Forecasting realized volatility of bitcoin returns: Tail events and asymmetric loss

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

This version: November 2020

Abstract

We use intraday data to construct measures of the realized volatility of bitcoin returns. We then construct measures that focus exclusively on relatively large realizations of returns to assess the tail shape of the return distribution, and use the heterogeneous autoregressive realized volatility (HAR-RV) model to study whether these measures help to forecast subsequent realized volatility. We find that mainly forecasters suffering a higher loss in case of an underprediction of realized volatility (than in case of an overprediction of the same absolute size) benefit from using the tail measures as predictors of realized volatility, especially at a short and intermediate forecast horizon. This result is robust controlling for jumps and realized skewness and kurtosis, and it also applies to downside (bad) and upside (good) realized volatility.

JEL classification: C22; F31; F37

Keywords: Bitcoin; Realized volatility; Forecasting; Tail events

[a] Corresponding author. Department of Management Science and Technology, University of Patras, Megalou Aleksandrou 1, Koukouli, 26334 Patras, Greece; Email address: gillask@upatras.gr.

[b] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa; Email address: rangan.gupta@up.ac.za.

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

1 Introduction

Bitcoin can be defined as "an online communication protocol that facilitates the use of a virtual currency, including electronic payment" (Bohme, et al., 2015). It was designed by Nakamoto (2008) as the first decentralized cryptocurrency based on blockchain technology, and until now remains the most popular among the 4,000 alternative altcoins (i.e., cryptocurrencies that have been launched since the introduction of bitcoin) currently available. Despite being anonymous and reducing transaction costs (Kim, 2017), bitcoin is believed to be mostly used for speculative purposes (Baek and Elbeck, 2015), resulting in extreme volatility and creating bubbles followed by market crashes (Bouri et al., 2018). As such, normality assumptions are untenable as bitcoin seems to have heavy tails (Osterrieder and Lorenz, 2017; Gkillas and Katsiampa, 2018; Gkillas and Longin, 2018). Meanwhile, there is evidence supporting the view that the inefficiency of the bitcoin market is quite strong (Urquhart, 2016). Taken as a whole, the above issues gives rise to many difficulties when attempting to model and predict the evolution of the risk of bitcoin investments. Understandably, from the perspective of an investor, it is of utmost importance to forecast the risk related to an investment in bitcoin, that is, to predict its volatility.

Over the past few years, a large body of literature has emerged that aims to predict (in- and out-of-sample) daily movements in the volatility of cryptocurrencies, and in particular of bitcoin. Most studies are based on same-frequency or mixed-frequency variants of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, using various types of financial and macroeconomic variables as predictors (see, e.g., Chu et al., 2017; Ardia et al., 2018; Conrad et al., 2018 and Kochling et al., 2019), while a few attempts have been conducted in the framework of realized volatility (see, Hattori, 2019, 2020). GARCH-type of models rely heavily on the underlying model assumptions. Volatility, however, is a latent variable and, thus, requires non-parametric estimation techniques. Any non-parametric estimator of volatility, such as the realized volatility, is based on quadratic variation, which is regarded as the best estimator of integrated (latent) volatility. The realized volatility captures the dynamics of two key elements of volatility, namely its continuous sample-path part and its jump part. A benchmark measure of realized volatility is the realized variance.[1] The measurement of daily realized volatility by means of realized variance has been studied by Andersen et al. (2007), among others. Under weak regularity conditions, realized variance is a consistent estimator of quadratic variation. Hansen and Huang (2016) show that realized variance is more accurately estimated at a daily frequency by employing high-frequency intraday data. High-frequency data reveal important information about intraday changes and the microstructure of the market not easily seen on a daily basis. As noted by Bekaert and Hoerova (2014), models based on high-frequency information dominate standard GARCH-class models (see, also Chen and Ghysels, 2012), as high-frequency models highlight the importance of persistence (by employing lagged realized variances as predictors), capture additional information contained in the most recent return variances (Corsi, 2009), shed light on the asymmetry between positive and negative returns shocks (the classic volatility asymmetry, Engle and Ng, 1993), and bring to the forefront potentially differing predictive information present in jumps and continuous volatility components (Andersen et al., 2007). Not surprisingly, there is a growing interest to incorporate high-frequency data into models for lower-frequency data. As for the modeling of the conditional second moment of the asset-return distribution, examples include Shephard and Sheppard (2010), Noureldin et al. (2012), Hansen et al. (2012), and Hansen et al. (2014), among others. Other studies attempt

[1] In this research, we use the term realized volatility to refer to the realized variance rather than the realized standard deviation of bitcoin returns. In Section 4.5, we shall present results for the the realized standard deviation of bitcoin returns.

to incorporate high-frequency data into models for the time-varying dependence in a copula framework (see, De Lira Salvatierra and Patton, 2015; Oh and Patton, 2016), while Bee et al. (2019) suggest modeling the tails of the conditional asset-return distributions with a class of extreme-value-theory models incorporating high-frequency information.

As mentioned in the previous paragraph, most attempts focus on the first and second moments of the asset-returns distribution by exploiting information contained in lagged realized variances and past daily returns. Higher-order moments, however, may add significant explanatory power because they are associated with events far away from the center of the distribution, which is consistent with empirical evidence that points towards the existence of market frictions causing phenomena such as jumps, inducing heaviness in the tails of the asset-returns distribution. For example, jumps present dynamics that are often found to be related to the level of volatility (see, Bandi and Reno, 2012; Bollerslev and Todorov, 2011). Motivated by these ideas, we find it promising to link the dynamics of large intraday price fluctuations to that of the daily realized variance. We accommodate all of these features in our study, yet we take a completely different perspective grounded on the results of Dacorogna et al. (2001) and Amaya et al. (2015). More specifically, our study is motivated by the research of Dacorogna et al. (2001) who note that it is possible to use intraday returns to construct measures related to the daily distribution of intraday returns beyond standard higher-order moments. From a methodological perspective, we construct measures focusing exclusively on relatively large intraday returns observations in order to assess the tail shape of the returns distribution based on the Hill tail-index estimator. This estimator is one of the most commonly used non-parametric estimators and first appears in Hill (1975). We name our measures realized downside and upside tail-index estimators, as we construct them exclusively from intraday data. We consider these estimators as measures of the weight of the tails of the daily the bitcoin-returns distribution. In doing so, we are able to study the tail behavior of bitcoin returns by taking into consideration only the relevant information on extremes. We can then examine whether these two measures help the accuracy of forecasts of the realized volatility of bitcoin returns. In other words, we link the insights derived from an extensive strand of econometric literature on modeling and forecasting volatility with one of the most appropriate statistical tools, i.e., extreme-value theory, for modeling extremes. Over the years, extreme-value theory has proven its practical value in many fields such as earth sciences, engineering, and finance (de Haan and Ferreira, 2006). Danielsson and de Vries (1997), Longin (2000), and Longin and Solnick (2001) offered some early applications of extreme-value theory in finance. Other more recent examples include Bhatti and Nguyen (2012), Nguyen and Bhatti (2012), Nguyen et al. (2016), Rahahleh and Bhatti (2017), Nguyen et al. (2017), Mansor et al. (2019), among others. Following Amaya et al. (2015), we also control for realized skewness and kurtosis as predictors of realized volatility of bitcoin returns, and we consider its bad (downside) and good (upside) components. In this regard, it must be pointed out that the motivation to look at the role of realized skewness and realized kurtosis as predictors for forecasting the volatility of bitcoin returns originates from the enormous literature that hypothesizes that heavy-tailed shocks alone with left-tail events play an important role in explaining asset-price behavior (on realized skewness and kurtosis, see Mei et al., 2017; see also Neuberger and Payne, 2018).

