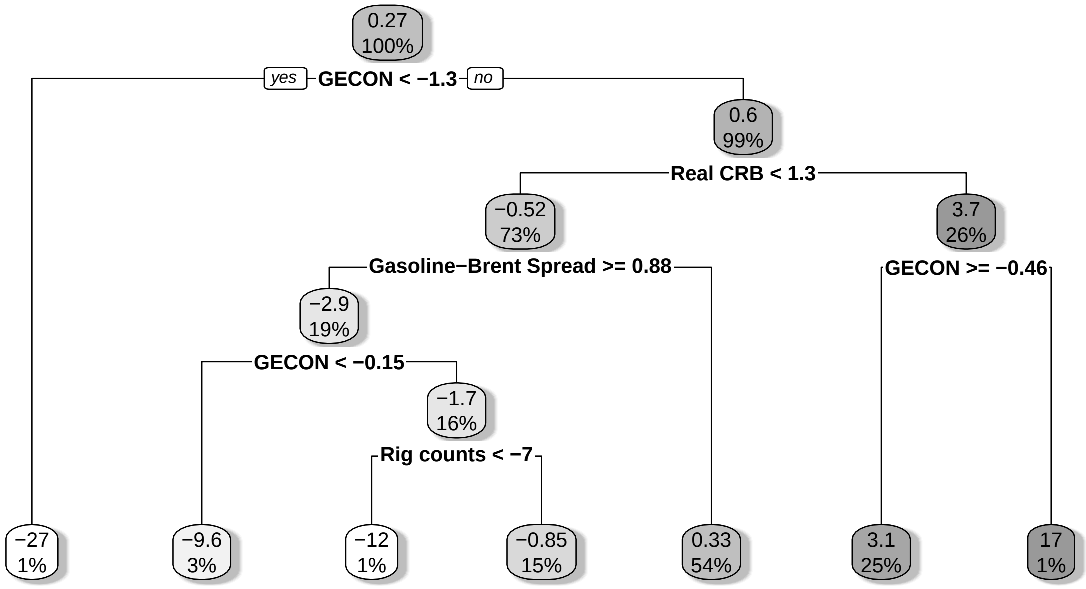
The tree structure $\mathcal{T}$ is comprised of six sequential decision rules. The binary rules show up at different levels of the tree. Starting with the root (level 0), the first rule checks whether GECON (in $t-1$) is smaller than the threshold of $c_{0,GECON} = -1.3$ which stems from the observed values covered by $\{\text{GECON}_t\}_{t=1}^{T}$. If this is the case, we end up with the first, so-called terminal node, and for observations that fulfill this criterion, we predict $\mu_1 = -27$, a 27 percent decline in the real price of oil. This is the case for one percent of all observations in our sample.

If GECON is larger than or equal to -1.3, we go down the right branch of the tree. The first splitting variable showing up in this branch and at level 1 of the tree is the real CRB index, with a threshold of $c_{1,CRB} = 1.3$. If the real CRB index is smaller than 1.3, the next splitting variable is the gasoline-Brent spread with a threshold of $c_{2,spread} = 0.88$. If the spread exceeds 0.88, we move further down the tree, and again, encounter GECON. If GECON is smaller than $c_{3,GECON} = -0.15$, we predict a $\mu_2 = -9.6$ percent decline in real oil prices. This set includes three percent of all observations.

If GECON is above -0.15, we end up in the fourth level of the tree with rig counts acting as the threshold variable. If the growth rate in rig counts is smaller than $c_{4,rigs} = -7$ percent, the tree predicts an oil price decline of around $\mu_3 = -12$ percent for one percent of all observations. If the growth rate of rig counts exceeds -7 percent instead, we predict a small decline of around $\mu_4 = -0.85$ percent. This happens for around 15 percent of all time periods. If the gasoline-Brent spread is smaller than 0.88, the model predicts a moderate increase of $\mu_5 = 0.33$ percent. This forms the biggest set with around 54 percent of all observations allocated to it.

If the real CRB index is greater than 1.3, we again encounter GECON as our threshold variable. If GECON exceeds $c_{2,GECON} = -0.46$, our model predicts an increase in the real price of oil of around $\mu_6 = 3.1$ percent. This is the case for around 25 percent of all sample observations.

