Forecasting realized oil-price volatility: The role of financial stress and asymmetric loss

Konstantinos Gkillas[a],*, Rangan Gupta[b] and Christian Pierdzioch[c]

[a]Department of Business Administration, University of Patras – University Campus, Rio, P.O. Box 1391, 26500 Patras, Greece
[b]Department of Economics, University of Pretoria, Pretoria 0002, South Africa
[c]Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

*Corresponding author. gillask@upatras.gr

Highlights

- We analyze the role of financial stress in forecasting oil-price volatility.
- We use various variants of the Heterogenous Autoregressive model of realized volatility.
- We find that indexes of financial stress help to improve forecasting performance.
- The shape of the forecaster loss function that used to evaluate performance is important.
- Alternative types of investors benefit from monitoring different regional sources of financial stress.

Abstract

We analyze the role of global and regional measures of financial stress in forecasting realized volatility of the oil market based on 5-min intraday data covering the period of 4th January, 2000 until 26th May, 2017. In this regard, we use various variants of the Heterogeneous Autoregressive (HAR) model of realized volatility (HAR-RV). Our main finding is that indexes of financial stress help to improve forecasting performance, with it being important to differentiate between regional sources of financial stress (United States, other advanced economies, emerging markets). Another key finding is that the shape of the forecaster loss function that one uses to evaluate forecasting performance plays an important role. More specifically, forecasters who attach a higher cost to an overprediction of realized volatility as compared to an underprediction of the same absolute size should pay particular attention to financial stress originating in the U.S. But, in case an underprediction is more costly than a comparable overprediction, then forecasters should closely monitor financial stress caused by developments in emerging-market economies. In sum, financial stress does have predictive value for realized oil-price volatility, with alternative types of investors benefiting from monitoring different regional sources of financial stress.

JEL classification: G17; Q02; Q47

Keywords: Oil price; Realized volatility; Financial stress; Forecasting; Asymmetric loss

1. Introduction

Oil-price volatility, when interpreted as a measure of uncertainty, has been shown to negatively impact economic activity (Elder and Serletis, 2010, Aye et al., 2014, van Eyden et al., 2019). Moreover, recent financialization of the oil market has resulted in increased participation of hedge funds, pension funds, and insurance companies in the market, and the oil market is now also considered as a profitable alternative investment in the portfolio decisions of financial institutions (Akram, 2009, Tang and Xiong, 2012, Silvennoinen and Thorp, 2013, Fattouh et al., 2013, Büyükşahin and Robe, 2014, Bahloul et al., 2018). Accurate forecasts of oil-price volatility, thus, are of paramount importance to policymakers, oil traders, and academics.

Not surprisingly, a large literature exists that has analyzed the forecastability of daily oil-price volatility using various types of univariate and multivariate models, primarily from the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family (see, e.g., Sadorsky, 2006, Sadorsky and McKenzie, 2008, Agnolucci, 2009, Kang et al., 2009, Wei et al., 2010, Nomikos and Pouliasis, 2011, Arouri et al., 2012, Hou and Suardi, 2012, Kang and Yoon, 2013, Chkili et al., 2014, Efimova and Serletis, 2014). In general, these studies find that the univariate GARCH-type models are able to produce more accurate forecasts than any other competing models. More recently, Lux et al. (2016) compare the Markov-switching multifractal (MSM) model with a battery of GARCH models, and show that the MSM model comes out as the framework that most often (across forecasting horizons and subsamples) cannot be outperformed by other models.

All the above-mentioned papers use daily oil price returns and forecast the conditional oil-price volatility. Recent empirical evidence, however, suggests that the rich information contained in intraday data can produce more accurate estimates and forecasts of daily volatility (Andersen et al., 2001, Andersen et al., 2003 Oomen, 2001, Hansen and Lunde, 2005, Engle and Sun, 2007, McAleer and Medeiros, 2008, Tay et al., 2009). Given this, Haugom et al., 2014, Sévi, 2014, Prokopczuk et al., 2015, Degiannakis and Filis, 2017 use variants of the Heterogeneous Autoregressive (HAR) model developed by Corsi (2009) to forecast the realized volatility (RV) of oil price returns (i.e., the sum of non-overlapping squared high-frequency oil returns observed within a day, Andersen and Bollerslev, 1998).[1] The popularity of the HAR model stems from its ability to capture important stylized facts of financial-market volatility such as long memory and multi-scaling behaviour.[2] In sum, barring Degiannakis and Filis (2017), all the intraday-based studies conclude that none of the alternative models is able to outperform the forecasting accuracy of a simple HAR model (also called HAR-RV model), which uses only the information embedded in the realized volatility to compute forecasts. Degiannakis and Filis (2017) show that it is possible to beat the HAR-RV model by incorporating information on the volatilities (assumed to be exogenous) of four different asset classes.

Against this backdrop, our paper aims to extend the limited literature on forecasting realized oil-price volatility (derived based on 5-min-interval intraday data) based on the HAR-RV model by incorporating the role of financial stress into the modeling framework, where the sample period covers the daily period from 4th January, 2000 until 26th May, 2017. As pointed out by Nazlioglu et al. (2015), financial stress can impact oil-market movements through its impact on both economic activity and investor behavior. On the one hand, increased financial stress would cause economic activity to slow down (through the bank lending channel via decreasing the amount of available credits and through financial leverage via changes in creditworthiness of borrowing businesses) and lead to lower energy demand and declining oil prices. On the other hand, as discussed above, investors see oil markets as alternative investment options relative to financial markets. Naturally, when investors adjust their portfolios with respect to oil-price movements there will be repercussions felt on financial-asset prices. At the same time, increased financial stress is likely to cause investors to change their portfolios and, hence, is likely to have an impact on oil markets. Moreover, the financialization of the oil market, as discussed above, is believed to have led to its increased comovement with financial markets and, hence, financial stress should affect the oil market. Given these lines of reasoning, in-sample evidence of financial stress in predicting oil-market volatility based on multivariate GARCH-based models has been reported by Nazlioglu et al. (2015) and Gupta et al. (2018). Because in-sample predictability does not necessarily ensure out-of-sample forecasting gains, with Campbell (2008) suggesting that the ultimate test of any predictive model is its out-of-sample performance, our paper makes the first attempt to forecast realized oil-price volatility based on metrics of regional (US, other advanced economies, and emerging economies) and global financial stress. In addition, oil-market participants are known to care not only about the nature of volatility, but also its level (Gkillas et al., 2018), with traders distinguishing between "good" and "bad" volatility. Thus, in a realized estimation framework, following Bollerslev et al. (2017), we also analyze the role of financial stress in forecasting good and bad realized volatility derived from the sum of squared positive and negative intraday returns over a day.

Another contribution of our research is that we use an asymmetric loss function to assess forecasting gains. Forecasters have an asymmetric loss function when the loss they incur in case of an overprediction of oil-price volatility differs from the loss they incur in case of an underprediction of the same (absolute) magnitude. The asymmetric loss function that we consider in our research nests the symmetric quadratic and absolute loss functions commonly studied in earlier research to evaluate forecasting performance and has been extensively studied in earlier research (Elliott et al., 2008, Pierdzioch et al., 2016a). An asymmetric loss function is a natural candidate to evaluate forecasts when one seeks to emulate a utility-function-based approach to evaluate forecasts (as studied in detail by West et al., 1993), when information asymmetries between forecasters and their customers result in biased forecasts (see, e.g., Ehrbeck et al., 1996), when risk-averse policymakers seek to gauge the potential impact of oil-price movements on the overall economy (for the case of inflation forecasts, see Kilian and Manganelli, 2007), and in a risk-management context when forecasters or their customers use predictions of oil-price volatility, for example, to implement option-trading strategies.

We organize the remainder of this research as follows: In Section 2, we describe the methods that we use in our research. In Section 3, we describe our data. In Section 4, we summary our findings. We find that extending the HAR-RV model to include measures of financial stress helps to improve forecasting performance, where we document the differential effect of regional sources of financial stress for forecasts of realized oil-price volatility. We also document how forecasting performance depends upon the shape of the forecaster loss function. In Section 5, we conclude.

2. Methods

We use the median realized variance (*MRV*) proposed by Andersen et al. (2012) as an estimator of the integrated variance of oil intraday returns. We focus on *MRV* for the following reasons: (i) *MRV* has better theoretical efficiency properties than the tripower variation measure. (ii) *MRV* is a jump-robust estimator of integrated variance, that is, it is a less biased estimator than other measures of *RV* in the presence of jumps. Using *MRV*, thus, allows information on jumps (the return variation caused by jumps) to be separated from the model (and especially from financial stress), keeping the forecasting model less complex. Although jumps have a strong impact on future volatility, they are extremely unpredictable and add a significant source of nondiversifiable risk in volatility, which is locally sourced and more difficult to predict (see Bollerslev et al., 2008, Todorov and Tauchen, 2011, Degiannakis and Floros, 2015, among others).[3] (iii) *MRV* attenuates the effect of market microstructure noise and exhibits better finite sample properties even in the presence of noise than other *RV* measures. (iv) *MRV* is not influenced by the sampling frequency compared with other jump-robust estimators and displays better finite-sample robustness to the occurrence of "zero" returns in the sample. The *MRV* is given by

$$MRV_t = \frac{\pi}{6-4\sqrt{3}+\pi} \frac{T}{T-2} \sum_{i=2}^{T-1} med\left(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|\right)^2, \tag{1}$$

Where $r_{t,i}$ denotes the intraday return *i* within day *t* and $i = 1, \ldots, T$ denotes the number of intraday observations within a day. We consider *MRV* as our measure of daily standard realized volatility (RV^S).[4]

Further, Barndorff-Nielsen et al. (2010) study downside and upside realized semi-variance (RV^B and RV^G) as measures based entirely on downward or upward movements of intraday returns. The realized down term corresponds to the variability of downside movements constructed using negative intraday returns, while the realized up term is constructed using the variability of upside price movements constructed only from positive intraday returns. Formally, as defined by Barndorff-Nielsen et al. (2010), RV_t^B and RV_t^G are computed as follows:

$$RV_t^B = \sum_{i=1}^{T} r_{t,i}^2 \; I_{[(r_{t,i})<0]}, \tag{2}$$

$$RV_t^G = \sum_{i=1}^{T} r_{t,i}^2 \; I_{[(r_{t,i})>0]}, \tag{3}$$

where $I_{\{.\}}$ denotes the indicator function. Understandably, $RV = RV^B + RV^G$, where $RV = \sum_{i=1}^{T} r_{t,i}^2$. Following Bollerslev et al. (2017), in a realized estimation framework, we consider daily as RV^B "bad" realized volatility and RV^G as "good" realized volatility. This distinction renders it possible to capture the sign asymmetry of the volatility process and, thus, to define more accurately the role of financial stress in various market phases such as booms and crashes. Furthermore, previous studies have identified the importance of downside risk in portfolio risk management. Unlike other proxies of downside risk (e.g., downside deviation, downside beta, value at risk, etc.), downside realized semivariance is constructed based on high-frequency data. As a result, it is a better proxy for downside risk because it is more accurately estimated at a daily frequency (Hansen and Huang, 2016). The main reason is that intraday data reveal important information about the market not easily seen at a daily frequency. Such an analysis is also important for traders and policymakers.

As far as the literature on modelling and forecasting realized volatility is concerned, Corsi (2009) proposes the HAR-RV model, which has become one of the most popular models in this line of research. The benchmark HAR-RV model, for *h*-days-ahead forecasting, can be described as follows:

$$RV^j_{t+h} = \beta_0 + \beta_d\ RV^j_t + \beta_w\ RV^j_{w,t} + \beta_m\ RV^j_{m,t} + \epsilon_{t+h}, \tag{4}$$

where (to simplify notation) *h* denotes the forecast horizon and *j* can be either S, B or *G* as described earlier. $RV^j_{w,t}$ denotes the average RV^j from day $t-5$ to day $t-1$, while $RV^j_{m,t}$ denotes the average RV^j from day $t-22$ to day $t-1$. The HAR-RV model, thus, uses volatilities from different time resolutions to forecast realized oil-price volatility and, thereby, captures the main idea motivating the so called heterogeneous market hypothesis (Müller et al., 1997). This hypothesis captures the idea that different classes of market participants populate the oil market, where the different classes can be formed according to the sensitivity of market participants to information flows at different time horizons (that is, short-term traders versus long-term traders).

We use the standard HAR-RV model as our benchmark model for predicting realized volatility. We then add to this benchmark model indexes of regional (U.S., other advanced economies, emerging market economies) and global financial stress, denoted by FSI^i, where *i* denotes the source of financial stress. To this end, we consider the following modified HAR-RV model (where the index *j* denotes the realized volatility series being studied, that is, standard, good, or bad volatility):

$$RV^j_{t+h} = \beta_0 + \beta_d\ RV^j_t + \beta_w\ RV^j_{w,t} + \beta_m\ RV^j_{m,t} + \theta\ FSI^i_t + \epsilon_{t+h}. \tag{5}$$

We let a hat denote forecasts, that is, $\widehat{RV}_{t+h|t}$ denotes the period-*t*forecast of realized volatility in period $t+h$. We consider three forecast horizons: a short forecasting horizon ($h=1$), a medium forecasting horizon ($h=5$), and a long forecasting horizon ($h=22$).[5]

We define the forecast error as $FE_{t+h} = RV_{t+h} - \widehat{RV}_{t+h|t}$, and use the loss function studied by Elliott et al., 2005, Elliott et al., 2008to evaluate forecasting performance. The loss function is given by

$$L(FE_{t+h}, \alpha) = [\alpha + (1-2\alpha)I_{[FE_{t+h}<0]}]|FE_{t+h}|^p. \tag{6}$$

This loss function nests a so-called lin–lin loss function, which obtains when we set $p=1$, and a quad–quad loss function when we set $p=2$. The parameter $\alpha \in (0,1)$ governs the shape of the loss function. A symmetric loss function obtains when we choose $\alpha = 0.5$. Hence, when we choose the parameter configuration $\alpha = 0.5$ and $p=1$, then we evaluate forecasts under the common absolute loss criterion, while the popular squared-error-loss criterion obtains when we set $\alpha = 0.5$ and $p=2$. When the shape parameter differs from its symmetric benchmark value of $\alpha = 0.5$, in turn, the loss function becomes asymmetric. Specifically, for $\alpha = 0.5$ the loss a forecaster incurs in case of an underprediction of realized oil-price volatility exceeds the loss from an overprediction of the same (absolute) magnitude. In contrast, or $\alpha = 0.5$ the loss from an overprediction of realized oil-price volatility exceeds the loss from a corresponding underprediction.

