Time-varying risk aversion and realized gold volatility

Riza Demirer[a], Konstantinos Gkillas[b], Rangan Gupta[c], Christian Pierdzioch[d]

Submission: December 2018
Resubmission: April 2019
2nd Resubmission: June 2019

Abstract

We study the in- and out-of-sample predictive value of time-varying risk aversion for realized volatility of gold returns via extended heterogeneous autoregressive realized volatility (HAR-RV) models. Our findings suggest that time-varying risk aversion possesses predictive value for gold volatility both in- and out-of-sample. Time-varying risk aversion is found to absorb the in-sample predictive power of economic uncertainty at a short forecasting horizon. We also study the out-of-sample predictive power of time-varying risk aversion in the presence of realized higher-moments, jumps, gold returns, a leverage effect as well as economic policy uncertainty in the forecasting model. In addition, we study the role of the shape of the loss function used to evaluate losses from forecast errors for the role of time-varying risk aversion as a predictor of realized volatility. Overall, our findings show that time-varying risk aversion often captures information useful for out-of-sample prediction of realized volatility not already contained in the other predictors.

Keywords: Gold-price returns; Realized volatility; Forecasting

[a] Department of Economics and Finance, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1102, USA; E-mail address: rdemire@siue.edu.

[b] Department of Business Administration, University of Patras – University Campus, Rio, P.O. Box 1391, 26500 Patras, Greece; Email address: gillask@upatras.gr.

[c] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; E-mail address: rangan.gupta@up.ac.za.

[d] Corresponding author. Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany; Email address: c.pierdzioch@hsu-hh.de.

1 Introduction

Recent research on global financial markets establishes a link between cycles in capital flows and the level of risk aversion (e.g. Rey 2018), showing that risk aversion has significant explanatory power over equity-market comovements (e.g., Xu 2017, Demirer et al. 2018). Clearly, capital flows across risky and relatively safer assets would be closely linked to the level of risk aversion in financial markets as utility maximizing investors assume investment positions based on their willingness to take on risks. To that end, given the role of gold as a traditional safe haven in which investors seek refuge during periods of uncertainty, one can argue that the role of risk aversion as a driver of return dynamics in financial markets is not necessarily limited to equities, but also extends to the market for gold. Interestingly, however, despite the multitude of studies that explore the role of gold as a potential safe haven (e.g., Baur and Lucey 2010, Lucey and Li 2015), the influence of time-varying risk aversion on the volatility of gold-price movements is largely understudied, partially due to the challenges in controlling for the time variation in macroeconomic uncertainty to estimate the time variation in risk aversion. The main contribution of this paper is to examine the predictive power of risk aversion over gold volatility by utilizing a recently developed measure of time-varying risk aversion which distinguishes the time variation in economic uncertainty from the time variation in risk aversion. By doing so, we provide new insight to the role of risk aversion in financial markets and volatility modeling in safe-haven assets.

Clearly, forecasting volatility of gold returns is of interest not only for investors in the pricing of related derivatives as well as hedging strategies for stock market-fluctuations (Poon and Granger, 2003), but also for policy makers given the evidence that commodities, in particular gold, possess predictive value over currency-market fluctuations (e.g., Chen and Rogoff 2003, Cashin et al. 2004, Apergis 2014), an issue that is particularly important for emerging economies that have high risk exposures with respect to currency fluctuations. Furthermore, given the evidence of significant volatility spillovers across gold and other commodities, particularly oil (e.g., Ewing and Malik 2013), and that precious metals served as sources of information transmission during financial crises (Kang et al. 2017), exploring the predictive role of risk aversion over gold volatility can provide valuable insight to whether the time-variation in risk aversion is the underlying fundamental factor driving the spillover effects across asset classes, particularly during periods of high uncertainty. Although the literature offers a limited number of studies relating various uncertainty measures to gold-return dynamics (e.g., Jones and Sackley 2016, Balcilar et al. 2016, Bouoiyour et al. 2018), these studies have not specifically examined the effect of the changes in the level of market risk aversion on safe-haven assets, particularly gold. To that end, the time-varying risk aversion measure recently developed by Bekaert et al. (2017) offers a valuable opening as it distinguishes the time variation in economic uncertainty (the amount of risk) from time variation in risk aversion (the price of risk), providing an unbiased representation for time-varying risk aversion in financial markets. To the best of our knowledge, ours is the first study to utilize this unbiased measure of risk aversion in the context of forecasting for safe-haven assets.

In recent studies, volatility is associated with a volatility risk premium, which can be regarded as the compensation that investors demand against unexpected market fluctuations beyond what can be computed based on the first moments of a price process (e.g. Li and Zinna 2018). Furthermore, there is a well established literature on the so-called leverage effect which refers to the empirical evidence that establishes a link between asset returns and volatility (e.g. Christie 1982). To that end, one can argue that the level of risk aversion that serves as a driver of a possible volatility risk premium embedded in the price of gold can be linked to expected volatility either from a risk-premium or a leverage-effect channel. As noted, gold has traditionally been considered an asset to preserve value, offering long-term stability and security in financial markets during volatile periods and against downside risks (Erb and Harvey 2013). Given this, one can argue that return and volatility dynamics in gold can be affected in different ways depending on the effect of different types of investors. The first type of investors are those looking for a safe haven during periods of uncertainty. The second type of investors sees gold as an investment just like other traditional assets (e.g., investors in mining firms). Clearly, during periods of turbulence in financial markets, there will be upward pressure on the price of gold, while during market booms the price of gold may (or may not) fall, depending on whether or not investors who consider gold as a traditional investment still see value in gold-related assets. When we consider gold related stocks whose prices generally move in tandem with the price of gold, a relatively small increase in gold price could lead to significant profits in some gold stocks, while a relatively small drop in gold prices could lead to significant losses in some gold stocks, especially in the case of firms with relatively weaker fundamentals and management structure. Therefore, depending on the market state and investors' appetite for risk, gold and gold related assets can exhibit highly volatile patterns, closely related to the degree of risk appetite among investors, which can be captured by time-varying risk aversion. To that end, insofar that risk aversion has predictive power on gold volatility, our findings can provide useful insights to understanding how risk aversion relates to the time variation in gold market fluctuations, and thus, to the gold volatility risk premium. Furthermore, estimating volatility via quadratic variation, which is regarded as the best estimator of integrated (latent) volatility, our findings can also provide useful insights into gold-jump risk premium documented in the literature (see Todorov and Tauchen 2011).

In our empirical analysis, we focus on the realized volatility of gold returns that we compute from intraday data. The use of intraday data allows us to control for higher moments including the realized skewness and kurtosis that have been shown to have predictive power in forecasting models in a number of different contexts including gold (Mei et al. 2017, Bonato et al. 2018, Gkillas et al. 2018). We also control for economic uncertainty in our models, allowing us to separately examine the impact of uncertainty and changes in risk aversion on realized volatility. This distinction is particularly important as risk aversion can fluctuate due to changes in wealth, background risk, and emotions that alter risk appetite (Guiso et al. 2018). To that end, given the unbiased nature of the risk-aversion measure we utilize in our tests, distinguishing the time variation in economic uncertainty from the time variation in risk aversion, our study provides new insight to the drivers of realized volatility of gold returns. We employ the heterogeneous autoregressive realized volatility (HAR-RV) model developed by Corsi (2009) to model and forecast the realized volatility of gold returns as this widely-studied model accounts for several stylized facts such as fat tails and the long-memory property of financial-market volatility, despite the simplicity offered by the model. By doing so, we extend the HAR-RV model to study the in- and out-of-sample predictive value of risk aversion, after controlling for various alternative predictors including realized higher-moments, realized jumps, gold returns, a leverage term as well as economic uncertainty.

