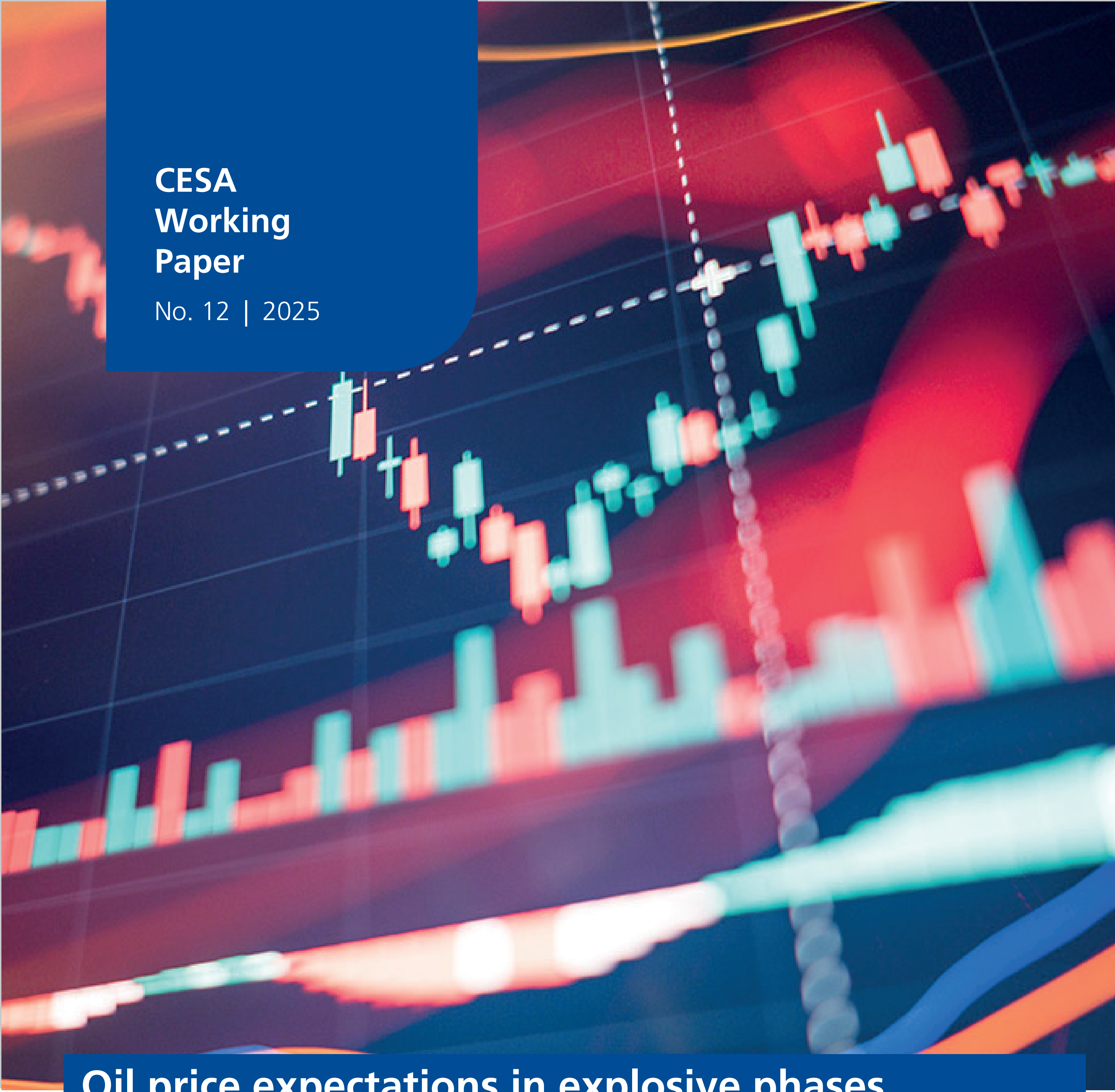


CESA Working Paper

No. 12 | 2025

Oil price expectations in explosive phases

Robinson Kruse-Becher[a] and Philip Letixerant[a]

a University of Hagen (FernUniversität in Hagen)

FernUniversität in Hagen

Oil price expectations in explosive phases*

Robinson Kruse-Becher† Philip Letixerant‡

January 14, 2025

Abstract

Accurate oil price expectations are of great importance for a variety of economic and financial applications. We find that state-of-the-art market-based expectations only weakly outperform a simple no-change benchmark. This gives rise to changing the perspective from an unconditional to a conditional evaluation method. This consideration is relevant when the forecasting methods potentially behave very differently conditional on certain time-varying economic states. Strikingly, it seems that the no-change benchmark outperforms market-based expectations systematically during turbulent market phases. The entertained conditional predictive ability framework allows us to study the role of important state variables for the time-varying performance explicitly. Among these are established variables from the related oil market literature, covering oil price change measures, volatility as well as supply and demand. Additionally, we suggest a novel and complementing indicator for oil price explosiveness. Our results robustly indicate the existence of conditional time-variation. Furthermore, they underline the importance of the new indicator reflecting temporary exuberance and subsequently collapsing oil prices. We find similar results when evaluating expectations obtained from the Energy Information Administration. Our findings may have consequences for a variety of economic and financial applications e.g. construction of expectational shocks and testing for speculative oil price bubbles.

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 three anonymous referees and the guest editor for helpful comments and suggestions. Moreover the authors are indebted to Marta Banbura, Joscha Beckmann, Jörg Breitung, 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, respectively.

†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), Kilian and Zhou (2022) and Baumeister (2023). Due to information frictions, market participans can hardly observe whether current oil prices are entirely consistent with fundamentals and hence, the role of expectations becomes crucial, see Sockin and Xiong (2015). Oil price forecasts are, for instance, important predictors for macroeconomic indicators, see e.g. Degiannakis and Filis (2023). 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).

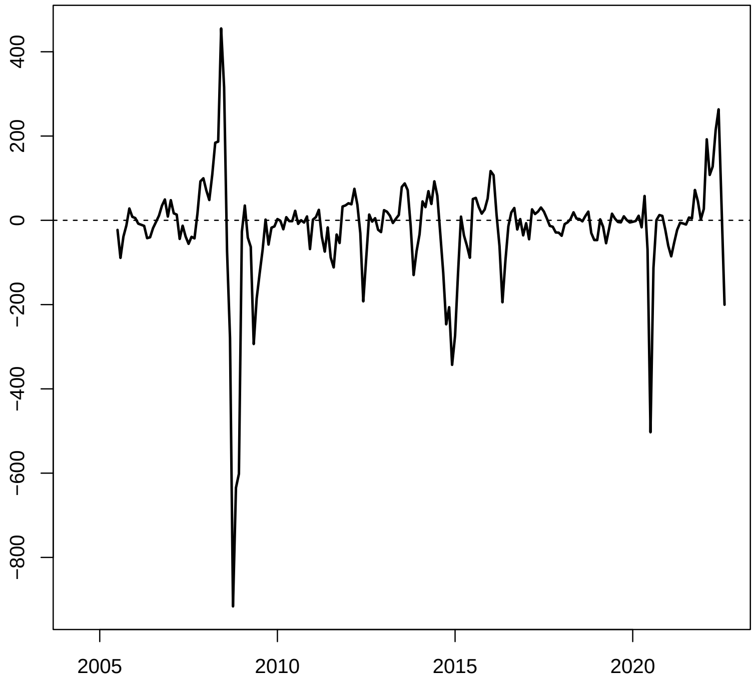
Forecast efficiency regressions have shown that futures prices are not unbiased predictors of future spot prices in the oil market (and elsewhere), see Baumeister (2023) for a recent and excellent survey article. One source, and possibly the main driver, is a time-varying risk premium. The arbitrage-free affine Gaussian term structure model proposed by Hamilton and Wu (2012, 2014) allows 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 [HW] 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 predictions in an unconditional evaluation against a large number of competing approaches, qualifying them for a wide field of application in economic analyses.[1]

Despite the improvement in mean squared error reduction, the results are not so clear cut when comparing the predictive accuracy with the Diebold and Mariano (1995) test. The test does not indicate significant outperformance of the no-change benchmark by the market-based expectation. Furthermore, 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] clearly outperforms market expectations.[2] This gives rise to

[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.

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

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



Notes: The loss differential between the squared losses of the Hamilton and Wu (2014) approach and the no-change forecasts for forecasting horizon $h = 3$. Values above zero indicate better forecasting performance of the no-change at that time.

investigating the time-varying economic circumstances which are related to the relative forecasting performance of the respective predictions. As Li et al. (2021) argue, this is of particular interest when they appear similar on average as indicated by the Diebold and Mariano (1995) test. 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.[3] As reliance on accurate forecasts is especially important during turbulent times and market-based expectations are systematically outperformed during this times, this further motivates to study the time-varying conditions connected to the loss differential.