Given the loss function as specified in Eq. (6), we use a rolling-estimation-window approach to estimate the models given in Eqs. (4), (5). We then use the version of the familiar Diebold and Mariano (1995) test proposed by Harvey et al. (1997) to assess the statistical significance of the loss differential between the benchmark HAR-RV model and the HAR-RV model extended to include financial stress.

3. Data

We use intraday data on West Texas Intermediate (WTI) oil futures that are traded at NYMEX over a 24 h trading day (pit and electronic), to construct daily measures of realized volatility, and the corresponding good and bad variants. The upper panel of Fig. 1 shows the resulting realized oil-returns volatility series. The futures price data, in continuous format, are obtained from www.disktrading.com and www.kibot.com. Close to expiration of a contract, the position is rolled over to the next available contract, provided that activity has increased. Daily returns are computed as the end of day (New York time) price difference (close to close). In the case of intraday returns, 1-min prices are obtained via last-tick interpolation (if the price is not available at the 1-min stamp, the previously available price is imputed), and following the realized volatility literature 5-min returns are then computed by taking the log-differences of these prices.

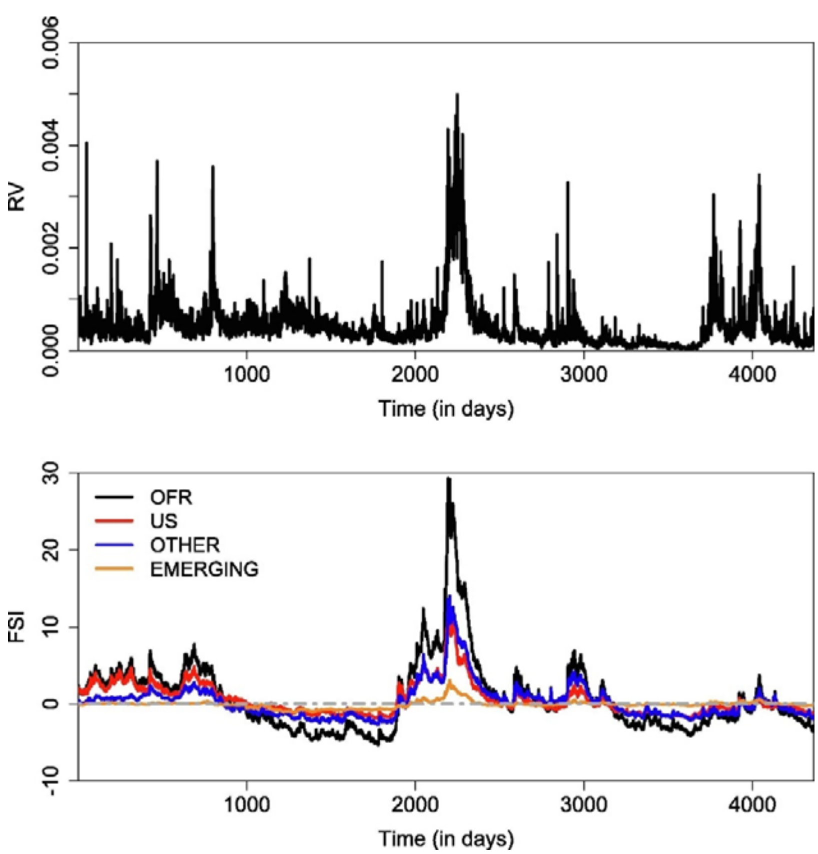


Fig. 1. The Data.

Even though theory suggests that intraday returns should be computed at the highest possible frequency (say, 1-min), in order that volatility estimators converge asymptotically towards the true conditional volatility following fixed domain asymptotics (also called infill asymptotics), we select a 5-min sampling frequency for several quantitative and qualitative reasons. In particular, such a sampling frequency is not too high to induce spurious jumps due to market frictions and also it is not too low to lead to poor data analysis. In other words, in order to avoid extreme high-frequency noise and no-activity periods in very small-time windows (as an extra noise is added when the sampling frequency converges on zero, see Andersen and Bollerslev, 1997, Andersen and Bollerslev, 1998, Taylor and Xu, 1997), following earlier literature, we select the highest sampling frequency as the optimal sampling frequency which minimizes the autocovariance bias (see Oomen, 2004, Degiannakis and Floros, 2016, Degiannakis and Filis, 2017, among others). The optimal sampling frequency takes into account the trade-off between the bias that is inserted in realized volatility estimators and their accuracy. A 5-min sampling frequency also is in line with the evidence provided by Liu et al. (2015). As pointed out by these authors, such a sampling frequency is adequate for liquid assets (as in the case of oil), while higher frequencies are more appropriate for a diversified set of asset classes. The authors also report that the standard 5-min realized variance measure is difficult to beat in forecasting exercises. Furthermore, because we analyze a widely used dataset, a 5-min sampling frequency makes our results comparable with results reported in the earlier oil-volatility forecasting literature (especially for oil futures), where it is common practice to study 5-min returns for crude oil futures (see Chevallier and Sévi, 2012, Sévi, 2014, Ma et al., 2017, Ma et al., 2018, Wen et al., 2019, among others). Finally, it should be noted that beyond MRV (which is considered to attenuate the effect of market-microstructure noise as shown by Andersen et al., 2012 and, thereby, should be robust to the specific choice of sampling frequency), we also study various extensions of the HAR-RV model to study the predictive value of FSI for realized volatility (including higher-order moments and jumps). Amaya et al. (2015) use a 5-min sampling frequency for realized skewness and realized kurtosis. This same sample frequency is also used by several studies for detecting jumps and studying their importance for realized-volatility forecasting (see Andersen et al., 2003, Andersen et al., 2007, Corsi et al., 2010, Duong and Swanson, 2015, among others).

As far as measures of financial stress are concerned, we use the Office of Financial Research (OFR) Financial Stress Indexes (FSIs), which provides daily market-based snapshot of stress in global financial markets. It is constructed from 33 financial market variables, such as yield spreads, valuation measures, and interest rates. For a detailed description, see Monin (2017). The indexes are positive when stress levels are above average, and negative when stress levels are below average. The FSI incorporates five categories of indicators: credit, equity valuation, funding, safe assets, and volatility. It shows stress contributions by three regions: United States (US), other advanced economies, and emerging markets, with the weighted average capturing global financial stress.[6] The lower panel of Fig. 1 shows the global and regional FSI indexes.

Because oil is a global market, we use the OFR FSI as its coverage is global as well, besides being available at the daily frequency for a longer period of time, unlike other alternatives, which are either short in data coverage (for example, the Bank of America Merrill Lynch global financial stress index) or restricted only to the US (for example, the financial stress index from the Federal Reserve Bank at Cleveland). The data sample in our study covers the period from 4th January, 2000 until 26th May, 2017, with the start-date defined by the FSIs and the end-date corresponding to the availability of intraday data on oil.[7]

4. Estimation results

4.1. Baseline results

Table 1 reports full-sample estimation results.[8] Results show that the estimated coefficients of the baseline HAR-RV model are significant at all three forecasting horizons. The fit of the estimated models, as captured by the adjusted R^2 statistic, decreases from approximately 0.77 for the short forecasting horizon to about 0.47 for the long forecasting horizon. When we add the various financial-stress indexes, the fit of the estimated models increases, where the adjusted R^2 statistic attains a maximum when we consider the US measure of financial stress. The estimated coefficients of the financial-stress indexes are all highly significant and positive. Hence, as one would have expected, realized oil-price volatility increases in times of financial stress at all three forecasting horizons.

Given that Table 1 shows that the baseline HAR-RV model works remarkably well in sample, it is interesting to test whether adding the financial-stress indexes to the model helps to improve out-of-sample forecast accuracy and, if so, which type of investor benefits from the financial-stress-based forecasts. To answer these questions, we report in Fig. 2, Fig. 3 results of the Diebold-Mariano test.[9] The figures show the p-values (shaded areas) of the test as a function of the asymmetry parameter of the forecaster loss function on the vertical axis and the length of the rolling-estimation window on the horizontal axis. We report results for rolling-estimation windows ranging from 500, 600,…to 3000 observations. Fig. 2 shows the results for a forecaster loss function of the lin–lin type and Fig. 3 shows the results for a quad-quad loss function.[10]

Results shows that using the overall OFR financial-stress index improves the accuracy of forecasts relative to the baseline HAR-RV model mainly for asymmetry parameters less than approximately 0.4–0.5. Hence, a forecaster who incurs a larger loss from overprediction of realized volatility as compared to an underprediction of the same (absolute) magnitude benefits from adding the OFR financial-stress index to the baseline HAR-RV model. The test results, however, for asymmetry parameters smaller than 0.4–0.5 depend on the length of the rolling-estimation window. For example, the test results become insignificant even in the case of a short forecast horizon, for which the area of insignificance of the test results is larger than for the medium and long forecast horizon, if the length of the rolling-estimation windows takes on values in the neighborhood of 1500 observations. From the viewpoint of an investor, the dependence of the test results on the length of the rolling-estimation window raises the rather important question which length is optimal and should be used for forecasting realized volatility. In order to provide some guidance in this regard, we use stars to indicate which length of the rolling-estimation window minimizes the ratio of the sum of losses of the baseline model divided by the sum of losses of the model extended to include a measure of financial stress, given the asymmetry parameter that characterizes an investor's loss function (that is, her preferences). For example, the stars in case of the OFR financial-stress index show that, in case of the short forecasting horizon, an investor with asymmetry parameters smaller than approximately 0.4 should use a rolling-estimation window that comprises 3000 observations if the aim is to minimize the relative loss of a model that features the OFR financial-stress index. An investor whose preferences can be characterized by a larger asymmetry parameter, in contrast, fares well when she uses a shorter rolling-estimation window that comprises about 1500 observations.

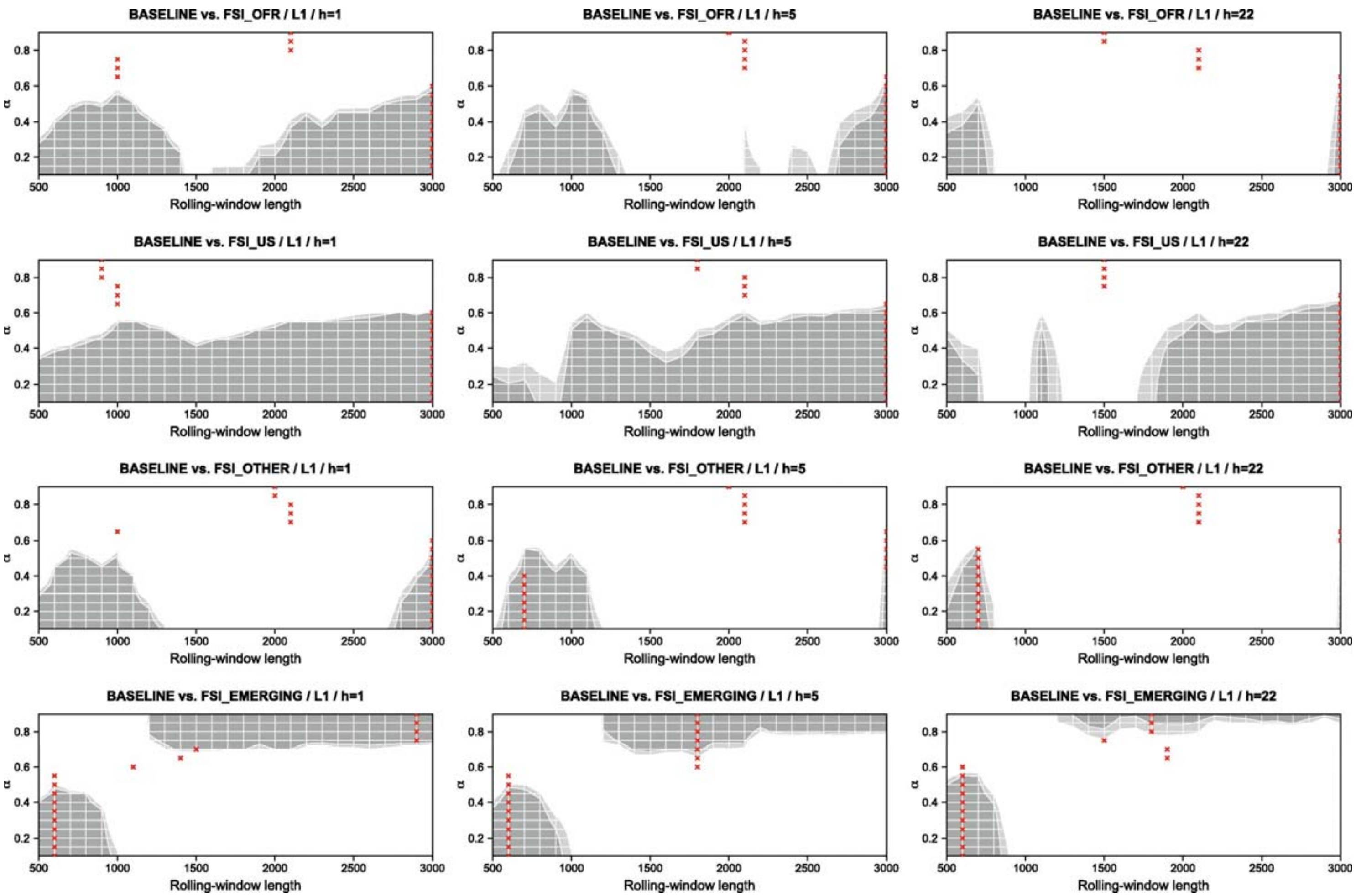


Fig. 2. Out-of-Sample Results (L1 Loss Function). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function.