Figure A1: Example of an estimated tree to explain the growth rate of the real Brent price



By contrast, if GECON is between -1.3 and 0.46, our model predicts a price increase of around $\mu_7 = 17$ percent. This only happens in one percent of all observations.

This results in a 7-dimensional vector $\boldsymbol{\mu}$, which contains the predicted values for the real price of oil at each of the terminal nodes: $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7)' = (-27, -9.6, -12, -0.85, 0.33, 3.1, 17)'$. As can be seen from this illustration, the structure of the tree is determined by the estimated thresholds, selected splitting variables, and their positions at different levels of the tree. The tree not only provides information about how different configurations in $\boldsymbol{z}_t$ impact Δp_t^{oil}, but also which variables are important for explaining the growth rate of the real price of oil in a nonlinear way.

B Prior Specifications and Bayesian Inference

Since we adopt a Bayesian approach to estimation and inference, we need to specify priors on all the parameters for our three model classes. Some sets of parameters are shared across models and the prior choices for those will be discussed jointly, while other sets of parameters are specific to one class. The best way to see which parameters are involved in each model is to write down the joint prior for each. Let $\boldsymbol{\Xi}_r$ denote a vector of hyperparameters that determine the prior variances associated with the prior on parameters r.

The joint prior for the parametric model (BVAR-SV) takes the form:

$$p(\boldsymbol{A}|\boldsymbol{\Xi}_A)p(\boldsymbol{\Xi}_A)p(\boldsymbol{\Lambda}|\boldsymbol{\Xi}_\Lambda)p(\boldsymbol{\Xi}_\Lambda) \prod_{i=1}^{Q+M} \{p(\mu_{h_i})p(\rho_{h_i})p(\sigma_{h_i})\}.$$

The joint prior for the nonparametric model (mixBART-SV) takes the form:

$$p(\boldsymbol{A}|\Xi_A)p(\Xi_A)p(\boldsymbol{\Lambda}|\Xi_\Lambda)p(\Xi_\Lambda)\prod_{s=1}^{S}\{p(\mathcal{T}_s)p(\boldsymbol{\mu}_s)\}\prod_{i=1}^{Q+M}\{p(\mu_{h_i})p(\rho_{h_i})p(\sigma_{h_i})\}.$$

The joint prior for the semiparameteric model (BQR) takes the form:

$$\prod_{\tau}\{p(\boldsymbol{\beta}_\tau|\Xi_{\boldsymbol{\beta},\tau})p(\Xi_{\boldsymbol{\beta},\tau})p(\sigma_\tau)\}$$

We now detail the prior distributions and their hyperparameters.

B.1 Priors for regression coefficients

We assume that the (conditional) prior on the rows of $\boldsymbol{A}$ is Gaussian, $p(\mathbf{a}_j|\Xi_{A,j}) \sim \mathcal{N}(\underline{\mathbf{a}}_j, \underline{\mathbf{V}}_{A,j})$, where $\underline{\mathbf{a}}_j$ contains elements $\underline{a}_{js}$ and $\Xi_{A,j}$ is a subset of Ξ_A pertaining to the j^{th} equation. Given the variable transformations to growth rates, we set the prior means to capture the notion that each series follows a stationary process by setting $\underline{a}_{js} = 0$ for all j, s.

For mixBART-SV, the prior covariance matrix $\underline{\mathbf{V}}_{A,j}$ is a diagonal matrix that depends on two shrinkage parameters; the first, labeled ζ_{js}^2, is a local shrinkage parameter, whereas the second, labeled ω_j^2, is an equation-specific shrinkage parameter. Thus, a typical diagonal element of $\underline{\mathbf{V}}_{A,j}$ is given by $\zeta_{js}^2\omega_j^2$. The two scaling parameters have their own priors: $\zeta_{js} \sim \mathcal{C}^+(0,1)$ and $\omega_j \sim \mathcal{C}^+(0,1)$ where $\mathcal{C}^+(0,1)$ denotes the standard half-Cauchy distribution. This hierarchy leads to the well-known horseshoe prior (Carvalho, Polson, and Scott, 2010). The key idea is that, a priori, we assume that elements in $\boldsymbol{A}$ that are not related to the first own lag of a particular equation are zero with high prior probability, whereas coefficients associated with the first own lag of a particular equation equal one. This is captured by setting ω_j^2 close to zero, thereby exerting 'global' shrinkage toward the prior mean. However, it could be that some of the coefficients are non-zero; in that case, the local shrinkage parameters drag prior mass away from zero. The horseshoe prior is a particular example of a global-local shrinkage prior that has been shown to have excellent empirical properties (see Huber and Feldkircher, 2019).