Our findings show that time-varying risk aversion possesses predictive value for realized gold volatility both in- and out-of-sample. While realized skewness stands out as a significant in-sample predictors for realized volatility, risk aversion has significant in-sample predictive value and is found to absorb the predictive power of economic uncertainty if investors predict realized volatility at a short forecasting horizon. Out-of-sample results show that the inclusion of risk aversion in the HAR-RV model yields better results for various model configurations in terms of forecast accuracy relative to alternative models that include realized higher-moments, jumps, gold returns, a leverage term, and economic uncertainty. We systematically document how the relative forecast accuracy of the HAR-RV-cum-risk-aversion model relates to the length of the forecast horizon and the loss function (absolute versus squared error loss, asymmetric loss) used to evaluate forecast errors. Moreover, considering that the prices of risky assets drop as investors demand greater compensation for risk when risk aversion is high, one can argue that the volatility impact on gold would be in the positive direction, captured by good realized volatility (computed from positive returns), while the opposite holds during good times. For this reason, we differentiate between "good" and "bad" realized volatility, allowing us to explore possible asymmetric effects of risk aversion on gold volatility. Overall, our findings show that time-varying risk aversion contains information useful for out-of-sample forecasting of ("good" and "bad") realized volatility over and above the information already embedded in other widely-studied predictors like higher-order moments, jumps, and economic uncertainty measured by the economic policy uncertainty index developed by Baker et al. (2016).

We present in Section 2 a brief review of the different strands of studies on gold. We describe in Section 3 the methods that we use in our empirical analysis. We present our data in Section 4, summarize our empirical results Section 5, and conclude in Section 6.

2 Literature Review

Given the potential safe-haven and hedging properties of gold investments, a growing number of studies has undertaken significant efforts to model and forecast return volatility in the gold market. One strand of research focuses on macroeconomic determinants of gold returns and volatility. For example, Tulley and Lucey (2007) estimate an asymmetric power GARCH model on monthly data and show that fluctuations in the value of the dollar have an impact on gold returns whereas major macroeconomic variables do not help to model volatility. On the other hand, Batten et al. (2010) highlight the role of fluctuations in monetary macroeconomic variables for modeling volatility in gold returns although their results suggest that the effect of macro variables is potentially unstable over time. In their empirical analysis, they show that the effect of macroeconomic fluctuations on gold volatility is stronger during an earlier sub-period (1986−1995), while the role played by the volatility of other financial variables strengthened in a later sub-period (1996−2006). Batten et al. (2010), therefore, conclude that gold behaved like an investment instrument in the later sub-period, suggesting that its links to monetary variables have loosened in recent years.

Another strand of research utilizes increasingly sophisticated GARCH models to model and forecast gold volatility (e.g., Hammoudeh and Yuan 2008). Using daily data to study the out-of-sample performance of various GARCH models, Bentes (2015) reports that a fractionally integrated GARCH model delivers the best forecasts of gold returns and volatility, based on widely studied forecast-accuracy criteria (including the mean-absolute and the mean-squared forecasting error). Similarly, Chkili et al. (2014) estimate several GARCH models to study the role of long memory and asymmetry for modeling and forecasting the conditional volatility and market risk of gold and other commodities (see also Demiralay and Ulusoy 2014).

A third strand of research focuses on the properties of the realized volatility of gold price fluctuations. For example, using a boosting approach, Pierdzioch et al. (2016) examine the time-varying predictive value of several financial and macroeconomic variables for out-of-sample forecasting the monthly realized gold-price volatility over the sample period from 1987 to 2015. Focusing on the role of the forecaster's loss function on forecast performance, Pierdzioch et al. (2016) find that a forecaster who encounters a larger loss when underestimating rather than overestimating gold-price volatility benefits from using the forecasts implied by their boosting approach. In an earlier study, using high-frequency, intra-daily gold data to construct measures of realized gold-price volatility, Neely (2004) shows that option implied volatility is a biased forecast of the realized volatility and that implied volatility tends to be informationally inefficient with respect to forecasts computed by means of competing econometric models, while econometric forecasts have no incremental value over implied volatility when a delta hedging tracking error is used to evaluate out-of-sample volatility forecasts.

The literature on realized gold volatility is directly relevant for our research. Specifically, we use the HAR-RV model developed by Corsi (2009) to model and forecast realized gold volatility. Variants of the HAR-RV model have been widely studied in recent research (see, for example, Haugom et al. 2014, Lyócsa and Molnár 2016) as it accounts for several stylized facts such as fat tails and the long-memory property of financial-market volatility, despite the simplicity the model offers. In our empirical analysis, we extend the core HAR-RV model to include measures of realized higher-moments including realized skewness and kurtosis that have been found to significantly improve model performance in the case of stock-market indexes (Mei et al. 2017). This is a remarkable finding given that, in the case of gold and silver, evidence suggests that it is difficult to beat the HAR-RV model in terms of forecasting performance by using versions of the univariate HAR-RV model extended to include semi-variances and jumps (Lyócsa and Molnár 2016). Finally, we control for realized jumps, gold returns, and the economic policy uncertainty and examine whether risk aversion possesses incremental in- and out-of-sample predictive power over gold market volatility beyond that of a number of predictors that researchers have studied in earlier literature.

3 Methods

We follow Andersen et al. (2012) who propose median realized variance (MRV) as a jump-robust estimator of integrated variance, computed using intraday data. MRV_t is considerably less biased than other measures of realized volatility in the presence of jumps and/or market-microstructure noise. It is given by MRV_t:[1]

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

where $r_{t,i}$ denotes the intraday return i within day t and $i = 1,..,T$ denotes the number of intraday observations within a day. We consider MRV as our measure of daily standard RV (RV^S) in order to attenuate the effect of microstructure noise on our empirical results. It is well-known that intraday data are contaminated by market-microstructure noise, the influence of which we try to avoid in our empirical analysis (Ghysels and Sinko 2011).

In a recent study, Bonato et al. (2018) show that realized moments, computed from intraday gold returns, can improve the predictive value of estimated forecasting models for gold returns. Given this, we supplement our benchmark HAR-RV model by including realized skewness and kurtosis as potential predictors. Building on the work of Barndorff-Nielsen et al. (2010), Amaya et al. (2015) compute higher-moments of realized skewness (RSK) and realized kurtosis (RKU) from intraday returns. Like Amaya et al. (2015), we define, RSK_t and RKU_t, standardized by the realized variance, as follows:

$$RSK_t = \frac{\sqrt{T}\ \sum_{i=1}^{T} r_{t,i}^3}{(\sum_{i=1}^{T} r_{t,i}^2)^{3/2}}, \tag{2}$$

$$RKU_t = \frac{T\ \sum_{i=1}^{T} r_{t,i}^4}{(\sum_{i=1}^{T} r_{t,i}^2)^2}. \tag{3}$$

We consider daily RSK as a measure of the asymmetry of the distribution of the daily returns, while RKU measures the extremes of the same.

As far as the literature on modeling and forecasting realized volatility is concerned, Corsi (2009) proposes the HAR-RV model, which in turn has become one of the most popular models in this strand of the literature. The HAR-RV model has been shown to capture "stylized facts" of long memory and multi-scaling behavior associated with volatility of financial markets. The

[1] It should be noted that researchers often use the term volatility to denote the standard deviation of asset-price movements. We use in this paper the term realized volatility to denote the realized variance of gold-price movements and the terms realized volatility and realized variance interchangeably.

benchmark HAR-RV model, for $h-$days-ahead forecasting, can be described as follows:

$$RV_{t+h}=\beta_0+\beta_d\,RV_t+\beta_w\,RV_{w,t}+\beta_m\,RV_{m,t}+\varepsilon_{t+h}, \tag{4}$$

where $RV_{w,t}$ denotes the average RV from day $t-5$ to day $t-1$, while $RV_{m,t}$ denotes the average RV from day $t-22$ to day $t-1$.