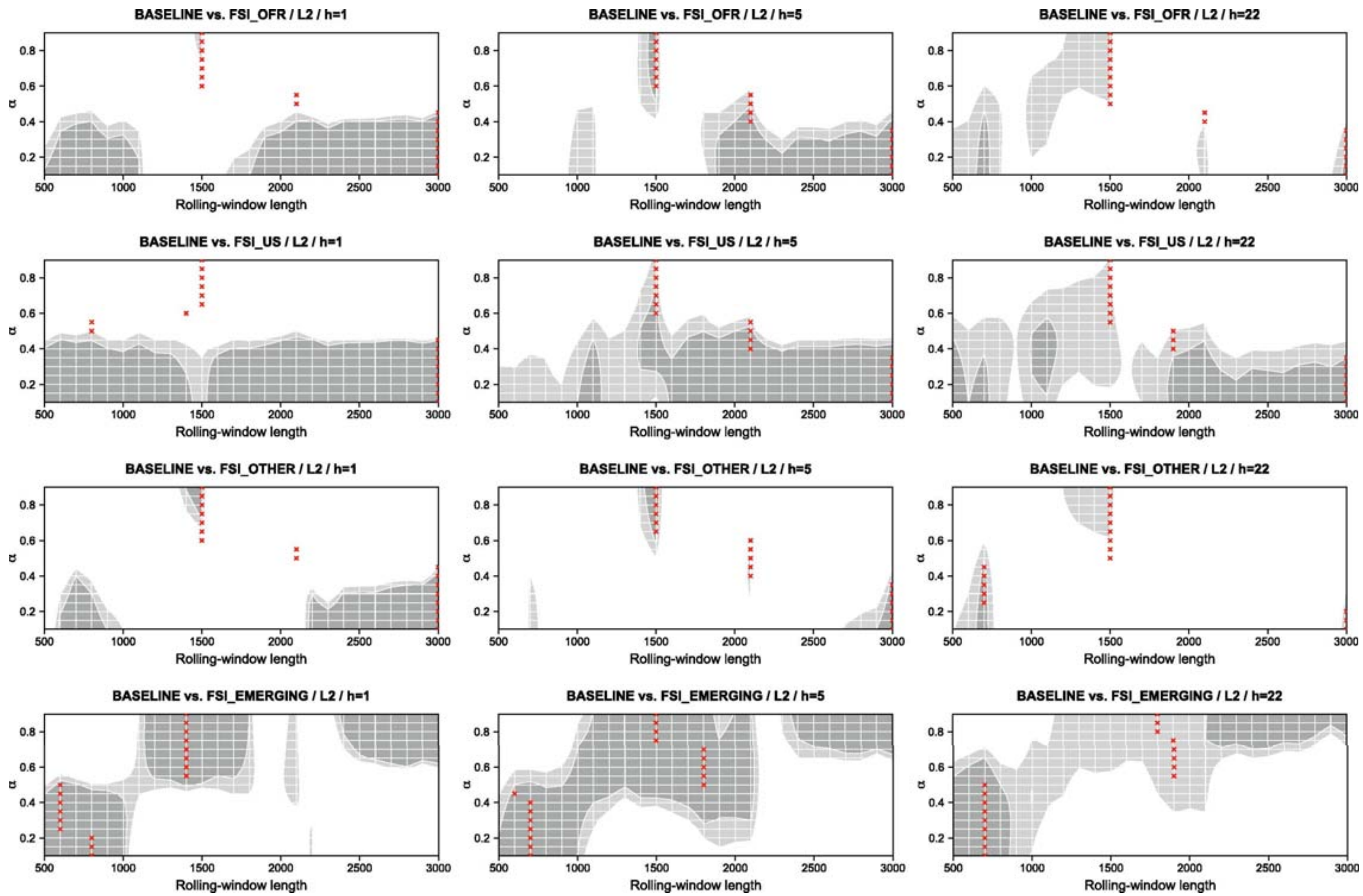


Fig. 3. Out-of-Sample Results (L2 Loss Function). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L2: Quad-quad loss function.

We further observe for the OFR financial-stress that, irrespective of the assumed type of the forecaster loss function, the test results get weaker (that is, the shaded areas tend to become smaller) when we consider the medium and long forecasting horizons. In other words, the range of combinations of the asymmetry parameter and the lengths of the rolling-estimation windows for which the Diebold-Mariano test yields significant test results becomes fragmented for the medium forecasting horizon and rather small for the long forecasting horizon. We also observe that the stars indicate, similar to what we observe for the short forecasting horizon, that an investor with an asymmetry parameter smaller than approximately 0.4–0.5 benefit from using the OFR financial-stress to forecast realized volatility provided the length of the rolling-estimation window is fixed at the upper boundary of lengths that we consider in our research. In contrast, an investor with an asymmetry parameter larger than roughly 0.5 also benefits from using the OFR financial-stress index provided the rolling-estimation window is of an intermediate length (about 1500 observations) and the loss function is of the quad-quad type.

Turning next to the test results for the US-based measure of financial stress, we find that the significant test results that we observe for the OFR financial-stress index in case of an asymmetry parameter less than approximately 0.4–0.5 are driven by the US-based index of financial stress. In this range of the asymmetry parameter, the test results are significant for all rolling-estimation windows for the short and medium forecasting horizons for both the lin–lin and the quad-quad forecaster loss function. We also observe significant test results for the long forecasting horizons, though the results get weaker and become insignificant (mainly lin–lin loss function) for intermediate lengths of the rolling-estimation window (roughly 700–1700 observations). If we focus on the preferred length of the rolling-estimation window as indicated by the stars, then we observe significant test results for the medium and long forecasting horizon for a lin–lin loss function and an asymmetry parameter of smaller than about 0.6 if the rolling-estimation window is long. For a quad-quad loss function, the preferred model also yields weakly significant results for larger asymmetry parameters and intermediate rolling-estimation window of about 1500 observations.

In contrast to the US results, the test results for the other advanced economies are significant mainly when we study some short and some rather long rolling-estimation windows. While the stars indicate significant test results for several of the preferred models, the overall impression is that financial stress that originates in other advanced economies contributes less to explain realized volatility than financial stress that reflects US financial-markets conditions. On balance, the test results for the forecasts derived from models that include financial stress fostered by developments in other other advanced economies are also weaker than the test results that we observe when we include emerging-markets financial stress in the HAR-RV models. Emerging-markets financial stress helps to improve forecast accuracy when we set the asymmetry parameter to approximately $\alpha < 0.5$ and the rolling-estimation window is not longer than roughly 1000 observations, where the stars indicate that an investor with such preferences benefits from using an even shorter rolling-estimation window. Interestingly, the stars show that a forecaster who incurs a much larger loss from an underprediction of realized volatility as compared to an overprediction of the same size (that is, when the loss function can be characterized by an asymmetry parameter of roughly $\alpha > 0.7$) also benefits from considering emerging-markets financial stress to predict realized volatility, but for such an investor it is advisable to use a somewhat longer rolling-estimation window. This result obtains especially in case of a lin–lin loss function. For a quad-quad loss function, the results are less clear-cut and there are also several significant test results for a

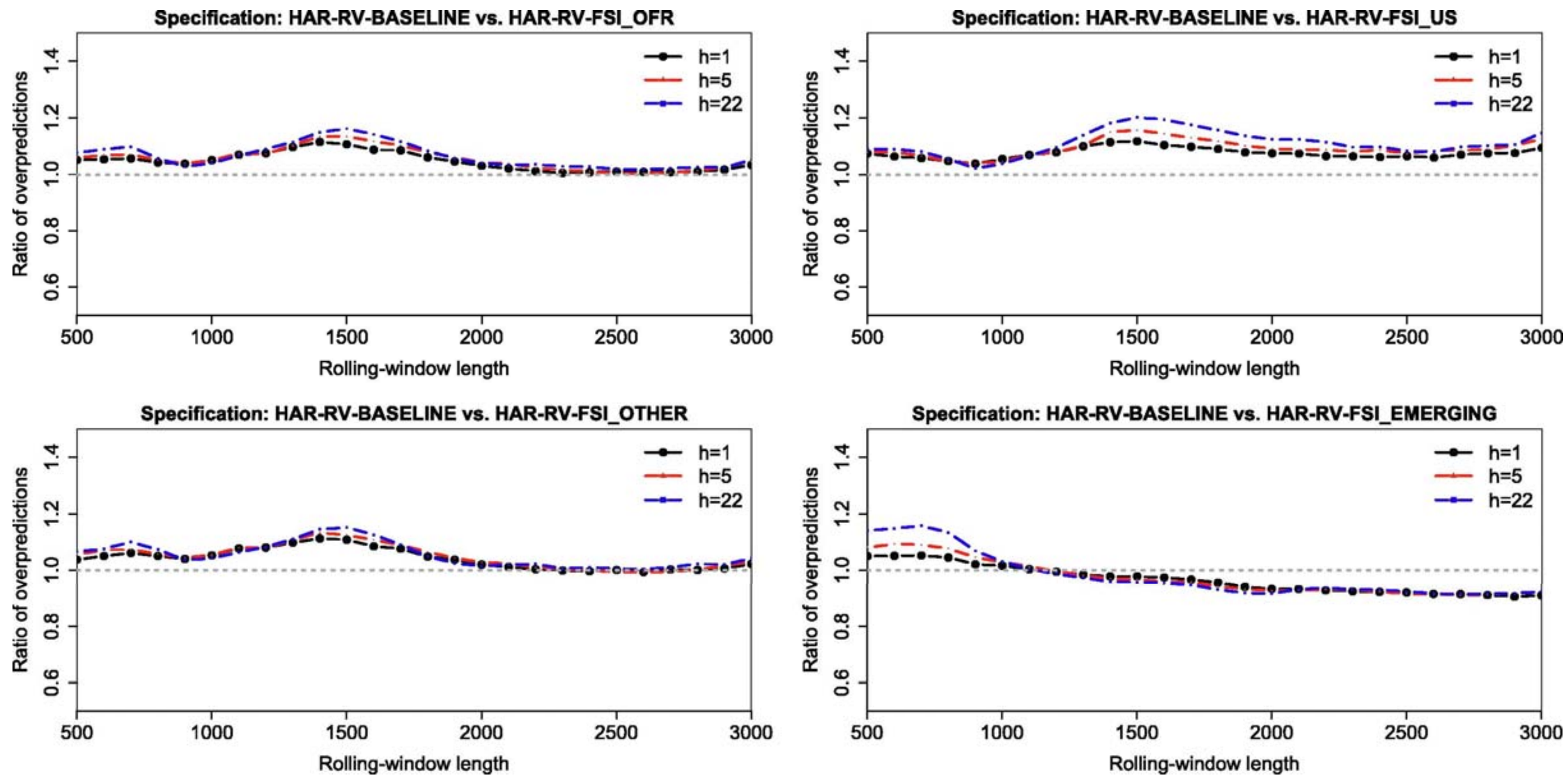


Fig. 4. Ratio of Overpredictions. Note: The ratio of overpredictions is defined as the sum of overpredictions implied by the baseline model divided by the sum of overpredictions implied by the model that features a financial stress measure. Overpredictions occur when the forecast exceeds the actual realized volatility. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window.

forecaster who incurs a larger loss from an over- rather than an underprediction of realized volatility, particularly for rolling-estimation windows of intermediate length.

Fig. 4 sheds light on the intuition behind our results. The figure shows the number of overpredictions of realized volatility implied by the baseline HAR-RV model divided by the number of overpredictions implied by the HAR-RV model extended to include financial stress. The resulting ratio takes on mainly positive values when we consider the OFR financial-stress index and the measure of financial stress originating in other advanced economies, and especially when we consider the measure that captures financial stress originating in the United States. Hence, the baseline HAR-RV model overpredicts actual realized volatility more often than the models that feature financial stress as a predictor and, as a result, forecasts computed by means of the baseline model perform poorly for those forecaster loss functions that attach a larger weight to overpredictions than to underpredictions. In contrast, the ratio takes on negative (positive) values for rolling-estimation windows comprising more (less) than approximately 1000 observations when we consider financial stress originating in emerging-market economies. In consequence, forecasts that we compute by means of the baseline HAR-RV model outperform the financial-stress-based forecasts when the shape of the forecaster loss function implies that underprediction (overprediction) is relatively more costly than a corresponding overprediction (underprediction) when the rolling-estimation window is comparatively long (short).

4.2. Extensions

Given the result that US financial stress tends to improve forecast performance for $\alpha < 0.5$, while financial stress originating in emerging-market economies also tends to be useful when $\alpha < 0.5$, it is natural to consider an extended HAR-RV model that features combinations of the financial-stress index in the vector of predictors. Fig. 5 reports results for two such models. The first model features US and emerging-markets financial stress but neglects financial stress originating in other advanced economies, while the second model features all three indexes of regional financial stress.

Results show that, when compared with the results for a model that features only the US financial stress, the area of significant Diebold-Mariano tests widens for a quad-quad loss function and for a medium and a long forecasting horizon. For example, for the model that features all three indexes of regional financial stress the range of asymmetry parameters for which the forecasts extracted from the extended model are more accurate than the forecasts extracted from the baseline HAR-RV model comprises (almost) all asymmetry parameters for a quad-quad loss function and a long forecasting horizon, though (as indicated by the stars) the preferred forecasting model differs across investor types. Similarly, for a medium forecasting horizon the range of asymmetry parameters that result in significant test results covers a broad range of rolling-estimation windows. For a short forecasting horizon, the widening of the areas of significance of the test results are less clear-cut relative to the US-only case, which suggests that US financial stress is a main driver of the results in these cases.