Another contribution of our research is that we use an asymmetric loss function to assess the predictive value of our two measures of tail behavior. Given that we study the predictive value of downside and upside measures of tail behavior, it is a natural research strategy in our setting to consider the possibility that a forecaster may have an asymmetric loss function. A forecaster has an asymmetric loss function when the loss of an overprediction of realized volatility differs from the loss of an underprediction of the same absolute size. We consider the asymmetric loss function which was also studied by Elliott et al. (2005, 2008), among others. This loss function has the advantage that it nests the symmetric quadratic and absolute loss functions widely studied in the forecasting literature. An asymmetric loss function is a natural modeling choice when one uses a utility-function-based approach to evaluate forecasts (West et al., 1993) and also in a risk-management context when a forecaster uses forecasts of realized volatility to implement certain option-trading strategies.

For our empirical analysis, we use an extensive sample based on (15-minute interval) intraday data, covering the period from 26th January 2015 to 11th November 2018. In order to forecast the daily realized volatility of bitcoin returns and, in particular, to evaluate the out-of-sample predictive power of the realized measures of tail behavior proposed in this study, we proceed as follows: First, we use the heterogeneous autoregressive realized volatility (HAR-RV) model proposed by Corsi (2009) as our benchmark model. Second, we study whether the downside and upside measures of tail behavior add explanatory power to the baseline HAR-RV model. Third, we study whether the predictive value of the these measures remains intact when we control for realized skewness and realized kurtosis. We show that our measures for the shape of the tails constructed by intraday data hold information about future volatility. In particular, our main finding is that mainly forecasters who suffer a higher loss in case of an underprediction of realized volatility than in case of an overprediction of the same absolute size benefit from using downside and upside tail-index estimators as predictors of realized volatility, especially at a short and an intermediate forecast horizon. Furthermore, this result is robust across several proxies of the daily variance and it is not affected by the inclusion of jumps (on the role of jumps, see, Andersen et al., 2007). We also consider "bad" and "good" realized volatility. Our findings show that forecasters who suffer a relatively higher loss in case of an overprediction of realized volatility benefit from using the tail indices as predictors only in few model configurations, and mainly if the length of the rolling window used for estimation is relatively short.

The remainder of the paper is organized as follows: In section 2, we present the theoretical considerations underlying our empirical research. In Section 3, we describe the methods that we use in our empirical research. In Section 4, we describe our data and summarize our empirical results. Section 5 concludes the paper.

2 Theoretical considerations

2.1 Tail index: A general framework

Let $X = \{X_1, X_2, ..., X_n\}$ be a sequence of independent and identically distributed random variables with a common continuous cumulative distribution function, F_X, for $j = 1, 2, ..., n$, defined on some probability space (Ω, F, P). Suppose that F_X is a heavy-tailed distribution, and let $\xi > 0$ be a parameter such that $P(X > x) = x^{-1/\xi} L(x)$, for $x \geq 0$, and $L(x)$ refers to a slowly varying function.[2] The parameter ξ is known as the tail index of the distribution. We can define the tail index as a measure of the weight of the tails of the distribution. A commonly used estimator of the parameter ξ is the Hill estimator (Hill, 1975). The Hill estimator for some k_n is defined by

$$\xi^H = \frac{1}{k_n} \sum_{j=n-k_n+1}^{n} \log \frac{X_{(j)}}{X_{(n-k_n+1)}}, \tag{1}$$

where $X_{(1)} \leq ... \leq X_{k_{(n)}} \leq ... \leq X_{(n)}$, considering only the k_n largest order statistics of the sequence $\{X_1, X_2, ..., X_n\}$. The Hill estimator can be interpreted as the average vertical exceeds of the log-transformed data above a given threshold defined

[2]With the term heavy-tailed, we refer to those distributions that do not have all their moments finite.

by a sample fraction k_n, where $k_n \ll n$. Note that, if $k_n \to \infty$ and $k_n/n \to 0$, as $n \to \infty$ then $\xi^H \to \xi$, and $\sqrt{k_n}(\xi^H - \xi) \to N(0, \xi^2)$.

2.2 Tail index: An intraday framework

Let $X_{j,i} = \{X_{1,i}, X_{2,i}, ..., X_{n,i}\}$ again be a sequence of independent and identically distributed random variables with a common continuous cumulative distribution function, F_{X_i}, for $j = 1, 2, ..., n$ and $i = 1, 2, ..., N$. Suppose also that F_{X_i} is a distribution that satisfies $P(X_i > x_i) = x_i^{-1/\xi_i} L(X_i)$, for $x_i \geq 0$, and $L(x_i)$ is a slowly varying function. For $\xi_i > 0$, we can also use the Hill estimator to estimate ξ_i. In this case, the Hill estimator for some $k_{i,N}$ is defined by

$$\xi_i^H = \frac{1}{k_{i,N}} \sum_{i=N-k_{i,N}+1}^{N} \log \frac{X_{(i)}}{X_{(N-k_{i,N}+1)}} \tag{2}$$

where $X_{(1,i)} \leq ... \leq X_{k_{(i,N)}} \leq ... \leq X_{(n,i)}$, considering only the $k_{i,N}$ largest order statistics of the sequence $X_i = \{X_{1,i}, X_{2,i}, ..., X_{n,i}\}$. The Hill estimator applied to intraday data can be interpreted as the average vertical exceeds of the log-transformed data above a given threshold defined by a sample fraction $k_{i,N}$, where $k_{i,N} \ll N$. As in the general framework, we have $k_{i,N} \to \infty$ and $k_{i,N}/N \to 0$, as $N \to \infty$, $\xi_i^H \to \xi_i$, and $\sqrt{k_{i,N}}(\xi_i^H - \xi_i) \to N(0, \xi_i^2)$.

3 Methods

3.1 Realized volatility

Before we present our realized volatility estimator, we first recall some basic definitions and notation. Time is discrete and measured in trading days, t. Within each day, there are $N+1$ intraday prices or N intraday returns. Within any day t, the observed prices concern these intraday time periods: $t_0 < t_1 < ... < t_{N+1}$.

Following the research by Diebold et al. (1999) and Andersen et al. (2001), intraday returns are retrieved as the logarithmic difference (log-returns) between two consecutive intraday observed prices.

Turning now our attention to the estimator being used, while there are many available realized volatility estimators, our main focus centers on the median realized variance, MRV, proposed by Andersen et al. (2012) as a jump-robust estimator of the realized variance. We primarily use the MRV as an estimator of daily realized volatility RV for bitcoin returns. For the implementation of our forecasting exercise, this measure exhibits some specific time-series properties that render it possible to compare and contrast the behavior of the suggested tail-index estimators and, more importantly, how they impact future realized volatility. In particular, MRV attenuates the effect of market-microstructure noise and exhibits better finite sample properties than other realized measures (see, Ghysels and Sinko, 2011 for further discussion of volatility forecasting in the presence of market-microstructure noise). It is also well recognized that MRV constructed from intraday data is not influenced by the sampling frequency as compared with other jump-robust estimators. Meanwhile, it displays better finite-sample robustness to the occurrence of “zero” returns in the sample.

Last but not least, as a jump-robust RV estimator, MRV is substantially less biased than other realized measures in the presence of jumps. Although jumps present dynamics that are often found to be related to the level of volatility, they are extremely unpredictable and hard to anticipate (see, e.g., Bollerslev et al., 2008). In parallel to the reasons discussed above, a less obvious but rather important gain from using MRV, therefore, is that this estimator enables us to shed light on aspects of the dynamics of bitcoin returns that reflect other elements of the data-generating process apart from price discontinuities -and that are, thus, not caused by jumps- such as the asymmetry in the tails of the daily bitcoin-returns distribution. Intuitively, by keeping the forecasting model parsimonious, we are able to separate the predictive information of jumps, and are, thus, in a position to focus on the predictive power of the proposed tail-index estimators. In other words, separating jumps from RV helps us to better understand the tail behavior of the bitcoin-returns distribution and the role played by the tail behavior for modeling and forecasting an extremely volatile variable. To put it differently, while jumps are associated "indirectly" with phenomena that may induce heaviness in the tails of the bitcoin-returns distribution, we model the tail behavior of this distribution in a more direct way by resorting to specific statistical tools that are tailored for such an analysis. In the case where our tail measures provide extra predictive value in a model that also features jumps as a predictor of realized volatility, we are capturing something "more complicated" associated with, for example, the clustering of extremes and the shape of the (upper and lower) tails of the bitcoins-returns distribution.