We use the standard HAR-RV model as our benchmark model to predict the realized-volatility and, as in Mei et al. (2017), we add realized skewness and realized kurtosis (or both) as additional predictors to the benchmark model. In addition, based on the question we are aiming to answer, we add the economic policy uncertainty (EPU) index of Baker et al. (2016), and the measure of time-varying risk aversion (RA) that Bekaert et al. (2017) have recently developed in order to explore whether risk aversion has incremental predictive information. To this end, we consider variantss of the following extended HAR-RV model:

$$RV_{t+h}=\beta_0+\beta_d\,RV_t+\beta_w\,RV_{w,t}+\beta_m\,RV_{m,t}+\sum_{j=1}^{n}\theta_j Z_{j,t}+\varepsilon_{t+h} \tag{5}$$

where n denotes the number of control variables, Z, used to predict RV. Depending on the model configuration being studied, the vector of control variables represents RSK, RKU, EPU, jumps, returns, a leverage effect, and/or RA.

4 Data

We use intraday data on gold to construct daily measures of standard realized volatility, the corresponding good and bad variants, realized skewness, and realized kurtosis. Gold futures are traded in NYMEX over a 24 hours trading day (pit and electronic). We focus on gold futures prices, rather than spot prices, due to the low transaction costs associated with futures trading, which makes the analysis more relevant for practical applications in the context of hedging and/or safe-haven analyses. Furthermore, one can expect price discovery to take place primarily in the futures market as the futures price responds to new information faster than the spot price due to lower transaction costs and ease of short selling associated with the futures contracts (Shrestha 2014). 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-minute prices are obtained via last-tick interpolation (if the price is not available at the 1-minute stamp, the previously available price is imputed). 5-minute returns are then computed by taking the log-differences of these prices and are then used to compute the realized moments.

Besides the intraday data, economic uncertainty is captured by the EPU index developed by Baker et al. (2016). This index covers a large selection of news sources tracking the frequency of various keywords related to policy uncertainty, regulatory changes as well as disagreement among economic forecasters. The EPU index is available for download from: http://www.policyuncertainty.com/. In addition, in order to measure risk aversion (RA), we utilize the risk aversion index of Bekaert et al. (2017), which is available for download from: https://www.nancyxu.net/risk-aversion-index. They develop a new measure of time-varying risk aversion that ultimately can be calculated from observable financial information at high (daily) frequencies. This measure relies on a set of six financial instruments, namely, the term spread, credit spread, a de-trended dividend yield, realized and risk-neutral equity return variance and realized corporate bond return variance. As discussed earlier, an important feature of this measure is that it distinguishes time variation in economic uncertainty (the amount of risk) from time variation in risk aversion (the price of risk) and, thus, provides an unbiased representation for time-varying risk aversion based on a utility function in the hyperbolic absolute risk aversion (HARA) class.

– Please include Table 1 about here. –

The sample period runs from December 2, 1997 to December 30, 2016 (reflecting data availability of the risk aversion index used as one of the predictors), giving us a total of 4,748 observations. Table 1 reports some summary statistics of the data. In our forecasting tests, we consider three forecasting horizons (short, $h = 1$; medium, $h = 5$; long $h = 22$) and construct the data matrix such that we have exactly the same number of observations (4,704 observations; computing RV for the long forecast horizon and computing RV_m each consumes 22 observations) for all three forecasting horizons.

5 Empirical Findings

5.1 In-Sample Findings

Table 2 summarizes the estimation results for realized volatility for the full sample period. Estimation results are computed using the R programming environment (R Core Team 2017). Newey-West robust standard errors are computed using the R packages "sandwich" (Zeileis 2004). We estimate, in the first step, the core HAR-RV model. In the second step, we add risk aversion as a predictor. In the third step, we include the other predictors and compare the forecasting performance of the competing models. We observe that the MRV component of the core HAR-RV model has significant predictive power at all forecasting horizons. The MRV_w component is significant at the short forecasting horizon, while the MRV_m component is significant mainly at the medium and long forecasting horizons. The models supplemented by the realized moments yield evidence partially in line with Mei et. al (2017) in that only realized skewness is found to have significant in-sample predictive power (and with a negative sign), across all forecast horizons.

Interestingly, we see that including risk aversion in the model adds predictive value in the case of the short forecasting horizon, as indicated by the significant estimated coefficients. At the samt time, adding economic policy uncertainty hardly affects the predictive power of the model. In the case of the medium and long forecast horizons, the estimated coefficients of risk aversion is found to be insignificant. As a robustness check, we also present the results for a model that features the first-differenced measure of risk aversion as a predictor in order to account for the potential effect of persistence of this regressor on the estimation results.[2] We observe that using the first-difference of risk aversion does not affect the overall inference, i.e., risk aversion adds

[2]The first four coefficients of autocorrelations of time-varying risk aversion are 0.83, 0.72, 0.66, and 0.58.

in-sample predictive value to the model only in case of the short forecasting horizon. Hence, we conclude that time-varying risk aversion has significant predictive power for realized volatility of gold-price movements at the short forecast horizon over and above the predictive power of economic policy uncertainty and the other predictors.

– Please include Table 2 about here. –

We also report results for the AIC and BIC of the estimated models. To this end, we divide the AIC (BIC) of the estimated models by the minimum of the AIC (BIC) computed across the estimated models, that is, the best performing model has an AIC (a BIC) equal to one. While the interpretation of these results should not bee stretched too far, they show that the core HAR-RV model is the top model only for the medium forecast horizon, and only when we use the BIC. For the short forecast horizon, the HAR-RV-RA-RSK model performs best according to both criteria. For the long forecast horizon, the HAR-RV-RA-RSK model performs best according to the BIC, while the AIC selects the HAR-RV-RA-RKU-RSK-EPU model.

Another noteworthy observation is that the sign of the coefficient for risk aversion switches from positive in the short forecast horizon to negative for the long forecast horizons. As a higher value for risk aversion is associated with market conditions during which investors have a greater tendency to move out of risky assets and into safer alternatives, possibly those considered as safe havens, the finding of a positive risk aversion effect on the short-term volatility in gold returns is not unexpected. The negative coefficients on risk aversion for the long forecast horizon, however, may be an indication of a correction effect to the initial, immediate reaction of the market to unexpected news.

5.2 Baseline Out-of-Sample Findings

Understandably, in-sample predictive value does not necessarily imply that a predictor also has out-of-sample predictive value. For our out-of-sample analysis, we use a rolling-estimation window. To this end, we use a three step approach. In a first step, we vary the length of the estimation window between 1,000 and 3,000 observations and then move the rolling estimation window forward in time on a daily basis until we reach the end of the sample period. In a second step, we identify the optimal length of the rolling-estimation window. For every length of the rolling-estimation window, we compute the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors for the model that features time-varying risk aversion and for its competitor model. We then compute the ratio of cumulated absolute (squared) forecast errors for both HAR-RV models. we select the minimum of this ratio to identify the optimal length of the rolling-estimation window. In a third step, we use the series of forecasts for the optimal rolling-estimation window to compute the Diebold and Mariano (1995) test. We use this test to compare forecast accuracy across the various model specifications. The test results are derived using the modified Diebold-Mariano test proposed by Harvey, Leybourne and Newbold (1997), where we report the p-values computed using the R package "forecast" (Hyndman 2017, Hyndman and Khandakar 2008). In order to account for heteroskedasticity, we study the scaled forecast error defined as actual realized volatility minus its forecast, divided by actual realized volatility (for such a heteroskedasticity adjustment in case of the root-mean-squared error, see Bollerslev and Ghysels 1996).

– Please include Table 3 about here. –

Table 3 presents the results for two different loss functions based on the absolute loss (L1 loss) and the squared error loss (L2). We compare the forecasts implied by the HAR-RV-RA model with the forecasts implied by three benchmark models. The first benchmark model is the core HAR-RV model without any additional predictors. The second benchmark model is the HAR-RV-EPU model. The third benchmark model is the HAR-RV-RSK-RKU model. We observe that the results for the L1 and L2 loss are similar. When we use the core HAR-RV model as our benchmark model, adding time-varying risk-aversion to the model improves out-of-sample forecast accuracy at all three forecast horizons. Turning next to the benchmark model that features economic policy uncertainty as a predictor, we observe that the HAR-RV-RA-model produces more accurate forecasts for the short forecast horizon. Finally, for the HAR-RV-RSK-RKU benchmark model, we observe improvements in forecast accuracy for L1 loss und, for the short forecast horizon, for L2 loss.