Next, we present results for bad and good realized volatility in Fig. 6, Fig. 7, where we focus again on models that combine the US-based measure of financial stress with other regional sources of financial stress. Results for the short forecasting horizon and realized bad volatility show that adding financial stress indexes to the HAR-RV model significantly improves

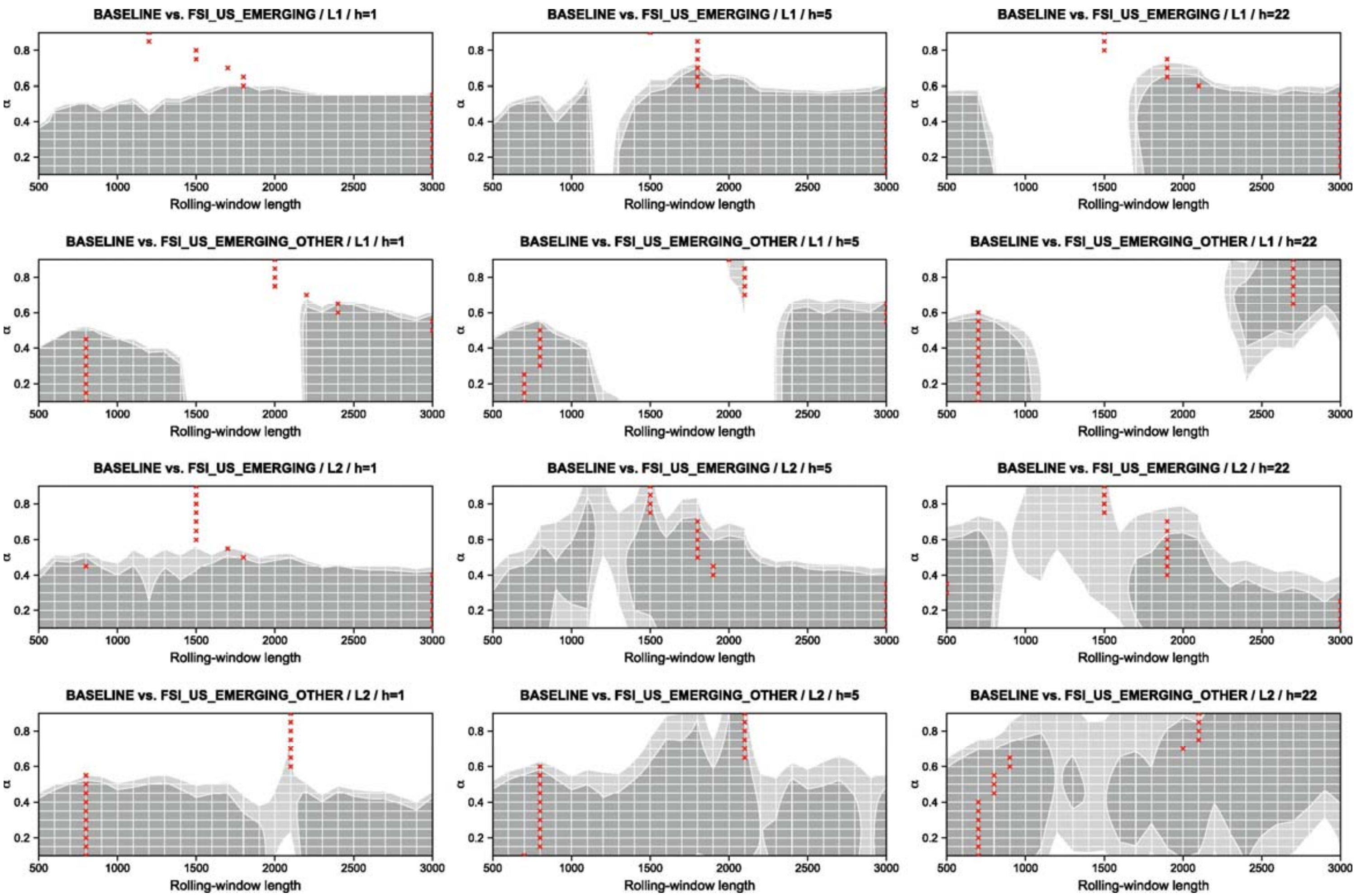


Fig. 5. Out-of-Sample Results (Models Feature Regional FIS Measures). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function. L2: Quad-quad loss function.

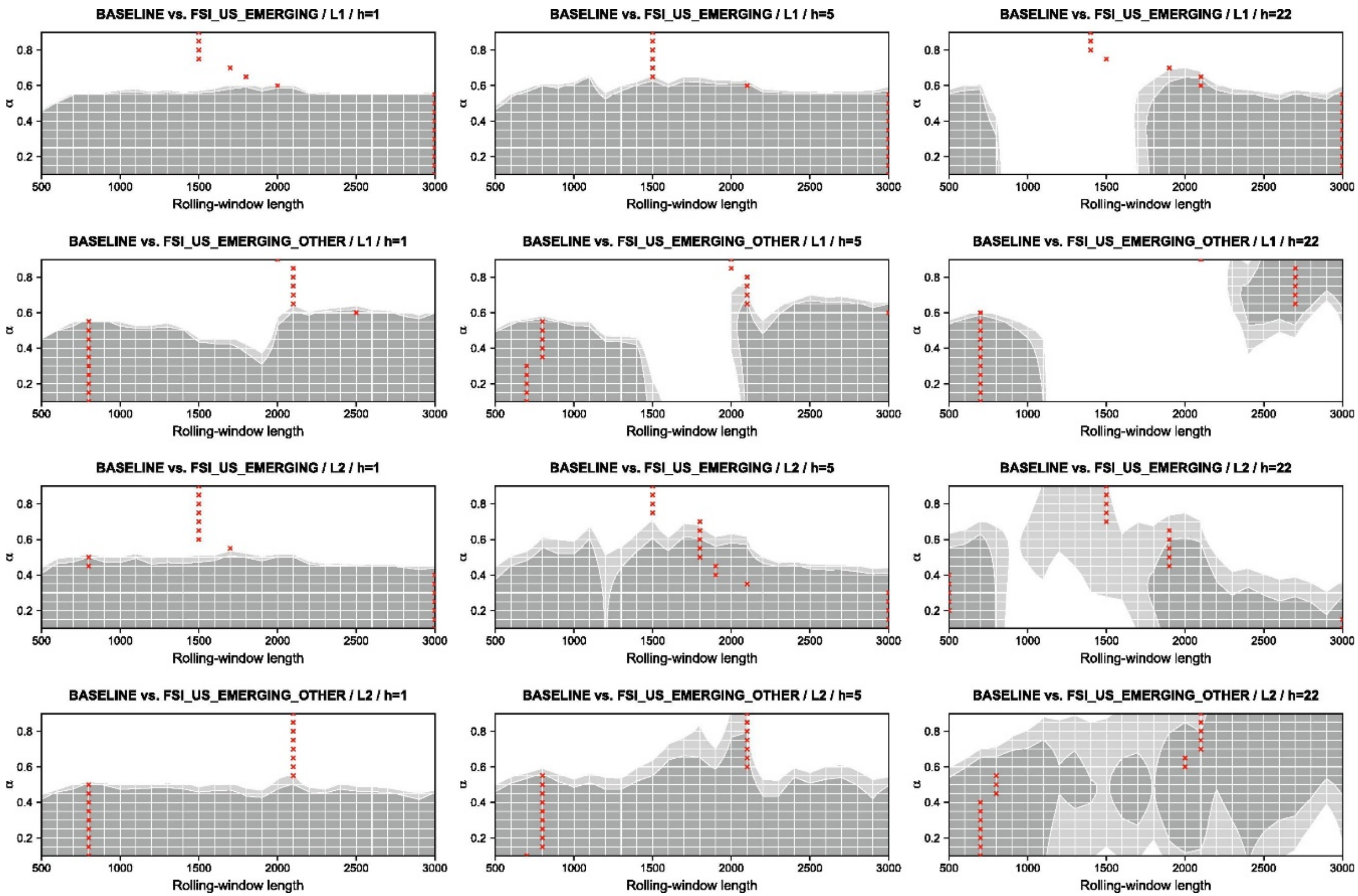


Fig. 6. Out-of-Sample Results (Realizd Bad Volatility). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function. L2: Quad-quad loss function.

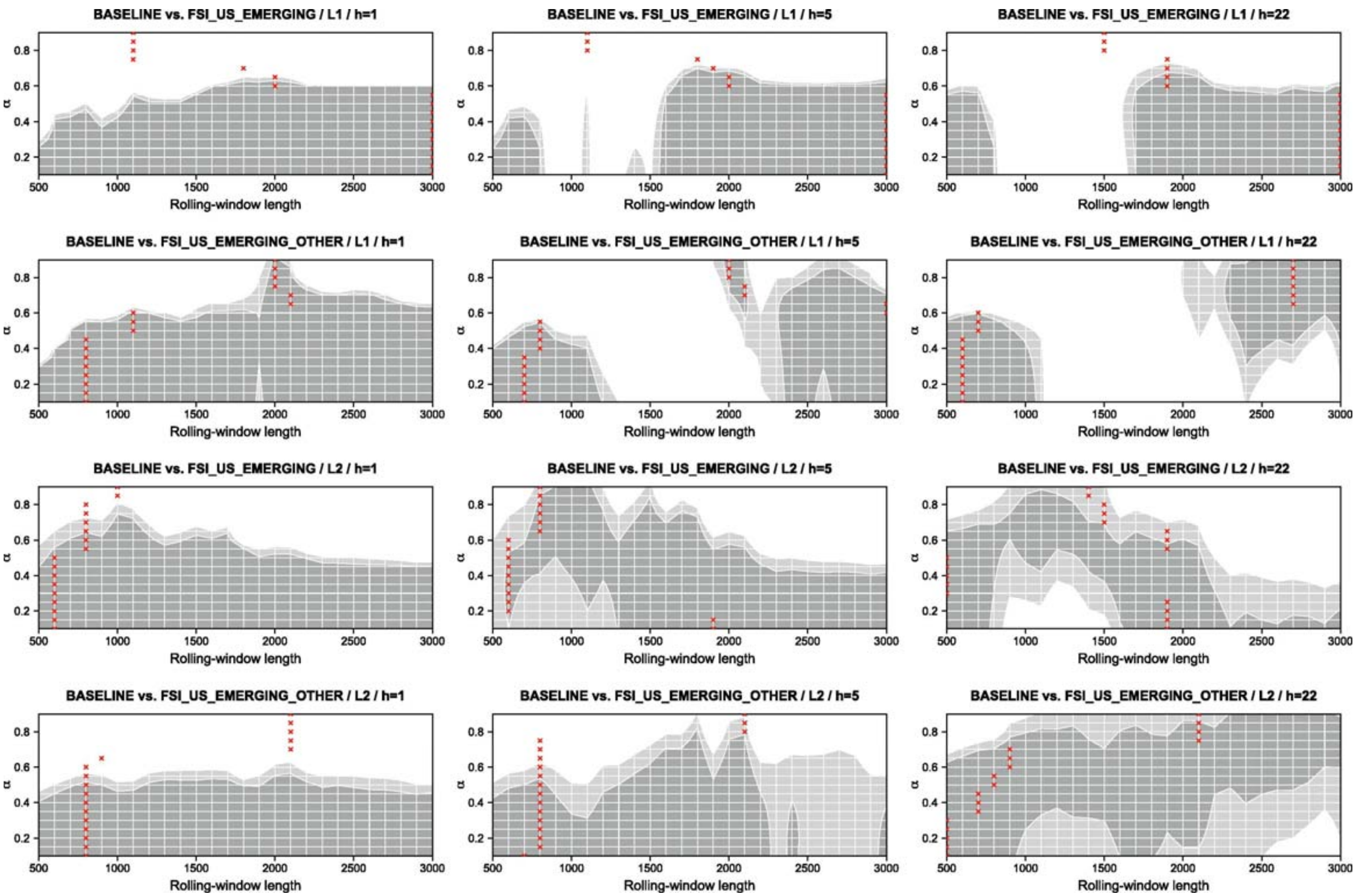


Fig. 7. Out-of-Sample Results (Realized Good Volatility). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function. L2: Quad-quad loss function.

forecasting performance mainly for an investor whose loss function implies that the costs of an overprediction are larger than the costs of an underprediction of comparable magnitude. The range of significant test results is somewhat wider in terms of the asymmetry parameter for realized good than for realized bad volatility, that is, an investor who attaches a higher cost to an underprediction than to a similarly seized overprediction can benefit from using financial stress for forecasting realized good volatility.

For the medium forecasting horizon, test results are again significant in case of realized bad volatility when the loss function can be characterized by an asymmetry parameter of roughly $\alpha < 0.6$. For some asymmetry parameter, the range of asymmetry parameters, however, is wider, and in case of the model that features all three indexes of financial stress results for rolling-estimation windows of length from about 1500 to 2000 observations are entirely insignificant for a lin–lin loss function. Insignificant results in this range, but also for the model that features only US and emerging-market-based financial stress, also obtain in case of realized good volatility and a medium forecasting horizon when we consider a lin–lin loss function. In contrast, the area of significant test result covers almost all asymmetry parameters for several rolling-estimation windows when we study a quad-quad loss function, especially in case of the US-emerging-market-based financial-stress model. Finally, for the long forecasting horizon, the area of significant test results becomes more fragmented, except when we evaluate the model that features all three indexes of financial stress under a quad-quad loss function, where the stars indicate that an investor who uses our simple metric of relative model performance to choose the length of the rolling-estimation window benefits even in these cases from using financial stress to forecast realized volatility.

As a further extension, we model and forecast the square root of RV, that is, the realized standard deviation of oil-price returns. Such an extension may be relevant, for example, for the pricing of options when traders plug forecasts of the realized standard deviation directly into the Black-Scholes-pricing formula. Fig. 8 gives the results, where we focus on models that combine the US index of financial stress with the other indexes of regional financial stress. One noticeable difference as compared to the results that we report in Fig. 5 is that the square-root forecasts give stronger evidence of insignificant test results for rolling-estimation windows of intermediate length. Another difference is that in some cases the significance of the results strengthens for asymmetry parameters exceeding its symmetric benchmark value. Despite these differences, however, the main results already documented in 5 carry over to Fig. 8 gives. The stars indicate that it is in general preferable (in terms of the ratio of cumulated losses) to opt for a long rolling-estimation window in case of the short forecasting horizon when the asymmetry parameter is smaller than its symmetric benchmark (except under a quad-quad loss function and the model that features all three measures of financial stress), and that the choice of the optimal rolling-estimation windows becomes less clear-cut in case of the medium and long forecasting horizons. It is important to note, however, that even for the longer forecasting horizons an investor whose preferences can be characterized by an asymmetry parameters larger than 0.5 can achieve significant gains in terms of forecast accuracy.