For the BVAR-SV model, we follow Baumeister, Korobilis, and Lee (2022) who select the hyperparameters for the diagonal elements of $\underline{\mathbf{V}}_{A,j}$ according to the Minnesota prior, originally proposed by Doan, Litterman, and Sims (1984). The elements of $\underline{\mathbf{V}}_{A,j}$ are set as follows:

$$[\underline{\mathbf{V}}_{A,j}]_{ii} = \begin{cases} \frac{\lambda_1}{l^2} & \text{for the } l^{th} \text{ lag of variable } j \\ \frac{\lambda_1\lambda_2}{l^2}\frac{\hat{\sigma}_j^2}{\hat{\sigma}_k^2} & \text{for the } l^{th} \text{ lag of variable } k \neq j \\ \lambda_3\hat{\sigma}_j^2 & \text{for the intercept term.} \end{cases}$$

Here, we set the hyperparameters $\lambda_1 = \lambda_2 = 0.5$, implying more shrinkage on cross-variable lags within a particular equation. Moreover, we assume a quadratic lag decay. $\hat{\sigma}_j^2$ is the OLS variance obtained by running an AR(6) model on variable $j = 1, \ldots, M$.

The (conditional) prior for the regression coefficients in the BQR model is also Gaussian, $p(\boldsymbol{\beta}_\tau|\boldsymbol{\Xi}_{\beta,\tau}) \sim \mathcal{N}(\underline{\boldsymbol{\beta}}_\tau, \underline{\mathbf{V}}_{\beta,\tau})$. Within a given quantile, we use the same horseshoe prior as in the case of $\boldsymbol{A}$. Notice that this implies that coefficients within each quantile are forced to zero and we do not rely on a global shrinkage parameter that pools information across quantiles.

B.2 Priors for factor loadings

For the loadings in the FSV model, we use a zero-mean Gaussian prior on the columns of $\boldsymbol{\Lambda} = (\lambda_1, ..., \lambda_Q)$, $\lambda_q|\boldsymbol{\Xi}_{\Lambda,q} \sim \mathcal{N}(\mathbf{0}, \underline{\mathbf{V}}_{\Lambda,j})$, with $\underline{\mathbf{V}}_{\Lambda,j}$ denoting an $M \times M$ diagonal prior covariance matrix. We use a horseshoe prior similar to the one for $\boldsymbol{A}$. The key difference, however, is that the equation-specific parameters are replaced with column-specific parameters. This implies that the prior variance pertaining to the i^{th} element in λ_q, λ_{iq}, is $[\underline{\mathbf{V}}_{\Lambda,j}]_{ii} = Var(\lambda_{iq}) = \phi_{iq}^2 \xi_q^2$, where ϕ_{iq} is again a local shrinkage parameter and ξ_q is a columwise shrinkage term, both of which follow a standard half-Cauchy distribution. The key idea of using a columnwise shrinkage term is that if a factor is irrelevant, ξ_q should be very close to zero and the corresponding column in $\boldsymbol{\Lambda}$ should be close to zero. Hence the factor drops out. This approach allows to select the number of factors in a data-dependent way and thus avoids overfitting (see, e.g., Bhattacharya and Dunson, 2011; Kastner, 2019).

B.3 Priors for stochastic volatilities

The priors on the parameters in the law of motion of the stochastic volatilities follow the setup proposed in Kastner and Frühwirth-Schnatter (2014). Specifically, for the unconditional mean parameter μ_{h_i}, we postulate a zero-mean Gaussian prior with variance 10^2; for the (transformed) persistence parameter ρ_{h_i}, we use a Beta-distributed prior $(\rho_{h_i} + 1)/2 \sim \mathcal{B}(25, 5)$ to push the log-volatilities toward unit roots; and for the state innovation variances, we use a Gamma prior $\sigma_{h_i}^2 \sim \mathcal{G}(1/2, 1/2)$.

