Oil price expectations in explosive phases*

Robinson Kruse-Becher† Philip Letixerant‡

January 30, 2024

Abstract

Expectations regarding future oil prices are of great importance for a variety of economic and financial applications. The state-of-the-art methodology based on a Gaussian affine term structure model provides monthly risk-adjusted financial market expectations (Hamilton and Wu, 2012, 2014). The term structure model outperforms a number of competitors, see Baumeister (2023). We investigate the potentially time-varying performance of these oil price expectations and those from the Energy Information Administration (see Garratt, Vahey, and Zhang, 2019). It turns out that the superiority of the term structure model over a simple no-change benchmark is not only time-varying, but also characterized by phases in which the ranking among the competing expectations are reversed. Importantly, these phases are characterized by locally explosive oil price behavior and subsequent crashes (see Pavlidis, Paya, and Peel, 2018). To this end, we study the role of established real-time oil price shock measures (see Kilian and Vigfusson, 2011, 2013) alongside a newly proposed indicator which is obtained from a recursive monitoring test statistic against explosiveness (see Phillips and Shi (2018)). From a perspective of conditional expected performance (see e.g. Granziera and Sekhposyan, 2019), we investigate the predictability of the loss differential by different shock series. Our results underline the importance of the new indicator reflecting temporary exuberance and subsequently collapsing prices. Our findings have consequences for the literature on oil price speculative bubble testing and beyond.

Keywords: Market-based expectations, Conditional predictive ability, Best subset selection, General-to-specific modelling

JEL classification: C22, C58, Q41, Q47

*The authors would like to thank Joscha Beckmann, Bent-Jesper Christensen, Robert Czudaj, Isabel Figuerola-Ferretti, Pascal Goemans, Marco Kerkemeier, Michal Rubaszek, Christoph Wegener and the participants of the Workshop on expectations and sentiments for energy price dynamics in Berlin 2023 for helpful comments and discussions. Christiane Baumeister and Yunyi Zhang kindly provided data on oil price market expectations and Energy Information Administration forecasts.

†University of Hagen, Faculty of Business Administration and Economics and Center for Economic and Statistical Analysis (CESA), `robinson.kruse-becher@fernuni-hagen.de`.

‡University of Hagen, Faculty of Business Administration and Economics and Center for Economic and Statistical Analysis (CESA), `philip.letixerant@fernuni-hagen.de`.

1 Introduction

Oil price expectations are of great importance in economics and finance not only to producers and consumers, but also to investors, regulators and policy makers, see e.g. Coibion, Gorodnichenko, Kumar, and Pedemonte (2020) and Kilian and Zhou (2022). Oil price forecasts are, for instance, important predictors for future economic activity, see e.g. Alquist and Kilian (2010) and Kilian and Vigfusson (2013). Generally, oil futures prices are informative to measure market expectations. Other sources, see Baumeister and Kilian (2016) for a general discussion, are survey expectations (e.g. Consensus Economics, see Alquist, Kilian, and Vigfusson, 2013), analysts forecasts (e.g. Bloomberg, see Figuerola-Ferretti, Rodríguez, and Schwartz, 2021) and institutions as the Energy Information Administration (EIA), see Garratt et al. (2019).

Forecast efficiency regressions have shown that futures prices are not unbiased predictors of future spot prices in the oil market (and elsewhere), see Baumeister (2023) for a recent and excellent survey article. One source, and possibly the main driver, is a time-varying risk premium. The arbitrage-free affine Gaussian term structure model proposed by Hamilton and Wu (2012, 2014) allows us to extract the market-based expectations from oil futures contracts and to estimate the time-varying risk premium. It uses a small number of dynamic stochastic and latent risk pricing factors to model payoffs on a long position for a futures contract with a maturity of h months. The oil price market expectations extracted via the Hamilton and Wu (2014) model are state-of-the-art as reviewed in Baumeister (2023). The extracted market-based expectations provide the strongest reduction of the mean squared prediction error against no-change in an unconditional evaluation against a large number of competing approaches.[1]

Diebold-Mariano test statistics clearly indicate that market-based expectations strongly outperform the benchmark no-change forecasts. Henceforth, the extracted market-based expectations are used in many applications, e.g. testing for speculative bubbles in the oil market, see Pavlidis et al. (2018). The key ingredient of the Diebold-Mariano test statistic is the loss differential, see Diebold and Mariano (1995). Under quadratic loss, the loss differential is given as the difference between squared prediction errors from the Hamilton and Wu (2014) [HW] approach and squared prediction errors obtained from no-change forecasts which also form the benchmark in our analysis.

We ask whether market expectations outperform benchmark forecasts throughout the whole sample and whether there are some state variables explaining possible time-variation.

[1] Under quadratic loss, as shown in Granger (1969) and Granger and Newbold (1986), the conditional expectation is the minimizer of the mean squared prediction error.

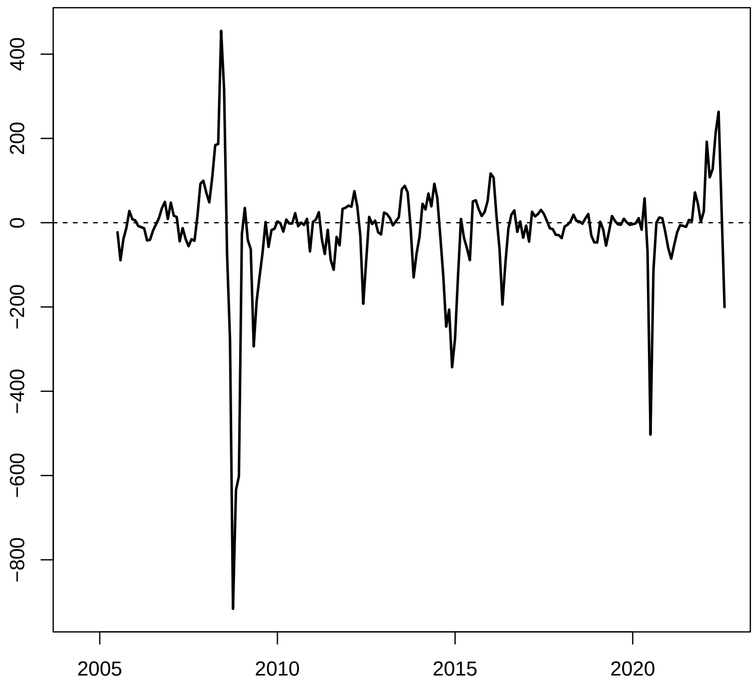

Figure 1: Loss differential (HW-NC), $h = 3$

A visual exploration of the loss differential in Figure 1, reveals potential 'pockets' in which the loss differential turns positive such that the no-change forecast [NC] outperforms market expectations.[2] The timing of these temporary phases are remarkably connected to episodes in which the oil price was subject to large fluctuations in 2007/2008, 2014, 2020 and 2022. As it turns out, these phases can be characterized by explosivity in the oil price. Explosiveness in oil prices itself has been investigated in a large number of contributions in the context of speculative bubbles, e.g. Fantazzini (2016), Gronwald (2016), Caspi, Katzke, and Gupta (2018), Pavlidis et al. (2018), Kruse and Wegener (2020), Figuerola-Ferretti et al. (2020) and Kruse-Becher (2023). Overall, the loss differential displays strong time-variation in the mean which is potentially related to a couple of variables in a dynamic way.

In this work, we switch the perspective from an unconditional evaluation to a conditional one. Such a framework enables us to investigate dynamic conditional predictability of the loss differential, see Giacomini and White (2006). In practice, we are not only interested

[2]The term 'pockets' is borrowed from the stock return predictability literature, see Farmer, Timmermann and Schmidt (2023).

in knowing whether one method is better than another on average, but we are also interested in which economic and financial circumstances, say boom or bust, this happens. While the unconditional Diebold-Mariano test is informative about the relative performance of forecasting methods on average, Giacomini and White (2006) propose testing the hypothesis of conditional equal predictive ability to investigate state-dependent forecasting performance. The question is, if there is any information (that is available at the time when the forecasts were made) which is able to explain the relative predictive performances of the methods. Thus, the null hypothesis in the framework of conditional equal predictive ability is that conditional expected squared loss of different forecasting methods are identical across all conditioning economic and financial states. Hence, we apply the Giacomini and White (2006) test to study how the relative predictive performances of the methods evolve in response to oil price shocks and a newly proposed real-time indicator for explosivity.