– Please include Table 4 about here. –

We next use an out-of-sample R^2 statistic similar to the one proposed by Campbell and Thomspson (2008) to assess the economic significance of the contribution of time-varying risk-aversion to the out-of-sample predictability of realized volatility. We define this statistic as $R^{L2} = 1 - \sum_t fe^2_{t,RA} / \sum_t fe^2_{t,B}$ for L2 loss and as $R^{L1} = 1 - \sum_t |fe_{t,RA}| \,/ \sum_t |\, fe_{t,B}|$ for L1 loss, where t indicates the out-of-sample periods, $fe_{t,RA}$ denotes the (scaled) forecast error of the model that features time-varying risk-aversion as a predictor, and $fe_{t,B}$ denotes the (scaled) forecast error of the benchmark model. A positive value of the statistic indicates that the model that features time-varying risk-aversion as a predictor yields better results than the benchmark model. Table 4 summarizes the results. The findings show that the out-of-sample performance of the HAR-RV-RA model is better than that of the competitor models, especially for the short forecast horizon. In additin, the results show that the competitive advantage of the HAR-RV-RA model over its rival models is larger for L2 loss than for L1 loss.

– Please include Figure 5 about here. –

Table 5 depicts the results of the modified Diebold-Mariano test (again applied to the scaled forecast errors) that we obtain when we use the first-difference of risk aversion as a predictor. We observe significant test results for the short forecast horizon. For the medium and long forecast horizon, the test results are significant when we use the HAR-RV-EPU model as the benchmark model (and they are almost significant for the HAR-RV benchmark model under L2 loss). Overall, we observe weaker results at the medium and long forecast horizon compared to the results for the model in which we do not difference the measure of time-varying risk-aversion as a predictor.

Next, we use Fair and Shiller (1990) type regressions to compare the informational content of the forecasts. We set up the Fair-Shiller regressions by regressing actual realized volatility on a constant, the forecast for the benchmark model, and the forecast of the alternative model that features time-varying risk-aversion as a predictor. The rationale behind the Fair-Shiller regression is that if the forecasts implied by the alternative model embed information over and above those already embedded in the forecasts of the benchmark model, then the coefficient associated with the forecasts implied by the alternative forecasts should be significantly different from zero, while the coefficient of the benchmark model should not be significantly different from zero. If, in contrast, the forecasts computed by means of the benchmark model completely encapsulate the information that the forecasts of the alternative model contain, then the coefficient of the benchmark (extended) model should (not) be significantly different from zero. If both models contain exactly the same information, then their coefficients in a Fair-Shiller regression are not separately identified. This means that if the forecasts of both the benchmark and the alternative model do not have predictive power for realized volatility, then both coefficients should be zero. Finally, both coefficients should be significantly different from zero when the benchmark and the alternative model contain independent information.

– Please include Table 6 about here. –

Because we are primarily interested in the significance of the coefficient associated with the alternative model (the one that features time-varying risk aversion as a predictor), able 6 depicts the p-values (based on Newey-West standard errors) of this coefficient for alternative benchmark models. The findings show that the forecasts extracted from the HAR-RV-RA model contain information not already embedded in the forecasts of the AR-RV model for the long forecast horizon. In addition, we observe that the HAR-RV-RA model produces forecasts that contain information not already encapsulated in the forecasts of the HAR-RV model for the short forecast horizon when we study the first-differenced measure of time-varying risk-aversion. When we use the HAR-RV-EPU model as our benchmark model then we observe significant test results for the medium and long forecast horizon. Finally, when the benchmark model features realized skewness and realized kurtosis as predictors the results for the short forecast horizon are significant.

5.3 Asymmetric Loss

The L1 and L2 loss functions that we use to set up the Diebold-Mariano test and the out-of-sample R^{L1} and R^{L2} statistics are special cases of a more general and potentially asymmetric loss function. Depending on the type of positions an investor holds, an asymmetric loss function arises naturally, for example, if underestimating volatility is more costly than an overestimation of the same magnitude. Specifically, we consider the following loss function Elliott et al. 2005, 2008):

$$L(\alpha, p) = [\alpha + (1-2\alpha)I(fe < 0]|fe_A|^p, \tag{6}$$

where fe denotes the forecast error, $\alpha \in (0,1)$, and $p \in \{1,2\}$. The parameter α governs the asymmetry of the loss function. A symmetric loss function obtains for $\alpha = 0.5$, while $\alpha > 0.5$ ($\alpha < 0.5$) implies that the loss from under-predicting (over-predicting) realized volatility exceeds the loss from an over-prediction (under-prediction) of the same magnitude. For $p = 1$, the loss function is of the lin-lin type, while $p = 2$ results in a quad-quad loss function. Specifically, the parameter configuration $\alpha = 0.5$ and $p = 1$ ($\alpha = 0.5$ and $p = 2$) results in the L1 (L2) loss functions.

– Please include Figure 1 about here. –

We use the time-series of (scaled) losses for a benchmark model and the model that features time-varying risk-aversion as a predictor as input for the Diebold-Mariano, which we compute for different parameters α and p. Figure 1 summarizes the results. When we use the core HAR-RV model as our benchmark model, we observe that the model that features time-varying risk aversion produces significantly more accurate forecasts for a wide range of numerical values of the asymmetry parameter. The test results become insignificant only when the asymmetry parameter approaches the upper boundary of its admissable range (especially in case of L1 loss). When we study the first-difference of time-varying risk aversion, the test results are significant for all asymmetry parameters and all three forecast horizons udner L2 loss. Under L2 loss, in contrast, the test results are significant only for the short and long forecast horizon. When we consider the HAR-RV-EPU model as the benchmark model, the test results are significant under L2 loss. The test results under L1 loss are significant for all three forecast horizons for approximately $\alpha < 0.8$. Finally, when we consider the model that features realized skewness and realized kurtosis as our benchmark model, the test results are significant for the short forecast horizon and both L1 and L2 loss. For the long forecsat horizon, the p-values increase when the asymmetry parameter increases and eventually the test results become insignificant. For the medium forecast horizon, in contrast, the p-values are decreasing in the asymmetry parameter. On balance, the test results corroborate that time-varying risk aversion often helps to improve forecast accuracy, and that this result, depending on the model configuration being studied, obtains for a wide range of asymmetry parameter.

5.4 “Bad” and “Good” Realized Volatility

Next, we study the role of time-varying risk aversion in a model of “bad” and “good” realized volatility. 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. Formally, as defined by Barndorff-Nielsen et al. (2010), we compute RV_t^B and RV_t^G as follows:

$$RV_t^B = \sum_{i=1}^{T} r_{t,i}^2 \, I_{[(r_{t,i})<0]}, \tag{7}$$

$$RV_t^G = \sum_{i=1}^{T} r_{t,i}^2 \, I_{[(r_{t,i})>0]}, \tag{8}$$

where $I_{\{\cdot\}}$ denotes the indicator function. Understandably, $RV^S = RV^B + RV^G$. We consider daily RV^B as “bad” realized volatility and RV^G as “good” realized volatility in order to capture the sign asymmetry of the volatility process.

Results of the Diebold-Mariano test for such a HAR-RV model are given in Table 7. Results for “bad” and “good” realized volatility are similar at the short and medium forecast horizons: the test yields significant results for the short and medium forecast horizons. We also observe that time-varying risk aversion improves forecast accuracy for the long forecast horizon in case of “bad” realized volatility. The test results for “good” realized volatility are insignificant. The tests produce p-values of about 11%.