In this work, we switch from the unconditional evaluation as in Diebold and Mariano (1995) to the conditional perspective presented in Giacomini and White (2006), con-

[3]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 (2024).

tributing in three ways. We identify time-variation in the loss differential, we determine which economic state variables are connected to this time-variation. As a by-product we explore to what extent this time-variation might be further utilized. 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. In contrast to other conditional frameworks, such as the fluctuation test by Giacomini and Rossi (2010) which are agnostic about the cause of the variation, the method at hand has the appeal of high interpretability and the identification of what drives the time-variation. The test allows to investigate which economic and financial states, say boom or bust, are connected to the relative predictive 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. 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.

As in Granziera and Sekhposyan (2019), we thus investigate whether the relative predictive performance can be explained. To this end, we study a number of established oil price change measures, see Kilian and Vigfusson (2013). In order to precisely analyse the role of explosivity in this context, we suggest a novel real-time indicator 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 variables 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 James, Witten, Hastie, and Tibshirani, 2021) 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 variables are net price increases, implied volatility of oil price changes and the newly proposed explosivity indicator. Overall, our results robustly indicate conditional time-variation and 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.

Various economic actors rely on accurate oil forecasts in their decision making. Especially in turbulent times, reliance on such forecasts is high. Nonetheless, we analyse circumstances in which the state-of-the-art HW forecast is systematically outperformed. Our results may have implications for applications in which oil price expectations play a decisive role, e.g. expectational oil price shocks commonly used in the VAR literate, regulation of oil markets or climate policy, but also consumers decisions, especially facing declining oil reserves (Baumeister, 2023). Oil price expectations are also useful in forecasting other macroeconomic variables such as inflation and industrial production (Degiannakis and Filis, 2023) underlining the importance in macroeconomic projections and therefore in monetary and fiscal policy. The literature on testing against speculative bubbles (see e.g. Pavlidis, Paya, and Peel, 2018) and analyses based on oil price shocks (see e.g. Baumeister, 2023) are potentially affected in particular. Their 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 expectations. Similarly, this holds for VAR analyses based on expectational oil price shocks (see e.g. Anderl and Caporale, 2024, Valenti, Bastianin, and Manera, 2023 or Degasperi, 2023).

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 the mean squared error, but in particular in testing the relative forecasting accuracy.

The popular unconditional test by Diebold and Mariano (1995) (DM) tests for equal expected loss 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). However, 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 carry explanatory power for the relative forecasting accuracy of the predictions at hand, leading to a test of conditional predictive ability, see Giacomini and White (2006). Latter test is of special interest when the DM test is non-conclusive. Even if there is no significant difference between the methods on average, there still might time-varying relative performance between the forecasts, see Li et al. (2021). Hence, one might investigate whether the relative forecasting performance can be explained itself. Under the null hypothesis, there is no such predictability of the zero-mean loss differential. Therefore, a rejection of the null hypothesis of the conditional predictive ability test is informative in various ways. The rejection might hint towards potential misspecification and non-optimality of the forecasts. Additionally, assuming there is conditional predictive ability, the framework is informative about the economic conditions related to the time-varying performance. By regressing the loss differential on the state variables, the framework allows to identify drivers of the relative predictive performance. The coefficients of the state variables tell about economic conditions under which changes in the relative performance occur. Therefore, it is possible to analyse possible economic mechanisms leading to changes in the relative forecasting performance.

2.1 Recovering market expectations

The Hamilton and Wu (2012, 2014) model allows to recover market expectations $E_t(S_{t+h})$ of the spot price S_t, as well as the risk premium RP_t^h where t denotes the time the future contract is bought and h denotes the maturity. These can be recovered from future prices F_t^h as well as a set of latent and observed variables/risk factors x_t and is well described in Baumeister (2023). In this framework the expectation of the spot price can be calculated from the futures price and a factor accounting for the risk premium:

$$E_t(S_{t+h}) = F_t^h(1 + a_h + b_h' x_t). \tag{1}$$

The time-varying risk premium is then calculated as difference between the future price and the expected spot price $RP_t^h = F_t^h - E_t(S_{t+h})$. After solving for the future contracts price at expiration and taking expectations, above Equation 1 and the parameters a_h and

b_h are derived from the return regression

$$\frac{F^1_{t+h-1} - F^h_t}{F^h_t} = a_h + b_t' X_t + \epsilon_{t+h}. \tag{2}$$

The fitted values of above equation would be interpreted as the risk premium, allowed to vary over time. Hamilton and Wu (2015) present how this return regression approach is represented in following equation:

$$f^{h-1}_{t+1} - f^h_t = \kappa_{h-1} + \delta'_{h-1} x_t + \epsilon^{h-1}_{t+1} \tag{3}$$

with f^h_t being the logarithmic future price. This equation not only nests the return regression idea, but also incorporates the Gaussian affine term structure approach and thus relates the two concepts. Regarding the return regression model, the parameters in the above equation can be interpreted as parameters in an unrestricted OLS-regression model specifically estimated for a maturity h. Typical variables contained in x_t which are known to the market participants are measures explaining the market behaviour such as stock indices or the slope of the future curve, i.e. proxies of the risk factors.[4]

The term structure model is connected to this as follows. It assumes that all asset prices are determined by the set of observed and unobserved variables x_t which can be modelled in a small VAR(1) with intercept c and coefficient matrix ρ:

$$x_{t+1} = c + \rho x_t + \Sigma u_{t+1}, \quad u_{t+1} \sim N(0, I). \tag{4}$$

Consequently futures prices can be modelled by these fundamental factors as $f^h_t = \alpha_h + \beta'_h x_t$ where $\alpha_h = \alpha_{h-1} + \beta'_{h-1}(c - \lambda) + \frac{1}{2}\beta'_{h-1}\Sigma\Sigma'\beta_{h-1}$ and $\beta'_h = \beta'_{h-1}(\rho - \Lambda)$, i.e. functions of the parameters from the Gaussian VAR(1) (Hamilton and Wu, 2014). In this framework, risk averseness is implied by the parameter $\lambda_t = \lambda + \Lambda X_t$ which is the market price of risk. Hamilton and Wu (2015) show that the parameters in Equation 3 are a combination of the term structure models factor loading β'_{h-1}, the scale matrix Σ and the risk parameters λ and Λ as $\kappa_{h-1} = \beta'_{h-1}\lambda - \frac{1}{2}\beta'_{h-1}\Sigma\Sigma'\beta'_{h-1}$, $\delta'_{h-1} = \beta'_{h-1}\Lambda$ and $\epsilon^{h-1}_{t+1} = \beta'_{h-1}\Sigma u_{t+1}$ which again can be interpreted as intercept, slope vector and error term from the return regression. In contrast to the return regression, the term structure model relies on parameter restrictions for the different maturities, ruling out arbitrage possibilities. Nonetheless the structural parameters can be inferred from the return regression given the future prices and x_t (Hamilton and Wu, 2015). The risk premium again follows as difference between the

[4] An extensive overview of possible and common proxy variables and factors can be found in Baumeister (2023).

future price and the market expectation of the spot price implied by the estimated model, assuming risk neutrality setting $\lambda = \Lambda = 0$.

2.2 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, which is the natural choice, as the conditional expectation minimizes the scared prediction error (Granger, 1969). The h-step forecast error for method $j = \{1, 2\}$ is

$$e_{j,t+h} = y_{t+h} - \widehat{y}_{j,t+h} \tag{5}$$

and the loss differential thus formulates as:

$$\Delta L_{t+h} = e^2_{1,t+h} - e^2_{2,t+h}. \tag{6}$$

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:

$$DM = \sqrt{P}\frac{\overline{\Delta L}}{\widehat{\sigma}_{\overline{\Delta L}}} \sim N(0, 1), \tag{7}$$

where $\overline{\Delta L}$ is the mean of the loss differential, $\widehat{\sigma}_{\overline{\Delta L}}$ is a HAC-type estimator of the loss differentials long-run standard deviation and P is the number of evaluated forecasts.[5] By rejecting the null hypothesis, 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 economic conditions under which one forecast produces smaller squared errors than the other. This leads us to the test of conditional predictive ability.

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

2.3 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.[6]

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 ($s_t = \emptyset$).