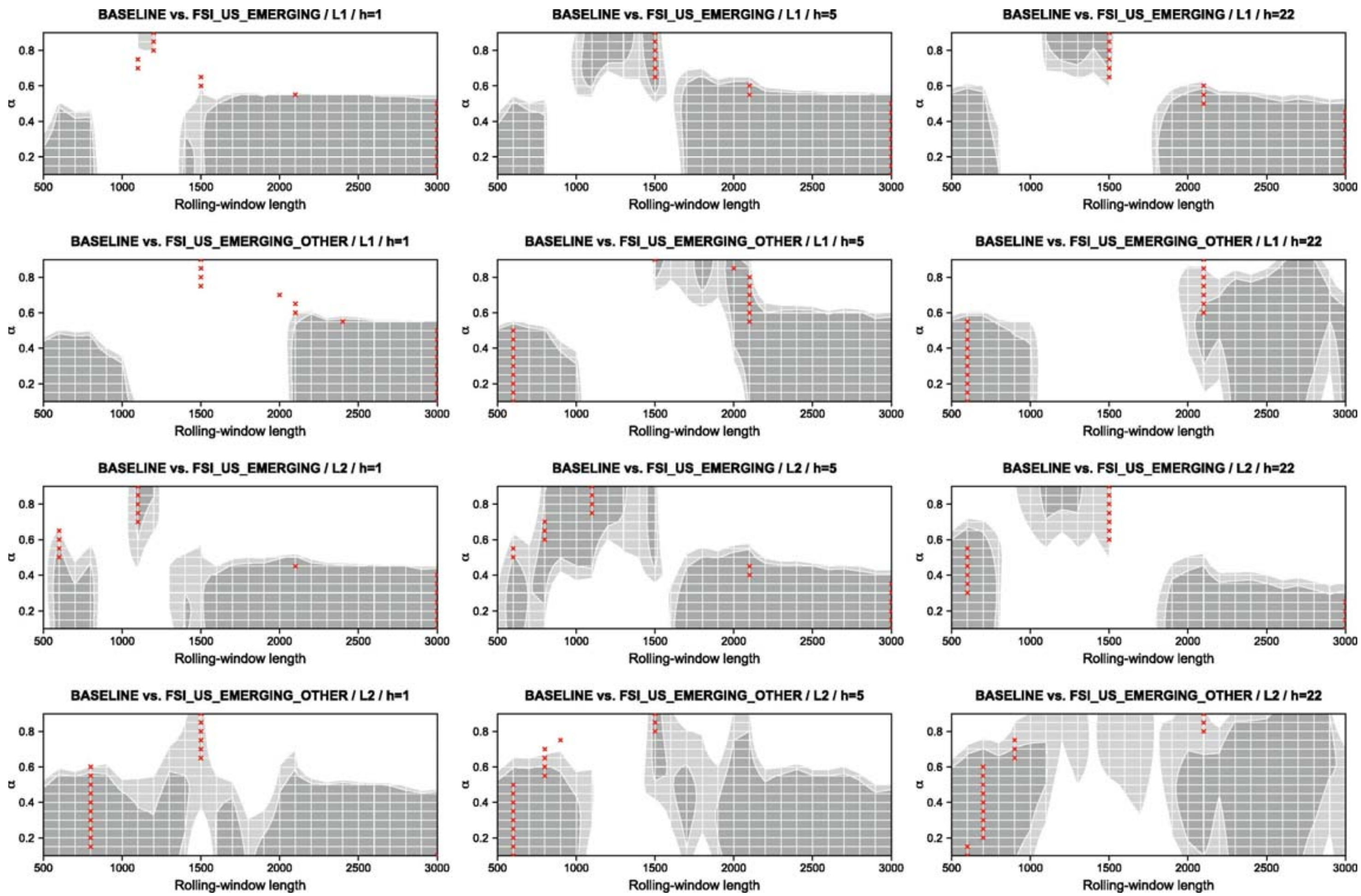


Fig. 8. Out-of-Sample Results ($RV^{0.5}$, Models Feature Regional FIS Measures). Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include financial stress is more accurate. Results are based on rolling-window estimates. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function. L2: Quad-quad loss function.

4.3. Statistical analysis of forecasts: supplementary results

Against the background that various extensions of the HAR-RV model have been studied in earlier research, we also consider HAR-RV models that include higher-order moments (realized skewness and realized kurtosis, see, e.g., Mei et al., 2017).[11] We also consider volatility jumps as in Sévi (2014).[12] While details of the results for these extended models differ from the baseline results, including higher-order moments or jumps does not affect our main conclusions. This means that there is no benefit by adding more complexity related to volatility decomposition to the forecasting models we study in this research, which in turn justifies the use of *MRV* as an estimator of the integrated variance (the main estimator used in this research). Such evidence has been also found and in other studies (see, e.g. Sévi, 2014, Degiannakis et al., 2019, among others). For the sake of brevity, we report these results and all other results that we summarize in this section as Supplementary Material.

The focus of this research is on the implications of the OFR financial-stress index and its regional components for forecasting realized oil-price volatility. As an alternative, we also consider financial stress decomposed along asset-class categories (credit, equity valuation, funding, and safe assets).[13] Results show that financial stress in funding and equity valuation are more important at short and medium forecasting horizons than financial stress in credit markets and especially safe assets, where the test tend to yield more often significant results when the asymmetry parameter is smaller than its symmetric benchmark.

Next, we estimate the HAR-RV model on the natural logarithm of the realized standard deviation of oil-price returns. Some authors (see, e.g., Andersen et al., 2007) have considered variants of the HAR-RV model that feature the logarithm of realized volatility/realized standard deviation. After having estimated the model in logs, we compute the anti-log of the forecasts derived from the estimated model and then add a Jensen-Ito term, as in Degiannakis and Filis (2017). Results for the short and the medium forecasting horizons indicate that, irrespective of the shape and the functional form (lin–lin vs. quad-quad) of the loss function, forecasts computed using US-based financial stress are more accurate than the forecasts from the baseline model when we consider a rolling-estimation window that comprises (not) more than about 1800 observations, and an investor who incurs a higher loss in case of an overprediction (underprediction) than in case of a corresponding underprediction (overprediction). For the long forecasting horizon, the region of significant test results for the US-based financial stress mainly comprises relatively short rolling-estimation windows and an asymmetry parameter of about $\alpha < 0.6$ (but we also observe some positive test results for some larger asymmetry parameters). For financial stress originating in emerging-market economies, the results are roughly a mirror-image of the US results. When we consider a short rolling-estimation window, then using the forecasts from the extended model yields more accurate forecasts for an investor who suffers more from overpredictions than from overpredictions, while the situation is reversed when the rolling-estimation window lengthens. Hence, for financial stress originating in emerging-market economies, results are similar to the results we observe for the baseline scenario (Fig. 2, Fig. 3).

As in Degiannakis and Filis (2017), we consider two additional forecasting horizons (44-days-ahead and 66-days-ahead). Results confirm that financial stress originating in the US mainly benefit a forecaster who incurs a larger loss in case of an overprediction rather than an underprediction of the same (absolute) magnitude. Similarly, for relatively short rolling-estimation windows such a forecaster also benefits from monitoring financial stress originating in emerging markets economies. In addition, a forecaster who incurs comparatively higher costs in case of an underprediction also benefits in a few cases. On balance, though, the results for emerging-market economies are weaker than in the baseline scenario as far as the underprediction scenario is concerned.

We also consider HAR-RV models that features also weekly and monthly averages of financial stress (in analogy to the MRV_w and MRV_m predictors). By and large, the results corroborate that investors whose preferences imply that they suffer more from an underprediction than a comparable overprediction benefit from focusing on US-based financial stress. Investors that suffer more from and overprediction (underprediction) than an underprediction (overprediction) should use a comparatively long (long) rolling-estimation window as far as financial stress originating in emerging market economies is concerned. Results are again provided as Supplementary Material.

Finally, we take into account that our identification of an optimal length of a rolling-estimation window is based on an ex-post analysis of our findings. As a further sensitivity check, we, therefore, consider as an alternative specification a recursively expanding estimation window. To this end, we use the first 500 data to train the model, and then successively expand the estimation window until we reach the end of the sample period. Our results do remain qualitatively similar to the results reported in Fig. 2, Fig. 3. More specifically, we find significant results for an asymmetry parameter smaller than roughly $\alpha < 0.5$ for the OFR model. Comparing this finding with the results for the models featuring financial stress originating in the U.S. and other advanced countries, we find that the U.S. is the main driver of the significant results for $\alpha < 0.5$. Finally, test results which are significant for the recursive window financial stress, they are significant for $\alpha > 0.8$ (L1 loss and $h = 1$) and $\alpha > 0.6$ (L2 loss and $h = 1, 5$). Hence, it is important to differentiate between regional sources of financial stress where types of investors - who differ with regard to their relative loss preferences - suffer from overprediction versus underprediction of realized volatility benefit by monitoring different regional sources of financial stress.

4.4. Assessing the benefits of forecasts

We next analyze the benefits of using the various FSI measures to forecast realized volatility. Our preferred approach to studying the benefits of using financial stress for forecasting realized volatility is based on a relative-loss-criterion (Pierdzioch et al., 2014, Pierdzioch et al., 2016b). The relative-loss-criterion builds on forecaster preferences as specified in Eq. (6), and it nests the out-of-sample R^2 criterion studied by Campbell and Thompson (2008) for the special case of a symmetric and quadratic loss function. The relative-loss criterion is defined, given a specified length of the rolling-estimation window, using the loss function given in Eq. (6) as follows:

$$\mathscr{R}(b, f, \alpha, h) = 1 - \sum_t L(FE^f_{t+h}, \alpha) / \sum_t L(FE^b_{t+h}, \alpha), \tag{7}$$

where b, f denote the benchmark and the FSI model, and the summation runs over all periods for which out-of-sample forecasts are available. The relative-loss criterion is positive when the cumulated loss a forecaster realizes from using the FSI model for forecasting realized volatility is smaller than the cumulated loss from using the benchmark model. We use the baseline HAR-RV model as our benchmark model. Fig. 9, Fig. 10 plot results for the relative-loss criterion for the L1 and the L2 loss functions for the baseline scenario (see also Fig. 2, Fig. 3).[14] The results show that for financial stress originating in the US the best models yield

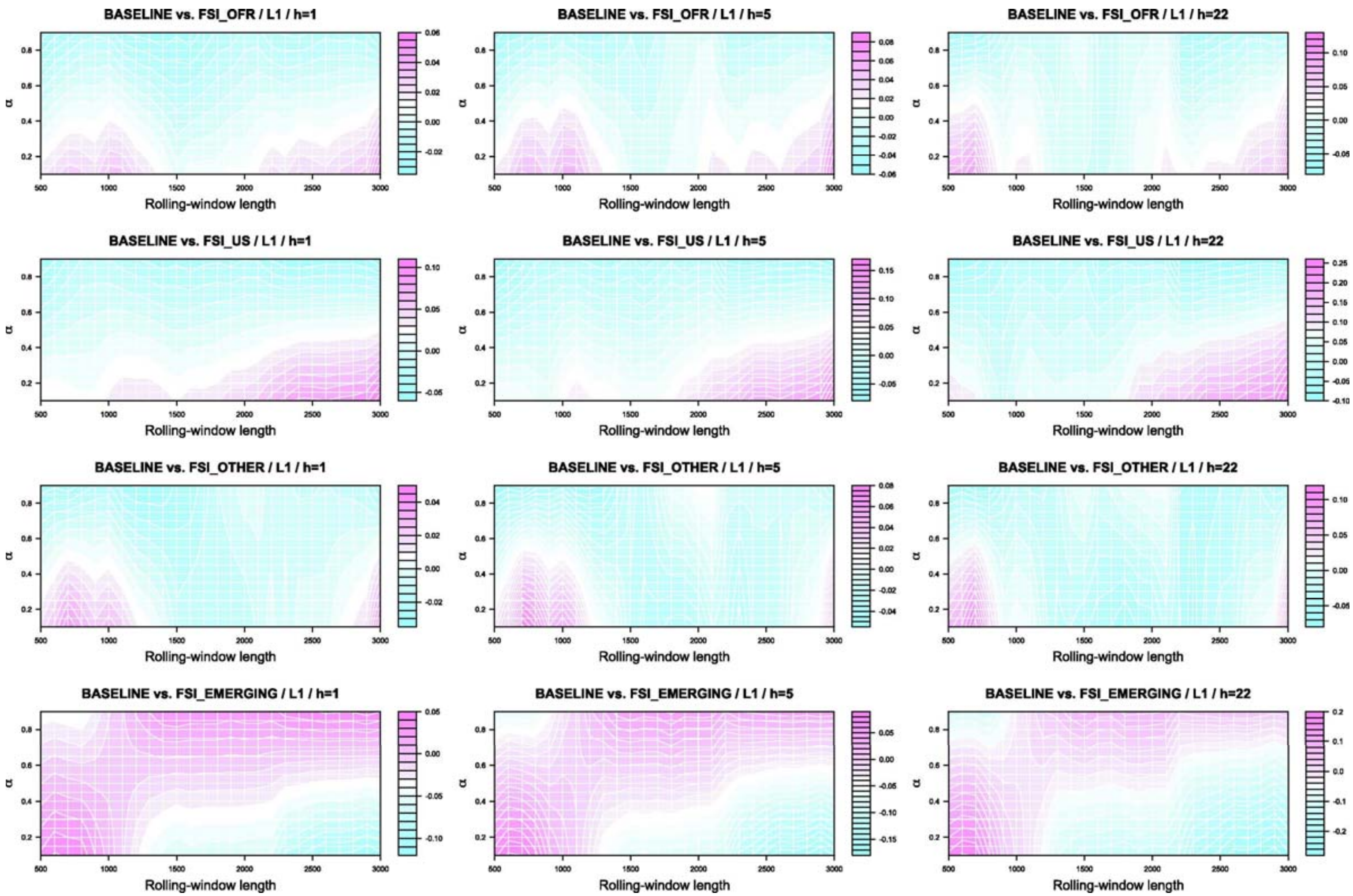


Fig. 9. Out-of-Sample Results (Relative-Loss Criterion, $\mathscr{R}$, L1 Loss Function). Note: Results for the out-of-sample relative-loss criterion, $\mathscr{R}$. The benchmark model is the model that does not feature financial stress. A positive value shows that the model that features FSI yields a lower loss for a given the shape of the loss function. L1: Lin–lin loss function.