I think we need to add some economic discussion here along the lines suggested by the reviewer. XXXX

5.5 Jumps and Other Extensions

In an application to stock-market volatility forecasting, Patton and Sheppard (2015) find that adding jumps and semi-variance improves the forecasting performance for longer forecast horizons relative to the HAR-RV benchmark. Motivated by this finding, we consider a model that features realized jumps. Andersen et al. (2012) noted that:

$$\lim_{M\to\infty} RV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{j=1}^{N_t} \kappa_{t,j}^2, \tag{9}$$

where N_t is the number of jumps within day t and $\kappa_{t,j}$ is the jump size. Thus RV_t is a consistent estimator of the integrated variance $\int_{t-1}^{t} \sigma^2(s)ds$ plus the jump contribution. Moreover, the results of Barndorff-Nielsen and Shephard (2004) imply that

$$\lim_{M\to\infty} BV_t = \int_{t-1}^{t} \sigma^2(s)ds, \tag{10}$$

where BV_t is the realized bipower variation defined as

$$BV_t = \mu_1^{-1}\left(\frac{N}{M-1}\right)\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}| = \frac{\pi}{2}\sum_{i=2}^{M} |r_{t,i-1}||r_{i,t}|, \tag{11}$$

where $\mu_a = E(|Z|^a)$, $Z \sim N(0,1)$, and $a > 0$. Therefore,

$$J_t = RV_t - BV_t \tag{12}$$

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:

$$JT_t = \frac{RV_t - BV_t}{(\nu_{bb} - \nu_{qq})\frac{1}{N}TP_t} \tag{13}$$

where, $\nu_{bb} = \left(\frac{\pi}{2}\right)^2 + \pi - 3$, $\nu_{qq} = 2$, and TP_t is the Tri-Power Quarticity defined as:

$$TP_t = M\mu_{4/3}^{-3}\left(\frac{M}{M-1}\right)\sum_{i=3}^{M} |r_{t,i-2}|^{4/3}|r_{t,i-}|^{4/3}|r_{t,i}|^{4/3} \tag{14}$$

which converges to

$$TP_t \to \int_{t-1}^{t} \sigma^4(s)ds \tag{15}$$

even in the presence of jumps. Note, for each t, $JT_t \overset{D}{\sim} N(0,1)$ as $M \to \infty$.

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 component as:

$$J_t = \max(RV_t - BV_t, 0) \tag{16}$$

We summarize results for a model that features jumps in Table 8. The results of the Diebold-Mariano test (based on the scaled forecast error) show that the forecasts implied by the HAR-RV-RA model are significantly more accurate than the forecasts implied by the HAR-RV model extended to include jumps for the short and medium forecast horizon. **ARE THE FOLLOWING SENTENCES STILL APPLICABLE??????? However, this result was quite expected. In this study, following Andersen et al. (2012), we use *MRV* as a jump-robust estimator, which is considerably less biased than other measures of realized volatility in the presence of jumps and/or market-microstructure noise. *MRV* is associated with the continuous and persistent part of volatility. Hence, it is not related to the discontinuous and jump part (see Giot et al. 2010).**

– Please include Figure **??** about here. –

As yet another extension, we consider a model that features "bad" and "good" realized volatility as predictors of standard realized volatility. The results (reported in Table 8 along with the results of the other extensions that we consider) are signficant at all three forecast horizons, where the evidence is somewhat weaker for the long forecast horizon. In addition, we consider a model that features gold returns as a predictor. The motivation for including gold returns in the list of predictors is that returns possibly capture the effects of other factors on the gold market that are not already captured by the other predictors in the model. Once again, we observe significant test results. We obtain a variant of the returns model when we consider the possibility that negative returns (that is, leverage) rather than returns per se affect realized volatility (see, e.g., Corsi and Renò 2012). To this end, we use $\min(0, r_t)$ as a predictor. The results are similar to those that we obtain from the model that features returns as a predictor. Overall, our findings show that time-varying risk aversion improves forecast accuracy even when we study several other widely-studied predictors of realized volatility.

6 Concluding Remarks

This paper examines the predictive power of risk aversion over gold-return volatility by utilizing a recently developed measure of time-varying risk aversion, which distinguishes the time variation in economic uncertainty from the time variation in risk aversion. We employ the popular heterogeneous autoregressive realized volatility (HAR-RV) model developed by Corsi (2009) to model and forecast the realized volatility of gold returns as this widely-studied model accounts for several stylized facts such as fat tails and the long-memory property of financial-market volatility, despite the simplicity offered by the model. We extend the HAR-RV model to study the in- and out-of-sample predictive value of risk aversion, after controlling for various alternative predictors including realized skewness, realized kurtosis, realized jumps, gold returns, a leverage term as well as the economic uncertainty as measured by the economic policy uncertainty index.

Our findings suggest that time-varying risk aversion possesses predictive value for gold realized volatility both in- and out-of-sample. While realized skewness is found to be a significant in-sample predictor, we find that time-varying risk aversion has significant in-sample predictive value at a short forecasting horizon. Moreover, time-varying risk aversion dominates the predictive power of EPU when we consider a short forecasting horizon as far as the in-sample results are concerned. Out-of-sample results show that the HAR-RV model that features time-varying risk aversion as a predictor often yields significantly better results in terms of forecast accuracy than reasonable rival models. Additional tests shed light on the role played by the shape of the loss function. Overall, our findings show that time-varying risk aversion captures information useful for predicting realized volatility not already contained in the other predictors, and often allows more accurate out-of-sample forecasts to be computed, where we have documented how the length of the forecast horizon and the shape of a forecaster's loss function affect our results.

Finally, future research may evaluate alternative predictors of gold realized volatility. Following, Zhang and Wei (2010) oil could be used as an alternative predictor. Furthermore, investor sentiment as proxied in Da et al. (2015), could also provide useful insights in predicting return volatility for gold.

Acknowledgments

We thank two anonymous reviewers for helpful and constructive comments. The usual disclaimer applies.

References

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118: 135−167.

Andersen, T.G., Dobrev, D., and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169: 75−93.

Andersen, T., Bollerslev, T., and Diebold, F. (2010). Parametric and nonparametric volatility measurement. *Handbook of Financial Econometrics: Tools and Techniques*, Chapter 2 Volume 1 in Handbooks in Finance: 67−137.

Apergis, N. (2014). Can gold prices forecast the Australian dollar movements? *International Review of Economics and Finance* 29: 75−82.

Balcilar, M., Gupta, R., and Piedzioch, C. (2016). Does uncertainty move the gold Price? New evidence from a nonparametric causality-in-quantiles test. *Resources Policy* 49: 74−80.

Barndorff-Nielsen, O.E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2: 1−37.

Barndorff-Nielsen, O.E. and Shephard, N. (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics* 4: 1−30.

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

Batten, J.A., Ciner, C., and Lucey, B.M. (2010). The macroeconomic determinants of volatility in precious metal markets. *Resoures Policy* 35: 65−71.

Baur, D.G., and Lucey, B.M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review* 45: 217−229.

Bekaert, G., Engstrom, E. C., and Xu, N. R. (2017). The time variation in risk appetite and uncertainty. Columbia Business School Research Paper No. 17-108.

Bekaert, G., and Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics* 183: 181−192.

Bentes, S.I. (2015). Forecasting volatility in gold returns under the GARCH, IGARCH and FIGARCH frameworks: new evidence. *Physica A* 438: 355−364.

Bollerslev, T., and Ghysels, E. (1996). Periodic Autoregressive Conditional Heteroscedasticity. *Journal of Business and Economics Statistics* 14: 139−151.

Bonato, M., R. Demirer, R. Gupta, and Pierdzioch, C. (2018). Gold futures returns and realized skewness: A forecasting experiment using a quantile-boosting approach. *Resources Policy* 57: 196−212.

Bouoiyour, J., Selmi, R., and Wohar, M. E. (2018). Measuring the response of gold prices to uncertainty: An analysis beyond the mean. *Economic Modelling* 75: 05−116.

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

Cashin, P., Caspedes, L. F., and Sahay, R. (2004). Commodity currencies and the real exchange rate. *Journal of Development Economics* 75: 239−268.

Chen, Y. C., and Rogoff, K. 2003. Commodity-currencies. *Journal of International Economics* 60: 133−160.

Chkili, W., Hammoudeh, S., and Nguyen, D.K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics* 41: 1−18.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. Journal of Financial Economics 10, 407−432.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Economics* 7: 174−196.