B.4 Priors for trees

We follow Chipman, George, and McCulloch (2010) and specify regularization priors on $\mathcal{T}_s$ and $\boldsymbol{\mu}_s$ for all s. For the tree structures, we use a tree-generating stochastic process prior. This process grows trees sequentially and, for d denoting the current depth of the tree, the probability that it keeps growing declines with d. This is achieved by specifying the probability that a given node at depth d is non-terminal as $\frac{\alpha}{(1+d)\pi}$, where $\alpha \in (0, 1)$ denotes the base probability. For instance, at the root of the tree the probability that we move away from having a trivial tree (that returns a constant) is α. Hence, larger values of α make more complex trees more likely. The parameter $\pi \geq 0$ is a penalty parameter. If $\pi = 0$, the probability of continuing to grow a tree is independent of its current complexity. We set $\alpha = 0.95$ and $\pi = 2$ which are the benchmark values used in Chipman et al. (2010) that have been shown to work well across different datasets including standard macro datasets (see, e.g., Huber and Rossini, 2022; Huber et al., 2023; Clark et al., 2023). Once we decide on growing the tree, we need to set up the decision rules. This is done as follows. We use a uniform prior on the possible covariates that enter a decision rule (*i.e.* the columns of $\boldsymbol{x}_t$) and after deciding on a particular covariate, we use again a uniform prior on the thresholds c.

On the terminal node parameters, we use independent Gaussian priors:

$$p(\boldsymbol{\mu}_s) = \prod_{i=1}^{N_s} p(\mu_{si}) \sim \mathcal{N}\left(0, \frac{1}{2\kappa\sqrt{S}}\right)$$

where κ denotes a shrinkage parameter that controls the prior variance. Note that the number of trees shows up in the denominator. This implies that the prior tends to increasingly shrink μ_{si} towards zero if many trees are used. Such a behavior captures the notion that if many trees are used, each tree is expected to explain only little variation in responses. A standard choice is to set $\kappa = 2$.

B.5 Prior for scale parameter at quantile τ

Within a particular quantile, we use an inverse Gamma prior on $\sigma_\tau \sim \mathcal{G}^{-1}(a_\tau, b_\tau)$. We set the hyperparameters equal to $a_\tau = 3/2$ and $b_\tau = 0.3/2$. This is the standard choice proposed in Kozumi and Kobayashi (2011).

B.6 Bayesian inference

These priors, when combined with the likelihood, give rise to joint posterior distributions that have no well-known form, which is why we resort to simulation-based methods to carry out posterior inference. The Bayesian quantile regressions are estimated using the Gibbs sampling algorithm of Kozumi and Kobayashi (2011); the Bayesian VAR model with factor stochastic volatility is estimated using the Gibbs sampling algorithm described in Kastner and Huber (2020); and BART uses a Metropolis-within-Gibbs sampler where the Metropolis step applies to sampling the tree structures coupled with the Bayesian backfitting strategy outlined in Chipman, George, and McCulloch (2010). For further details on the precise form of the conditional distributions and the implementation of the estimation algorithms, the reader is referred to Clark, Huber, Koop, Marcellino, and Pfarrhofer (2023). We repeat our algorithms $15,000$ times and discard the first $5,000$ draws as burn-ins. For each draw from the posterior, we simulate the predictive distribution through Monte Carlo integration.

C Data Sources and Backcasting

The three main sources for oil-related data are the Monthly Energy Review (MER), the International Energy Statistics (IES), and the Short-Term Energy Outlook (STEO) compiled by the U.S. Energy Information Administration (EIA). Monthly data for global liquid fuels consumption measured in million barrels per day are available from the STEO database. Given that this time series only starts in January 1990, we backcast it to January 1982 based on the growth rate of OECD refined petroleum products consumption and further back until January 1975 based on the growth rate of global crude oil production; both series are taken from the IES. We seasonally adjust global petroleum consumption using the X13-ARIMA procedure. Monthly data for OECD inventories of crude oil and petroleum products measured in millions of barrels are provided in the IES from January 1988 onward. We extend the data back to January 1975 using the growth rate of U.S. total petroleum stocks which are reported in the MER (Table 3.4). The monthly spot price for Brent crude oil quoted in dollars per barrel has been published by the EIA since 1987.5;[13] for the period until 1975.1, we use the growth rate of the monthly refiner acquistion cost of imported crude oil obtained from the MER (Table 9.1) to extend Brent prices backward. Monthly prices for unleaded regular gasoline expressed in dollars per gallon are reported in the MER (Table 9.4) from 1976.1 onward which we backcast to 1975.1 with the growth rate of monthly prices for leaded gasoline available from the same source. To compute the spread between gasoline and Brent prices, we convert the gasoline price into dollars per barrel by multiplying it by 42. Monthly data on international rig counts start in January 1975 and weekly counts for rotary drilling rigs in the US and Canada go back to January 2000; they are released by Baker Hughes.[14] Worldwide oil rig counts are seasonally adjusted using the X13-ARIMA procedure. The CRB/BLS spot price index of raw industrial materials is obtained from Global Financial Data. The U.S. consumer price index for all urban consumers is taken from the FRED database. The global economic conditions (GECON) indicator is available on Christiane Baumeister's webpage.[15]