As in Granziera and Sekhposyan (2019), we thus investigate whether the relative predictive performance is predictable itself. To this end, we study a number of established oil price shocks, see Kilian and Vigfusson (2013). In order to precisely analyze the role of explosivity in this context, we suggest a novel real-time predictor generated from a monitoring statistic against explosiveness in oil prices, see Phillips and Shi (2018) and Phillips and Shi (2020). In order to select the predictors for the conditional predictive ability regressions, we compare different strategies. Among these are best subset selection (see e.g. Bertsimas, King, and Mazumder, 2016 and Witten and James, 2013) based on different selection criteria, i.e. Mallows criterion and Schwarz criterion, and general-to-specific modelling (see e.g. Hendry and Clements, 2003, Campos, Ericsson, and Hendry, 2005 and Pretis, Reade, and Sucarrat, 2018). Our results show clear evidence against the null hypothesis of no conditional predictive ability. Most important predictors are net price increases, implied volatility of oil price changes and the newly proposed explosivity indicator.

A rejection of the conditional predictive ability null hypothesis indicates that the loss differential depends on additional available information which is not included in the methods. Rejections can be interpreted as evidence of misspecification of the original methods and of non-optimality of the resulting forecasts. One direct implication is that forecasts might be improved by exploiting the additional available information in an appropriate way. However, in practice, it could be very difficult to include such information in existing specifications. A practical alternative is to consider dynamic conditional rotation in which the researches switches between two forecasts.

In an exercise of dynamic conditional rotation (Zhu and Timmermann, 2022), we study the possibility of switching between market-based expectations and no-change forecasts based on a prediction of the loss differential itself. Our results from an ex-post dynamic conditional rotation approach show that some, partially major (relative to the maximal achievable improvement), significant gains are possible at short and medium horizons to improve market expectation measures especially in times of large fluctuations in the oil price as these are the phases in which the performance of market-based expectations deteriorates relative to no-change forecasts. Depending on the particular application, either ex-post rotation is relevant (e.g. speculative bubble testing) or dynamic prediction of the loss differential (e.g. real-time analysis).

Overall, our results imply that market-based expectations neglect information in the explosive episodes. For EIA forecasts, which are widely used by policymakers and the energy industry in their decision making, we obtain similar findings. Our findings thus hold for opinion- and market-based expectations. Both, risk premia and information rigidities, may play a role. Hence, uses of oil price expectations, in particular testing against speculative bubbles, Pavlidis et al. (2018) might be directly affected. Their testing procedures directly rely on (explosive) spot and expected oil prices. Obviously, the (ex-post) measurement of expectations is key to the outcomes of and conclusions from such an analysis and these are affected by potential misspecification and non-optimality of oil price forecasts. Our results thus have important implications for the literature on testing against speculative bubbles and many other applications in which oil price expectations play a decisive role, e.g. regulation of oil markets as well as general macroeconomic projections.

The remainder of the paper is organized as follows. The econometric framework is given in Section 2. Our data is described in Section 3. Section 4 covers the empirical results and discussions. Conclusions are drawn in Section 5. The Appendix contains additional Tables and Figures.

2 Econometric Framework

Suppose there are sequences of forecasts $\widehat{y}_{j,t+h}$ formed in period $t = 1, ..., P$ for the target variable (oil price) y_{t+h} that shall be evaluated. The forecast horizons are $h = \{3, 6, 9, 12\}$ months. The benchmark prediction is labeled as $j = 1$ and the competing forecast is signified by $j = 2$. In the evaluation, we are not only interested in metrics like the mean absolute/squared error, but in particular in testing the relative forecasting accuracy. Furthermore, a criterion is established to switch between forecasts at specific points in time instead of combining forecasts (by averaging over them or estimating combination weights).

The popular unconditional test as in Diebold and Mariano (1995) investigates whether two competing forecasts differ over a whole evaluation period. Such a test has two possible extensions. First, a researcher or practitioner may be interested in time-varying unconditional predictive ability as in Giacomini and Rossi (2010). Such a test does not reveal insights about the drivers of possible differences in the predictive ability. Second, given a set of state or conditioning variables, one may be interested whether these hold predictive power for the relative forecasting accuracy of the predictions at hand, leading to a test of conditional predictive ability, see Giacomini and White (2006).

A rejection of such a test can be informative about potential misspecification and non-optimality of the forecasts. Theoretically the forecasts can be improved, by conditioning the loss differential on a set of state variables and then rotating between the predictions depending on the effect of the conditioning variables. This leads to the dynamic conditional rotation approach. The main idea is to predict the sign of the loss differential based on the conditioning variables and to switch between the two forecasts accordingly. Zhu and Timmermann (2022) derive theoretical conditions under which gains in forecasting accuracy are achievable.

2.1 Unconditional predictive ability

A starting point for our analysis is the well known test of unconditional predictive ability proposed by Diebold and Mariano (1995). It is based on the loss differential ΔL_t which is calculated via the errors of two competing forecasts. While principally allowing for other loss functions, we follow most of the literature and use the quadratic loss. The forecast errors for the h-step prediction are:

$$e_{j,t+h} = y_{t+h} - \widehat{y}_{j,t+h} \tag{1}$$

and the loss differential for the h-step forecast thus formulates as:

$$\Delta L_{t+h} = e^2_{1,t+h} - e^2_{2,t+h}. \tag{2}$$

A positive value of the loss differential thus indicates superiority of the competing forecast ($j = 2$), while a negative loss differential suggests, that the benchmark produces smaller loss.

The test statistic for the null hypothesis of equal forecast accuracy, i.e. $H_0: \ E[\Delta L_{t+h}] = 0$

formulates as:

$$DM = \sqrt{P}\frac{\Delta\overline{L}}{\widehat{\sigma}_{\Delta\overline{L}}} \sim N(0,1), \tag{3}$$

where $\Delta\overline{L}$ is the mean of the loss differential, $\widehat{\sigma}_{\Delta\overline{L}}$ is a HAC-type estimator of the loss differentials long-run standard deviation and P is the number of evaluated forecasts.[3] By rejecting the null, it can be concluded that the forecasts do not exhibit equal predictive ability. However, it does not inform us about time-variation, nor about phases in or conditions under which one forecast produces smaller squared errors than the other. This leads us to the test of conditional predictive ability.

2.2 Conditional predictive ability

Crucial for the following analysis is the test of conditional predictive ability which originates in the seminal work of Giacomini and White (2006). This setup assesses the relative performance of two forecasts by conditioning the loss differential on a set of state variables s_t. Within this framework, the relative predictive ability is not only compared over the whole sample, but also time periods or states of the conditioning variables are taken into consideration.

Advantages of this test are the possibility to not only compare the forecasts, but the null hypothesis evaluates the complete forecasting setup in form of the estimation method or window size and allows for nested as well as non-nested setups. Furthermore, the unconditional test (e.g. Diebold and Mariano, 1995) is a special case of the Giacomini and White (2006) setup, when the set of conditioning state variables is empty.

As well as the Diebold-Mariano test, the test of conditional predictive accuracy can be integrated into a regression framework:

$$\Delta L_{t+h} = \theta' s_t + \varepsilon_t, \tag{4}$$

where θ is the parameter vector of dimension p to be estimated and s_t is the vector of p conditioning variables including an intercept. The innovation term is labeled as ε_t.[4] The Wald-type test statistic for the null hypothesis of $E[\Delta L_{t+h}|s_t] = E[s_t \Delta L_{t+h}] = 0$ can be

[3]Clearly, P might depend on the forecast horizon h under investigation, but for notational convenience, we suppress this dependence.

[4]We allow for autocorrelation and heteroskedasticity and use HAC covariance matrix estimates accordingly.

calculated as:

$$GW = P\left(P^{-1}\sum_{t=1}^{P} s_t \Delta L_{t+h}\right)' \widehat{\Omega}^{-1}\left(P^{-1}\sum_{t=1}^{P} s_t \Delta L_{t+h}\right) \sim \chi_p^2, \tag{5}$$

with $\widehat{\Omega}$ being a HAC-type estimator of the variance of $P^{-1/2}\sum_{t=1}^{P} s_t \Delta L_{t+h}$. Alternatively, the test could be implemented as a Wald-type test with $H_0: \theta = 0$ resulting from Equation 4.

Following Granziera and Sekhposyan (2019), we do not only report the Wald-type statistic GW, but also the adjusted R^2 from the dynamic linear regression in Equation 4 and the individual t-statistics for each included conditioning variable. This is of utmost importance for analyzing the relevance of the respective variable. Moreover, significant variables can be interpreted with respect to their coefficients direction, allowing economic interpretation of the loss differentials predictability depending on the respective state variable.

2.3 Dynamic conditional rotation

When rejecting the null hypothesis of equal conditional predictive ability, this has actual consequences for selecting a forecast. Following Giacomini and White (2006) the rejection is due to the predictive ability the state variables hold for the loss differentials. Consequently, this means that not only the forecast of the target variable itself is of particular interest, but also the loss differential of selected forecasts.

Furthermore, Granziera and Sekhposyan (2019) state a case of major interest for practical use. In case the test of unconditional predictive ability does not reject the null, but the conditional test does, the forecasting performance can be treated as equal on average, yet the relative performance of the forecasts differs and can be predicted with help of s_t.