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

Corsi, F., and Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economics Statistics* 30: 368−380.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *Review of Financial Studies* 28: 1−132.

Demiralay, S., and Ulusoy, V. (2014). Non-linear volatility dynamics and risk management of precious metals. *North American Journal of Economics and Finance* 30: 183−202.

Demirer, R., Omay, T., Yuksel, A., and Yuksel, A. (2018). Global risk aversion and emerging market return comovements. *Economics Letters* 173:118−121.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253−263.

Duong, D., and Swanson, N. R. (2015). Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics* 187: 606−621.

Elliott, G., Komunjer, I., and Timmermann, A. (2005) Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies* 72: 1107−1125.

Elliott, G., Komunjer, I., and Timmermann, A (2008) Biases in macroeconomic forecasts: Irrationality or asymmetric loss? *Journal of the European Economic Association* 6: 122−157.

Erb, C., and Harvey, C. (2013). The Golden Dilemma. *Financial Analysts Journal*, 69: 10−42.

Ewing, B.T., and Malik, F. (2013). Volatility transmission between gold and oil futures under structural breaks. *International Review of Economics and Finance* 25: 113−121.

Fair, R.C., and Shiller, R.J. (1990). Comparing information in forecasts from econometric models. *American Economic Review* 80: 375−389.

Ghysels, E., and Sinko, A. (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics* 160: 257−271.

Giot, P., Laurent, and S., Petitjean, M. (2010). Trading activity, realized bolatility and jumps. *Journal of Empirical Finance*. 17: 168−175.

Gkillas, K., Gupta, R., and Pierdzioch, C. (2018). Forecasting (good and bad) realized exchange-rate variance: Is there a role for realized skewness and kurtosis? Working paper 2018-79. University of Pretoria.

Guiso, L., Sapienza, P., and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics* 128: 403−421.

Hammoudeh, S., and Yuan, Y. (2008). Metal volatility in presence of oil and interest rate shocks. *Energy Economics* 30: 606−620.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13: 281−291.

Haugom, E., Langeland, H., Molnár, P., and Westgaard, S. (2014). Forecasting volatility of the U.S. oil market. *Journal of Banking and Finance* 47: 1−14.

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., and Khandakar, Y (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26: 1−22.

Jones, A.T., and Sackley, W.H., (2016). An uncertain suggestion for gold-pricing models: the effect of economic policy uncertainty on gold prices. *Journal of Economics and Finance* 40: 367−379.

Kang, S. H., McIver, R., and Yoon, S.-M. (2017). Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics* 62: 19−32.

Li J., and Zinna, G. (2018) The variance risk premium: components, term structures, and stock return predictability. *Journal of Business and Economic Statistics*. 36:411?425.

Lyócsa, Ś., and Molnár, P. (2016). Volatility forecasting of strategically linked commodity ETFs: gold-silver. *Quantitative Finance* 16: 1809−1822.

Lucey, B.M., and Li, S. (2015). What precious metals act as safe havens, and when? Some US evidence. *Applied Economics Letters* 22: 35−45.

Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help? *Physica A* 481: 153−159.

Neely, C.j. (2004). Implied colatility from options on gold futures: Do econometric forecasts add value or simply paint the lilly? Working Paper 2003-018C, Federal Reserve Bank of St. Louis.

Patton, A. J., and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97: 683−697.

Poon, S-H., and Granger, C.W.J., (2003). Forecasting volatility in financial markets: a review. *Journal of Economic Literature* 41: 478−539.

Pierdzioch, C., Risse, M., and Rohloff, S. (2016a). A boosting approach to forecasting the volatility of gold-price fluctuations under flexible loss. *Resources Policy* 47: 95−107.

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

Rey, H. (2018). Dilemma not trilemma: The global financial cycle and monetary policy independence. NBER Working Paper: 21162 (Feb. 2018).

Shrestha, K. (2014). Price discovery in energy markets. *Energy Economics* 45: 229?233.

Todorov V, and Tauchen, G. (2011) Volatility jumps. *Journal of Business and Economic Statistics*. 29: 356−371.

Tulley, E., and Lucey, B.M. (2007). A power GARCH examination of the gold market. *Research in International Business and Finance* 21: 316−325.

Xu, N. R. (2017). Global risk aversion and international return comovements. Working paper (Dec. 11, 2017).

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software* 11: 1−17.

Zhang, Y.-J., and Wei, Y.-M. (2010). The crude oil market and the gold market: evidence for cointegration, causality and price discovery. *Resources Policy* 35: 168−177.

Zhou, H., and 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 Finance Journal* 20: 857−880.

Table 1: Summary Statistics

Statistic	MRV	RKU	RSK	EPU	RA
Min	0.01	-10.30	-9.67	3.32	2.23
Mean	0.13	4.00	6.42	96.96	2.76
Median	0.09	0.24	5.05	78.78	2.578
Max	4.38	150.10	382.77	719.07	27.15

Note: MRV was multiplied by the factor 10^3.

Table 2: In-Sample Forecasting Results for Realized Volatility

Panel A: Forecast horizon $h = 1$

results.table	Intercept	MRV	MRV_w	MRV_m	RA	RKU	RSK	EPU	Adj. R2	AIC	BIC
HAR-RV	2.7336***	9.2715***	4.5742***	1.7444*	–	–	–	–	0.6035	0.9996	0.9997
p-value	0.0063	0.0000	0.0000	0.0811	–	–	–	–	–	–	–
HAR-RV-RA	-2.0590**	8.6267***	4.2719***	1.2964	3.6596***	–	–	–	0.6057	0.9999	1
p-value	0.0395	0.0000	0.0000	0.1948	0.0003	–	–	–	–	–	–
HAR-RV-RA-RKU	-1.6237	8.0016***	1.7663*	1.6895*	3.2072***	0.3078	–	–	0.6057	0.9999	0.9999
p-value	0.1044	0.0000	0.0773	0.0911	0.0013	0.7583	–	–	–	–	–
HAR-RV-RA-RSK	-1.5243	8.5823***	4.0156***	1.2672	3.7328***	–	-5.6126***	–	0.6062	1	1
p-value	0.1274	0.0000	0.0001	0.2051	0.0002	–	0.0000	–	–	–	–
HAR-RV-RA-EPU	-1.5850	7.0993***	1.9456*	1.7171*	3.1621***	–	–	0.9636	0.6058	0.9999	0.9999
p-value	0.1130	0.0000	0.0517	0.0860	0.0016	–	–	0.3352	–	–	–
HAR-RV-RA-RKU-RSK-EPU	-1.4379	9.1500***	2.9587***	1.6137	3.2934***	0.1439	-3.1605***	1.0379	0.6063	1	0.9998
p-value	0.1505	0.0000	0.0031	0.1066	0.0010	0.8856	0.0016	0.2993	–	–	–
HAR-RV-d(RA)-RKU-RSK-EPU	1.7511**	9.3209***	2.7413***	2.1558**	2.1433**	0.1588	-3.2881***	1.8597*	0.6069	0.9998	0.9998
p-value	0.0799	0.0000	0.0061	0.0311	0.0321	0.8739	0.0010	0.0629	–	–	–