D Results for the Full Hold-Out Period

Here we report the average quantile score ratios and the average quantile-weighted CRPS ratios, both relative to the no-change forecast, for the entire hold-out period that includes four additional years that were characterized by some major turmoil in the oil market and the global economy. Tables A1 and A2 show that our results are robust to this turbulent period and real-time data constraints; if anything, adding these features further improves the tail forecasting performance of our models at longer horizons.

[13] See https://www.eia.gov/dnav/pet/hist/rbrteM.htm
[14] See https://rigcount.bakerhughes.com/
[15] See https://sites.google.com/site/cjsbaumeister/datasets

Table A1: Average quantile score ratios relative to no-change forecast of real Brent price
Hold-out period: 1992.1-2024.1

Monthly horizon	Model	Quantiles						
		0.05	0.10	0.25	0.50	0.75	0.90	0.95
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
$h = 1$	BVAR-SV	**0.87**	**0.95**	**0.99**	1.00	1.04	1.02	**0.99**
	mixBART-SV	**0.89**	**0.96**	1.00	**0.99**	1.01	**0.99**	**0.95**
	BQR	2.26	1.61	1.08	**0.97**	1.27	2.10	3.21
$h = 3$	BVAR-SV	1.03	**0.97**	**0.98**	1.03	**0.98**	**0.86**	**0.85**
	mixBART-SV	1.03	**0.96**	**0.98**	1.02	**0.97**	**0.84**	**0.83**
	BQR	2.35	1.55	1.07	1.02	1.26	1.80	2.72
$h = 6$	BVAR-SV	**0.86**	**0.80**	**0.83**	1.01	**0.81**	**0.68**	**0.70**
	mixBART-SV	**0.84**	**0.78**	**0.83**	1.01	**0.80**	**0.65**	**0.67**
	BQR	1.94	1.28	**0.92**	1.01	1.03	1.37	2.07
$h = 9$	BVAR-SV	**0.73**	**0.67**	**0.74**	**0.99**	**0.68**	**0.56**	**0.59**
	mixBART-SV	**0.70**	**0.66**	**0.73**	**0.99**	**0.66**	**0.53**	**0.54**
	BQR	1.61	1.07	**0.80**	**0.99**	**0.87**	1.11	1.67
$h = 12$	BVAR-SV	**0.64**	**0.60**	**0.67**	**0.99**	**0.60**	**0.50**	**0.54**
	mixBART-SV	**0.62**	**0.59**	**0.67**	**0.99**	**0.59**	**0.46**	**0.49**
	BQR	1.42	**0.94**	**0.72**	1.00	**0.78**	**0.98**	1.47

Notes: Boldface indicates improvements relative to the no-change forecast. Green shaded cells indicate the best-performing model for a particular forecast horizon.

Table A2: Average quantile-weighted CRPS ratios relative to no-change forecast of real Brent
Hold-out period: 1992.1-2024.1

Monthly horizon	Model	Quantile weights			
		left tail	center	right tail	both tails
		(1)	(2)	(3)	(4)
$h = 1$	BVAR-SV	**0.98**	1.00	1.01	**0.98**
	mixBART-SV	**0.98**	1.00	**0.99**	**0.97**
	BQR	1.26	1.12	1.44	1.73
$h = 3$	BVAR-SV	1.00	1.00	**0.97**	**0.96**
	mixBART-SV	**0.99**	**0.99**	**0.96**	**0.95**
	BQR	1.28	1.14	1.41	1.66
$h = 6$	BVAR-SV	**0.88**	**0.90**	**0.84**	**0.80**
	mixBART-SV	**0.87**	**0.90**	**0.83**	**0.78**
	BQR	1.13	1.03	1.21	1.37
$h = 9$	BVAR-SV	**0.78**	**0.82**	**0.73**	**0.68**
	mixBART-SV	**0.77**	**0.81**	**0.71**	**0.66**
	BQR	1.00	**0.93**	1.05	1.15
$h = 12$	BVAR-SV	**0.72**	**0.76**	**0.67**	**0.61**
	mixBART-SV	**0.71**	**0.75**	**0.65**	**0.59**
	BQR	**0.91**	**0.86**	**0.96**	1.03

Notes: See Table A1.