The test of conditional predictive accuracy can be integrated into a linear regression framework ensuring high interpretability of the resulting OLS coefficients, see Granziera and Sekhposyan (2019). The regression equation is:

$$\Delta L_{t+h} = \theta' s_t + \varepsilon_t, \tag{8}$$

where θ is the parameter vector of dimension p to be estimated via least squares and s_t is the vector of p conditioning variables including an intercept. The innovation term is labeled as ε_t.[7] 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 calculated as:

$$GW = P \left(P^{-1} \sum_{t=1}^{P} s_t \Delta L_{t+h} \right)' \widehat{\Omega}^{-1} \left(P^{-1} \sum_{t=1}^{P} s_t \Delta L_{t+h} \right) \sim \chi^2_p, \tag{9}$$

with $\widehat{\Omega}$ being a HAC-type estimator of the variance of $P^{-1/2} \sum_{t=1}^{P} s_t \Delta L_{t+h}$. Therefore, the test can be implemented as a regression based Wald-type test with $H_0: \ \theta = 0$ resulting from Equation 8, which we denote as GW_W.

Following Granziera and Sekhposyan (2019), we do not only report the Wald-type statistic GW_W, but also the individual t-statistics for each included conditioning variable and the adjusted R^2 from the dynamic linear regression in Equation 8. This is of importance

[6] Conditional evaluation becomes increasingly more relevant as a recent publication by Odendahl et al. (2023) shows. The authors investigate the relative forecasting performance under state-dependence and discover pockets of predictability.

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

in order to analyse the relevance of the respective variable. Moreover, significant variables can be interpreted with respect to their coefficients' direction, allowing an economic interpretation of the loss differentials predictability depending on the respective state variable.

2.4 Oil price change measures

Due to the relevance of the oil price for the overall economy, the literature on oil price change measures 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 changes are considered as well, e.g. in Mork (1989) who finds asymmetric effects on the GDP growth depending on the kind of oil price change. Before introducing oil price change 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 change 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 change measures are an obvious choice. For this we rely on the following oil price change 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 price change measures are introduced, focusing on positive and negative changes, 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 $oil_t^{\max} = \max\{oil_{t-1}, \ldots, oil_{t-36}\}$.[8] 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:

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

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

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 compute $large_t$ 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 change measures serve as explanatory variables in the conditional ability testing framework.

Apart from conventional oil price change measures, we include variables covering supply and demand fundamentals for robustness. These are relevant to forecasting the oil price and its volatility itself and potentially covering non-linearities (see e.g Zhang, Chen, and Bouri, 2024). Following Wang et al. (2017) and Salisu et al. (2022) we include the index of global economic activity (*gea*) as proposed in Kilian (2009) and the monthly global economic conditions indicator (*gecon*) as proposed in Baumeister et al. (2022) to cover demand side. With respect to the supply side we include the changes in OECD oil inventory (*oic*). In contrast to the above oil price change measures, these variables are not reported in real-time. We therefore apply our methodology in two distinct settings (i) a purely real-time setting excluding supply and demand factors and (ii) a setting including these variables.[9] The results based on the second setting are reported in subsection 4.4 and the Appendix, not changing the real-time results of the conditional predictive ability test.

2.5 Monitoring explosive oil prices

We construct a new real-time variable indicating explosiveness in the oil market. To this end, we make use of the established econometric methodology developed in Phillips,

[9] Alternatively the respective variables would have to be replaced with nowcasts as in Wang et al. (2017).

Shi, and Yu (2015a,b), Phillips and Shi (2018) and Phillips and Shi (2020).[10] 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).[11] Positive values of the $mBSADF_t$ statistic[12] 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 censoring 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 variables and is also based on recent and lagged oil prices.

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 explosiv-

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

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

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

ity. 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 indicator 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 censoring). These features are advantageous to extract the most relevant information regarding explosivity from the BSADF statistic. Moreover, our measure is constructed in real-time and does not rely on information from subsequent periods. This is an important feature in the comparison to other real-time variables, as we want to study the time-variation in relative forecast performance based on information that was available at the time the forecast is generated.

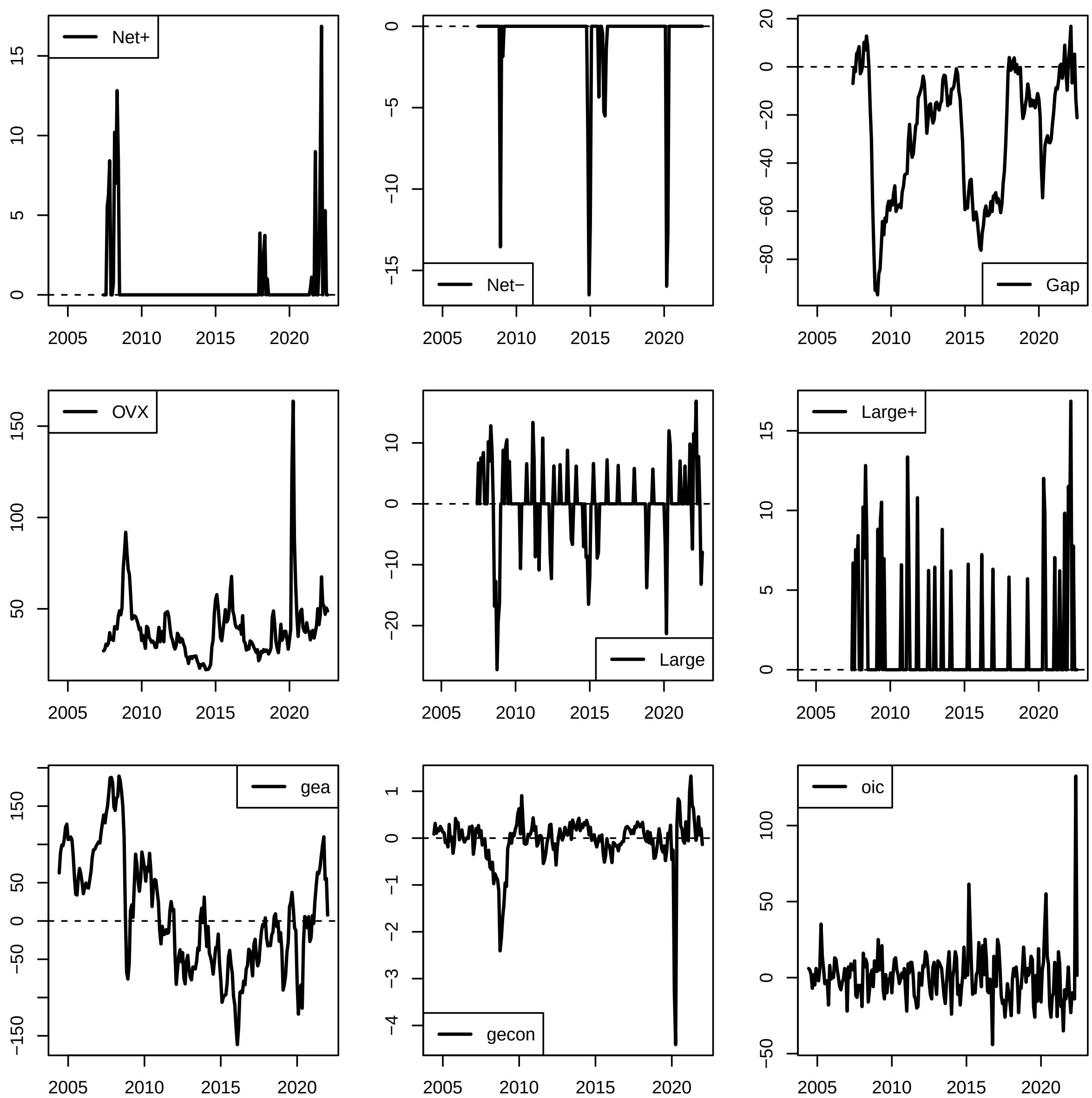
3 Data

Our sample runs from July 2005 to August 2022 yielding $T = 206$ monthly observations.[13] The sample start and end are restricted to the availability of EIA forecast which are essential to the analysis. The data is obtained from Yunyi Zhang's homepage (updated and structured EIA forecasts; OECD oil inventory), the FRED data base (WTI spot oil price and CBOE crude oil ETF volatility index), the Federal Reserve Bank of Dallas (global economic activity index) and Christiane Baumeister's homepage. Latter directly provides the oil price market expectations for the horizons $h = \{3, 6, 9, 12\}$ which can be found in the file "Monthly WTI Oil Price Expectations". We therefore obtain readily available estimates for the HW approach by Christiane Baumeister and data for the Monthly Global Economic Conditions (GECON) indicator, resulting in a sample of forecasts and explanatory variables for the same period. The established oil price change measures are constructed as outlined in the previous section, see Figure 2. In Figure 3, we display the newly proposed explosivity indicator, the underlying monitoring statistic $mBSADF_t$ and the re-scaled oil price series.[14] 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 expec-

[13]Note that the CBOE crude oil ETF volatility index is available from June 2007 onwards and the variables based on oil_t^{max} need an initialisation of 36 months. Consequently the unconditional predictive ability analysis begins in June 2007, resulting in $T = 183$ observations. Therefore conditional predictive ability regressions including these variables lose observations depending on the forecast horizon, i.e. $T = 183 - h$ observations.