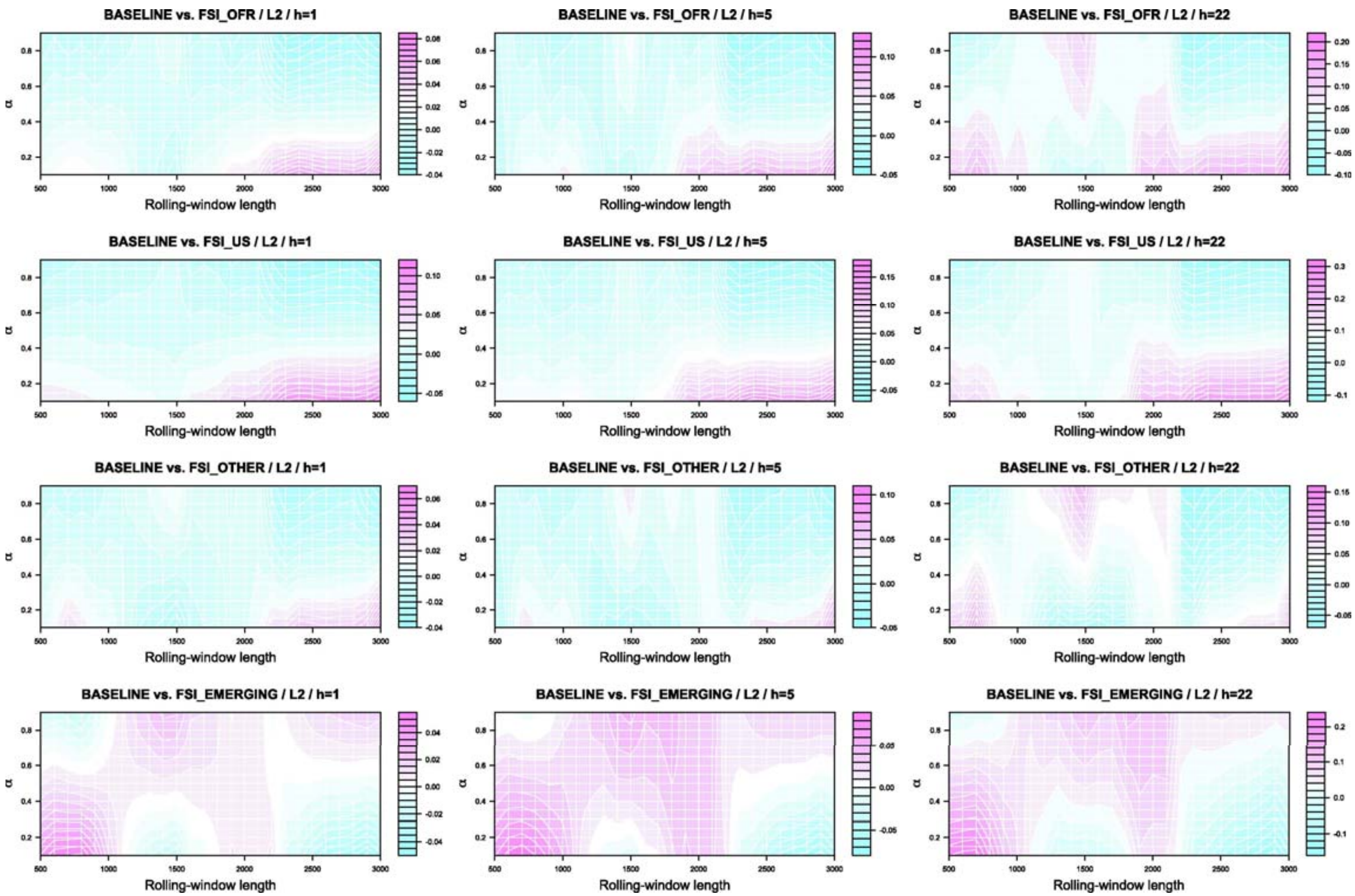


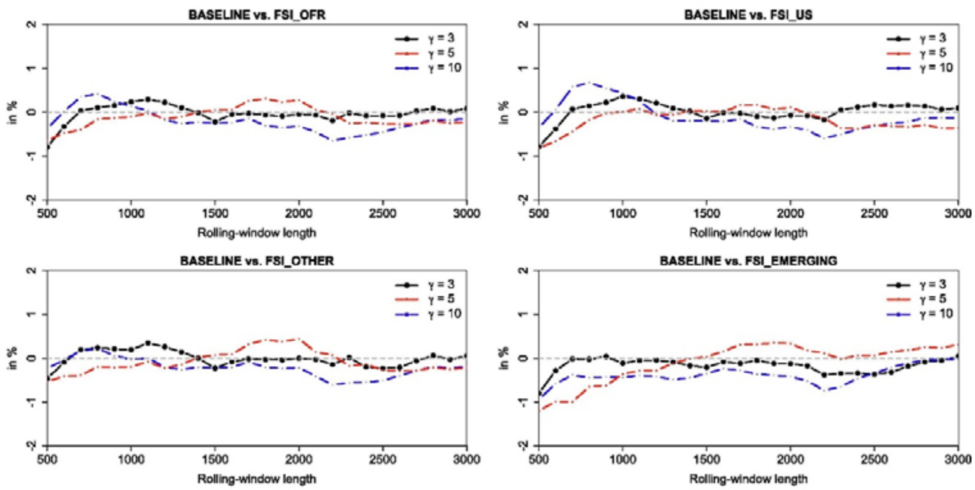
Fig. 10. Out-of-Sample Results (Relative-Loss Criterion, $\mathscr{R}$, L2 Loss Function). Note: Results for the out-of-sample relative-loss criterion, $\mathscr{R}$. The benchmark model is the model that does not feature financial stress. A positive value shows that the model that features FSI yields a lower loss for a given the shape of the loss function. L2: Quad-quad loss function.

a relative-loss criterion between 0.10 and 0.25 for a for the L1 loss function (up to roughly 0.3 for a L2 loss function and a long forecast horizon) when the asymmetry parameter of the loss function is smaller than its symmetric benchmark value and the rolling-estimation window is relatively long, depending on the forecast horizon being studied.[15] For financial stress originating in emerging market economies, the relative loss criterion takes on values in the range between 0.05 and 0.2 for the L1 loss function, and in the range of approximately 0.05 to 0.22 for the L2 loss function (again depending on the forecast horizon being studied). While the maximum benefit of using financial stress to forecast realized volatility is somewhat smaller in case of the emerging market economies than in case of the US, the range of combinations of the asymmetry parameter and rolling-estimation window for which the relative-loss criterion indicates a superior performance of the model featuring financial stress is larger for emerging markets than for the other sources of financial stress (United States, other advanced economies). At the end of the paper (Technical Appendix), we plot simulation results that show that the relative-loss criterion often indicates a significant benefit of using financial stress for forecasting realized volatility.[16]

Another common approach to study the benefits of forecasts is to consider a forecaster who uses forecasts to invest in a portfolio consisting of a riskless asset (assumed here, for simplicity, to yield zero interest income) and in oil futures. Like Cenesizoglu and Timmermann (2012), we abstract from transaction costs and assume away intertemporal hedging considerations (and we focus on the short forecast horizon). The forecaster, in other words, rebalances his or her portfolio in every period of time. We assume that a forecaster has preferences that can be described in terms of a power utility function (constant relative risk aversion) and three levels of risk aversion (risk-aversion parameter, $\gamma = 3, 5, 10$). We further assume that a forecaster uses the returns he or she observes in the period in which a forecast is to be formed to forecast returns.[17]

Results (Fig. 11) show that adding financial stress to the vector of predictors leads in several cases to relatively higher certainty equivalent returns (Panel A) for the lower and high risk-aversion parameter when the rolling-estimation window is relatively short. A forecaster with a medium risk-aversion parameter benefits from tracing financial stress mainly for intermediate rolling-estimation windows, where tracing financial stress originating in emerging markets also leads to a better portfolio performance for relatively long rolling-estimation windows. A forecaster who measures portfolio performance in terms of Sharpe's ratio, in contrast, often benefits from taking emerging-markets financial stress into consideration as a predictor of realized volatility when the rolling-estimation window is not too short, especially for a medium and high risk-aversion parameter. Financial stress originating in the US, in turn, reduces portfolio performance as measured in terms of Sharpe's ratio, while financial stress originating in other advanced economies improves portfolio performance for a medium risk-aversion parameter for some intermediate rolling-estimation windows.

Panel A: Certainty Equivalent Returns



Panel B: Sharpe's Ratio

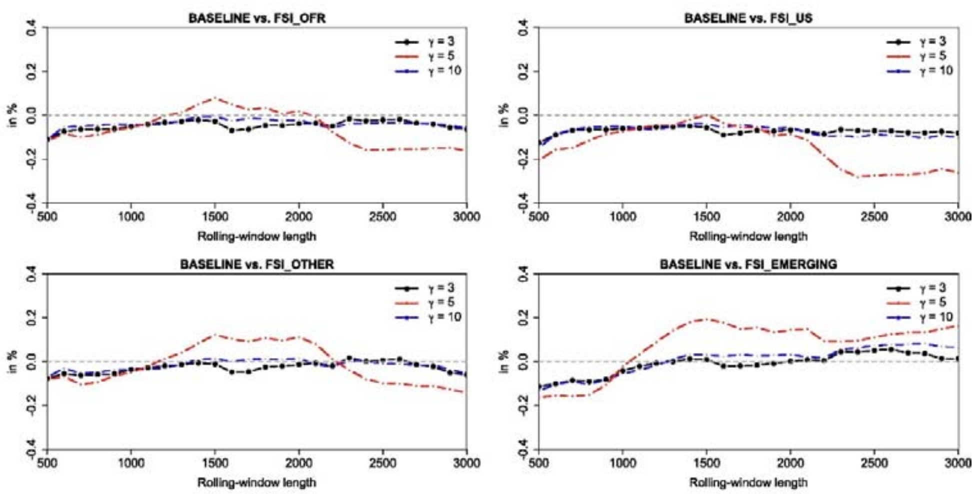


Fig. 11. Certainty Equivalent Returns and Sharpe's Ratio. Note: Results for certainty equivalents and Sharpe's ratios based on out-of-sample data and a constant relative risk aversion (CRRA) utility function with parameter γ. Portfolio weights are restricted to the interval $[0, 1.5]$. Results are expressed in percent (e.g., for certainty equivalent returns $CER = 100 \times (CER_{FIS} - CER_{BASELINE})/|CER_{BASELINE}|$) such that positive numbers indicate a superior performance of the model that features financial stress in the vector of predictors.

Taken together, the results are a reminder that the results of assessments of the benefits of using financial stress for forecasting realized volatility depend on the preference (or loss) function being used by a researcher. As we have already indicated at the beginning of this section, in the context of our analysis the relative- loss-criterion is our preferred approach to analyzing the benefits of accounting for the impact of financial stress on forecasts of realized volatility. The relative loss-criterion accounts in a simple and straightforward way for the potential asymmetry of a forecaster loss function and, thereby, builds a bridge between widely studied forecast evaluation criteria (such as the out-of-sample R2 criterion) and more behavioral approaches that form the foundation of much of the recent research on asymmetric loss functions for forecast evaluation (see, e.g., Elliott et al., 2008).

5. Concluding remarks

Timely and accurate forecasting of oil-price returns volatility is essential for academics, policy makers, and oil traders alike. In this regard, a large number of studies has used daily oil-price returns to forecast the conditional oil-price returns volatility. However, recent empirical evidence suggests that the rich information contained in intraday data can produce more accurate estimates and forecasts of daily volatility. Given this, we have used several variants of the popular HAR-RV model to forecast the realized volatility of oil-price fluctuations based on 5-min intraday data covering the period of 4rd January, 2000 until 26th May, 2017. Our focus has been on the role of global and regional (US, other advanced economies, and emerging economies) financial stress, given that financial market turmoil can impact oil-market movements through its impact on both economic activity and investor behavior.

Our main finding is that extending the model to include indexes of financial stress helps to improve forecasting performance, where our results suggest that it is important to differentiate between regional sources of financial stress (United States, other advanced economies, emerging markets). Another key finding is that the shape of the forecaster loss function that one uses to evaluate forecasting performance plays an important role. Specifically, our findings indicate that, on the one hand, forecasters who attach a higher cost to overpredictions of realized volatility as compared to an underprediction of the same size should pay particular attention to financial stress originating in the US when using the HAR-RV model to forecast realized volatility. On the other hand, in case an underprediction is more costly than a comparable overprediction then forecasters should closely monitor financial stress caused by developments in emerging-market economies. We also have documented how these baseline results have to be qualified when we consider alternative specifications of the HAR-RV model. The key message to be taken home from the alternative specifications, however, is the same as for the baseline scenario: Financial stress does have predictive value for realized oil-price volatility, where different types of investors benefit from monitoring different regional sources of financial stress.