In order to further validate the importance of the new estimators, however, we also study their predictive value along with jumps. Jumps are associated with the discontinuo us part of the price process and can explicitly improve the overall fit of volatility models (see, e.g., Duffie et al., 2000; Andersen et al., 2007). Also, as we already noted, the empirical literature points toward the existence of market frictions causing phenomena such as jumps, inducing heaviness in the tails of asset-returns distributions (see, Bandi and Reno, 2012; Bollerslev and Todorov, 2011). Against the background of the results documented in this literature, the result that the predictive information of our tail measures for the subsequent realized volatility of bitcoin returns is not affected by the inclusion of jumps in the forecasting model provides further insights that potentially complicated tail-event-dynamics have predictive value for realized volatility.

A daily point estimate of MRV is given by

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

where $r_{t,i}$ is the intraday return i within day t and $i = 1,..,N$ is the total number of intraday observations within a day. We consider MRV as our measure of daily RV.

3.2 Realized downside and upside tail indices

We present now the estimators of the realized downside and upside tail indices of the bitcoin-returns distribution. However, before doing that it is instructive to again consider the general theoretical framework that our idea is based on.

The increasing availability of intraday data has prompted researchers to study realized measures of the daily distribution of the returns of asset prices. Realized measures are non-parametric estimators mainly using high-frequency intraday data to model conditional moments of the asset-returns distribution. Further information on realized measures can be found in the survey articles by Barndorff-Nielsen and Shephard (2007) and Andersen et al. (2009). The most well-known example of such measures is realized volatility (Andersen and Bollerslev, 1998; Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2001). A benchmark and widely used estimator for realized volatility is the realized variance as it is a consistent estimate of actual volatility based on the theory of quadratic variation. Realized variance estimates the second moment of the daily asset-return distribution and characterizes the dispersion around the mean or the dispersion risk. As such, it has unobservable behavior, and events during a trading day can affect its behavior in different ways.

Moreover, as we already noted, the empirical literature points toward the existence of market frictions causing "extreme" price changes inducing heaviness in the tails of the asset-returns distribution. As discussed by Dacorogna et al. (2001), it is possible to use intraday data to construct additional summary measures of the daily asset-return distribution. Indeed, Amaya et al. (2015) construct a measure of ex-post realized daily skewness based on intraday returns to capture with more precision statistical forces that affect the variations of the price path. They also construct daily realized kurtosis as a measure that characterizes the flatness of the distribution as it takes into account the existence of extremes in the distribution of returns. While the latter measures in one way or another captures the higher than normal daily kurtosis, no consensus has emerged in the literature as to the exact amount of probability mass in the tails of daily asset-returns distributions and, thus, about the most suitable tool to use in modeling large intraday returns. We approach this issue in a novel way by explicitly focusing on the tails of the daily bitcoin-returns distribution. Further, we propose a framework for non-parametrically estimating the shape of the tails of the daily bitcoin-returns distribution. As the tail shape of the distribution is essential for determining the frequency of large returns, a precise estimation of the tail shape of the bitcoin-returns distribution is crucial for proper risk assessment during extremely volatile periods (see Jansen and de Vries, 1991; Danielsson and de Vries, 1997).

Our framework is based on the idea that aggregation of multiple jump events over a fixed time interval (e.g., a trading day) will result in heavy-tailed asset-returns distribution (see Bollerslev et al., 2013). In light of this, we construct two measures to analyze the tail shape of the daily bitcoin-returns distribution by building on the Hill tail-index estimator presented in Section 2. This estimator is a sufficient statistic for the tail shape of the asset-return distribution, while it does not depend on a particular probability model. Going one step further, another advantage of our measures is that they render it possible, in the context of our forecasting analysis, to distinguish days with high levels of volatility that can result in seemingly large price fluctuations, even though the returns are drawn from a light-tailed distribution, yet with large variance, from volatile days where the frequency of market events can induce heaviness in the tails of the bitcoin-returns distribution. To put it differently, it is interesting to investigate whether our measures are useful predictors of realized volatility and/or whether the frequency of events in the tails affects future volatility. In what follows, we present the estimation framework of our estimators.

Within a non-parametric framework, as our estimators are based on intraday data to model the shape of the distribution tails, we refer to them as realized estimators. The estimators are denoted by TN_t and TP_t, where TN is defined as the downside tail-index estimator, while TP as the upside tail-index. The former is referring to negative intraday returns and the left tail, and the latter to positive intraday returns and the right tail. A daily point estimate for TN_t and TP_t may given by

$$TN_t = \sqrt{\frac{1}{p^-}} \log \left(\frac{1}{k^-_{t,N}} \sum_{i=N-k^-_{t,N}+1}^{N} \log \frac{r_{(t,i)^-}}{r_{(N-k^-_{t,N}+1)^-}} \right), \tag{4}$$

$$TP_t = \sqrt{\frac{1}{p^+}} \log \left(\frac{1}{k^+_{t,N}} \sum_{i=N-k^+_{t,N}+1}^{N} \log \frac{r_{(t,i)^+}}{r_{(N-k^+_{t,N}+1)^+}} \right), \tag{5}$$

where $r_{(t,i)^-}$ is the ordered intraday return, i, within day t for $i = 1,..,N$, for the sequence $r_{(t,1)^-} \leq ... \leq r_{(t,k^-_{t,N})^-} \leq ... \leq r_{(t,N)^-}$. Similarly, $r^+_{(t,i)}$ is the ordered intraday return, i, within day t, $i = 1,..,N$, for the sequence $r_{(t,1)^+} \leq ... \leq r_{(t,k^+_{t,N})^+} \leq ... \leq r_{(t,N)^+}$. N denotes the total number of intraday observations within a day. Moreover, p refers to the tail probability obtained from $k_{t,N}/N$, while $k_{t,N}$ is the number of upper order statistics or the sample fraction to be included in the estimation out of a sample of size N. For our empirical analysis, we select for $k^-_{t,N}$ ($k^+_{t,N}$) the 15% lowest (highest) observations of the sample of size N.[3]

[3]In order to find the appropriate sample fractions $k_{t,N}$, we apply a failure-to-reject method based on the idea of Choulakian and Stephens (2012). As such, we conduct a sensitivity analysis by selecting a number of possible sample fraction values. We then we select the lowest sample fraction for which the average value of the daily realized estimates (considering all the available trading days in the sample) stabilizes. Very low sample fraction values are not selected to avoid small sample errors which can give a false sense of accuracy. In any case, results suggest that the exact choice of the optimal sample fraction $k_{t,N}$ does not matter much.

In addition, we use two modifications to improve the efficiency of estimators introduced in Section 2.2. First, the scaling by p^- and p^+ (for TN_t and TP_t ensures that the magnitudes of the estimators correspond to daily estimates. Second, the logarithmic transformation gives better finite sample properties.