More precisely, the loss differential ΔL_{t+h} is regressed on the state variables s_t as in Zhu and Timmermann (2022). The ex-post predictions resulting from the estimated regression coefficients and s_t can be used to forecast the loss differential itself. In order to do so, the fitted values are computed: $E[\Delta L_{t+h}|s_t] = \theta' s_t$ or the out-of-sample prediction: $E[\Delta L_{T+h}|s_T] = \theta' s_T$. To decide which forecast is chosen in period t, a threshold c has to be chosen. The natural choice would be $c = 0$ such that the benchmark is chosen when $\theta' s_t \leq 0$ and the competing forecast is selected when $\theta' s_t > 0$. The final prediction via dynamic rotation [DR], $\widehat{y}_{DR,t+h}$ then results as:

$$\widehat{y}_{DR,t+h} = \widehat{y}_{1,t+h} \cdot \mathbb{1}[\theta' s_t \leq 0] + \widehat{y}_{2,t+h} \cdot \mathbb{1}[\theta' s_t > 0]. \tag{6}$$

To asses the usefulness of the conditioning variables, the rotation as well as the competing forecast, we apply three metrics: The relative MSE between the dynamic conditional rotation and the benchmark and what we call the relative performance (RP) and the success ratio (SR). If the relative MSE takes values below unity, the dynamic rotation offers an improvement over the benchmark in terms of MSE accuracy and if it exceeds unity, the benchmark alone should be preferred with regard to the MSE.

We define the relative performance as the percentage of times in which the competing forecast is chosen over the benchmark in the rotation which is the case when the fitted value exceeds zero:

$$RP = \frac{1}{P}\sum_{t=1}^{P} \mathbb{1}[\theta' s_t > 0] \;\in\; [0,1]. \tag{7}$$

Larger values of RP indicate a stronger relevance of the benchmark forecast.

The success ratio is defined as the percentage of times the dynamic conditional rotation has correctly predicted the sign of the loss differential, i.e. the forecast with the smaller squared forecast error based on the conditioning variable:

$$SR = \frac{1}{P}\sum_{t=1}^{P} \mathbb{1}[\theta' s_t > 0] \cdot \mathbb{1}[\Delta L_{t+h} > 0] + \mathbb{1}[\theta' s_t \leq 0] \cdot \mathbb{1}[\Delta L_{t+h} \leq 0] \;\in\; [0,1]. \tag{8}$$

The reference value for SR equals 0.5. A value exceeding this threshold indicates a better performance and vice versa.

Obviously, to make the preceding metrics feasible, we replace the parameter vector θ by its least squares estimate.

2.4 Oil price shocks

Due to the relevance of the oil price for the overall economy, the literature on oil price shocks is large. Early literature, e.g. Hamilton (1983) focused especially on positive increases in the oil price and finds a strong connection between oil prices and GDP growth. Negative shocks are considered as well, e.g. in Mork (1989) who finds asymmetric effects on the GDP growth depending on the kind of oil shock. Before introducing oil price shock measures, we define an oil shock according to Hamilton (2003) who considers oil price shocks primarily as reactions to exogenous political or military shocks, rather than endogenous responses to the economy. Nonetheless, Hamilton (2003) finds oil price shock measures that are able to filter out movements in the oil price which do not stem from exogenous shocks. This motivates the introduction of the following established measures.

When choosing conditioning variables for oil price forecasts, conventional oil price shock measures are an obvious choice. For this we rely on the following oil shock measures as they can be found in Kilian and Vigfusson (2013) and Nonejad (2021). In order to take care of the asymmetries and non-linearities, several oil shock measures are introduced, focusing on positive and negative shocks, and also large or net changes.

The first measure is the three-year net oil price increase net_t^+, as proposed in Hamilton (1996) and extensively studied in Kilian and Vigfusson (2013). In order to define net_t^+, we first introduce $oil_t^{\max}$ as the three-year maximum oil price $oil_t^{\max} = \max\{oil_{t-1}, \ldots, oil_{t-36}\}$.[5] It follows

$$net_t^+ = \max\{0, oil_t - oil_t^{\max}\}.$$

Analogously, we define the counterpart net_t^- as the three-year net oil price decrease to account for asymmetry: $net_t^- = \min\{0, oil_t - oil_t^{\min}\}$ with $oil_t^{\min} = \min\{oil_{t-1}, \ldots, oil_{t-36}\}$. Instead of taking the net change into account, the change from the highest price in recent history can be calculated:

$$gap_t = oil_t - oil_t^{\max}.$$

In order to also account for non-linearities, large changes are considered via the measures $large_t$ and $large_t^+$. First, we look at ovx_t which is the CBOE Crude Oil ETF Volatility Index measuring forward-looking oil price volatility. From this we derive $large_t$ as a case in which the absolute oil price change exceeds the implied standard deviation resulting from ovx_t:

$$large_t = \Delta oil_t \cdot \mathbb{1}\left(|\Delta oil_t| > \sqrt{ovx_t}\right).$$

Analogously, we define large positive changes in which the oil price change exceeds the implied standard deviation:

$$large_t^+ = \Delta oil_t \cdot \mathbb{1}\left(\Delta oil_t > \sqrt{ovx_t}\right).$$

Due to the use of ovx_t instead of the full sample standard deviation, the construction of $large_t$ and $large_t^+$ are essentially measured in real-time. This collection of established oil price shocks serve as predictors in the conditional ability testing framework.

[5] Please note that we deviate from the original notation $\dot{oil}_t$ in order to increase readability.

2.5 Monitoring explosive oil prices

We construct a new real-time predictor indicating explosiveness in the oil market. To this end, we make use of the established econometric methodology developed in Phillips, Shi, and Yu (2015a,b), Phillips and Shi (2018) and Phillips and Shi (2020).[6] The main idea is to investigate explosiveness in prices via recursions on a right-tailed unit root statistic, i.e. the augmented Dickey-Fuller (ADF) statistic. This procedure allows for multiple temporary explosive phases and has been further developed in a monitoring context, see Phillips and Shi (2018). Let us denote the monitoring backward supremum augmented Dickey-Fuller statistic as $mBSADF_t$. The monitoring procedure is initialized after $s_0 = \lfloor T(0.01+1.8/\sqrt{T}) \rfloor$ months, see Phillips and Shi (2018) and henceforth updated in a recursive way until the sample is completed. Our newly proposed indicator extracts information on the explosiveness of oil prices and its strength in real-time. It consists of two parts and is constructed as

$$expl_t = mBSADF_t \cdot \mathbb{1}\left(mBSADF_t > 0\right) \;\in\; [0, \infty)$$

with $mBSADF_t$ being the real-time monitoring statistic against explosiveness in the oil price at time t (initialized at s_0).[7] Positive values of the $mBSADF_t$ statistic[8] indicate explosive behavior and are captured by the second part. Hence, a positive value is due to an autoregressive coefficient exceeding unity which indicates an explosive root in the underlying autoregressive process.

Now, we isolate the information on explosivity by truncating the monitoring statistic at zero (from below) and keeping the positive values only. Hence, the newly proposed indicator exploits not only information on explosivity in real-time, but also reflects its strength. The larger the indicator deviates from zero, the stronger is the explosivity signal.

Importantly, our measure is not subject to a subjective nominal significance level as we are not interested in testing against explosiveness, but rather aim at an indicator for explosive behavior. This goal is achieved by the newly constructed variable $expl_t$. Furthermore, this construction is in line with the other previously discussed predictors and is also based on recent and lagged oil prices.

[6]The general idea of monitoring explosive bubble processes was first suggested by Homm and Breitung (2012) in this context.

[7]For a detailed exposition and asymptotic properties, the interested reader is referred to Phillips and Shi (2018).

[8]The well-known Dickey-Fuller statistic is given as a t-ratio of the autoregressive coefficient centered at unity, divided by its standard error.

Anundsen (2015) considers a bubble indicator constructed by the p-value of a cointegration statistic in the context of housing prices, see also Mikhed and Zemčík (2009b,a). Their bubble indicator is thus constrained to take values in the $[0, 1]$-interval and is used in a Granger causality analysis. The similarity to our measure is limited to the usage of the recursive BSADF statistic and its p-value is related to the strength of explosivity. Main differences are, however, that we focus in particular on positive values of the BSADF statistic only, thereby excluding all other irrelevant information. Second, we do not restrict the predictor to the unit interval by a nonlinear p-value transformation, but take the information in the BSADF statistic directly into account in a linear fashion (after truncation). These features are advantageous to extract the most relevant information regarding explosivity from the BSADF statistic. Moreover, our measure is constructed in real-time which is an important feature in the comparison to other real-time predictors.