References

Aastveit, Knut Are, Jamie L. Cross, and Herman K. van Dijk (2023). "Quantifying Time-Varying Forecast Uncertainty and Risk for the Real Price of Oil," *Journal of Business and Economic Statistics* 41(2): 523-537.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019). "Vulnerable Growth," *American Economic Review* 109: 1263-1289.

Aguilar, Omar, and Mike West (2000). "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business and Economic Statistics* 18: 338-357.

Alquist, Ron, Lutz Kilian, and Robert Vigfusson (2013). "Forecasting the Price of Oil," in: Graham Elliott and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2A, Amsterdam: North-Holland, 427-507.

Baumeister, Christiane (2022). "Recent Developments in (Modeling) Energy Market Dynamics," keynote at the workshop on "Energy and Climate: Macroeconomic Implications," BI Norwegian Business School, Oslo, August 18.

Baumeister, Christiane (2023). "Pandemic, War, Inflation: Oil Markets at a Crossroads?" NBER Working Paper 31496.

Baumeister, Christiane, and Pierre Guérin (2021). "A Comparison of Monthly Global Indicators for Forecasting Growth," *International Journal of Forecasting* 37(3): 1276-1295.

Baumeister, Christiane, Pierre Guérin, and Lutz Kilian (2015). "Do High-Frequency Financial Data Help Forecast Oil Prices? The MIDAS Touch at Work," *International Journal of Forecasting* 31(2): 238-252.

Baumeister, Christiane, and Lutz Kilian (2012). "Real-Time Forecasts of the Real Price of Oil," *Journal of Business and Economic Statistics* 30(2): 326-336.

Baumeister, Christiane, and Lutz Kilian (2014a). "What Central Bankers Need to Know About Forecasting Oil Prices," *International Economic Review* 55(3): 869-889.

Baumeister, Christiane, and Lutz Kilian (2014b). "Real-Time Analysis of Oil Price Risks Using Forecast Scenarios," *IMF Economic Review* 62(1): 119-145.

Baumeister, Christiane, and Lutz Kilian (2015). "Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach," *Journal of Business and Economic Statistics* 33(3): 338-351.

Baumeister, Christiane, Lutz Kilian, and Xiaoqing Zhou (2018). "Are Product Spreads Useful for Forecasting Oil Prices? An Empirical Evaluation of the Verleger Hypothesis," *Macroeconomic Dynamics* 22(3): 562-580.

Baumeister, Christiane, Dimitris Korobilis, and Thomas K. Lee (2022). "Energy Markets and Global Economic Conditions," *Review of Economics and Statistics* 104(4): 828-844.

Baumeister, Christiane, and Gert Peersman (2013). “The Role of Time-Varying Elasticities in Accounting for Volatility Changes in the Crude Oil Market,” *Journal of Applied Econometrics* 28(7): 1087-1109.

Bernard, Jean-Thomas, Lynda Khalaf, Maral Kichian, and Clement Yelou (2018). “Oil Price Forecasts for the Long Term: Expert Outlooks, Models, or Both?” *Macroeconomic Dynamics* 22(3): 581-599.

Bhattacharya, Anirban, and David B. Dunson (2011). “Sparse Bayesian Infinite Factor Models,” *Biometrika* 98: 291-306.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016). “Common Drifting Volatility in Large Bayesian VARs,” *Journal of Business and Economic Statistics* 34: 375-390.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019). “Large Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors,” *Journal of Econometrics* 212(1): 137-154.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2022). “Nowcasting Tail Risk to Economic Activity at a Weekly Frequency,” *Journal of Applied Econometrics* 37: 843-866.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (forthcoming). “Capturing Macroeconomic Tail Risks with Bayesian Vector Autoregressions,” *Journal of Money, Credit and Banking.*

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2022). “Addressing COVID-19 Outliers in BVARs with Stochastic Volatility,” *Review of Economics and Statistics.*

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010). “The Horseshoe Estimator for Sparse Signals,” *Biometrika* 97: 465-480.

Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (2010). “BART: Bayesian Additive Regression Trees,” *Annals of Applied Statistics* 4: 266-298.

Clark, Todd E. (2011). “Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility,” *Journal of Business and Economic Statistics* 29(3): 327-341.

Clark, Todd E., Florian Huber, Gary Koop, Massimiliano Marcellino, and Michael Pfarrhofer (2023). “Tail Forecasting with Multivariate Bayesian Additive Regression Trees,” *International Economic Review* 64: 979-1022.

Clark, Todd E., and Francesco Ravazzolo (2015). “Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility,” *Journal of Applied Econometrics* 30(4): 551-575.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984). “Forecasting and Conditional Projection Using Realistic Prior Distributions,” *Econometric Reviews* 3(1): 1-100.

Dossani, Asad, and John Elder (2024). "Uncertainty and Investment: Evidence from Domestic Oil Rigs," *Journal of Futures Markets* 44(2): 323-340.

Foroni, Claudia, Massimiliano Marcellino, and Dalibor Stevanovic (2022). "Forecasting the Covid-19 Recession and Recovery: Lessons from the Financial Crisis," *International Journal of Forecasting* 38(2): 596-612.

Friedman, Jerome H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 1189-1232.

Garratt, Anthony, Shaun P. Vahey, and Yunyi Zhang (2019). "Real-Time Forecast Combinations for the Oil Price," *Journal of Applied Econometrics* 34(3): 456-462.

Giacomini, Raffaella, and Ivana Komunjer (2005). "Evaluation and Combination of Conditional Quantile Forecasts," *Journal of Business and Economic Statistics* 23: 416-431.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt (2016). "Systemic Risk and the Macroeconomy: An Empirical Evaluation," *Journal of Financial Economics* 119: 457-471.

Gneiting, Tilmann, and Roopesh Ranjan (2011). "Comparing Density Forecasts Using Threshold and Quantile-weighted Scoring Rules," *Journal of Business and Economic Statistics* 29: 411-422.

Hamilton, James D. (2003). "What Is an Oil Shock?" *Journal of Econometrics* 113: 363-398.

Hamilton, James D. (2009). "Understanding Crude Oil Prices," *The Energy Journal* 30(2): 179-206.

Huber, Florian, and Martin Feldkircher (2019). "Adaptive Shrinkage in Bayesian Vector Autoregressive Models," *Journal of Business and Economic Statistics* 37: 27-39.

Huber, Florian, Gary Koop, Luca Onorante, Michael Pfarrhofer, and Josef Schreiner (2023). "Nowcasting in a Pandemic Using Non-Parametric Mixed Frequency VARs," *Journal of Econometrics* 232: 52-69.

Huber, Florian, and Luca Rossini (2022). "Inference in Bayesian Additive Vector Autoregressive Tree Models," *Annals of Applied Statistics* 16: 104-123.

Kastner, Gregor (2019). "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions," *Journal of Econometrics* 210: 98-115.

Kastner, Gregor, and Sylvia Frühwirth-Schnatter (2014). "Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models," *Computational Statistics and Data Analysis* 76: 408-423.

Kastner, Gregor, and Florian Huber (2020). "Sparse Bayesian Vector Autoregressions in Huge Dimensions." *Journal of Forecasting* 39(7): 1142-1165.

Koenker, Roger, and Gilbert Bassett, Jr. (1978). "Regression Quantiles," *Econometrica* 46(1): 33-50.

Kozumi, Hideo, and Genya Kobayashi (2011). "Gibbs Sampling Methods for Bayesian Quantile Regression," *Journal of Statistical Computation and Simulation* 81: 1565-1578.

Pitt, Michael K., and Neil Shephard (1999). "Time-Varying Covariances: A Factor Stochastic Volatility Approach," in: Bernardo, Jose M, James O. Berger, A. Philip Dawid, and Adrian F.M. Smith (eds.), *Bayesian Statistics – Proceedings of the Sixth Valencia International Meeting*, Oxford University Press, 547-570.

Plante, Michael D., and Kunal Patel (2019). "Breakeven Oil Prices Underscore Shale's Impact on the Market," Dallas Fed Economics, May 21.

Primiceri, Giorgio E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72(3): 821-852.