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

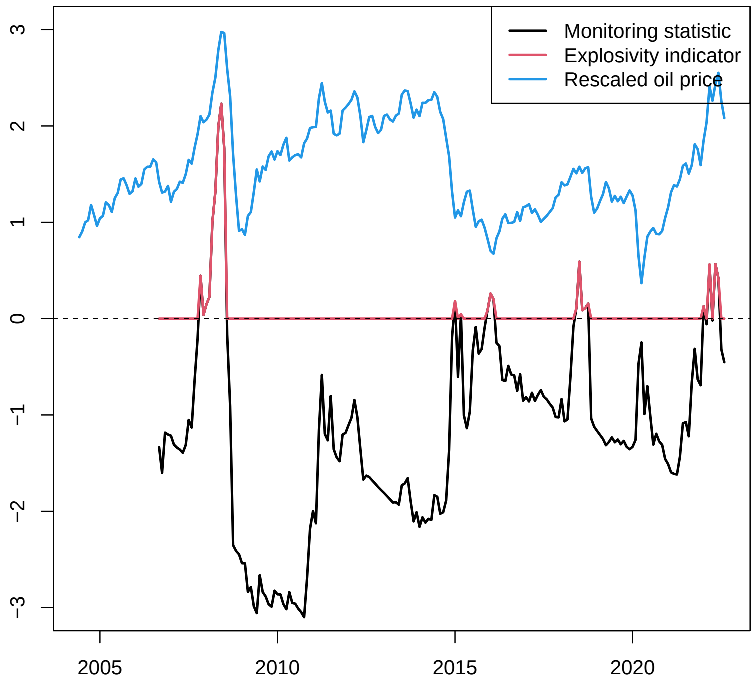
Figure 2: State variables.



Notes: Variables used in the conditional predictive ability regression setting. The top two rows show the real-time oil price change measures and the bottom row the additional variables used in the further analysis. Information on the construction of the real-time oil price change measures is reported in section 2.

tations and no-change forecasts and (ii) EIA forecasts and no-change forecasts. In total, we analyse eight different loss differential series in the (un)conditional predictive ability setting.

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



Notes: The newly proposed explosivity indicator (red) which derives from the monitoring statistic (black). When the monitoring statistic exceeds zero, the explosivity indicator and the monitoring statistic are identical. Otherwise, if the monitoring statistic is smaller than zero, the indicator is censored at zero. In addition the rescaled oil price (blue) is plotted for comparison.

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[15] (DM-stat) for the evaluation period from June 2007 to August 2022 for the Hamil-

[15]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 et al. (2022).

ton and Wu (2014) forecasts against the benchmark no-change forecast. We evaluate the forecasts considering quadratic loss. This is the common metric in order to measure the accuracy of market expectations due to the minimisation of the quadratic loss by the conditional expectation (Granger 1969). In addition, we report the relative MSEs given in Baumeister (2023) for the evaluation period from August 1997 until December 2018 for comparison, revealing a strong similarity between those results and ours.[16] Overall, the relative MSE values decrease with longer horizons, this relationship is also reported in Wang et al. (2017) where forecasts from a time-varying parameter model are compared to the no-change forecast. 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 reduction in the mean squared error, while the reduction at the three-month horizon is not that pronounced. The accompanying Diebold-Mariano statistics only suggest mild significance at longer horizons. At the nominal significance level of five percent, a rejection is only obtained for $h = 12$.

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

Period	Statistic	$h = 3$	$h = 6$	$h = 9$	$h = 12$
2007M6-2022M8	rMSE	0.905	0.821	0.742	0.670
	DM-stat	-1.331	-1.677	-1.814	-2.097
1997M8-2018M12	rMSE	0.896	0.829	0.762	0.697

Notes: Relative MSE values and Diebold-Mariano statistics for market-based expectations (HW) compared to the NC forecast for $h = \{3, 6, 9, 12\}$. Values below one for the rMSE indicate better performance of the HW forecast. The upper evaluation period is the sample later evaluated by the conditional predictive ability framework. The lower period is the evaluation from Baumeister (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 said sources, 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.

[16] Due to data availability of an important state variable (ovx), we are restricted to the sample between 2007 and 2022.

Regarding the selection of explanatory variables in the auxiliary dynamic regression Equation 8, we proceed with the best subset selection via Mallows (C_p) criterion, see Mallows (1973).[17] 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 regressors. As a further robustness check, we compare our findings to those obtained from the general-to-specific approach (GETS) in Pretis et al. (2018).

In more detail, the algorithm of the best subset selection approach first fixes the number of regressors and then the following procedure is repeated for all remaining possible numbers of regressors. For a fixed number of variables, the best subset of regressors is chosen via the residual sum of squares (or equivalently the coefficient of determination). As the number of regressors is fixed in this step, no penalty is needed for model complexity. Thirdly, the optimal number of regressors 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 regressors) with different levels of model complexity are compared. The finally chosen set of regressors 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 regressors 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 and the adjusted R^2.

The null hypothesis of no conditional predictive ability is clearly rejected by Giacomini-White statistic in all cases. Most important variables of the loss differential are the explosivity indicator, net price increases and implied volatility.[18] 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).[19] Additionally, predictive power is relatively high with

[17] 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.

[18] 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.

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

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

Regressor (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
GW_W	18.892	104.526	326.956	603.895
Adjusted R^2	0.259	0.521	0.592	0.428

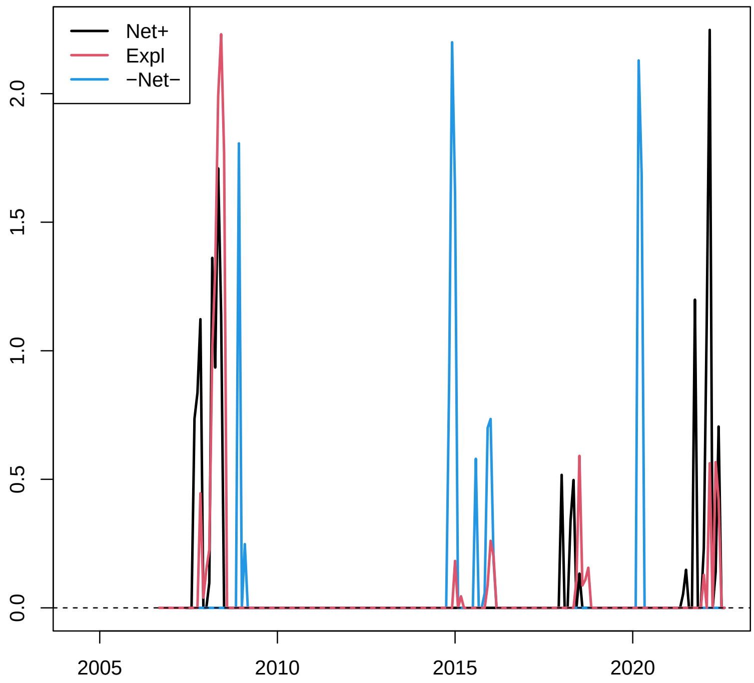
Notes: The results for the conditional predictive ability regression, comparing the HW forecast and the NC forecast. The regressors are chosen via best subset selection using Mallows criterion. The table lists the t-statistics for selected regressors, the Wald statistic GW_W on which the test of conditional predictive ability is based on, as well as the adjusted R^2 of the regression

adjusted R^2-values ranging from 25.9% ($h = 3$) to 59.2% ($h = 9$).

Moreover, we find that the newly constructed explosivity indicator complements, rather than substitutes, the established net positive oil price change series during explosive episodes. In particular, we find a strongly significant negative effect in predictability with increasing importance of the horizon. The explosivity indicator 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 explanatory variables covering strong price fluctuations in the oil price, we display both series

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



Notes: Comparison of net price increases, the negative of the net price decreases and explosivity indicator. It becomes obvious how the explosivity indicator complements the other two oil price change measures, spiking at similar times.