3 Data

Our sample runs from July 2005 to August 2022 yielding $T = 206$ monthly observations. The sample start and end are restricted to the availability of EIA forecast which are essential to the analysis. The data is obtained from Christiane Baumeister's homepage (market-based expectations), Yunyi Zhang's homepage (updated and structured EIA forecasts) and the FRED data base (WTI oil price and CBOE crude oil ETF volatility index). The established predictors are constructed as outlined in the previous section, see Figure 2. In Figure 3, we display the newly proposed explosivity predictor, the underlying monitoring statistic $mBSADF_t$ and the re-scaled oil price series.[9] The loss differential series are studied at the horizons $h = \{3, 6, 9, 12\}$. Figure 1 (located in the introduction) and Figures 5-7 in the Appendix show the different loss differentials for market-based expectations. In our empirical analysis, we consider the loss differential between (i) market-based expectations and no-change forecasts and (ii) EIA forecasts and no-change forecasts. In total, we analyze eight different loss differential series in the (un)conditional predictive ability setting.

[9] We use two lags for the BSADF monitoring statistic to capture the dynamics in the oil price and to generate the explosivity predictor. The monitoring procedure is initialized after $s_0 = 28$ months according to the rule in Section 2.5.

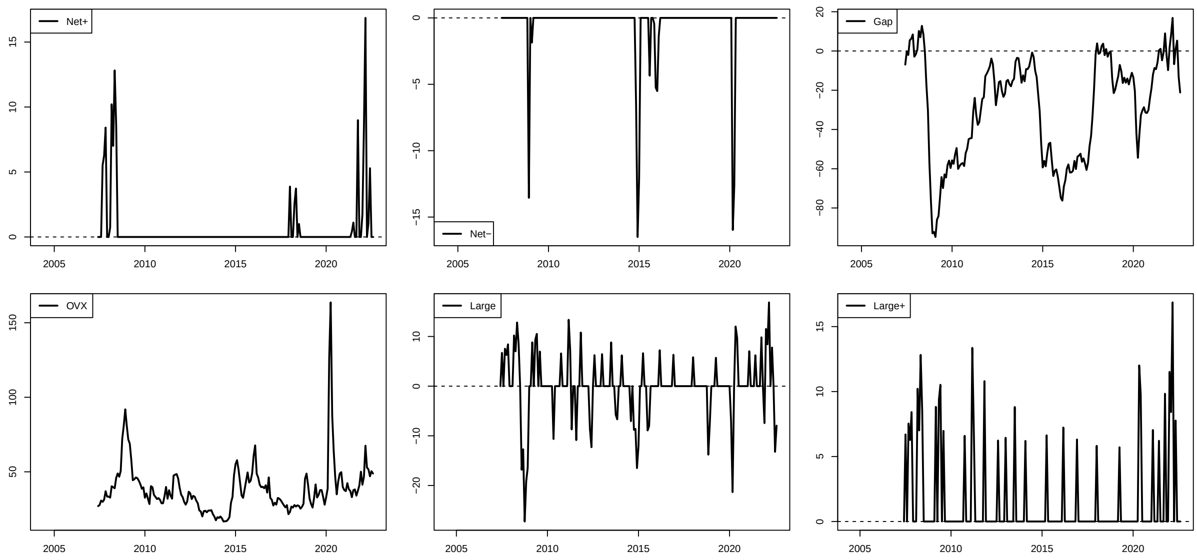

Figure 2: Predictors.

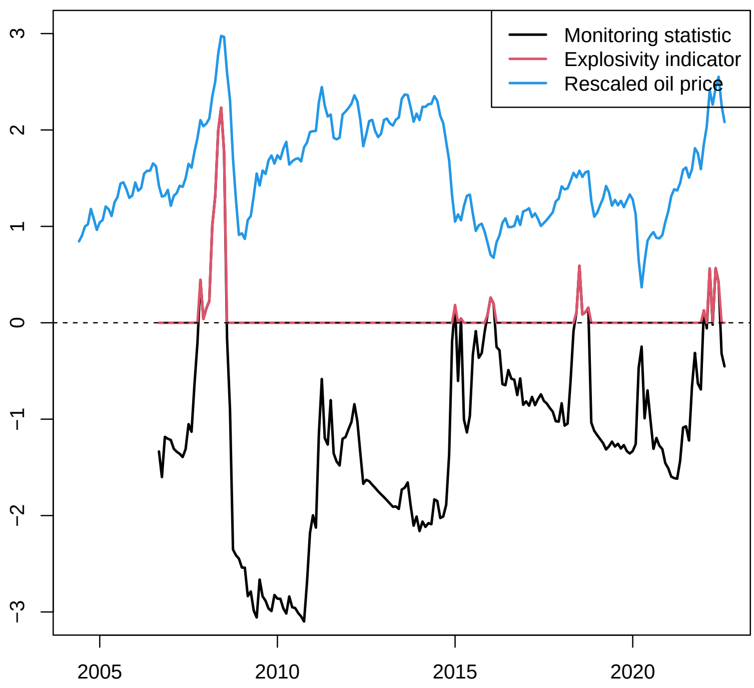

Figure 3: Explosivity indicator, monitoring statistic and rescaled oil price.

4 Empirical results

In this section, we report our empirical results. In Subsection 4.1, we start out with an unconditional evaluation of predictive ability for market-based oil price expectations. Subsection 4.2 continues with the conditional evaluation perspective. EIA forecasts are treated in Subsection 4.3., while robustness checks are summarized in Subsection 4.4.

4.1 Unconditional predictive ability

The following Table 1 reports relative MSE values (rMSE) and Diebold-Mariano statistics[10] (DM-stat) for the evaluation period from July 2005 to August 2022 for the Hamilton and Wu (2014) forecasts against the benchmark no-change forecast. In addition, we report the relative MSEs given in Baumeister (2023) for the evaluation period from August 1997 until December 2018 for comparison. The results reveal a strong similarity between our results and those reported in Baumeister (2023). Overall, the relative MSE values decrease with longer horizons. This suggests increasing accuracy of market-based expectations (relative to the no-change benchmark) for longer horizons. At a one-year horizon, the ratio takes the value of 0.676 indicating a strong reduction in the mean squared error, while the reduction at the three-month horizon is not that pronounced. The accompanying Diebold-Mariano statistics further suggest mild, but monotonically increasing significance at longer horizons. At the nominal significance level of five percent, a rejection is obtained for $h = 12$ in a two-sided setting, while the null is rejected for $h = 6$, 9 and 12 against a one-sided alternative. For the shortest horizon of $h = 3$, no rejection is obtained.

Period	Statistic	$h = 3$	$h = 6$	$h = 9$	$h = 12$
2005M7-2022M8	rMSE	0.904	0.821	0.746	0.676
	DM-stat	-1.381	-1.698	-1.841	-2.133
1997M8-2018M12	rMSE	0.896	0.829	0.762	0.697

Table 1: Relative MSE values and Diebold-Mariano statistics for market-based expectations.

[10]Throughout the analysis, we employ robust HAC standard errors. Newey-West robust standard errors provide consistent estimates also under strong forms of heteroskedasticity, see e.g. Demetrescu, Hanck, and Kruse-Becher (2023).

4.2 Conditional predictive ability

We now turn from the unconditional perspective to the conditional one which allows us to deepen the investigation and to study time-varying effects. Of course, it would be possible to test for time-varying unconditional predictive ability as in Giacomini and Rossi (2010) and Demetrescu, Hanck, and Kruse-Becher (2022). But, such an analysis would not reveal any insights on the dynamic sources of time-varying predictive ability. As our main interest lies in the evaluation of predictors, the conditional predictive ability framework is a natural choice in our context. Besides, the framework is well suited also because of the unconditional evaluation results which are being far from clear-cut.

Regarding the selection of predictors in the auxiliary dynamic regression Equation 4, we proceed with the best subset selection via Mallows (C_p) criterion, see Mallows (1973).[11] As robustness checks, we also consider the Schwarz criterion. In the best subset selection approach, the algorithm searches for the specification with smallest Mallows criterion or Schwarz criterion out of all combinations of predictors. As a further robustness check, we compare our findings to those obtained from the general-to-specific approach in Pretis et al. (2018).

In more detail, the algorithm of the best subset selection approach first fixes the number of predictors and then the following procedure is repeated for all remaining possible number of predictors. For a fixed number of predictors, the best subset of predictors is chosen via the residual sum of squares (or equivalently the coefficient of determination). As the number of predictors is fixed in this step, no penalty is needed for model complexity. Thirdly, the optimal number of predictors is chosen by means of Mallows criterion or an alternative measure balancing goodness-of-fit and model complexity. In this stage, only the best performing models (one model for each possible number of predictors) with different levels of model complexity are compared. The finally chosen set of predictors is the one belonging to the selected model in the final stage. All compared models include an intercept as required for the conditional predictive ability test.