3.3 Realized downside and upside volatility

We compute RV^- and RV^+, as defined by Barndorff-Nielsen et al. (2010), to capture the sign asymmetry of the price process. We consider RV^- and RV^+ as measures based entirely on downward or upward moves of intraday returns. The daily RV_t^- and RV_t^+ are given by

$$RV_t^- = \sum_{i=1}^{N} r_{t,i}^2 I_{[(r_{t,i})<0]}, \quad (6)$$

$$RV_t^+ = \sum_{i=1}^{N} r_{t,i}^2 I_{[(r_{t,i})>0]}, \quad (7)$$

where $r_{t,i}$ is the intraday return i within day t, and $i = 1,..,N$ is the total number of intraday observations within a day. Earlier studies have identified the importance of downside risk in portfolio risk assessment and management. Unlike other proxies of downside risk (e.g., downside deviation or downside beta), downside realized semivariance is constructed entirely using high-frequency data, and hence, it is considered to be a better proxy for downside risk (see, Hansen and Huang, 2016).

3.4 Realized skewness and realized kurtosis

We compute realized skewness, RSK, and realized kurtosis, RKU, as measures of the higher moments of the daily asset-return distribution computed from intraday returns. These measures allow us to consider with more precision statistical forces that may affect the variations of the price process. Mei et al. (2017) investigate the importance of RSK and RKU for forecasting realized volatility for the stock markets of China and the U.S. (for exchange rates, see Gkillas et al., 2019). In our research, like Amaya et al. (2015), we consider RSK as a measure of the asymmetry of the daily asset-return distribution. The interpretation of this measure is straightforward. Positive values of RSK indicate that the distribution has a right tail that is fatter than the left tail, and negative values indicate the opposite. Because we are interested in extremes of the bitcoin-returns distribution more generally, we also construct RKU as a measure of kurtosis. The daily RSK_t, and RKU_t (standardized by RV_t) are given by

$$RSK_t = \frac{\sqrt{N}\sum_{i=1}^{N} r_{t,i}^3}{(\sum_{i=1}^{N} r_{t,i}^2)^{3/2}}, \tag{8}$$

$$RKU_t = \frac{N\sum_{i=1}^{N} r_{t,i}^4}{(\sum_{i=1}^{N} r_{t,i}^2)^{2}}. \tag{9}$$

The scaling of RSK_t and RKU_t by $\sqrt{N}$ and N makes sure that their magnitudes correspond to daily skewness and kurtosis.

3.5 HAR-RV models

We use the standard HAR-RV model originally proposed by Corsi (2009) as the benchmark model for realized-volatility forecasting. The HAR-RV model is one of the most often studied models in the literature on realized-volatility forecasting because, despite its simplicity, it is flexible enough to to account for standard features of realized volatility such as long memory and multi-scaling behavior. The benchmark HAR-RV model for $h-$days-ahead forecasting is given by

$$RV_{t+h}^{j} = \beta_0 + \beta_d RV_t^j + \beta_w RV_{w,t}^j + \beta_m RV_{m,t}^j + \varepsilon_{t+h}, \tag{10}$$

where RV_t^j can be either RV_t^s, RV_t^- or RV_t^+, i.e., overall (standard), bad and good realized volatility, with $RV_t^s = RV_t^- + RV_t^+$. $RV_{w,t}^j$ refers to the average RV^j from day $t-5$ to day $t-1$. $RV^j_{m,t}$ refers to the average RV^j from day $t-22$ to day $t-1$.

The above model can be extended in various ways. In the context of our analysis, we use as additional explanatory variables our downside and upside realized tail-index estimators (or both) as additional predictors. The extended HAR-RV models model are given by

$$RV^j_{t+h} = \beta_0 + \beta_d RV^j_t + \beta_w RV^j_{w,t} + \beta_m RV^j_{m,t} + \gamma_- TN_t + \varepsilon_{t+h}, \tag{11}$$

$$RV^j_{t+h} = \beta_0 + \beta_d RV^j_t + \beta_w RV^j_{w,t} + \beta_m RV^j_{m,t} + \gamma_+ TP_t + \varepsilon_{t+h}, \tag{12}$$

$$RV^j_{t+h} = \beta_0 + \beta_d RV^j_t + \beta_w RV^j_{w,t} + \beta_m RV^j_{m,t} + \gamma_- TN_t + \gamma_+ TP_t + \varepsilon_{t+h}, \tag{13}$$

In addition, we examine whether adding realized skewness, RSK_t, realized kurtosis, RKU_t, or both, as to the benchmark model affects the contribution of TN_t and TP_t to forecast performance. To this end, we add as additional covariates RSK_t and/or RKU_t to the three models above.

Finally, we consider an alternative benchmark HAR-RV model that features jumps (HAR-RV-J; Andersen et al., 2007). As we already mentioned, in order to further validate the importance of our estimators for realized-volatility forecasting, we also study their predictive value along with jumps. Jumps are important elements in volatility modeling as they can explicitly improve the overall fit of volatility models (see, e.g., Duffie et al., 2000; Andersen et al., 2007). For this alternative benchmark, we use the realized variance constructed by summing up all successive intraday squared returns as an RV estimator. Under weak regularity conditions, this estimator is a consistent estimator of quadratic variation as the sampling frequency increases. Quadratic variation, in turn, is regarded as the best estimator of integrated (latent) volatility. In order to separate the price increments into jumps and continuous price moves, we use MRV for the part of the total variation due to continuous price moves, while we estimate the total variation by realized variance. For the detection of jump components, we apply the jump test statistic of Barndorff-Nielsen and Shepherd (2006), and Huang and Tauchen (2005) using the median realized quarticity ($MedRQ$). We refer to the study by Andersen et al. (2012) for how MRV can be used for detecting jumps, and for how to construct $MedRQ$.

3.6 Forecast evaluation

We define the forecast error as $FE^{j}_{t+h} = RV^{j}_{t+h} - \hat{RV}^{j}_{t+h|t}$ (j = s, $-$, $+$), and use the loss function studied by Elliott et al. (2005, 2008) to evaluate forecasting performance. The loss function is given by (dropping the j index)

$$L(FE_{t+h}, \alpha) = [\alpha + (1-2\alpha) I_{[FE_{t+h}<0]}] |FE_{t+h}|^{s}, \tag{14}$$

This loss function is of the lin-lin type when we set $s = 1$, and of the quad-quad type when we set $s = 2$. The parameter $\alpha \in (0, 1)$ governs the shape of the loss function. A symmetric loss function is obtained when $\alpha = 0.5$. As a result, when we set $\alpha = 0.5$ and $s = 1$, then we use the common absolute loss criterion to evaluate forecasts. In contrast, the widely-studied squared-error-loss criterion is obtained when we set $\alpha = 0.5$ and $s = 2$. When the shape parameter differs from $\alpha = 0.5$, the loss function becomes asymmetric. For $\alpha > 0.5$, the loss a forecaster incurs in case an underprediction of realized volatility is larger than the loss from an overprediction of the same absolute size. For $\alpha < 0.5$, the loss from an overprediction of realized volatility is larger than the loss from a corresponding underprediction.

4 Empirical analysis

4.1 Data

We use high-frequency (intraday) bitcoin prices. The sample starts on 26th January 2015 and ends on 11th November 2018 incorporating various booms and crashes in the bitcoin market. We select prices every fifteen minutes (15-min) and construct 15-min log-returns. The sample period and the data frequency are selected in order to avoid liquidity issues from no-activity periods during very short time windows.

More specifically, earlier empirical evidence suggests that intraday returns should be computed at the highest possible frequency so that volatility estimators converge asymptotically towards the true conditional volatility following fixed domain asymptotics (also called infill asymptotics). At the same time, the sampling frequency should be not too high to induce spurious jumps due to market frictions and not too low to lead to poor data analysis. In order to put it differently, as the noise increases when the sampling frequency converges on zero (see Andersen and Bollerslev 1997, 1998; Taylor and Xu, 1997), we select the highest sampling frequency as the optimal sampling frequency at which the autocovariance bias term disappears (see, Oomen, 2001; Oomen, 2004; Degianakis and Floros, 2016; Degiannakis and Filis, 2017; Sévi, 2014, among others). The optimal sampling frequency takes into consideration the trade-off between the bias of the estimators and their accuracy. Also, it should be noted that, while in the literature on traditional assets a 5-min sampling frequency has been found adequate for forecasting experiments (Liu et al., 2015), bitcoin is a new asset, and liquidity is one of many factors that leads to sudden movements in its price (see, Scaillet et al., 2020).