Panel B: Forecast horizon $h = 5$

results.table	Intercept	MRV	MRV_w	MRV_m	RA	RKU	RSK	EPU	Adj. R2	AIC	BIC
HAR-RV	2.3493**	2.6541***	0.2726	3.9363***	–	–	–	–	0.322	0.9998	1
p-value	0.0188	0.0080	0.7852	0.0001	–	–	–	–	–	–	–
HAR-RV-RA	1.8262*	2.7786***	0.2921	3.1633***	0.1034	–	–	–	0.3219	0.9998	0.9999
p-value	0.0678	0.0055	0.7702	0.0016	0.9176	–	–	–	–	–	–
HAR-RV-RA-RKU	1.7771*	2.9458***	0.2578	2.5582**	0.1370	0.7081	–	–	0.322	0.9998	0.9998
p-value	0.0756	0.0032	0.7966	0.0105	0.8910	0.4789	–	–	–	–	–
HAR-RV-RA-RSK	2.2179**	2.8001***	0.2670	3.0674***	0.0793	–	-3.2953***	–	0.3235	0.9999	0.9999
p-value	0.0266	0.0051	0.7895	0.0022	0.9368	–	0.0010	–	–	–	–
HAR-RV-RA-EPU	1.4574	2.9538***	0.2389	2.3733**	-0.1297	–	–	1.5759	0.3231	0.9999	0.9999
p-RA	0.1450	0.0031	0.8112	0.0176	0.8968	–	–	0.1150	–	–	–
HAR-RV-RA-RKU-RSK-EPU	1.9403*	2.9661***	0.2300	2.5274**	-0.2053	0.3403	-4.1012***	1.6711*	0.3247	1	0.9998
p-value	0.0523	0.0030	0.8181	0.0115	0.8373	0.7336	0.0000	0.0947	–	–	–
HAR-RV-d(RA)-RKU-RSK-EPU	2.4421**	2.9386***	0.2408	2.4947**	0.7668	0.3362	-4.0188***	1.6109	0.3248	0.9998	0.9998
p-value	0.0146	0.0033	0.8097	0.0126	0.4432	0.7367	0.0001	0.1072	–	–	–

Panel C: Forecast horizon $h = 22$

results.table	Intercept	MRV	MRV_w	MRV_m	RA	RKU	RSK	EPU	Adj. R2	AIC	BIC
HAR-RV	4.3306***	3.1522***	0.6856	2.0230**	–	–	–	–	0.147	0.9996	0.9999
p-value	0.0000	0.0016	0.4929	0.0431	–	–	–	–	–	–	–
HAR-RV-RA	4.2436***	2.6961***	0.6905	2.6920***	-0.7395	–	–	–	0.1485	0.9997	0.9999
p-value	0.0000	0.0070	0.4899	0.0071	0.4596	–	–	–	–	–	–
HAR-RV-RA-RKU	4.6980***	2.7494***	0.6009	2.0504**	-0.8485	0.6073	–	–	0.1483	0.9997	0.9998
p-value	0.0000	0.0060	0.5479	0.0403	0.3962	0.5437	–	–	–	–	–
HAR-RV-RA-RSK	4.7869***	2.7857***	0.6581	2.3823**	-0.8387	–	-2.9262***	–	0.1509	0.9999	1
p-value	0.0000	0.0053	0.5105	0.0172	0.4016	–	0.0034	–	–	–	–
HAR-RV-RA-EPU	4.1327***	2.7119***	0.6964	2.9165***	-0.9520	–	–	1.1907	0.1503	0.9998	0.9999
p-value	0.0000	0.0067	0.4861	0.0035	0.3411	–	–	0.2338	–	–	–
HAR-RV-RA-RKU-RSK-EPU	4.4778***	2.7918***	0.6672	2.8171***	-1.0249	-0.2295	-2.9419***	1.2567	0.1528	1	0.9999
p-value	0.0000	0.0052	0.5047	0.0048	0.3054	0.8185	0.0033	0.2089	–	–	–
HAR-RV-d(RA)-RKU-RSK-EPU	3.1168***	3.1482***	0.6948	2.2013**	0.3969	-0.2611	-2.704***	0.9140	0.1501	0.9989	0.9997
p-value	0.0018	0.0016	0.4872	0.0277	0.6914	0.7940	0.0069	0.3607	–	–	–

Note: *** (**, *) denotes significance at the 1% (5%, 10%) level. p-values are based on Newey-West robust standard errors. Estimated coefficients are scaled by their estimated standard error. Adj. R2 = adjusted coefficient of determination. d(RA) = First-differenced risk aversion. AIC = Akaike information criterion. BIC = Bayesian information criterion. The results for the AIC (BIC) are expressed relative to the AIC (BIC) of the best model.

Table 3: Out-of-Sample Forecast Performance (Baseline Results)

Models	Loss	$h=1$	$h=5$	$h=22$
HAR-RV vs. HAR-RV-RA	$L1$	0.0000	0.0000	0.0346
HAR-RV vs. HAR-RV-RA	$L2$	0.0000	0.0000	0.0097
HAR-RV-EPU vs. HAR-RV-RA	$L1$	0.0000	0.1819	0.4799
HAR-RV-EPU vs. HAR-RV-RA	$L2$	0.0000	0.1310	0.2816
HAR-RV-RSK-RKU vs. HAR-RV-RA	$L1$	0.0000	0.0041	0.0063
HAR-RV-RSK-RKU vs. HAR-RV-RA	$L2$	0.0184	0.2233	0.3036

Note: p-values of Diebold-Mariano tests for alternative rolling-window lengths and three different forecast horizons. Null hypothesis: the series of forecasts from the two models are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features time-varying risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the ratio of cumulated absolute (squared) forecast errors for the HAR-RV model that features time-varying risk aversion and the HAR-RV model that does not feature time-varying risk aversion were computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window.

Table 4: Estimates of the Out-of-Sample R^{L1} and R^{L2} Statistics

Models	Loss	$h=1$	$h=5$	$h=22$
HAR-RV vs. HAR-RV-RA	$L1$	0.0324	0.0258	0.0199
HAR-RV vs. HAR-RV-RA	$L2$	0.0826	0.0398	0.0365
HAR-RV-EPU vs. HAR-RV-RA	$L1$	0.0620	0.0094	0.0009
HAR-RV-EPU vs. HAR-RV-RA	$L2$	0.1430	0.0331	0.0225
HAR-RV-RSK-RKU vs. HAR-RV-RA	$L1$	0.0496	0.0171	0.0294
HAR-RV-RSK-RKU vs. HAR-RV-RA	$L2$	0.1984	0.0378	0.0217

Note: The out-of-sample is defined as $R^{L2} = 1 - \sum_t fe^2_{t,RA} / \sum_t fe^2_{t,B}$ for L2 loss and as $R^{L1} = 1 - \sum_t |fe_{t,RA}| / \sum_t | fe_{t,B}|$ for L1 loss, where t indicates the out-of-sample periods, $fe_{t,RA}$ denotes the (scaled) forecast error. A positive value of the statistic indicates that the HAR-RV-RA model outperforms the benchmark (B) model. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the R^{L2} and R^1 statistics were computed for every length of the rolling-estimation window. The results reported in this table are the corresponding maximum values of the R^{L2} and R^1 statistics.

Table 5: Out-of-Sample Forecast Performance (First Difference of Risk Aversion)

Models	Loss	$h=1$	$h=5$	$h=22$
HAR-RV vs. HAR-RV-d(RA)	$L1$	0.0382	0.6356	0.8593
HAR-RV vs. HAR-RV-d(RA)	$L2$	0.0001	0.1093	0.1058
HAR-RV-EPU vs. HAR-RV-d(RA)	$L1$	0.0000	0.0000	0.0102
HAR-RV-EPU vs. HAR-RV-d(RA)	$L2$	0.0000	0.0005	0.0132
HAR-RV-RSK-RKU vs. HAR-RV-d(RA)	$L1$	0.0009	0.9352	0.9213
HAR-RV-RSK-RKU vs. HAR-RV-d(RA)	$L2$	0.0080	0.4831	0.8351

Note: p-values of Diebold-Mariano tests for alternative rolling-window lengths and three different forecast horizons. Null hypothesis: the series of forecasts from the two models are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features the first-difference of time-varying risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the ratio of cumulated absolute (squared) forecast errors for the HAR-RV model that features the first difference of time-varying risk aversion and the HAR-RV model that does not feature the first difference of time-varying risk aversion were computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window.