Results are reported in Table 2. Here, we provide the results for all four different horizons and the case of market-based expectations versus no-change forecasts. In the upper panel of Table 2, we give individual t-statistics (based on HAC standard errors) for the selected predictors via the best subset algorithm based on Mallows criterion. Moreover, we report the Giacomini-White Wald statistic for conditional predictive ability which tests the joint nullity of all parameters including the intercept. As evaluation statistics, we report the

[11] This criterion is directly related to the Shibata criterion which in turn can be approximated by the well-known AIC. Asymptotically these criteria are identical, but different from the Schwarz criterion.

Predictor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	1.739	0.471	0.352	0.179
net^+	3.202	2.538	2.398	-
net^-	-	-	-	-
gap	-	-	-	-0.777
$large$	-	-	-	-
$large^+$	-	-	-	-
ovx	-3.164	-1.063	-0.929	-1.659
$expl$	-2.498	-6.819	-11.316	-11.468
Wald statistic	18.892	104.526	326.956	603.895
Evaluation				
Adjusted R^2	0.259	0.521	0.592	0.428
Relative performance	0.183	0.102	0.052	0.000
Success ratio	0.556	0.508	0.494	0.585
MSE(DR)/MSE(HW)	0.982	1.032	1.033	1.000

Table 2: Conditional predictive ability results, HW, best subset selection via Mallows criterion.

adjusted R^2, the relative performance ratio RP, the success ratio SR and the relative MSE of the ex-post dynamically conditionally rotated expectations in the lower panel of Table 2. While larger numbers indicate a better performance (of the dynamic conditional rotation) for the first three metrics, the opposite holds for the latter one.

The null hypothesis of no conditional predictive ability is clearly rejected by Giacomini-White statistic in all cases. Most important predictors of the loss differential are the explosivity indicator, net price increases and implied volatility.[12] The results are very similar across different horizons. We find that net price increases positively affect future realizations of the loss differential suggesting that times of strongly rising oil prices are followed by periods in which the market-based expectations perform worse than the no-change forecasts (ceteris paribus).[13]

Moreover, we find that the newly constructed explosivity indicator complements, rather than substitutes, the established net positive oil price shock series during explosive episodes. In particular, we find a strongly significant negative effect in predictability with increasing

[12] The correlation coefficients between (i) net^+ and $expl$ equals 0.6, (ii) ovx and $expl$ equals 0.07 and (iii) net^+ and ovx is 0.06.

[13] We also investigate the composition of net^+ and net^- as an aggregated predictor of both directions of price movements (Net), but it turns out that the decomposition into positive and negative price changes is much more informative.

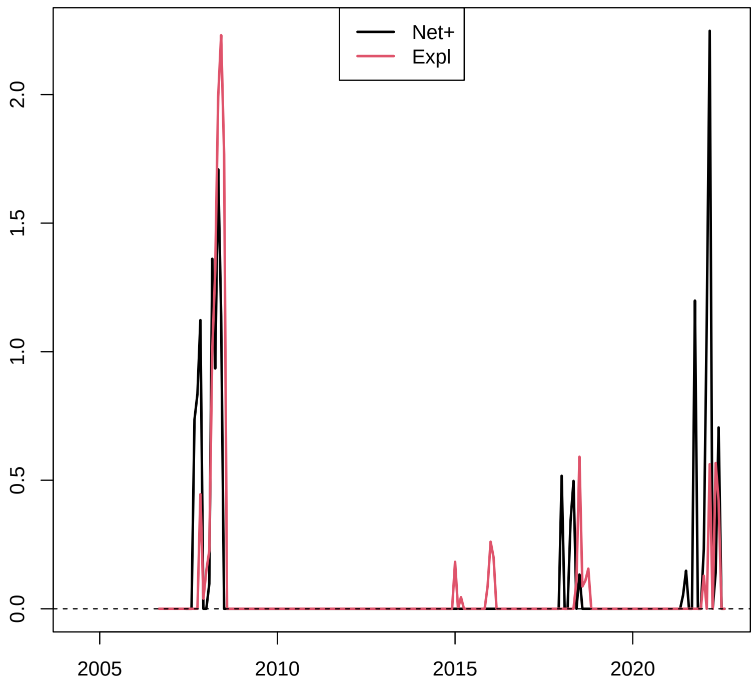

Figure 4: Comparison of net price increases and explosivity indicator.

importance of the horizon. The explosivity predictor is by far the most important one across the set of considered candidates. It mainly reflects the phase around the peak of a locally explosive episode and the subsequent collapse during the market downturn. Its negative effect on the future loss differential resembles the phenomenon that market-based expectations strongly recover (in terms of relative performance against the no-change forecasts) around the timing of the local peak and during the downward market adjustment phase. In fact, the positive effect of net price increases reflect the start of the explosive regime in which no-change forecasts significantly improve in their relative performance and may even outperform the market-based expectations. However, this picture is reversed after a few months of temporary explosiveness in the oil market as market-based expectations start to improve significantly over the no-change forecasts around the date of the peak.

In order to shed further light on the complementary nature of the two important predictors covering strong price fluctuations in the oil price, we display both series in Figure 4. It is clearly visible that the explosivity series is lead by the net positive indicator. Moreover, by construction the net positive indicator drops back to zero during downward oil market price corrections, while the explosivity indicator is still active. The technical explanation is the typical and inherent delay in monitoring procedures per se, see e.g. Chu, Stinchcombe, and White (1996). This is due to the fact that observations from the pre-explosive phase are uninformative about the regime change towards explosivity. Hence, it takes a few steps until the signal dominates the accumulated noise, see also the discussion in Homm and Breitung (2012) and Breitung and Kruse (2013). Notably, observations from an explosive regime have a relatively fast divergence rate, see e.g. Phillips and Magdalinos (2007), see also Kurozumi (2021) who provides a novel study on the asymptotic behaviour of delays in monitoring explosive processes.

In summary, we see that the real-time explosivity predictor is somewhat shifted in time in comparison to the net positive net^+ predictor. To this end, the latter one captures the strong upswings in the oil price (relative to its recent historical evolution), while the former one also captures the phase of market correction, i.e. the downturn of oil prices after the peak. Hence, both predictors can be seen as complements rather than substitutes. Large positive price changes are also measured by the predictor $large^+$. However, due to its construction it is less focused on explosive phases, but can be seen as a more general measure of oil price increases. The net negative net^- predictor focuses solely on strong downturns and is therefore also partly related to the explosivity predictor.

As a third selected predictor, implied volatility turns out to be of relevance for the three-month horizon, but not so much thereafter. The effect is negative and thus further supporting the notion that market-based expectations tend to be more accurate then no-change forecasts in more volatile states of the oil market. The *gap* measure is included in the dynamic regression for the one-year horizon, but appears to play a minor role as opposed to the explosivity indicator.

Overall, the predictive power is relatively high with adjusted R^2-values ranging from 25.9% ($h = 3$) to 59.2% ($h = 9$). The relative performance measure (RP) indicates that the no-change forecasts is selected more often at short horizons, but essentially never for the longest horizon of one year. Consequently, the conditional dynamic rotation approach only offers some gains at the shortest horizon of three month.[14] As shown in Section 4.3 covering EIA forecasts and also in Section 4.4 containing the robustness results, this picture is reversed. The infeasible, minimally possible, MSE(DR)/MSE(EIA) ratios are 0.872 ($h = 3$), 0.874 ($h = 3$), 0.861 ($h = 3$) and 0.851 ($h = 12$). These calculations are based on the full information of realized signs of the loss differential instead ot its predictions and thus provide a lower empirical bound. The success ratio is close to fifty

[14]The Clark and West (2007) statistic equals -1.274 is very close to the critical value at the ten percent level of significance.

percent for $h = \{6, 9\}$, but we observe significantly larger values at $h = 3$ and $h = 12$. The Anatolyev and Gerko (2005) statistic rejects the null hypothesis of no sign predictability at all horizons at the one percent level. The relatively high degree of predictability via the adjusted R^2 suggests that the level of the loss differential is indeed quite predictable.

4.3 EIA forecasts

We continue the analysis for the oil price forecasts obtained by the Energy Information Administration (EIA). Updated data from Garratt et al. (2019) is obtained from Yunyi Zhang's website. Figures 8-11 in the Appendix show the corresponding different loss differential.

Period	Statistic	$h = 3$	$h = 6$	$h = 9$	$h = 12$
2005M7-2022M8	rMSE	0.896	1.030	0.960	0.887
	DM-stat	−2.143	0.394	−0.649	−1.645

Table 3: Relative MSE and Diebold-Mariano statistics for EIA forecasts.