We define a trading day from Monday to Sunday from 00:00 EST to 23:55 EST. This definition also implies that we have enough observations to avoid poor data analysis (Andersen, 2000). The intraday prices for bitcoin are collected from the website Bitcoincharts (https://www.bitcoincharts.com), which offers data on a number of liquid bitcoin markets.

We clean the sample following the suggestions by Barndorff-Nielsen et al. (2009) in order to reduce the effect of market micro-structure noise and to avoid misleading results related to sample bias. Such a procedure is also required to deal with higher-order moments so as to ensure that these moments can effectively capture specific features like asymmetry and fat-tailedness (Amaya et al., 2015). More specifically, we proceed in the following steps. First, we exclude fixed and moving holidays, including Christmas, New Year's Day etc. from the sample. Second, we remove days with infrequent trades (days with recorded prices for less than 70 percent of the expected observations on operating time) from the sample. Third, we apply the estimators to the series of mid-quotes after filtering out spread outliers (less than 0.1 percent in each distribution tail).

– Please include Table 1 about here. –

The final sample includes 1381 trading days. The data on returns and realized volatility (RV^s) are plotted in Figure 1, and descriptive statistics of our key predictor variables are summarized in Table 1.

4.2 Baseline results

We use a rolling-estimation window to estimate the variants of the HAR-RV models that we study in our empirical research. We then use the variant of the Diebold and Mariano (1993) test studied by Harvey et al. (1997) to analyze the incremental predictive value of the realized tail-index estimators. We consider a short ($h = 1$), a medium ($h = 5$), and a long ($h = 22$) forecast horizon.[4] We mainly focus on the quad-quad version of the loss function and present results for the lin-lin loss function at the end of this section.

[4]We arrange the data matrix such that we have exactly the same number of observations for all three forecasting horizons. We compute all estimation results that we report in this research using the R programming environment (R Core Team 2017). We compute the p-values of the Diebold-Mariano test using the R package "forecast" (Hyndman, 2017; Hyndman and Khandakar, 2008), with the code changed to account for the asymmetry of the loss function.

– Please include Figure 2 about here. –

Figure 2 summarizes the results for a baseline scenario in which we compare the HAR-RV baseline model with an extended model that features one or both of the tail-index estimators. For the short forecast horizon (specifications TP and $TP+TN$) and for the medium forecast horizon (all three specifications), we observe a relatively large area of significant test results for the quad-quad loss function for $\alpha > 0.5$ and a rolling-estimation window that comprises more than approximately 400 data. For the long forecast horizon, we only observe occasionally significant test results. Moreover, we observe relatively small areas of significant test results for $\alpha << 0.5$ in case of the medium forecast horizon, and only for a rather short rolling-estimation window.[5]

While *MRV* is a jump-robust estimator, results for a HAR-RV-J model that features the jump component and estimated on a standard measure of *RV* are similar to the results we report in Figure 2. We report the results for such a model at the end of the paper (Appendix, Figure A1). It is important to mention that such results are expected and that they suggest that our main evidence is robust to the choice of alternative realized volatility estimators with and without the presence of jumps.

– Please include Figure 3 about here. –

Figure 3 summarizes the results that we obtain when we control for *RSK*, *RKU*, or both. We observe for the $RKU/TN+TP$ and $RSK/TN+TP$ models relatively

[5]While we report results for the medium and long forecast horizon based on multiple-step forecasts of y_{t+h}, results for forecasts of mean$(y_{t+1}+...+y_{t+h})$ are similar. Moreover, as suggested by an anonymous reviewer, we also consider as an extension HAR-RV models that includes returns as an additional control variable. Results for these extended HAR-RV models are similar to those reported in Figure 2. Complete results are available upon request.

large areas of significant test results for both the short and the medium forecast horizon when we assume that a forecaster incurs a larger loss from an underprediction than from a corresponding overprediction and the length of the rolling-estimation window is not too short. Such a forecaster benefits from considering the tail-index estimators also in case of the $RKU + RSK/TN + TP$ model, but the areas of significant test results are smaller than in the case of the other two models (short- and medium-forecast horizon). As compared to the baseline scenario, the areas of significant test results that we observe for $\alpha << 0.5$ are somewhat larger in case of the medium-forecast horizon. For the long-forecast horizon, the areas of significant test results are small and fragmented.

– Please include Figure 4 about here. –

Figure 4 summarizes the results for bad and good volatility. Results resemble the results for the standard realized volatility. A forecaster who suffers more from an underprediction than from a corresponding overprediction benefits from considering the tail-index estimators as predictors for medium and short forecast horizons (quad-quad loss function) and when the rolling-estimation window is not too short. For the long forecast horizon, test results are significant only for a few specific constellations of the asymmetry parameter and the rolling-estimation window.

– Please include Figure 5 about here. –

Figure 5 summarizes the results for a lin-lin loss function, where we focus on the baseline model (without RKU and RSK). The results show that a forecaster who suffers more from an overprediction than an underprediction ($\alpha < 0.5$) tends to benefit from accounting for the information embedded in the tail-index estimators when we study for a short- and medium-forecast horizon, and a short rolling-estimation window. In contrast, when the length of the rolling-estimation window increases, a forecaster who suffers more from an underprediction than an overprediction (where α is substantially larger than its symmetric benchmark value of 0.5) benefits from considering the tail indices for forecasting realized volatility, a result that we observe for all three forecast horizons. For the short and medium forecast horizon, however, the test results are significant only for a somewhat small range of very large asymmetry parameters than in the case of a quad-quad loss function, a result that was to be expected. Because the loss a forecaster suffers in case of a large forecast error is larger under a quad-quad than under a lin-lin loss function, it is not surprising that we observe for these two forecast horizons larger areas of significant test results (for values of the asymmetry parameter $\alpha > 0.5$) in case of a quad-quad loss function.

4.3 Simulation results

The Diebold-Mariano test has a degenerate asymptotic distribution under the null hypothesis in case one tests nested models. Given that we use a rolling-estimation window, it is likely that the asymptotic properties of the Diebold-Mariano test are not a major issue in the case of our research. In order to inspect this issue further, however, we now present the results of a simulation study that serves to assess the reliability of the results of the Diebold-Mariano test. The simulation study is based on the simulation design described by Rapach et al. (2005).

In order to set up the simulation study, we estimate in a first step the baseline HAR-RV model along with models for the realized tail-index estimators on the full sample of data. As for the models for the tail-index estimators, we assume AR(5) models to capture longer-range dependence in a stylized way. It is important to note that the baseline HAR-RV model does not feature the measures of tail-index estimators, that is, the null hypothesis is that the tail-index estimators do not add any value to the baseline HAR-RV model. We store the coefficients and the residuals we obtain from the baseline HAR-RV model and the models estimated for the measures of tail indices.

In a second step, we sample with replacement from the residuals, where we preserve the contemporaneous correlation of the residuals from the different models. We draw sampled time series of residuals for 1000 simulation runs. We sample in such a way that after the initialization of RV_m and MRV_m and computing leads (22 days-ahead-forecasts), 100 transitory data are available. The transitory data are not used for the computation of the Diebold-Mariano test but are rather used to initialize the models to be estimated.