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 indicator is somewhat shifted in time in comparison to the net positive net^+ oil price change measure. 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 indicators can be seen as complements rather than substitutes. Large positive price changes are also measured by the variable $large^{+}$. 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^{-} variable focuses solely on strong downturns and is therefore also partly related to the explosivity indicator as it can also be seen in Figure 4.

As a third selected measure, 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.

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.

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

Period	Statistic	$h=3$	$h=6$	$h=9$	$h=12$
2007M6-2022M8	rMSE	0.896	1.038	0.959	0.877
	DM-stat	-2.078	0.484	-0.637	-1.772

Notes: Relative MSE values and Diebold-Mariano statistics for the EIA forecast compared to the NC forecast for $h = \{3, 6, 9, 12\}$. Values below one for the rMSE indicate better performance of the EIA forecast.

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 again indicates mild significance at the three-month horizon. Remarkably, the relative MSE exceeds unity for the six-month horizon, resulting in a reversed sign of the DM statistic, excluding a rejection in favour of EIA forecast superiority. 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.

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

Regressor (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
GW_W	16.565	21.365	19.037	63.936
Adjusted R^2	0.052	0.229	0.128	0.091

Notes: For more information see Table 2.

Overall, more diverse regressors are selected compared to the HW forecast. 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. 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 variables are included. We still find net^+ to have a significant positive effect, while most other regressors 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.

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 sizeable 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 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 Further analysis and robustness checks

In this subsection we present additional results and various robustness checks. First, we include supply and demand variables in the conditional predictive ability regression. These variables are not measured in real-time, nonetheless, it is of interest, whether these hold explanatory power for the loss differential. We apply the same variable selection procedures in these additional regressions. Next, we use the Schwarz criterion which has a different penalty term in the best subset selection for the dynamic conditional predictive ability regressions in order to control for the variable selection. Furthermore, we study the general-to-specific approach (GETS) 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. 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.

4.4.1 Additional variables

Up to now the conditional predictive ability regression solely included real-time regressors. This is our main focus, as it is of special interest to predict the loss differential with information available when the forecast is made. This is in line with the definition of forecast efficiency, minimizing the loss using all available information (Nordhaus, 1987). Nonetheless, there might be other drivers of the oil price than previously used oil price change measures such as supply and demand variables.[20] Including such regressors is of threefold interest: (i) such variables have been useful in the context of forecasting oil prices or volatility, often connected to non-linearities. This might transmit to the loss differential; (ii) with respect to the forecast efficiency, it would be of interest whether such fundamental factors are included in the information set, if so they should not hold predictive power for the loss differential; (iii) even if not available as real-time variable, the effect of including new variables on the significance of other regressors applies as further robustness check.

Results including supply (*oic*) and demand variables (*gea*, *gecon*) are reported in Tables 9 to 14 in the Appendix. We apply the same robustness checks in form of variable selection as before. Regarding the HW approach, the results do not change much. The null hypothesis of no conditional predictive ability is rejected for all forecast horizons and all variable selection algorithms. The demand variables *gea* and *gecon* are also selected, especially at the $h = 3$ and $h = 12$ month horizon. Despite that, the other evaluation metrics are subject to minor changes not changing the overall results.

With regard to the EIA forecasts, the effect of additional variables is stronger. While the null hypothesis of no conditional predictive ability is still rejected with the exception of $h = 3$ when using Mallows criterion, the effect of the newly proposed explosivity indicator decreased. Meanwhile, the demand variables are selected in several settings, even though having a minor effect on the Wald statistics which is in line to the main analysis.

4.4.2 Variable selection

In order to robustify the variable selection we apply different selection algorithms to all previous setting. The results can be found in the Appendix in Tables 5 to 14. 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 regressor. 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 the Giacomini-White statistic. Turning to the general-to-specific approach, we find a lot more variables to be included at all forecast horizons, see Table 6. Besides, the results are quite similar to the previous findings and the interpretation of results is consistent with the main analysis emphasizing the importance of the newly proposed indicator.

[20] Additionally, we included the nominal advanced foreign economies U.S. dollar index which was barely selected and not significant. The results are available upon request.

When considering the results for EIA forecasts based on the Schwarz criterion (Table 7), we find that fewer variables 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. 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.

Overall the robustness checks lead to similar results compared to the main analysis. They underline the importance of the newly proposed explosivity indicator when predicting the loss differential, especially when including the market-based expectations. With only small changes to the Wald statistics, the results hold, when changing the variable selection algorithm.

4.4.3 Dynamic rotation

When rejecting the null hypothesis of equal conditional predictive ability, this has actual consequences for selecting a forecast (Giacomini and White, 2006). Granziera and Sekhposyan (2019) state an interesting case 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 chosen based on s_t. The fitted values from the conditional predictive ability regression can be computed in order to decide which forecast is chosen in period t. Depending on whether the fitted value $E[\Delta L_{t+h}|s_t] = \theta' s_t$ is larger or smaller than one, either the competing forecast or the benchmark is chosen respectively.

We employ this dynamic rotation approach for all previous regressions, e.g. rotating between either the HW and the NC forecast or the EIA and the NC forecast. The results can be found in the Appendix in Table 15 based on real-time predictors and Table 16 including supply and demand variables.[21]

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

[21]The dynamic rotation results are evaluated by following metrics: The rMSE between the HW/EIA forecast and the rotated forecast, the relative performance, indicating the percentage of times, the benchmark was chosen in the rotation and the success ratio, stating how often the fitted value predicted the correct sign of the loss differential.

horizon of three month.[22] 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 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 and the rMSE statistics show that some minor improvements can be achieved.

For the EIA forecasts 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. [23] 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 and West (2007) statistic is significant at the five percent level for $h = 3$ and $h = 6$ and at the ten percent level for $h = 9$. For either HW and EIA forecasts, results including supply and demand variables barely change.

Overall, improvements to the rMSE are achievable, with major reductions in some cases. Note that this methodology can serve as a starting point for a deeper analysis. The regression framework is essential to the GW test and offers high interpretability of the parameters allowing economic interpretation. The variable selection from our analysis may serve as build-up to a conditional forecast combination approach (see e.g. Gibbs and Vasnev, 2024) or a more sophisticated dynamic conditional rotation based on non-linear or machine learning models.[24]

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

[23] For this forecast horizon, the high improvement in the rMSE is especially driven by the no-change forecast outperforming the EIA forecast. Furthermore, the R^2 of the respective regression is considerably higher compared to the other forecast horizons, despite lower t-statistics of the oil change regressors. Nonetheless, the intercept explains lots of the variation of the loss differential for $h = 6$.

[24] As an alternative to dynamic rotation, changes to the respective method would be of interest as well. A possible technical explanation why the HW market-based expectations relative forecasting performance is worse during explosive phases may relate to the model persistence parameter matrix ρ. In case of local-to-unity or even exact unit roots, the log-likelihood surface becomes flat, resulting in an identification problem (Hamilton and Wu, 2012). It could be helpful to address this problem in future considerations.

4.5 Economic implications

Accurate and reliable expectations are of utmost importance in decision-making during uncertain times and highly relevant in forward-looking models such as the New-Keynesian Phillips Curve, see e.g. Mavroeidis et al. (2014). An expectational shock can be defined as the difference between the expected value and the actual value. Such shocks are of major interest when assessing economic models or policy (see e.g. Baumeister and Kilian, 2016 and Baumeister, 2023). As our robust results show, market-based and survey-based expectations are not unconditionally well-suited expectation measures. Therefore, analyses based on these (i) expectations and (ii) implied expectational shocks can be prone to biased conclusions. Especially, as the relative performance of such expectations is worst when the oil price undergoes explosive phases, the subsequent shocks are not located in the centre of the distribution, but rather in the tails. In particular the left tail imposes the highest risk for most economic agents (Baumeister et al., 2024).

In general, expectational shocks are commonly used in VAR analyses (see e.g. Barsky and Sims, 2012 or Clements and Galvão, 2021). For example, Anderl and Caporale (2024) investigate the effect of expectational oil shocks on inflation. Further, Känzig (2021) studies the macroeconomic effects of expectational oil price shocks and Bruns and Lütkepohl (2023) investigate the effect of revisions to price expectations. Similarly, Valenti et al. (2023) include an expectational shock in a structural VAR model, acknowledging the role of HW expectations. For a non-zero risk premium, the HW expectation is a more suitable measure than the futures price alone, see Baumeister (2023). Shocks derived from survey expectations are also common in the literature, see for example Prat and Uctum (2024) who analyse such expectations as a risk premium measure in the context of fundamental and speculative effects on risk pricing.