We first start with unconditional predictive ability results which are reported in Table 3. We find mostly similar results with a few noticeable differences. There is no clear monotonic pattern in the relative MSE over the different horizons. However, the best performance is observed for the shortest and longest horizons. The Diebold-Mariano statistics indicate significance at the three-month horizons with a borderline result at the one-year horizon. Remarkably, the relative MSE exceeds unity for the six-month horizon indicating that competing no-change forecasts are not outperformed.

We continue our analysis with the conditional predictive ability analysis based on the best subset selection according to Mallows criterion. Results are reported in Table 4.

Overall, more diverse predictors are selected. In all cases, the Giacomini-White statistic is highly significant. The adjusted R^2 in the dynamic regressions are considerably lower than for market-based expectations. These results indicate lower conditional predictability of the loss differential overall. Interestingly, the relative performance measure takes much larger values and also the success ratio is increased. This results in considerably lower relative MSE values for the dynamically rotated expectations with a minimal value of 0.915 ot $h = 6$ which is consistent with the previous unconditional rMSE statistic. However, at $h = 12$, the dynamic rotations do not pay off. The infeasible, minimally possible, MSE(DR)/MSE(EIA) ratios are 0.854 ($h = 3$), 0.826 ($h = 3$), 0.855 ($h = 3$) and 0.876 ($h = 12$). In particular, the achieved improvement at $h = 6$ (0.915) is notable. The Clark

Predictor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	0.439	-1.505	-1.376	0.644
net^{+}	2.132	2.246	2.394	-
net^{-}	-	1.390	2.001	-
gap	-	-	-	-
$large$	-	-1.024	-1.642	-2.237
$large^{+}$	-	1.196	1.614	1.976
ovx	-1.682	-	-	-1.272
$expl$	-	1.514	-2.739	-3.904
Wald statistic	16.565	21.365	19.037	63.936
Evaluation				
Adjusted R^2	0.052	0.229	0.128	0.091
Relative performance	0.072	0.232	0.161	0.175
Success ratio	0.583	0.565	0.534	0.626
MSE(DR)/MSE(EIA)	0.967	0.915	0.959	1.034

Table 4: Conditional predictive ability results, EIA, best subset selection via Mallows criterion.

and West (2007) statistic is significant at the five percent level for $h = 3$ and $h = 6$ and at the ten percent level for $h = 9$.

Results obtained for longer horizons, i.e. $h = 9$ and $h = 12$, are broadly consistent with (and comparable to) the ones for market-based expectations. At these longer horizons, the explosivity indicator enters the dynamic regressions with a negative significant effect, while the net^{+} and $large^{+}$ indicator positively affect future values of the loss differential. At the shortest horizon of $h = 3$, the net^{+} indicator is positively significant and the volatility index enters in a negative way. Both effects are consistent with the findings for the market-based expectations. At $h = 6$, the interpretation is not that clear-cut as a couple of predictors are included. We still find net^{+} to have a significant positive effect, while most other predictors have weaker t-values.

Overall, we find remarkably less predictability (ranging from 5.2% to 22.9%) in the loss differential involving EIA forecasts than for market-based expectations. On the contrary, improvements due to dynamic conditional rotation are more pronounced for EIA forecasts. This is no contradiction as one looks at the level of the loss differential, while the other is about its sign. The success ratio exceeds 0.5 in all cases. The Anatolyev and Gerko (2005) statistic rejects the null hypothesis of no sign predictability for all horizons mostly at the one percent level.

One possible economic explanation relates to the scapegoat approach by Bacchetta and Van Wincoop (2013) originally put forward in the context of exchange rate fluctuations. The scapegoat theory states that fundamentals become a scapegoat if the size of the deviation from its equilibrium value is large and there is a sizable shock to unobservable fundamentals, see Engel and West (2005). It could possibly be argued that large variations in the relationship between the oil price and fundamentals (whose information is included in forward-looking futures prices) naturally evolve when structural parameters in the oil market are time-varying and unknown to the market participants. During explosive periods and in particular after the market experiences a transition from the local peak to the downward (or even collapse) phase, market participants are likely to focus much more on fundamentals (so-called 'scapegoats') rendering the market-based expectations decisively more accurate.

4.4 Robustness checks

In this subsection we present various robustness checks. First, we use the Schwarz criterion and the adjusted R^2 which have different penalty terms in the best subset selection for the dynamic conditional predictive ability regressions. Furthermore, we study the general-to-specific approach as employed in Pretis et al. (2018). As in the other procedures, we keep the intercept in the regression to enable the conditional predictive ability test which includes a testable zero restriction on the intercept.

Starting with market-based expectations and the best subset selection approach, we find quite similar results for the Schwarz criterion (Table 5). Two differences are found. First, at $h = 6$ the implied volatility series is not chosen as a predictor. Remarkably, the relative MSE ratio of the dynamically rotated expectations is now below unity and equals 0.978 in comparison to the value of 1.032 for Mallows criterion (cf. Table 2). Second, at $h = 12$, the *gap* series is not included, but without any noteworthy consequences. Overall, the null hypothesis of no conditional predictive ability is clearly rejected by Giacomini-White statistic. Turning to the general-to-specific approach, we find a lot more predictors to be included at all horizons, see Table 6. There are improvements in the relative MSE at all horizons. Notably, at the shortest horizon of $h = 3$, the Clark and West (2007) statistic turns significant at the ten percent level. Besides, the results are quite similar to the previous findings and the interpretation of results is consistent with the main analysis.

When considering the results for EIA forecasts based on the Schwarz criterion (Table 7), we find that much less predictors are selected in comparison to Mallows criterion. Mainly, the positive net price increases and the explosivity indicator are selected. In one case, i.e.

$h = 12$, the null hypothesis of the conditional predictive ability test can only be rejected at the nominal significance level of ten percent. Overall, predictability deteriorates somewhat in this setting. Turning to the general-to-specific analysis in Table 8, we find quite similar results as in the main setting where Mallows criterion is used. For $h = 3$, the conditional predictive ability test does not lead to a rejection at conventional significance levels. However, the notably best (and significant) results for dynamic conditional rotation are obtained by the general-to-specific approach for the horizons $h = 6$ and $h = 9$. Also, the relative performance and success ratio measures are remarkably high and the sign predictability statistics are significant.

5 Conclusions

We investigate oil price expectations obtained from a Gaussian affine term structure model based on futures prices and the Energy Information Administration in the US. Strikingly, the loss differentials against no-change forecasts show time-variation mainly related to explosive oil price episodes. While an unconditional analysis over the full sample suggests the superiority of market-based expectations using futures prices overall, changing the perspective towards a conditional one reveals a couple of new insights. Our results suggest that there is clear evidence against the null hypothesis of no conditional predictive ability. The applied best subset selection approach suggests that besides a couple of established oil price shock measures (e.g. net oil price increases and implied volatility), our newly constructed real-time indicator for explosiveness has strong predictive power for the loss differential. Results turn out to be robust with respect to a number of variations in the econometric methodology. The main results are also established for the EIA forecasts. Our results may have immediate implications for applications of oil price expectations, e.g. testing for speculative bubbles, see e.g. Pavlidis et al. (2018) and beyond. A detailed analysis is currently under investigation by Kruse-Becher (2023).

The result that both opinion- and market-based forecasts fail to beat simple benchmarks in explosive price periods also deserves further attention from a theoretical perspective. Opinion or survey-based forecasts often do adjust to new information with a significant delay, for example due to the fact that the acquisition and incorporation of new information can be costly. On the other hand, market-based expectations might not be able to sufficiently capture risk premia in the spot market in times of explosive periods.

References

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

Alquist, R., L. Kilian, and R. Vigfusson (2013). Forecasting the price of oil. Volume 2, Chapter Chapter 8, pp. 427–507. Elsevier.

Anatolyev, S. and A. Gerko (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics 23*(4), 455–461.

Anundsen, A. K. (2015). Econometric regime shifts and the us subprime bubble. *Journal of Applied Econometrics 30*(1), 145–169.

Bacchetta, P. and E. Van Wincoop (2013). On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics 91*(1), 18–26.

Baumeister, C. (2023). Chapter 14 - measuring market expectations. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, pp. 413–441. Academic Press.

Baumeister, C. and L. Kilian (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives 30*(1), 139–160.

Bertsimas, D., A. King, and R. Mazumder (2016). Best subset selection via a modern optimization lens.

Breitung, J. and R. Kruse (2013). When bubbles burst: econometric tests based on structural breaks. *Statistical Papers 54*, 911–930.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). General-to-specific modeling: an overview and selected bibliography. *FRB International Finance Discussion Paper* (838).

Caspi, I., N. Katzke, and R. Gupta (2018). Date stamping historical periods of oil price explosivity: 1876–2014. *Energy Economics 70*, 582–587.

Chu, C.-S. J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica 64*(5), 1045–65.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics 138*(1), 291–311.