Acknowledgements

We thank the editor Menzie Chinn and two anonymous reviewers for helpful comments. The usual disclaimer applies. Pierdzioch thanks the German Science Foundation (Deutsche Forschungsgemeinschaft) for financial support (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

Appendix A. Technical appendix

In order to study the significance of the results for the relative-loss criterion, we run a simulation experiment. We use the simulation experiment to compare the following model: Baseline HAR-RV vs. HAR-RV-FSI-OFR, Baseline HAR-RV vs. HAR-RV-FSI-US, Baseline HAR-RV vs. HAR-RV-FSI-OTHER, Baseline HAR-RV vs. HAR-RV-FSI-EMERGING. The details of the simulation experiment are as follows:

Step 1:

We estimate by the ordinary-least-squares technique the baseline HAR-RV model and the HAR-RV-FSI models on the full sample of data. For the FSI models, we assume an AR(1) model. The HAR-RV model and the HAR-RV-FSI models are estimated independently from each other. The components of FSI do not enter the benchmark HAR-RV model, that is, the null hypothesis is that FSI does not have any predictive value for realized volatility. We store the coefficients and the residuals from these regressions. The residuals are stored in a matrix such that sampling can be done row-wise. This preserves the contemporaneous correlation of the residuals from the different models.

Step 2:

We sample row-wise from the matrix that contains the residuals. Samples for 1000 simulation runs are drawn. Sampling is done with replacement. Sampling is done such that, after taking into account that initialization of realized volatility and its weekly and monthly realizations consumes some data, 100 transitory data are available. The transitory data are used to initialize the baseline HAR-RV and the HAR-RV-FSI models, and are then deleted.

Step 3:

Based on the samples drawn in Step 2 and the matrix of residuals as well as the coefficient estimates computed in Step 1, we simulate artificial time series of realized volatility and FSI. Initial values (for the AR(1) models for FSI and for RV_w and RV_m) are set to zero. This assumption does not affect the results because we delete 100 transitory observations. We use the artificial realized volatility series are to create RV_w and MRV_m and the leads of RV. We then estimate the HAR-RV and HAR-RV-FSI models on the simulated data, where we let the length of the rolling-estimation window vary from 500, 600, to 3000 observations. We store forecasts and actuals.

Step 4:

We estimate the HAR-RV and HAR-RV-FSI models on the original data for rolling-estimation windows ranging from 500, 600, to 3000 observations.

Step 5:

We compute the relative-loss criterion for the models estimated on the original data and the simulated data. We compute p-values by checking how often the relative-loss criterion for the simulated data exceeds the relative-loss criterion for the original data (one-sided test). that is, we compute sum(simulated > original)/length(simulated). We compute the p-value for every length of the rolling-estimation window, every asymmetry parameter, and the three different forecast horizons (1, 5, 22 days ahead) (see Fig. A1, Fig. A2).

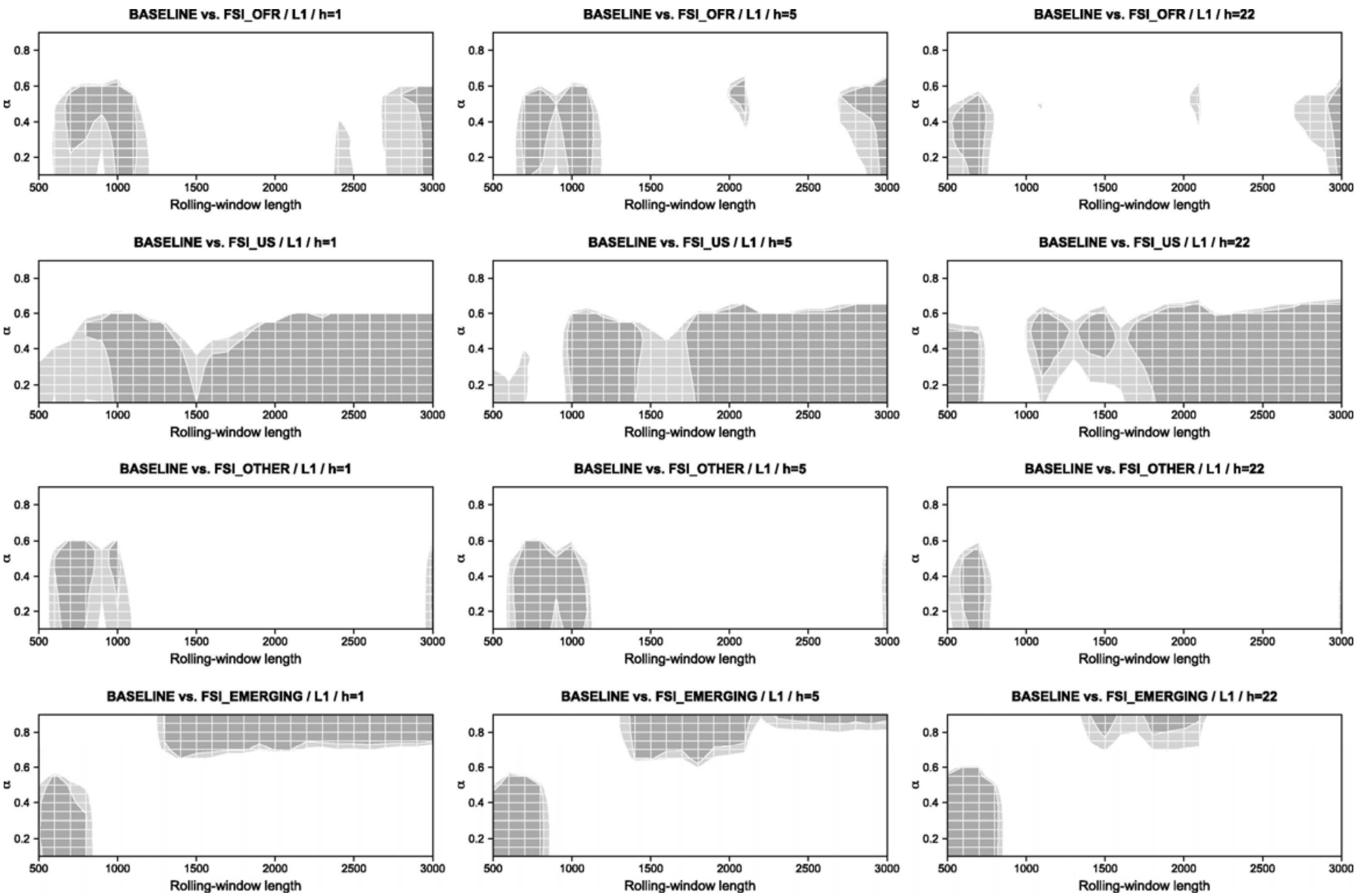


Fig. A1. Simulation Results (Relative Loss Criterion/L1 Loss). Note: Black area: Relative-loss criterion is significant at the 5% level. Gray area: Relative-loss criterion is significant at the 10% level. One sided test. Results are based on rolling-window estimates of the baseline HAR-RV model/the HAR-RV model extended to include a measure of financial stress. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L1: Lin–lin loss function.

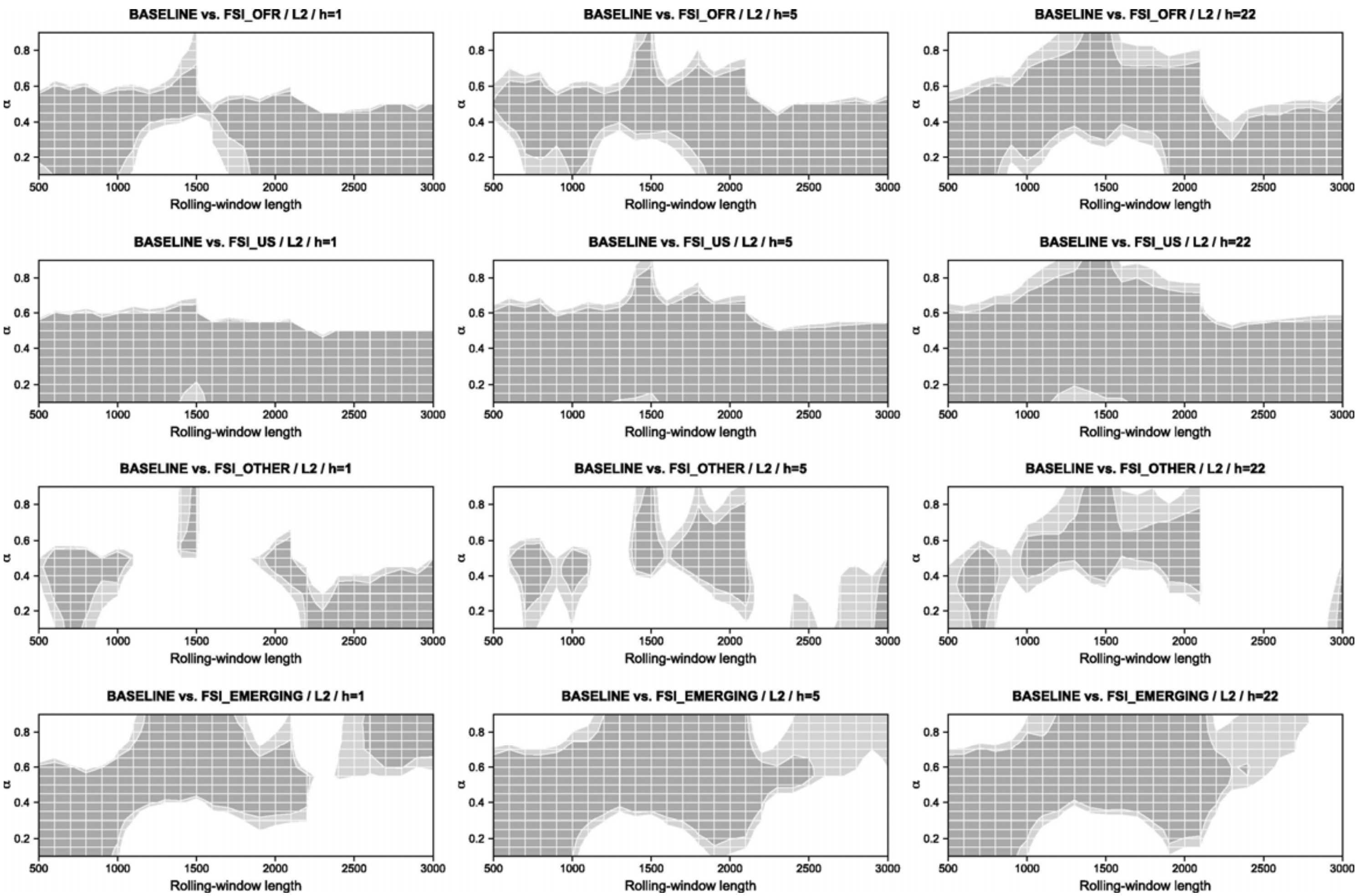


Fig. A2. Simulation Results (Relative Loss Criterion/L2 Loss). Note: Black area: Relative-loss criterion is significant at the 5% level. Gray area: Relative-loss criterion is significant at the 10% level. One sided test. Results are based on rolling-window estimates of the baseline HAR-RV model/the HAR-RV model extended to include a measure of financial stress. The horizontal axis displays the length of a rolling window. The vertical axis displays the asymmetry parameter of the loss function. L2: Quad-quad loss function.

References

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Econ. 31, 316–321.

Akram, Q.F., 2009. Commodity prices, interest rates and the dollar. Energy Econ. 31, 838–851.

Amaya, D., Christoffersen, P., Jacobs, K., Vasquez, A., 2015. Does realized skewness predict the cross-section of equity returns? J. Financ. Econ. 118, 135–167.

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in?nancial markets. J. Empirical Financ. 4, 115–158.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev., 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Vega, C., 2003. Micro effects of macro announcements: real-time price discovery in foreign exchange. Am. Econ. Rev. 93 (1), 38–62.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89, 701–720.

Andersen, T.G., Bollerslev, T., Huang, X., 2011. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. J. Econometrics 160, 176–189.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Am. Stat. Assoc. 96, 42–55.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169, 75–93.

Arouri, M.E.H., Lahiani, A., Lévy, A., Nguyen, D.K., 2012. Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. Energy Econ. 34, 283–293.

Aye, G.C., Dadam, V., Gupta, R., Mamba, B., 2014. Oil Price uncertainty and manufacturing production. Energy Econ. 43, 41–47.

Bahloul, W., Balcilar, M., Cunado, J., Gupta, R., 2018. The role of economic and financial uncertainties in predicting commodity futures returns and volatility: Evidence from a nonparametric causality-in-quantiles test. J. Multinatl. Financ. Manage. 45, 52–71.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. J. Financ. Econometrics 2, 1–37.

Barndorff-Nielsen, O.E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. J. Financ. Econometrics 4, 1–30.

Barndorff-Nielsen, O.E., Kinnebrouk, S., Shephard, N., 2010. Measuring downside risk: realised semivariance. In: Bollerslev, T., Russell, J., Watson, M. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, pp. 117–136.

Bollerslev, T., Law, T.H., Tauchen, G., 2008. Risk, jumps, and diversification. J. Econometrics 144 (1), 234–256.

Bollerslev, T., Li, S.Z., Zhao, B., 2017. Good volatility, bad volatility, and the cross section of stock returns. J. Financ. Quant. Anal., 1–57.

Büyüksßahin, B., Robe, M.A., 2014. Speculators, commodities and cross-market linkages. Journal of International Money and Finance 42, 38–70.

Campbell, J.Y., 2008. Viewpoint: estimating the equity premium. Canadian Journal of Economics 41, 1–21.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average. Review of Financial Studies 21, 1509–1531.

Cenesizoglu, T., Timmermann, S., 2012. Do return prediction models add economic value? Journal of Banking and Finance 36, 2974–2987.

Chevallier, J., Sévi, B., 2012. On the volatility–volume relationship in energy futures markets using intraday data. Energy Econ. 34, 1896–1909.

Chatrath, A., Miao, H., Ramchander, S., Wang, T., 2015. The forecasting efficacy of risk-neutral moments for crude oil volatility. Journal of Forecasting 34, 177–190.

Chkili, W., Hammoudeh, S., Nguyen, D.K., 2014. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Econ. 41, 1–18.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econometrics 7, 174–196.