Strikingly, we have shown that the relative performance of the HW expectations, but also survey expectations, strongly deteriorate in turbulent market phases. Next to the construction of surprise shocks in the first step, disentangling the supply and demand components in a second step as e.g. in Baumeister (2023), can be affected by an inappropriate expectation measure used in the first place. Therefore, our results may have implications for the literature of identifying pure expectational oil shocks, orthogonalized from supply and demand components, as in e.g. Degasperi (2023). Our results imply that analyses based on expectational oil shocks should be conducted carefully, as different underlying expectation measures are likely to be subject to time-varying performance.

Similarly, this holds for other applications relying on accurate oil price expectations. The literature on testing against speculative bubbles, Pavlidis et al. (2018) might be directly affected. Their testing procedures immediately 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. Further, modelling of storage demands (see e.g. Baumeister and Kilian, 2014) or individual market participants' decisions (see e.g. Allcott and Wozny, 2014 on how consumers base vehicle purchase decisions of expected future gasoline prices) rely on accurate forecasts. The same holds for producers decisions regarding investments in production capacities (see e.g. Anderson, Kellogg, and Salant, 2018) and therefore regulatory decisions. In all these cases, accurate and reliable oil price expectations are key to decision-making under uncertainty.

5 Conclusions

We investigate oil price expectations obtained from (i) a Gaussian affine term structure model based on futures prices and (ii) 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 only weakly 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 indicate that there is clear evidence against the null hypothesis of no conditional predictive ability. The applied best subset selection approach reveals that besides a couple of established oil price change measures (e.g. net oil price increases and implied volatility), our newly constructed real-time indicator for explosiveness has strong explanatory power for the loss differential. Results turn out to be robust with respect to a number of variations in the evaluation 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 or fiscal and monetary, but also climate policy.

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

Allcott, H. and N. Wozny (2014). Gasoline prices, fuel economy, and the energy paradox. *The Review of Economics and Statistics 96*(5), 779–795.

Alquist, R., L. Kilian, and R. J. Vigfusson (2013). Forecasting the Price of Oil. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2 of *Handbook of Economic Forecasting*, 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.

Anderl, C. and G. M. Caporale (2024). Functional oil price expectations shocks and inflation. *Journal of Futures Markets 44*(10), 1662–1693.

Anderson, S. T., R. Kellogg, and S. W. Salant (2018). Hotelling under pressure. *Journal of Political Economy 126*(3), 984–1026.

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.

Barsky, R. B. and E. R. Sims (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review 102*(4), 1343–77.

Baumeister, C. (2023). 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., F. Huber, and M. Marcellino (2024). Risky oil: It's all in the tails. Working Paper 32524, National Bureau of Economic Research.

Baumeister, C. and L. Kilian (2014). A general approach to recovering market expectations from futures prices with an application to crude oil. *CFS Working Paper*.

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.

Baumeister, C., D. Korobilis, and T. K. Lee (2022). Energy Markets and Global Economic Conditions. *The Review of Economics and Statistics 104*(4), 828–844.

Bertsimas, D., A. King, and R. Mazumder (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics 44*(2), 813 – 852.

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

Bruns, M. and H. Lütkepohl (2023). Have the effects of shocks to oil price expectations changed? evidence from heteroskedastic proxy vector autoregressions. *Economics Letters 233*, 111416.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). General-to-specific modeling: an overview and selected bibliography. International Finance Discussion Papers 838, Board of Governors of the Federal Reserve System (U.S.).

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.

Clements, M. P. and A. B. Galvão (2021). Measuring the effects of expectations shocks. *Journal of Economic Dynamics and Control 124*, 104075.

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

Degasperi, R. (2023). Identification of Expectational Shocks in the Oil Market using OPEC Announcements. The Warwick Economics Research Paper Series (TWERPS) 1464, University of Warwick, Department of Economics.

Degiannakis, S. and G. Filis (2023). Oil price assumptions for macroeconomic policy. *Energy Economics 117*, 106425.

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.

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.

Farmer, L. E., L. Schmidt, and A. Timmermann (2023). Pockets of predictability. *The Journal of Finance 78*(3), 1279–1341.

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.

Gibbs, C. G. and A. L. Vasnev (2024). Conditionally optimal weights and forward-looking approaches to combining forecasts. *International Journal of Forecasting 40*(4), 1734–1751.

Granger, C. W. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society 20*(2), 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.

Hamilton, J. D. and J. C. Wu (2015). Effects of index-fund investing on commodity futures prices. *International Economic Review 56*(1), 187–205.

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.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2021). *An introduction to statistical learning: With applications in R* (2 ed.). Springer.

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

Kilian, L. and 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.

Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from opec announcements. *American Economic Review 111*(4), 1092–1125.

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

Kruse-Becher, R. (2024). Let's switch again! Testing for speculative oil price bubbles based on dynamically rotated market expectations. *mimeo*.

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

Li, J., Z. Liao, and R. Quaedvlieg (2021, 06). Conditional superior predictive ability. *The Review of Economic Studies 89*(2), 843–875.

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

Mavroeidis, S., M. Plagborg-Møller, and J. H. Stock (2014). Empirical evidence on inflation expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature 52*(1), 124–88.

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

Nordhaus, W. D. (1987). Forecasting efficiency: Concepts and applications. *The Review of Economics and Statistics 69*(4), 667–674.

Odendahl, F., B. Rossi, and T. Sekhposyan (2023). Evaluating forecast performance with state dependence. *Journal of Econometrics 237*(2, Part C), 105220.

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

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

Prat, G. and R. Uctum (2024). Risk premium, price of risk and expected volatility in the oil market: Evidence from survey data. *Energy Economics 140*, 107930.

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.

Salisu, A. A., R. Gupta, E. Bouri, and Q. Ji (2022). Mixed-frequency forecasting of crude oil volatility based on the information content of global economic conditions. *Journal of Forecasting 41*(1), 134–157.

Sockin, M. and W. Xiong (2015). Informational frictions and commodity markets. *The Journal of Finance 70*(5), 2063–2098.

Valenti, D., A. Bastianin, and M. Manera (2023). A weekly structural VAR model of the US crude oil market. *Energy Economics 121*, 106656.

Wang, Y., L. Liu, and C. Wu (2017). Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. *Energy Economics 66*, 337–348.

Zhang, L., Y. Chen, and E. Bouri (2024). Time-varying jump intensity and volatility forecasting of crude oil returns. *Energy Economics 129*, 107236.

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

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

Regressor (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
GW_W	18.892	43.241	326.956	485.134
Adjusted R^2	0.259	0.514	0.592	0.422

Notes: For more information see Table 2.

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

Regressor (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
GW_W	21.208	220.094	1441.283	5267.865
Adjusted R^2	0.256	0.517	0.586	0.424

Notes: For more information see Table 2.

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

Regressor (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	-
GW_W	10.283	36.168	18.552	6.312
Adjusted R^2	0.041	0.213	0.095	0.074

Notes: For more information see Table 2.

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

Regressor (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
GW_W	4.351	26.494	30.079	85.041
Adjusted R^2	0.034	0.219	0.108	0.074

Notes: For more information see Table 2.

Table 9: Conditional predictive ability results including supply and demand variables, HW, best subset selection via Mallows criterion.

Regressor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	-1.835	0.471	0.352	0.270
net^{+}	2.934	2.538	2.398	-
net^{-}	-	-	-	-
gap	-	-	-	-
$large$	-	-	-	-
$large^{+}$	-	-	-	-
ovx	-	-1.063	-0.929	-1.017
$expl$	-	-6.819	-11.316	-15.890
oic	-	-	-	-
gea	-	-	-	-1.850
$gecon$	4.119	-	-	-
GW_W	20.443	104.526	326.956	611.880
Adjusted R^2	0.265	0.521	0.592	0.442

Notes: The results for the conditional predictive ability regression, comparing the HW forecast and the NC forecast. The regressors are chosen via best subset selection using Mallows criterion. The table lists the t-statistics for selected regressors, the Wald statistic GW_W on which the test of conditional predictive ability is based on, as well as the adjusted R^2 of the regression

Table 10: Conditional predictive ability results including supply and demand variables, HW, best subset selection via Schwarz criterion.

Regressor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	-1.835	-1.352	0.352	0.270
net^+	2.934	3.506	2.398	-
net^-	-	-	-	-
gap	-	-	-	-
$large$	-	-	-	-
$large^+$	-	-	-	-
ovx	-	-	-0.929	-1.017
$expl$	-	-6.348	-11.316	-15.890
oic	-	-	-	-
gea	-	-	-	-1.850
$gecon$	4.119	-	-	-
GW_W	20.443	43.241	326.956	611.880
Adjusted R^2	0.265	0.514	0.592	0.442

Notes: For more information see Table 9.