Coibion, O., Y. Gorodnichenko, S. Kumar, and M. Pedemonte (2020). Inflation expectations as a policy tool? *Journal of International Economics 124*, 103297.

Demetrescu, M., C. Hanck, and R. Kruse-Becher (2022). Robust inference under time-varying volatility: A real-time evaluation of professional forecasters. *Journal of Applied Econometrics 37*(5), 1010–1030.

Demetrescu, M., C. Hanck, and R. Kruse-Becher (2023). Robust fixed-b inference in the presence of time-varying volatility. *Econometrics and Statistics*.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Engel, C. and K. D. West (2005). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517.

Fantazzini, D. (2016). The oil price crash in 2014/15: Was there a (negative) financial bubble? *Energy Policy 96*, 383–396.

Figuerola-Ferretti, I., J. R. McCrorie, and I. Paraskevopoulos (2020). Mild explosivity in recent crude oil prices. *Energy Economics 87*, 104387.

Figuerola-Ferretti, I., A. Rodríguez, and E. Schwartz (2021). Oil price analysts' forecasts. *Journal of Futures Markets 41*(9), 1351–1374.

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

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*(4), 595–620.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Granger, C. W. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society 20*, 199–207.

Granger, C. W. J. and P. Newbold (1986). *Forecasting economic time series, 2nd edition*. Academic press.

Granziera, E. and T. Sekhposyan (2019). Predicting relative forecasting performance: An empirical investigation. *International Journal of Forecasting 35*(4), 1636–1657.

Gronwald, M. (2016). Explosive oil prices. *Energy Economics 60*, 1–5.

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

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics 38*(2), 215–220.

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

Hamilton, J. D. and J. C. Wu (2012). Identification and estimation of gaussian affine term structure models. *Journal of Econometrics 168*(2), 315–331.

Hamilton, J. D. and J. C. Wu (2014). Risk premia in crude oil futures prices. *Journal of International Money and Finance 42*(C), 9–37.

Hendry, D. and M. Clements (2003). Economic forecasting: some lessons from recent research. *Economic Modelling 20*(2), 301–329.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Kilian, L. and R. J. Vigfusson (2011). Are the responses of the u.s. economy asymmetric in energy price increases and decreases? *Quantitative Economics 2*(3), 419–453.

Kilian, L. and R. J. Vigfusson (2013). Do oil prices help forecast u.s. real gdp? the role of nonlinearities and asymmetries. *Journal of Business & Economic Statistics 31*(1), 78–93.

Kilian, L. and X. Zhou (2022). Oil prices, gasoline prices, and inflation expectations. *Journal of Applied Econometrics 37*(5), 867–881.

Kruse, R. and C. Wegener (2020). Time-varying persistence in real oil prices and its determinant. *Energy Economics 85*, 104328.

Kruse-Becher, R. (2023). Testing for speculative oil price bubbles based on futures market data. *mimeo*.

Kurozumi, E. (2021). Asymptotic behavior of delay times of bubble monitoring tests. *Journal of Time Series Analysis 42*(3), 314–337.

Mallows, C. L. (1973). Some comments on cp. *Technometrics 15*(4), 661–675.

Mikhed, V. and P. Zemčík (2009a). Do house prices reflect fundamentals? aggregate and panel data evidence. *Journal of Housing Economics 18*(2), 140–149.

Mikhed, V. and P. Zemčík (2009b). Testing for bubbles in housing markets: A panel data approach. *The Journal of Real Estate Finance and Economics 38*, 366–386.

Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: An extension of hamilton's results. *Journal of Political Economy 97*(3), 740–744.

Nonejad, N. (2021, November). Crude oil price point forecasts of the Norwegian GDP growth rate. *Empirical Economics 61*(5), 2913–2930.

Pavlidis, E., I. Paya, and D. Peel (2018). Using market expectations to test for speculative bubbles in the crude oil market. *Journal of Money, Credit and Banking 50*(5), 833–856.

Phillips, P. C. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Phillips, P. C. and S. Shi (2020). Real time monitoring of asset markets: Bubbles and crises. In *Handbook of statistics*, Volume 42, pp. 61–80. Elsevier.

Phillips, P. C., S. Shi, and J. Yu (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International economic review 56*(4), 1043–1078.

Phillips, P. C., S. Shi, and J. Yu (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review 56*(4), 1079–1134.

Phillips, P. C. and S.-P. Shi (2018). Financial bubble implosion and reverse regression. *Econometric Theory 34*(4), 705–753.

Pretis, F., J. J. Reade, and G. Sucarrat (2018). Automated general-to-specific (gets) regression modeling and indicator saturation for outliers and structural breaks. *Journal of Statistical Software 86*(3), 1–44.

Witten, D. and G. James (2013). *An introduction to statistical learning with applications in R.* springer publication.

Zhu, Y. and A. Timmermann (2022). Conditional rotation between forecasting models. *Journal of Econometrics 231*(2), 329–347. Special Issue: The Econometrics of Macroeconomic and Financial Data.

Appendix

Predictor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	1.739	-1.352	0.352	0.224
net^{+}	3.202	3.506	2.398	-
net^{-}	-	-	-	-
gap	-	-	-	-
$large$	-	-	-	-
$large^{+}$	-	-	-	-
ovx	-3.164	-	-0.929	-0.891
$expl$	-2.498	-6.348	-11.316	-20.006
Wald statistic	18.892	43.241	326.956	485.134
Evaluation				
Adjusted R^2	0.259	0.514	0.592	0.422
Relative performance	0.183	0.051	0.052	0.000
Success ratio	0.556	0.525	0.494	0.585
MSE(DR)/MSE(HW)	0.982	0.978	1.033	1.000

Table 5: Conditional predictive ability results, HW, best subset selection via Schwarz criterion.

Predictor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	1.295	-0.305	0.311	0.345
net^{+}	2.986	2.225	2.305	-
net^{-}	-	-	-	-1.604
gap	-0.645	0.347	-0.297	-0.777
$large$	0.610	-	0.105	-0.855
$large^{+}$	-0.395	-	-0.301	0.793
ovx	-2.492	1.092	-1.304	-1.804
$expl$	-2.478	-6.675	-10.721	-11.082
Wald statistic	21.208	220.094	1441.283	5267.865
Evaluation				
Adjusted R^2	0.256	0.517	0.586	0.424
Relative performance	0.239	0.130	0.034	0.012
Success ratio	0.577	0.525	0.511	0.596
MSE(DR)/MSE(HW)	0.973	0.987	0.984	0.993

Table 6: Conditional predictive ability results, HW, General-to-Specific.

Predictor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	2.955	-0.751	-0.957	0.663
net^{+}	1.999	2.598	2.733	-
net^{-}	-	-	-	-
gap	-	-	-	-
$large$	-	-	-	-2.292
$large^{+}$	-	-	-	-
ovx	-	-	-	-1.275
$expl$	-	2.561	-3.198	-
Wald statistic	10.283	36.168	18.552	6.312
Evaluation				
Adjusted R^2	0.041	0.213	0.095	0.074
Relative performance	0.072	0.153	0.063	0.146
Success ratio	0.583	0.542	0.552	0.608
MSE(DR)/MSE(HW)	0.967	0.943	0.974	1.039

Table 7: Conditional predictive ability results, EIA, best subset selection via Schwarz criterion.

Predictor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	-2.014	0.275	0.276	0.596
net^{+}		2.006	3.025	0.648
net^{-}	-	-	-	-
gap	-	0.733	0.253	
$large$	1.318	-0.954	-1.387	-1.749
$large^{+}$	-	1.218	1.649	-
ovx	-	-0.751	-0.664	-1.091
$expl$	-	1.051	-3.102	-3.141
Wald statistic	4.351	26.494	30.079	85.041
Evaluation				
Adjusted R^2	0.034	0.219	0.108	0.074
Relative performance	0.189	0.322	0.236	0.175
Success ratio	0.577	0.577	0.540	0.602
MSE(DR)/MSE(HW)	0.981	0.914	0.958	1.042

Table 8: Conditional predictive ability results, EIA, General-to-Specific.