In a third step, we use the sampled residuals along with the estimated coefficients (step one), to simulate artificial time series of realized volatility and the measures of realized tail-index estimators. Initial values are set to zero (which is innocuous given that we delete 100 transitory observations). We then estimate the baseline HAR-RV model and the HAR-RV models extended to include tail-index estimators, where we consider rolling-estimation windows of length ranging from 50, 100, ..., to 800 observations. Forecasts and actuals obtained from these models are stored.

In a fourth step, we estimate the baseline HAR-RV model and its extensions on the original data for the various rolling-estimation windows. We then estimate the Diebold-Mariano test for the original data and for the simulated data. We compute the p-values for the Diebold-Mariano test for every length of the rolling-estimation window, every asymmetry parameter, and for the three different forecast horizons. We compute the p-values as #(simulated DM tests $>$ original DM test) / #(simulated DM tests).[6]

---

[6]In few cases, the Diebold-Mariano test produces a negative estimate of the variance of the

– Please include Figure 6 about here. –

Figure 6 summarizes the simulation results. The simulation results corroborate our main finding that mainly forecasters who suffer a larger loss in case of an underprediction of realized volatility than in case of an overprediction of the same absolute size benefit from using measures of realized tail indices to compute forecasts. This main finding is obtained mainly for the short- and medium-forecast horizon.

4.4 The length of the rolling-estimation window

Our simulation study helps to assess the significance of the results of the Diebold-Mariano test, given the length of a rolling-estimation-window. A different question concerns how a forecaster should choose among the alternative lengths of the rolling-estimation windows. This choice is of crucial importance, as our results demonstrate. One approach to address this question is to ask whether an optimal rolling-estimation window can be identified.[7] The choice of an optimal rolling-estimation window involves a trade off that arises because a short rolling-estimation window reduces the risk of computing biased forecasts when the data-generating process underlying the dynamics of realized volatility changes over time but, at the same time, may result in a higher forecast-error variance than forecasts computed based on a long rolling-estimation window. A second approach is to use a unified test statistic that compares the accuracy of forecasts across the

loss differential. These cases are not used for the computation of the p-values.

[7]One way to identify the optimal length of the rolling-estimation window is to compute, given the asymmetry parameter and the forecast horizon, the cumulated loss for the models that feature the realized tail indices and divide it by the cumulated loss obtained for the baseline model. The minimum of this ratio gives the optimal length of the rolling-estimation window for every asymmetry parameter and every forecast horizon (results are available upon request).

whole array of different rolling-estimation windows. Such approaches have been studied in recent research by Rossi and Inoue (2012), Inoue et al. (2017) and and Pesaran and Timmermann (2007). The latter also consider, as a third approach, procedures for combining forecasts from models estimated using different lengths for the rolling-estimation window into a single forecast. We adopt in our research, as a robustness check of our results, this third approach.

– Please include Figure 7 about here. –

The specific forecast-combination procedure that we use requires that a forecaster chooses in every out-of-sample period the median forecast computed across rolling-estimation windows, where the length of the rolling-estimation windows varies from 50, 100, ..., to 800 observations.[8] We summarize the results in Figure 7. The results corroborate that the results of the Diebold-Mariano test get significant when the asymmetry parameter approaches the upper bound of its admissible range, where this effect is stronger for the short and the medium forecast horizon than for the long forecast horizon.

4.5 Nonlinear transformations

As two important variants of our baseline models, we consider two common variants of the HAR-RV model that make use of nonlinear transformations of realized volatility. In the first model, we replace realized volatility with its square root (that is, the realized standard deviation). Modeling the square-root of realized volatility may be particularly relevant in risk-management applications. In the second

[8]As one would have expected, the variance of forecasts increases as the rolling-estimation window gets shorter. For this reason, we do not to consider a mean forecast (on which the forecasts from the short rolling-estimation window would have a relatively large effect) but rather a median forecast.

model, we study the natural logarithm of the realized standard deviation (see, e.g., Andersen et al., 2007). After having estimated the model in logs, we compute the anti-log of the forecasts deriving from the estimated model and then add a Jensen-Ito term, as in Degiannakis and Filis (2017).

– Please include Figures 8 about here. –

Figure 8 summarizes the results (quad-quad loss function). For both the square-root model and the logarithmic model, we find that for not too short rolling-estimation windows a forecaster with an asymmetric loss function that attaches a larger loss to an underprediction relative to a corresponding overprediction benefits from using the tail-index estimators in the forecasting model. We obtain this finding not only for the short- and medium- forecast horizons, but also (albeit to a lesser extent) for the long-forecasting horizon.

5 Concluding remarks

We conduct in our research a predictive analysis of realized volatility obtained from intraday data of bitcoin returns. We use the HAR-RV model and study measures focusing exclusively on relatively large returns realizations to assess the tail shape of the bitcoin-returns distribution. We study whether these measures help to forecast the realized volatility of bitcoin returns more accurately. We construct our tail measures based on the widely used Hill tail-index estimator for investigating the tail behavior of asset returns. We also control for the impact of realized skewness and realized kurtosis as predictors of realized volatility of bitcoin returns, and we consider the bad (downside) and good (upside) component of realized volatility.

We find that our measures help to improve forecasts of daily realized bitcoin realized volatility mainly for forecasters who have a loss function that attaches a greater weight to losses from an underprediction of realized volatility than from an overprediction of the same absolute size. Test results for forecasters who suffer from a relatively high loss from an overprediction of realized volatility are significant only for few model configurations. Given that the bitcoin price is highly volatile, and that price movements can be sudden and disruptive, this main result of our empirical analysis is important for investors who rely on the highly popular HAR-RV model to forecast realized bitcoin volatility and to manage the risk of their bitcoin investments, mainly at a short and medium forecast horizon. We show that our main result also obtains when we consider variants of the HAR-RV model that feature realized skewness and/or realized kurtosis, and when we consider models that feature a nonlinear transformation of realized volatility (that is, the square-root model and the logarithmic model). Moreover, the significance of the predictive value added by our measures is somewhat stronger, as one would have expected, when we consider a quad-quad rather than a lin-lin loss function.

Acknowledgments

We thank an associate editor and two anonymous reviewers for helpful comments. The usual disclaimer applies.

References

Al Rahahleh, N., and Bhatti, M. I. (2017). Co-movement measure of information transmission on international equity markets. Physica A: Statistical Mechanics and its Applications 470: 119−131.

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., Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4: 115−158.

Andersen, T., and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39: 885−905.

Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. Journal of Business & Economic Statistics 18(2): 146−153.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. Journal of Financial Economics 61: 43−76.

Andersen, T. G., and Bollerslev, T., and Diebold, F.X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics 89: 701−720.

Andersen, T. G., Bollerslev, T., Diebold, F.X. and Labys, P. (2009). Parametric and Nonparametric Volatility Measurement. In Y. Ait-Sahalia and L. Hansen (eds), Handbook of Financial Econometrics, Vol. 1. p. 67−138. North Holland, Amsterdam.

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

Ardia, D., Bluteau, K., and Rede, M. (2018). Regime changes in bitcoin GARCH volatility dynamics. Finance Research Letters. DOI: `https://doi.org/`

10.1016/j.frl.2018.08.009.

Baek, C., Elbeck, M.A. (2015). Bitcoins as an investment or speculative vehicle? A first look. Applied Economics Letters 22: 30−34.

Bandi, F. M., and R. Reno. (2012). Time-Varying Leverage Effects. Journal of Econometrics 169: 94−113.

Barndorff-Nielsen, O.E., 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., and Shephard, N. (2007). Variation, Jumps and High Frequency Data in Financial Econometrics. In R. Blundell, T. Persson, and W. Newey (eds), Advanced in Economics and Econometrics: Theory and Applications, Ninth World Congress, Econometric Society Monograph, 328−72. Cambridge, UK: Cambridge University Press.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. Econometrics Journal 12: C1−C32.