Table 11: Conditional predictive ability results including supply and demand variables, HW, General-to-Specific.

Regressor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	-0.887	-0.342	-0.589	-1.889
net^+	2.435	2.084	2.766	-
net^-	-0.536	0.757	-0.299	-
gap	0.009	0.300	0.066	-
$large$	0.394	-	-0.420	-
$large^+$	-0.664	-	-	-
ovx	-	-	-	
$expl$	-3.585	-7.192	-9.831	-13.710
oic	-	-	-	-
gea	-0.236	-0.003	-0.302	-1.706
$gecon$	2.547	-	2.197	2.334
GW_W	167.328	453.096	527.886	494.645
Adjusted R^2	0.247	0.509	0.582	0.439

Notes: For more information see Table 9.

Table 12: Conditional predictive ability results including supply and demand variables, EIA, best subset selection via Mallows criterion.

Regressor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	-2.629	-0.353	-1.333	-2.325
net^{+}	1.580	1.775	2.343	-
net^{-}	-	-	-	-
gap	-	-	-	-
$large$	-	-0.790	-1.856	-3.305
$large^{+}$	-	-	1.805	1.913
ovx	-	-	-	-
$expl$	-	0.660	-2.472	-
oic	-	-	-	-
gea	1.711	1.732	1.328	-
$gecon$	1.679	-	1.951	-
GW_W	16.438	22.954	44.511	26.662
Adjusted R^2	0.069	0.249	0.143	0.147

Notes: For more information see Table 9.

Table 13: Conditional predictive ability results including supply and demand variables, EIA, best subset selection via Schwarz criterion.

Regressor (t-stats)	$h=3$	$h=6$	$h=9$	$h=12$
Intercept	-2.955	-0.035	-0.957	-2.325
net^{+}	1.999	1.748	2.733	-
net^{-}	-	-	-	-
gap	-	-	-	-
$large$	-	-	-	-3.305
$large^{+}$	-	-	-	1.913
ovx	-	-	-	-
$expl$	-	-	-3.198	-
oic	-	-	-	-
gea	-	1.926	-	-
$gecon$	-	-	-	-
GW_W	10.283	13.56	18.522	26.662
Adjusted R^2	0.041	0.233	0.095	0.147

Notes: For more information see Table 9.

Table 14: Conditional predictive ability results including supply and demand variables, EIA, General-to-Specific.

Regressor (t-stats)	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Intercept	-2.004	0.080	-0.63	-1.223
net^+	-	1.911	2.576	0.169
net^-	-	-	1.430	-
gap	-	0.295	-	-0.469
$large$	-	-0.865	-1.564	-2.828
$large^+$	-	-	-	1.682
ovx	-	-	-	-
$expl$	-	-	-2.415	-
oic	-	-	-0.627	-
gea	-	1.936	1.143	-
$gecon$	-	-	1.032	4.652
GW_W	4.017	16.638	29.865	33.468
Adjusted R^2	0.000	0.235	0.131	0.140

Notes: For more information see Table 9.

Table 15: Dynamic rotation results based on real-time predictors

		$h = 3$		$h = 6$		$h = 9$		$h = 12$	
		HW	EIA	HW	EIA	HW	EIA	HW	EIA
Mallows	Relative performance	0.183	0.072	0.102	0.232	0.052	0.161	0.000	0.175
	Success ratio	0.556	0.583	0.508	0.565	0.494	0.534	0.585	0.626
	Rotated rMSE	0.982	0.967	1.032	0.915	1.033	0.959	1.000	1.034
Schwarz	Relative performance	0.183	0.072	0.051	0.153	0.052	0.063	0.000	0.146
	Success ratio	0.556	0.583	0.525	0.542	0.494	0.552	0.585	0.608
	Rotated rMSE	0.982	0.967	0.978	0.943	1.033	0.974	1.000	1.039
GETS	Relative performance	0.239	0.189	0.130	0.322	0.034	0.236	0.012	0.175
	Success ratio	0.577	0.577	0.525	0.577	0.511	0.540	0.596	0.602
	Rotated rMSE	0.973	0.981	0.987	0.914	0.984	0.958	0.993	1.042

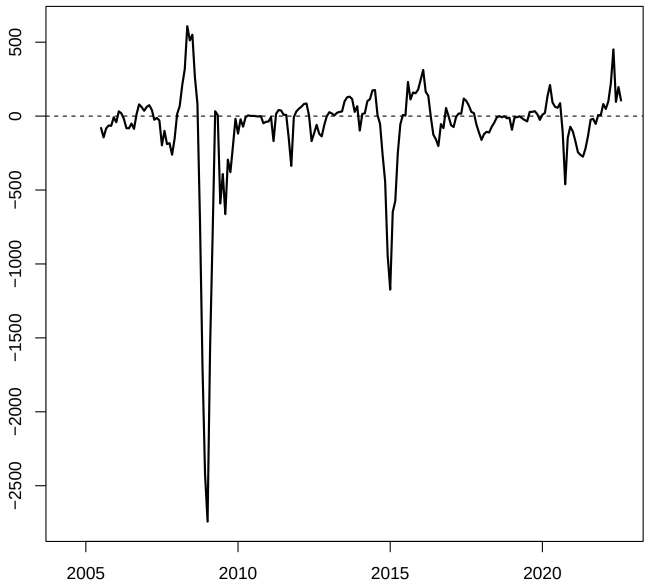
This table contains the dynamic rotation results based on real-time predictors. The results are based on the variables selected by the best subset selection (Mallows/Schwarz) and the general-to-specific (GETS) approach. The table contains results for a dynamic rotation between the HW approach and the NC forecast and the rotation between the EIA forecast and the NC prediction. The relative performance indicates the percentage of how often the NC was selected in the rotation. The success ratio indicates, how often the selected forecast actually produced a smaller squared error. And the rotated rMSE depicts the ratio between the rotated forecast and either the HW or the EIA forecast, thus values below unity indicate an improvement due to the rotation.

Table 16: Dynamic rotation results based on real-time predictors and supply and demand variables

		$h = 3$		$h = 6$		$h = 9$		$h = 12$	
		HW	EIA	HW	EIA	HW	EIA	HW	EIA
Mallows	Relative performance	0.144	0.172	0.102	0.452	0.052	0.322	0.070	0.170
	Success ratio	0.572	0.628	0.509	0.548	0.494	0.558	0.597	0.620
	Rotated rMSE	0.963	0.958	1.032	0.910	1.033	0.944	1.057	0.983
Schwarz	Relative performance	0.144	0.072	0.051	0.407	0.052	0.063	0.070	0.170
	Success ratio	0.572	0.583	0.525	0.537	0.494	0.552	0.597	0.620
	Rotated rMSE	0.963	0.967	0.978	0.912	1.033	0.974	1.057	0.983
GETS	Relative performance	0.172	0.000	0.057	0.446	0.075	0.385	0.233	0.170
	Success ratio	0.556	0.544	0.531	0.462	0.494	0.586	0.561	0.632
	Rotated rMSE	0.973	1.000	0.978	0.908	0.989	0.944	1.025	0.976

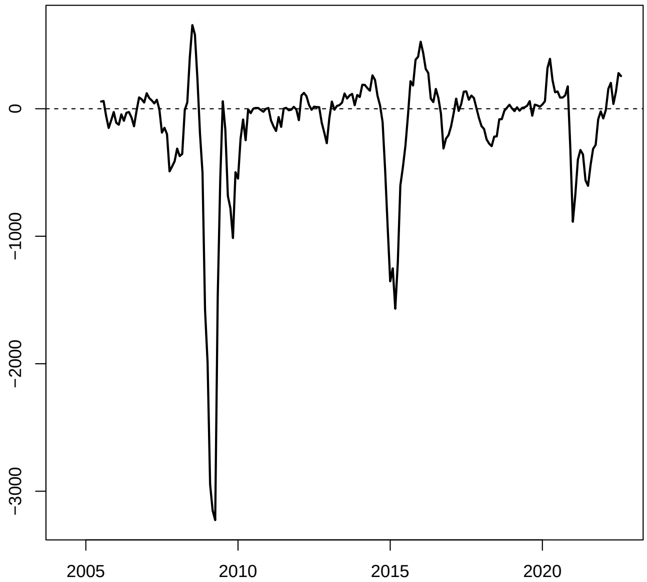
See Table 15 for more details. These results are based on the real-time predictors and the supply and demand variables.