Corsi, F., Pirino, D., Reno, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159 (2), 276–288.

Degiannakis, S., Floros, C., 2015. Modelling and Forecasting High Frequency Financial Data. Palgrave Macmillan UK, London, pp. 58–109.

Degiannakis, S., Floros, C., 2016. Intra–day realized volatility for European and USA stock indices. Global Finance Journal 29, 24–41.

Degiannakis, S., Filis, G., 2017. Forecasting oil price realized volatility using information channels from other asset classes. J. Int. Money Financ. 76, 28–49.

Degiannakis, S., Filis, G., Klein, T., Walther, T., 2019. Forecasting realized volatility of agricultural commodities. Int. J. Forecast. https://doi.org/10.2139/ssrn.3446748 (September 2, 2019). SSRN.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–263.

Di Matteo, T., 2007. Multi-scaling in finance. Quant. Financ. 7 (1), 21–36.

Duong, D., Swanson, N.R., 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. J. Econometrics 187 (2), 606–621.

Efimova, O., Serletis, A., 2014. Energy markets volatility modelling using GARCH. Energy Econ. 43, 264–273.

Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasts biased? Agency versus behavioral explanations. Quart. J. Econ. 111, 21–40.

Elder, J., Serletis, A., 2010. Oil price uncertainty. J. Money Credit Bank. 42, 1137–1159.

Elliott, G., Komunjer, I., Timmermann, A., 2005. Estimation and testing of forecasting rationality under flexible loss. Rev. Econ. Stud. 72, 1107–1125.

Elliott, G., Komunjer, I., Timmermann, A., 2008. Biases in macroeconomic forecasts: irrationality or asymmetric loss? J. Eur. Econ. Assoc. 6, 122–157.

Engle, E., Sun, Z., 2007. When is noise not noise? A microstructure estimate of realized volatility. NYU Working Paper No. FIN-07-047.

Fattouh, B., Kilian, L., Mahadeva, L., 2013. The role of speculation in oil markets: what have we learned so far? Energy J. 34, 7–33.

Gkillas, K., Gupta, R., Wohar, M.E., 2018. Volatility jumps: the role of geopolitical risks. Financ. Res. Lett. 27, 247–258.

Gupta, R., Kanda, P.T., Tiwari, A.K., Wohar, M.E., 2018. Time-varying predictability of oil market movements over a century of data: The role of US financial stress. University of Pretoria, Department of Economics, Working Paper No. 201848.

Hansen, P.R., Lunde, A., 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1). J. Appl. Econometrics 20, 873–889.

Hansen, P.R., Huang, Z., 2016. Exponential GARCH modeling with realized measures of volatility. J. Bus. Econ. Stat. 34 (2), 269–287.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. Int. J. Forecast. 13, 281–291.

Haugom, E., Langeland, H., Molnár, P., Westgaard, S., 2014. Forecasting volatility of the US oil market. J. Bank. Financ. 47, 1–14.

Hou, A., Suardi, S., 2012. A nonparametric GARCH model of crude oil price return volatility. Energy Econ. 34, 618–626.

Hyndman, R.J., 2017. Forecast: Forecasting functions for time series and linear models. R package version 8.0. URL: http://github.com/robjhyndman/forecast.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Softw. 26, 1–22.

Kang, S.H., Kang, S.M., Yoon, S.M., 2009. Forecasting volatility of crude oil markets. Energy Econ. 31, 119–125.

Kang, S.-H., Yoon, S.-M., 2013. Modelling and forecasting the volatility of petroleum futures prices. Energy Econ. 36, 354–362.

Kilian, L., Manganelli, S., 2007. Quantifying the risk of deflation. J. Money Credit Bank. 39, 561–590.

Lux, T., Segnon, M., Gupta, R., 2016. Forecasting crude oil price volatility and value-at-risk: evidence from historical and recent data. Energy Econ. 56, 117–133.

Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311.

Ma, F., Wahab, M.I.M., Huang, D., Xu, W., 2017. Forecasting the realized volatility of the oil futures market: a regime switching approach. Energy Econ. 67, 136–145.

Ma, F., Wei, Y., Liu, L., Huang, D., 2018. Forecasting realized volatility of oil futures market: a new insight. J. Forecast. 37 (4), 419–436.

McAleer, M., Medeiros, M.C., 2008. Realized volatility: a review. Econometric Rev. 27, 10–45.

Mei, D., Liu, J., Ma, F., Chen, W., 2017. Forecasting stock market volatility: do realized skewness and kurtosis help? Physica A 481, 153–159.

Monin, P., 2017. The OFR Financial Stress Index. Working Paper 17–04. Office of Financial Research.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V., 1997. Volatilities of different time resolutions – analyzing the dynamics of market components. J. Empirical Financ. 4, 213–239.

Nazlioglu, S., Soytas, U., Gupta, R., 2015. Oil prices and financial stress: a volatility spillover analysis. Energy Policy 82, 278–288.

Nomikos, N.K., Pouliasis, P.K., 2011. Forecasting petroleum futures markets volatility: the role of regimes and market conditions. Energy Econ. 33, 321–337.

Oomen, R.C., 2001. Using High Frequency Data to Calculate, Model and Forecast Realized Volatility (No. 75). Society for Computational Economics.

Oomen, R.C., 2004. Modelling realized variance when returns are serially correlated (No. SP II 2004-11). WZB Discussion Paper.

Phan, D.H.B., Sharma, S.S., Narayan, P.K., 2016. Intraday volatility interaction between the crude oil and equity markets. J. Int. Financ. Markets Inst. Money 40, 1–13.

Pierdzioch, C., Risse, M., Rohloff, S., 2014. The international business cycle and gold-price fluctuations. Quart. Rev. Econ. Financ. 54, 292–305.

Pierdzioch, C., Risse, M., Rohloff, S., 2016a. A boosting approach to forecasting the volatility of gold-price fluctuations under flexible loss. Resour. Policy 47, 95–107.

Pierdzioch, C., Risse, M., Rohloff, S., 2016b. Fluctuations of the real exchange rate, real interest rates, and the dynamics of the price of gold in a small open economy. Empirical Econ. 51, 1481–1499.

Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2015. Do jumps matter for volatility forecasting? Evidence from energy markets. J. Futures Markets 36, 758–792.

Rapach, D.E., Wohar, M.E., Rangvid, J., 2005. Macro variables and international stock return predictability. Int. J. Forecast. 21, 137–166.

R Core Team, 2017. R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/. R version 3.3.3.

Sadorsky, P., 2006. Modeling and forecasting petroleum futures volatility. Energy Econ. 28, 467–488.

Sadorsky, P., McKenzie, M.D., 2008. Power transformation models and volatility forecasting. J. Forecasting 27, 587–606.

Sévi, B., 2014. Forecasting the volatility of crude oil futures using intraday data. Eur. J. Oper. Res. 235, 643–659.

Silvennoinen, A., Thorp, S., 2013. Financialization, crisis and commodity correlation dynamics. J. Int. Financ. Markets Inst. Money 24, 42–65.

Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financ. Anal. J. 68, 54–74.

Tay, A., Ting, C., Tse, Y.K., Warachka, M., 2009. Using high-frequency transaction data to estimate the probability of informed trading. J. Financ. Econometrics 7, 288–311.

Taylor, S.J., Xu, X., 1997. The incremental volatility information in one million foreign exchange quotations. J. Empirical Financ. 4, 317–340.

Todorov, V., Tauchen, G., 2011. Volatility jumps. J. Bus. Econ. Stat. 29 (3), 356–371.

van Eyden, R., Difeto, M., Gupta, R., Wohar, M.E., 2019. Oil price volatility and economic growth: evidence from advanced OECD countries using over one century of data. Appl. Energy 233 (234), 612–621.

Wei, Y., Wang, Y., Huang, D., 2010. Forecasting crude oil market volatility: further evidence using GARCH-class models. Energy Econ. 32, 1477–1484.

Wen, F., Zhao, Y., Zhang, M., Hu, C., 2019. Forecasting realized volatility of crude oil futures with equity market uncertainty. Appl. Econ., 1–17.

West, K.D., Edison, H.J., Cho, D., 1993. A utility-based comparison of some models of exchange rate volatility. J. Int. Econ. 35, 23–46.

Zeileis, A., 2004. Econometric computing with HC and HAC covariance matrix estimators. J. Stat. Softw. 11, 1–17.

Zhou, H., Zhu, J.Q., 2012. An empirical examination of jump risk in asset pricing and volatility forecasting in China's equity and bond markets. Pacific-Basin Financ. J. 20, 857–880.

Footnotes

[1]Chatrath et al. (2015) and Phan et al. (2016) also forecast realized oil-price volatility derived using intraday data, but do not use the HAR framework. Rather, they use regression and GARCH-based models.
[2]Volatility estimated from high frequency data exhibits patterns that are repeated at different time scales (scaling laws). Such scaling-type regularities can provide useful information for volatility forecasting (also known as the "scaling concept"), as outlined in Matteo (2007).
[3]For robustness purposes, given the importance of jumps in volatility forecasting, we also study the role of jumps for forecasting the realized volatility of oil returns in Section 4.2.
[4]Researchers often use the term volatility to denote the standard deviation of asset-price returns. To avoid any risk of confusion, we use the term realized volatility to denote the realized variance of oil-price returns, and explicitly denote the square root of the realized volatility as the realized standard deviation of oil-price returns.
[5]The approach we adopt is quite standard in the literature in terms of forecasting realized volatility (see, e.g., Andersen et al., 2007). We use (as we shall describe at the end of this section) rolling-estimation windows to produce multiple-steps forecasts for the two longer forecast horizons. It should also be noted that, in Section 4.3, we shall briefly report results for two longer forecast horizons ($h = 44$ and $h = 66$).
[6]The data is available for download from: https://www.financialresearch.gov/financial-stress-index/.
[7]We constructed the data matrix such that we have exactly the same number of observations (4320 observations; after computing realized volatility for the long forecast horizon and computing RV_m each consumes 22 observations) for all three forecasting horizons.
[8] We computed all estimation results that we report in this research using the R programming environment (R Core Team, 2017), where we used the R packages "sandwich" (Zeileis, 2004) to compute the full-sample Newey-West robust standard errors reported in Table 1.

[9]The p-values were computed using the R package “forecast” (Hyndman, 2017, Hyndman and Khandakar, 2008). We modified the source code to account for the asymmetry of the loss function.
[10] Forecasting average realized volatility in case of the medium and long forecasting horizon gives qualitatively similar results (see the Supplementary Material).
[11] Following Amaya et al. (2015), $RSK_t = \frac{\sqrt{T}\ \sum_{i=1}^{T} r_{t,i}^3}{(\sum_{i=1}^{T} r_{t,i}^2)^{3/2}}$, and $RKU_t = \frac{T\ \sum_{i=1}^{T} r_{t,i}^4}{(\sum_{i=1}^{T} r_{t,i}^2)^2}$.
[12]Andersen et al. (2011) noted that $\lim_{T\to\infty} RV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=1}^{N_t} \kappa_{t,j}^2$, where N_t is the number of jumps within day t and $\kappa t, j$ is the jump size. Thus $RV_t = sum_{i=1}^{T} r_{t,i}^2$ is a consistent estimator of the integrated variance $\int_{t-1}^{t} \sigma^2(s)\, ds$ plus the jump contribution. On the other hand, the results of Barndorff-Nielsen and Shephard (2004) imply that $\lim_{T\to\infty} BV_t = \int_{t-1}^{t} \sigma^2(s)\, ds$, where BV_t vis the realized bipower variation defined as $BV_t = \mu_1^{-1}\left(\frac{T}{T-1}\right)\sum_{i=2}^{T} |r_{t,i-1}||r_{t,i}| = \frac{\pi}{2}\sum_{i=2}^{T} |r_{t,i-1}||r_{t,i}|$, where $\mu_a = E(|Z|^a), \quad ZN(0,1), a > 0$. Therefore, $J_t = RV_t - BV_t$ is a consistent estimator of the pure jump contribution and can form the basis of a test for jumps. For a formal test for jumps, we follow Barndorff-Nielsen and Shephard (2006), such that: $JT_t = \frac{RV_t - BV_t}{(\nu_{bb}-\nu_{qq})\frac{1}{T}TP_t}$, where, $\nu_{bb} = \left(\frac{\pi}{2}\right)^2 + \pi - 3, \nu_{qq} = 2$, and TP_t is the Tri-Power Quarticity defined as: $TP_t = T\mu_{4/3}^{-3}\left(\frac{T}{T-1}\right)\sum_{i=3}^{T} |r_{t,i-2}|^{4/3}|r_{t,i-1}|^{4/3}|r_{t,i}|^{4/3}$ which converges to $TP_t \to \int_{t-1}^{t} \sigma^4(s)\, ds$ even in the presence of jumps. Note, for each $t, JT_t \overset{D}{\sim} N(0,1)$ as $T \to \infty$. The jump contribution to RV_t is either positive or null. Therefore, to avoid having negative empirical contributions, following Zhou and Zhu (2012), we re-define the jump measure as: $J_t = max(RV_t - BV_t;0)$.
[13]We leave out the volatility-based financial-stress index because it contains measures of implied and realized volatility from commodity markets (among other measures; see Monin, 2017).
[14] It should be noted that the spectrum of colours used to compute the contour plots is the same across figures, but the scaling differs across figures. In all cases, however, a dark red colour indicates a “large” positive relative-loss criterion while a dark blue colour indicates a “large” negative relative-loss criterion.
[15]The relative-loss criterion, of course, assumes relatively large values for the “preferred” models shown in Fig. 2, Fig. 3 (see the stars in those figures). In addition to the information already conveyed by the results summarized in Fig. 2, Fig. 3, however, the relative-loss criterion informs about the magnitude of the benefits of using financial stress for forecasting realized volatility.
[16]The results are based on a simulation experiment that closely follows the simulation experiment described by Rapach et al. (2005). See the Technical Appendix for details on the simulation design.
[17]The results of the utility-based analysis, therefore, also reflect the implications of the forecasts of mean returns for portfolio composition. This is in contrast to the results for the relative-loss criterion.