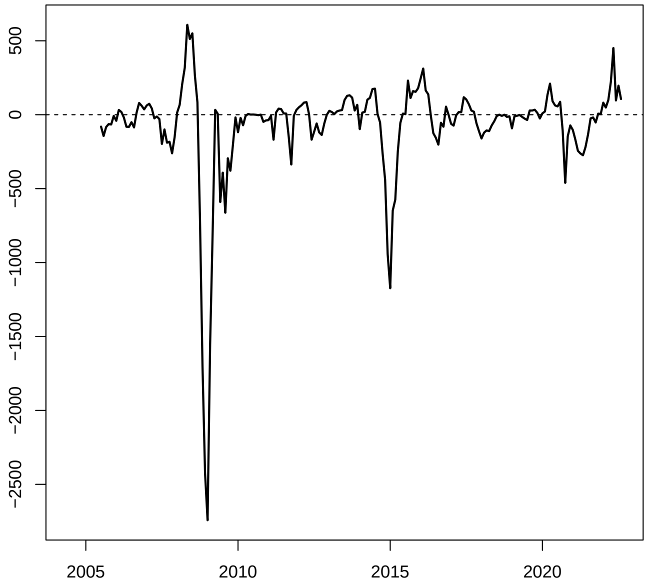

Figure 5: Loss differential (HW-NC), $h = 6$

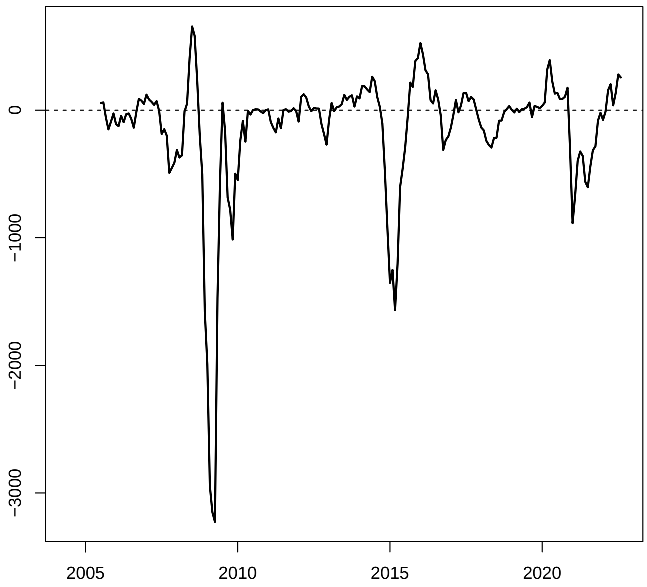

Figure 6: Loss differential (HW-NC), $h = 9$

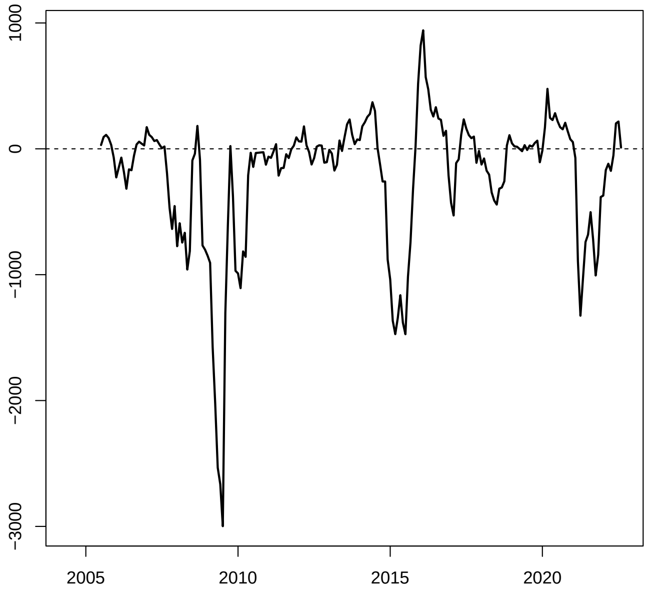

Figure 7: Loss differential (HW-NC), $h = 12$

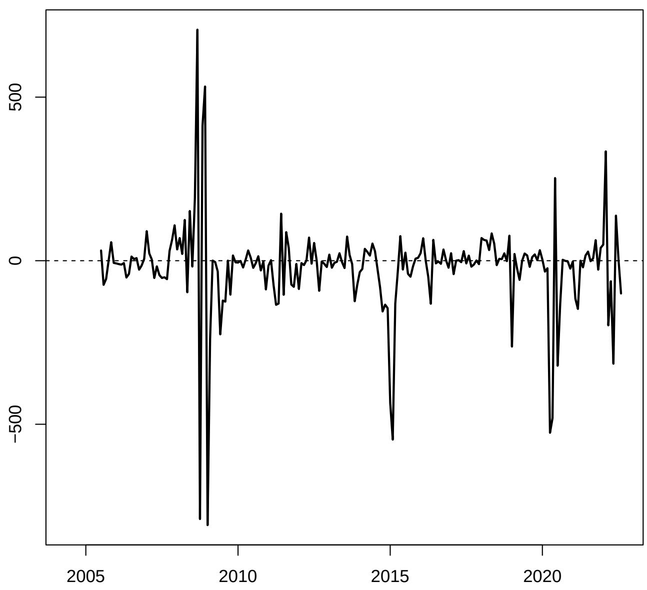

Figure 8: Loss differential (EIA-NC), $h = 3$

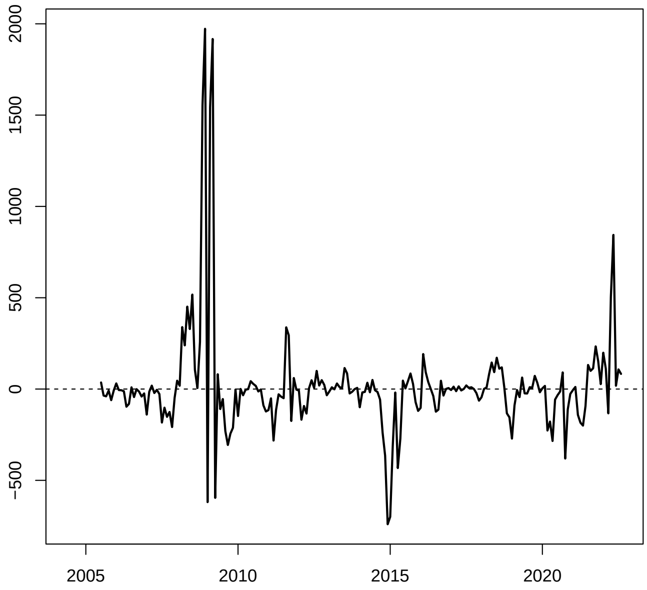

Figure 9: Loss differential (EIA-NC), $h = 6$

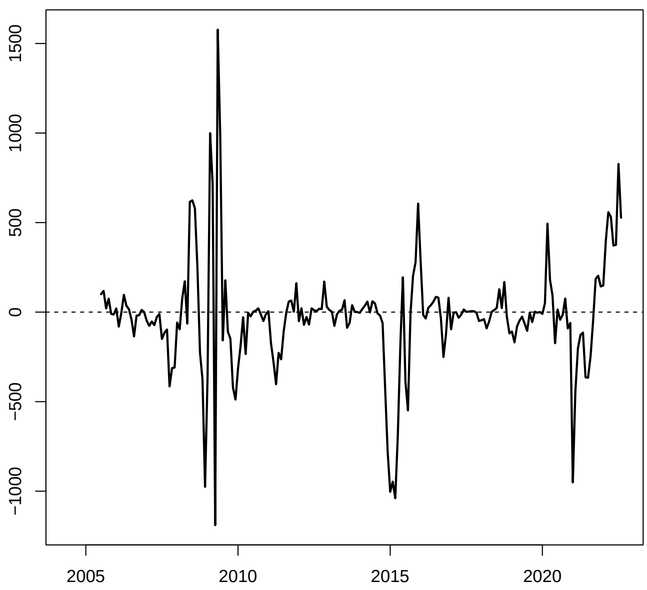

Figure 10: Loss differential (EIA-NC), $h = 9$

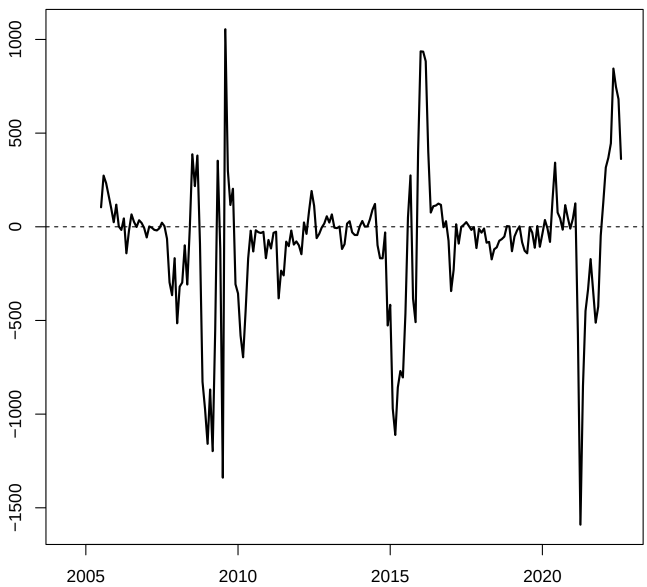

Figure 11: Loss differential (EIA-NC), $h = 12$