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



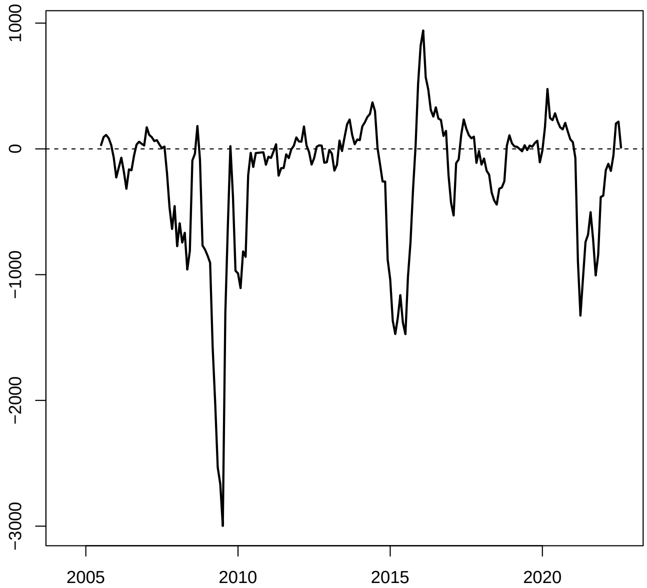
Notes: For more information see Figure 1.

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



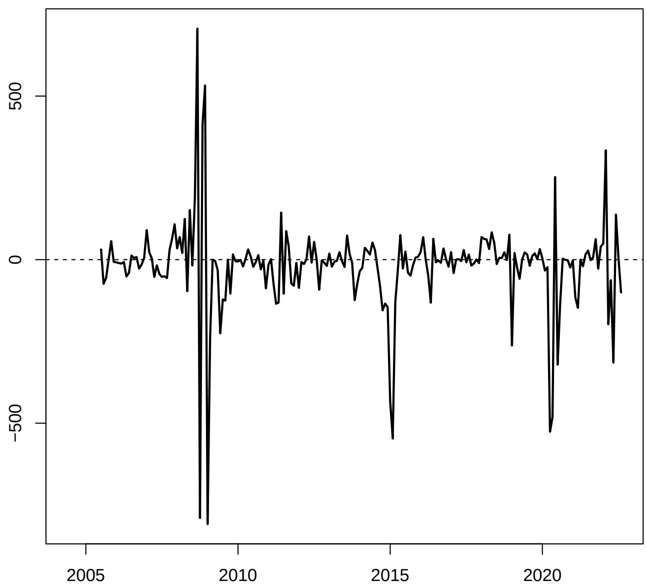
Notes: For more information see Figure 1.

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



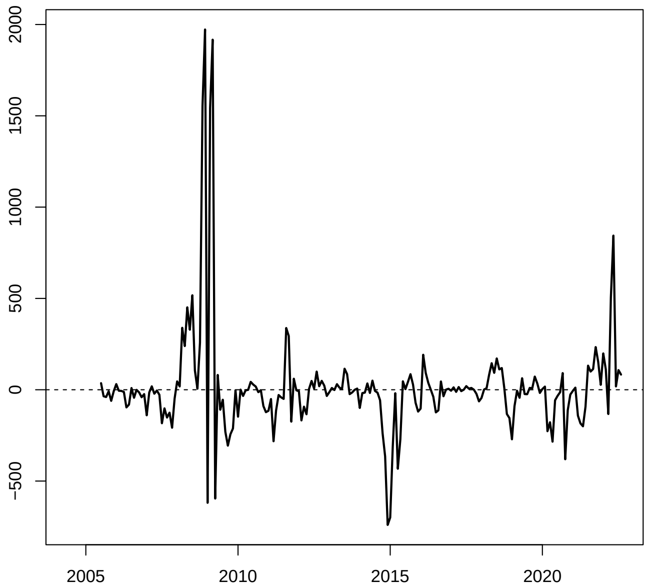
Notes: For more information see Figure 1.

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



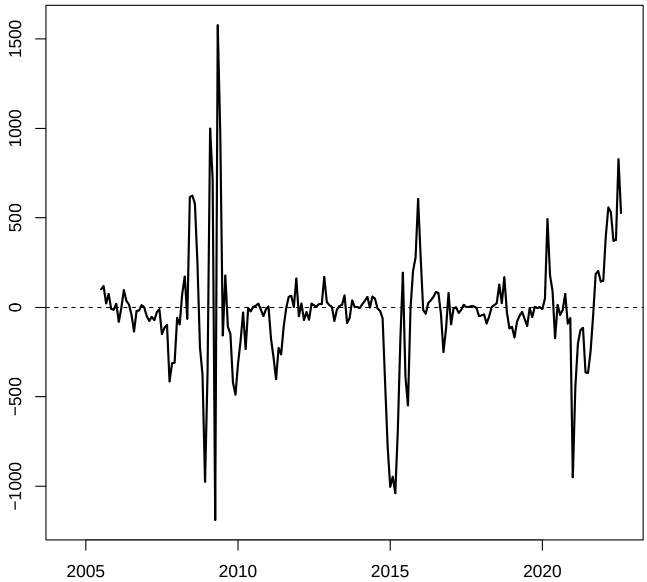
Notes: For more information see Figure 1.

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



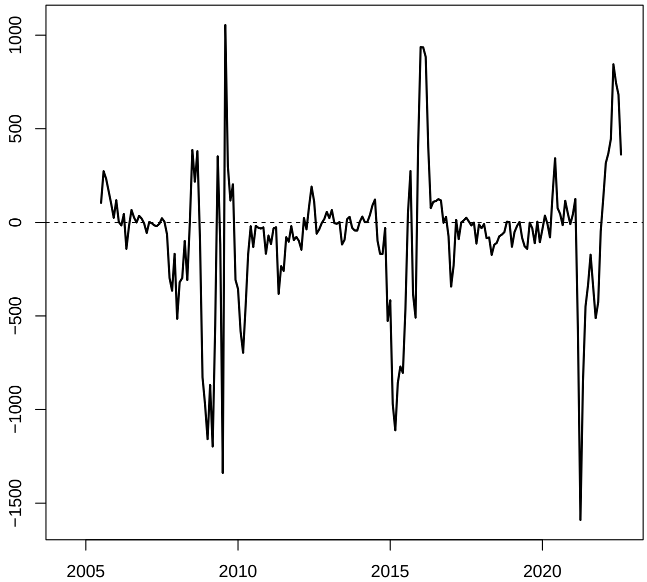
Notes: For more information see Figure 1.

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



Notes: For more information see Figure 1.

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



Notes: For more information see Figure 1.

CESA Working Paper

Series

CESA Working Paper No. 1 | 2023

Andreas Hauptmann, Benjamin Schwanebeck and Hans-Jörg Schmerer
Plant-level adjustments to imports and exports at the extensive margin

CESA Working Paper No. 2 | 2023

Joscha Beckmann, Marco Kerkemeier and Robinson Kruse-Becher
Regime-specific exchange rate predictability

CESA Working Paper No. 3 | 2023

Joscha Beckmann, Timo Heinrich and Jennifer Rogmann
Inflation expectations and cognitive uncertainty

CESA Working Paper No. 4 | 2024

Daniel Monsees and Matthias Westphal
Disruptions in Primary Care: Can Resigning GPs Cause Persistently Negative Health Effects?

CESA Working Paper No. 5 | 2024

Giulio Callegaro, Mario Lackner and Hendrik Sonnabend
The Napoleon complex revisited: New evidence from professional soccer

CESA Working Paper No. 6 | 2024

Mario Lackner and Hendrik Sonnabend
When performance melts away: Heat causes mental errors in high-stakes competitions

CESA Working Paper No. 7 | 2024

Johannes Hollenbach, Hendrik Schmitz and Matthias Westphal
Gene-environment interactions with essential heterogeneity

CESA Working Paper No. 8 | 2024

Kamila Cygan-Rehm and Matthias Westphal
School Starting Age and the Gender Pay Gap over the Life Cycle

CESA Working Paper

Series

CESA Working Paper No. 9 | 2024
Marco Kerkemeier
(Co-)Explosiveness of corporate credit spreads

CESA Working Paper No. 10 | 2024
Christoph Wegener, Robinson Kruse-Becher and Tony Klein
EU ETS Market Expectations and Rational Bubbles

CESA Working Paper No. 11 | 2025
Anna Krumme and Matthias Westphal
Monetary returns to upper secondary schooling, the evolution of unobserved heterogeneity, and implications for employer learning

CESA Working Paper No. 12 | 2025
Robinson Kruse-Becher and Philip Letixerant
Oil price expectations in explosive phases

FernUniversität in Hagen