Barndorff-Nielsen, O.E., Kinnebrouk, S., and Shephard, N. (2010). Measuring downside risk: realised semivariance. In: 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.

Bee, M., Dupuis, D. J., and Trapin, L. (2019). Realized Peaks over Threshold: A Time-Varying Extreme Value Approach with High-Frequency-Based Measures. Journal of Financial Econometrics 17(2): 254−283.

Bekaert, G., and Hoerova, M. (2014). The VIX, the Variance Premium and Stock Market Volatility. Journal of Econometrics 183: 181−192.

Bhatti, M. I., and Nguyen, C. C. (2012). Diversification evidence from international equity markets using extreme values and stochastic copulas. Journal of International Financial Markets, Institutions and Money 22(3): 622−646.

Bohme, R., Christin, N., Edelman, B., and Moore, T. (2015). Bitcoin: Economics, technology, and governance. Journal of economic Perspectives 29(2): 213−38.

Bollerslev, T., Law, T. H., and Tauchen, G. (2008). Risk, jumps, and diversification. Journal of Econometrics 144: 234−256.

Bollerslev, T., and V. Todorov. (2011). Estimation of Jump Tails. Econometrica 79: 1727−1783.

Bollerslev, T., Todorov, V., and Li, S. Z. (2013). Jump tails, extreme dependencies, and the distribution of stock returns. Journal of Econometrics 172(2): 307−324.

Bouri, E., Shahzad, S.J.H., and Roubaudb, D. (2018). Co-explosivity in the cryptocurrency market. Finance Research Letters. DOI: `https://doi.org/10.1016/j.frl.2018.07.005`.

Chen, X., and E. Ghysels (2012). News −4 Good or Bad −4 and its Impact on Volatility Predictions over Multiple Horizons. Review of Financial Studies 24(1): 46−481.

Choulakian, V., and Stephens, M. A. (2012). Goodness-of-fit tests for the generalized pareto distribution. Technometrics 43: 478−484.

Chu, J., Chan, S., Nadarajah, S., and Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. Journal of Risk and Financial Management 10: 17. DOI: `10.3390/jrfm10040017`.

Conrad, C., Custovic, A., and Ghysels, E. (2018). Long- and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. Journal of Risk and Financial Management 11: 23. DOI: 10.3390/jrfm11020023.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7: 174−96.

Dacorogna, M.M., Genay, R., Muller, U.A., Olsen, R.B. and Pictet, O.V. (2001). An introduction to high-frequency finance. Academic Press, San Diego.

Danielsson, J., and De Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. Journal of Empirical Finance 4(2−3): 241−257.

De Haan, L., and Ferreira, A. (2006). Extreme Value Theory: An Introduction, Springer. Springer. DOI: 10.1007/0-387-33477-7.

Duffie D., Pan J., Singleton K. (2000). Transform analysis and asset pricing for affine jump−diffusions. Econometrica 68:1343−1376.

Degiannakis, S., and Floros, C. (2016). Intraday realized volatility for European and USA stock indices. Global Finance Journal 29: 24−41.

Degiannakis, S., and Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. Journal of International Money and Finance 76: 28−49.

De Lira Salvatierra, I. A., and A. J. Patton. (2015). Dynamic Copula Models and High Frequency Data. Journal of Empirical Finance 30: 120−135.

Diebold, F.X., and Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3): 253−263.

Diebold, F. X., Hahn, J., and Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. Review of Economics and Statistics 81: 661−673.

Engle, R., and V. Ng (1993). Measuring and Testing the Impact of News and Volatility. Journal of Finance. 48(5): 1749−1778.

Elliott, G., Komunjer, I. and Timmermann, A. (2005). Estimation and testing of forecasting 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.

Ghysels, E., and Sinko, A. (2011). Volatility forecasting and microstructure noise. Journal of Econometrics, 160(1): 257−271.

Gkillas, K., and Katsiampa, P. (2018). An application of extreme value theory to cryptocurrencies. Economics Letters 164: 109−111.

Gkillas, K., and Longin, F. (2018). Is bitcoin the new digital gold? Evidence from extreme price movements in financial markets. SSRN Electronic Journal. URL: `https://doi.org/10.2139/ssrn.3245571`.

Gkillas, K., Gupta, R., and Pierdzioch, C. (2019). Forecasting (downside and upside) realized exchange-rate volatility: Is there a role for realized skewness and kurtosis?. Physica A: Statistical Mechanics and its Applications 532: 121867.

Hansen, P. R., Z. Huang, and H. H. Shek. (2012). Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. Journal of Applied Econometrics 27: 877−906.

Hansen, P. R., A. Lunde, and V. Voev. (2014). Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility. Journal of Applied Econometrics 29: 774−799.

Hansen, P. R., and Huang, Z. (2016). Exponential GARCH modeling with realized measures of volatility. Journal of Business & Economic Statistics 34: 269−287.

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

Hattori, T. (2019). Does 5-minute RV Outperform Other Realized Measures in the Cryptocurrency Market?. SSRN Electronic Journal. URL: `http://dx.doi.org/10.2139/ssrn.3416106`.

Hattori, T. (2020). A Forecast Comparison of Volatility Models Using Realized Volatility: Evidence from the Bitcoin Market. Applied Economics Letters 27: 591−595.

Hill, B. M. (1975). A simple general approach to interface about the tail of a distribution. Annals of Statistics 3: 1163−1174.

Huang, X., and Tauchen, G. (2005). The relative contribution of jumps to total price variance. Journal of Financial Economics 3: 456−499.

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.

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. Journal of Econometrics 196: 55−67.

Jansen, D. W., and de Vries, C. G. (1991). On the frequency of large stock returns: Putting booms and busts into perspective. Review of Economics and Statistics 73: 18−24.

Kim, T. (2017). On the transaction cost of bitcoin. Finance Research Letters 23: 300−305.

Kochling, G., Schmidtke, P., and Posch, P. N. (2019). Volatility Forecasting Accuracy for Bitcoin. Economics Letters. DOI: `://doi.org/10.1016/j.econlet.2019.108836`.

Mansor, F., Al Rahahleh, N, and Bhatti, M. I. (2019). New Evidence on Fund Performance in Extreme Events. International Journal of Managerial Finance, 2019:15(4), 511−532.

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

Nakamoto, S. (2008). Bitcoin: a peer to peer electronic cash system. Available for download at: `https://bitcoin.org/bitcoin.pdf`.

Neuberger, A., and Payne, R.G. (2018). The skewness of the stock market at long horizons. Available at SSRN: `https://ssrn.com/abstract=3173581`.

Nguyen, C. C., and Bhatti, M. I. (2012). Copula model dependency between oil prices and stock markets: Evidence from China and Vietnam. Journal of International Financial Markets, Institutions and Money 22(4): 758−773.

Nguyen, C., Bhatti, M. I., Komorníková, M., and Komorník, J. (2016). Gold price and stock markets nexus under mixed-copulas. Economic Modelling 58: 283−292.

Nguyen, C., Bhatti, M. I., and Henry, D. (2017). Are Vietnam and Chinese stock markets out of the US contagion effect in extreme events?. Physica A-Statistical Mechanics and its Applications 480: 10−21.

Noureldin, D., Shephard, N., and Shephard, K. (2012). Multivariate High−Frequency−Based Volatility (HEAVY) Models. Journal of Applied Econometrics 27: 907−933.

Liu, L. Y., Patton, A. J. and Sheppard, K. (2015), Does Anything Beat 5-minute RV? A Comparison of Realized Measures across Multiple Asset Classes. Journal of Econometrics 187: 293−311.

Longin, F. (2000). From Value at Risk to Stress Testing: The Extreme Value Approach. Journal of Banking & Finance 24: 1097−1130.