Table 6: Results of Fair-Shiller Regressions

Models	$h=1$	$h=5$	$h=22$
HAR-RV vs. HAR-RV-RA	0.4584	0.1635	0.0014
HAR-RV vs. HAR-RV-d(RA)	0.0565	0.3052	0.3511
HAR-RV-EPU vs. HAR-RV-RA	0.7761	0.0435	0.0008
HAR-RV-RSK-RKU vs. HAR-RV-RA	0.0000	0.7694	0.8088

Note: p-values of Fair-Shiller regressions for three different forecast horizons. The Fair-Shiller regressions were set up by regressing actual realized volatility on a constant, the forecast for the benchmark model, and the forecast of the model that features time-varying risk aversion as a predictor. This table depicts results for the coefficient of the model that features time-varying risk aversion. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of the squared out-of-sample forecast errors was computed for the series of forecasts from the two models. Then the ratio of cumulated squared forecast errors for the HAR-RV model that features the first difference of time-varying risk aversion and the HAR-RV model that does not feature the first difference of time-varying risk aversion was computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window.

Table 7: Out-of-Sample Forecast Performance for Bad and Good Realized Volatility

Models	Loss	$h=1$	$h=5$	$h=22$
HAR-RVB vs. HAR-RVB-RA	$L1$	0.0000	0.0000	0.0794
HAR-RVB vs. HAR-RVB-RA	$L2$	0.0000	0.0000	0.0914
HAR-RVG vs. HAR-RVG-RA	$L1$	0.0000	0.0000	0.1178
HAR-RVG vs. HAR-RVG-RA	$L2$	0.0000	0.0000	0.1125

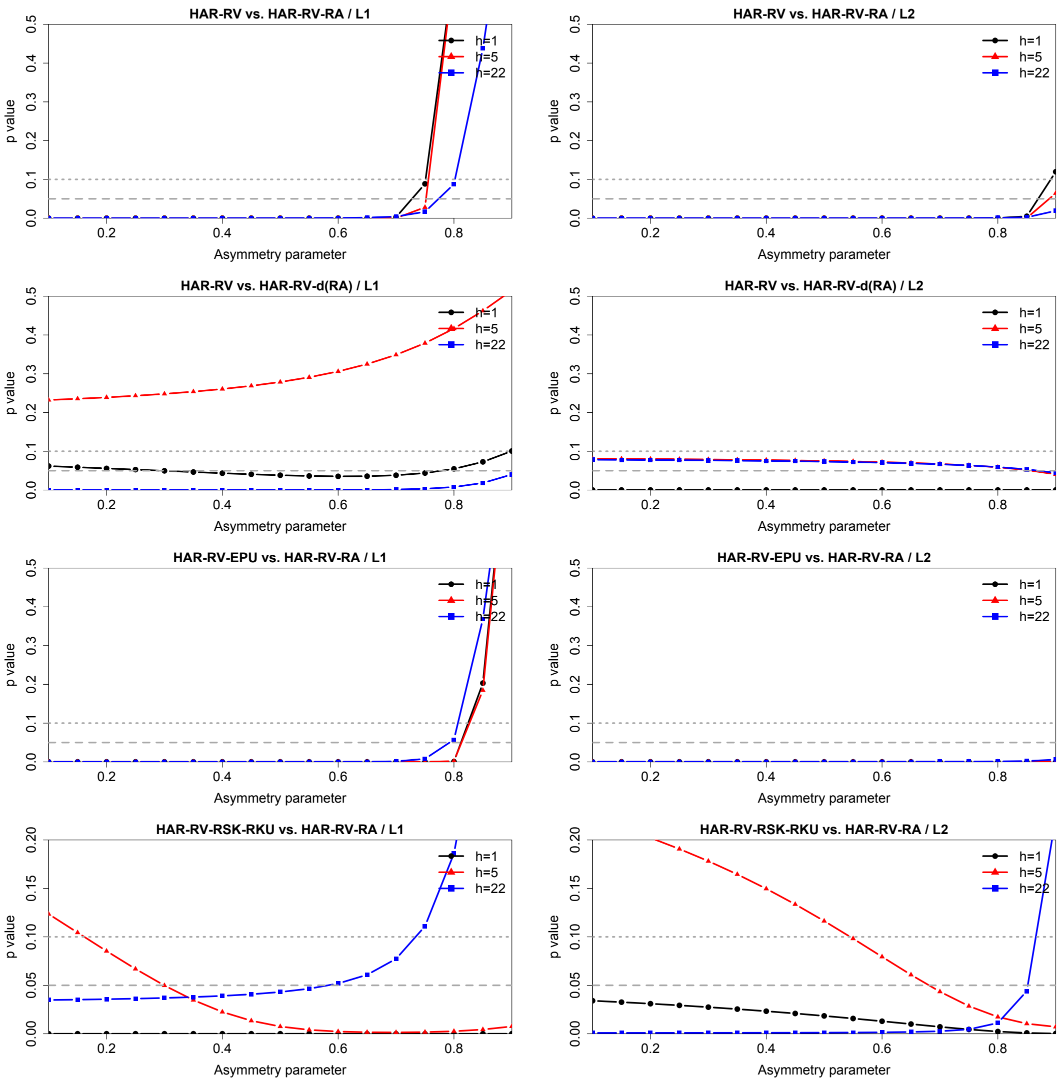
Note: p-values of Diebold-Mariano tests for alternative rolling-window lengths and three different forecast horizons. Null hypothesis: the series of forecasts from the two models are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features the first-difference of time-varying risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the ratio of cumulated absolute (squared) forecast errors for the HAR-RV model that features the first difference of time-varying risk aversion and the HAR-RV model that does not feature the first difference of time-varying risk aversion were computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window.

Table 8: Out-of-Sample Forecast Performance Using Alternative Predictors

Models	Loss	$h = 1$	$h = 5$	$h = 22$
HAR-RV-JUMP vs. HAR-RV-RA	$L1$	0.0000	0.0000	0.1054
HAR-RV-JUMP vs. HAR-RV-RA	$L2$	0.0000	0.0000	0.0465
HAR-RV-RVB-RVG vs. HAR-RV-RA	$L1$	0.0000	0.0000	0.0211
HAR-RV-RVB-RVG vs. HAR-RV-RA	$L2$	0.0006	0.0000	0.0885
HAR-RV-RETURNS vs. HAR-RV-RA	$L1$	0.0000	0.0000	0.0233
HAR-RV-RETURNS vs. HAR-RV-RA	$L2$	0.0000	0.0000	0.0026
HAR-RV-LEVERAGE vs. HAR-RV-RA	$L1$	0.0000	0.0000	0.0196
HAR-RV-LEVERAGE vs. HAR-RV-RA	$L2$	0.0000	0.0000	0.0013

Note: p-values of Diebold-Mariano tests for alternative rolling-window lengths and three different forecast horizons. Null hypothesis: the series of forecasts from the two models are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features the first-difference of time-varying risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the ratio of cumulated absolute (squared) forecast errors for the HAR-RV model that features the first difference of time-varying risk aversion and the HAR-RV model that does not feature the first difference of time-varying risk aversion were computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window.

Figure 1: Diebold-Mariano-Test (Asymmetric Loss)



Note: p-values of Diebold-Mariano tests for alternative rolling-window lengths and three different forecast horizons. Null hypothesis: the series of forecasts from the two models are equally accurate. Alternative hypothesis: the forecasts from the HAR-RV model that features the first-difference of time-varying risk aversion are more accurate. L1: absolute loss. L2: quadratic loss. The models were estimated using rolling-estimation windows. The length of the rolling-estimation windows ranged from 1000, 1100,..., 3000. For every length of the rolling-estimation window the sum of absolute (L1 loss) and squared (L2 loss) out-of-sample forecast errors were computed for the series of forecasts from the two models. Then the ratio of cumulated absolute (squared) forecast errors for the HAR-RV model that features the first difference of time-varying risk aversion and the HAR-RV model that does not feature the first difference of time-varying risk aversion were computed. The minimum of this ratio was used to select the optimal length of the rolling-estimation window. The p-values were computed based on the series of forecasts derived for the optimal length of the rolling-estimation window. The horizontal lines depict the 10% and 5% levels of significance. Horizontal axis: Asymmetry parameter of the loss function.