Longin, F., and Solnik, B. (2001). Extreme correlation of international equity markets. The journal of finance 56(2): 649−676.

Oh, D. H., and A. J. Patton. (2016). High Dimension Copula-Based Distributions with Mixed Frequency Data. Journal of Econometrics 193: 349–366.

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.

Osterrieder, J., and Lorenz, J. (2017). A statistical risk assessment of Bitcoin and its extreme tail behavior. Annals of Financial Economics 12: 1750003−1 −1750003−19.

Patton, J. A. (2011). Volatility Forecast Comparison Using Imperfect Volatility Proxies. Journal of Econometrics 160: 246−256.

Pesaran, H., and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics 137: 134−161.

Rapach, D.E., Wohar, M.E., and Rangvid, J. (2005). Macro variables and international stock return predictability. International Journal of Forecasting 21: 137−166.

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/.Rversion3.3.3`.

Rossi, B., and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. Journal of Business & Economic Statistics 30: 432−453.

Scaillet, O., Treccani, A., and Trevisan, C. (2020). High−frequency jump analysis of the bitcoin market. Journal of Financial Econometrics 18(2): 209−232.

Sevi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. European Journal of Operational Research 235: 643−659.

Shephard, N., and Shephard, K. (2010). Realizing the Future: Forecasting with High-Frequency-Based Volatility (HEAVY) Models. Journal of Applied Econometrics 25: 197−231.

Taylor, S. J., and Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance 4(4): 317−340.

Urquhart, A. (2016). The inefficiency of bitcoin. Economics Letters 150: 1−7.

West, K.D., Edison, H.J. and Cho, D. (1993). A utility-based comparison of some models of exchange rate volatility. Journal of International Economics 35: 23−46.

Figure 1: Bitcoin Returns and Realized Volatility

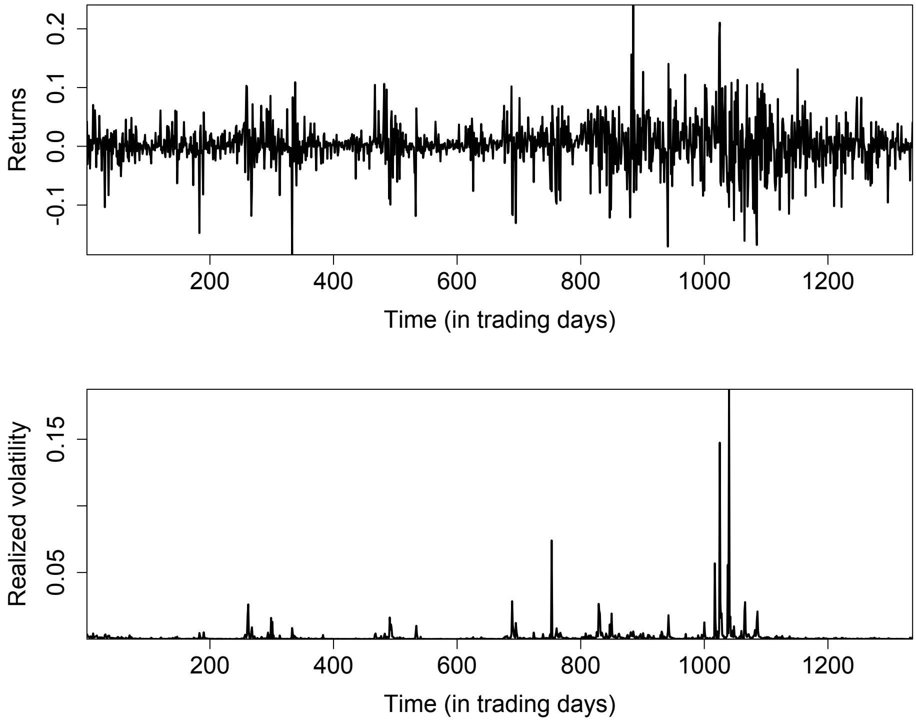
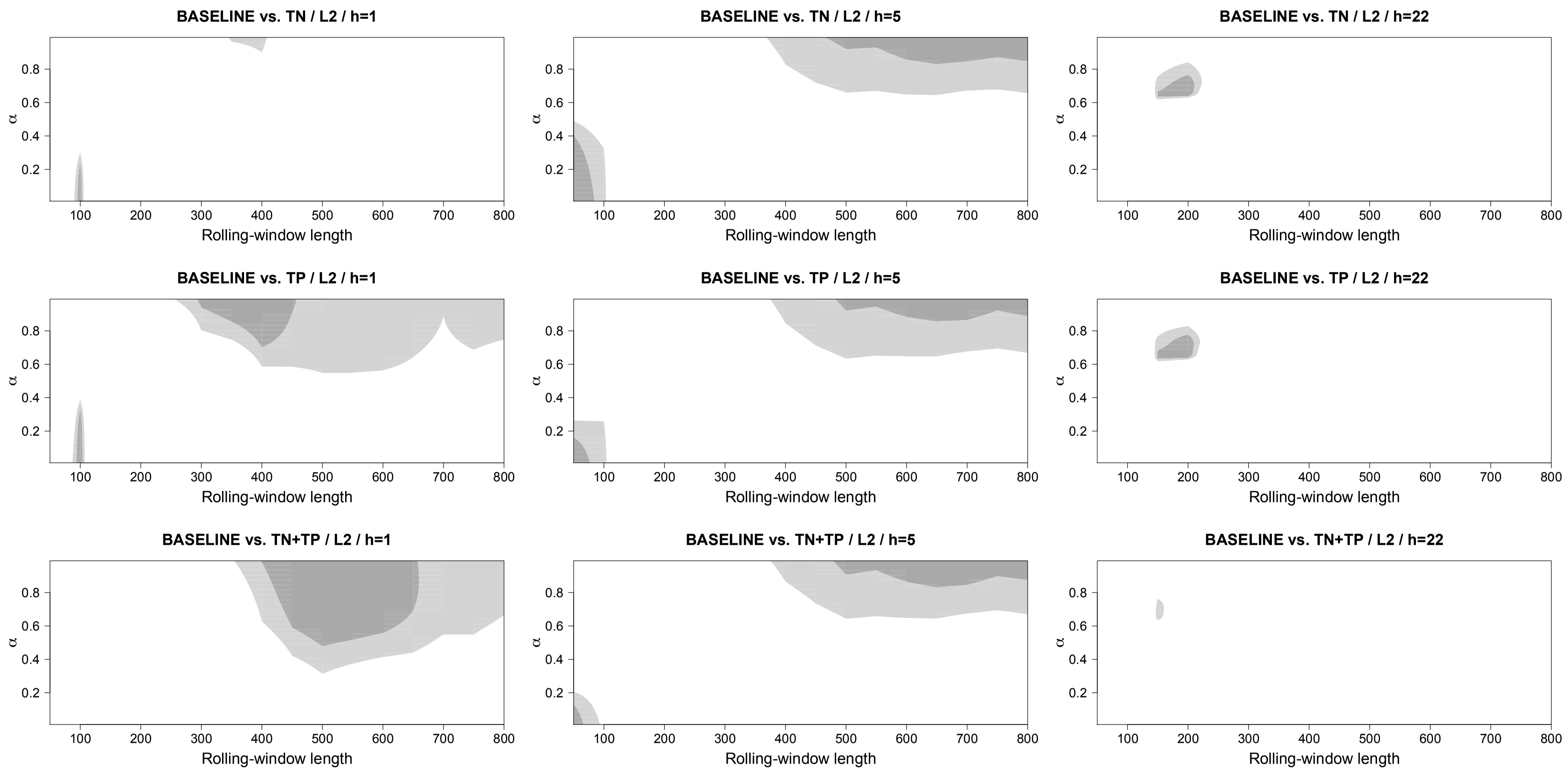
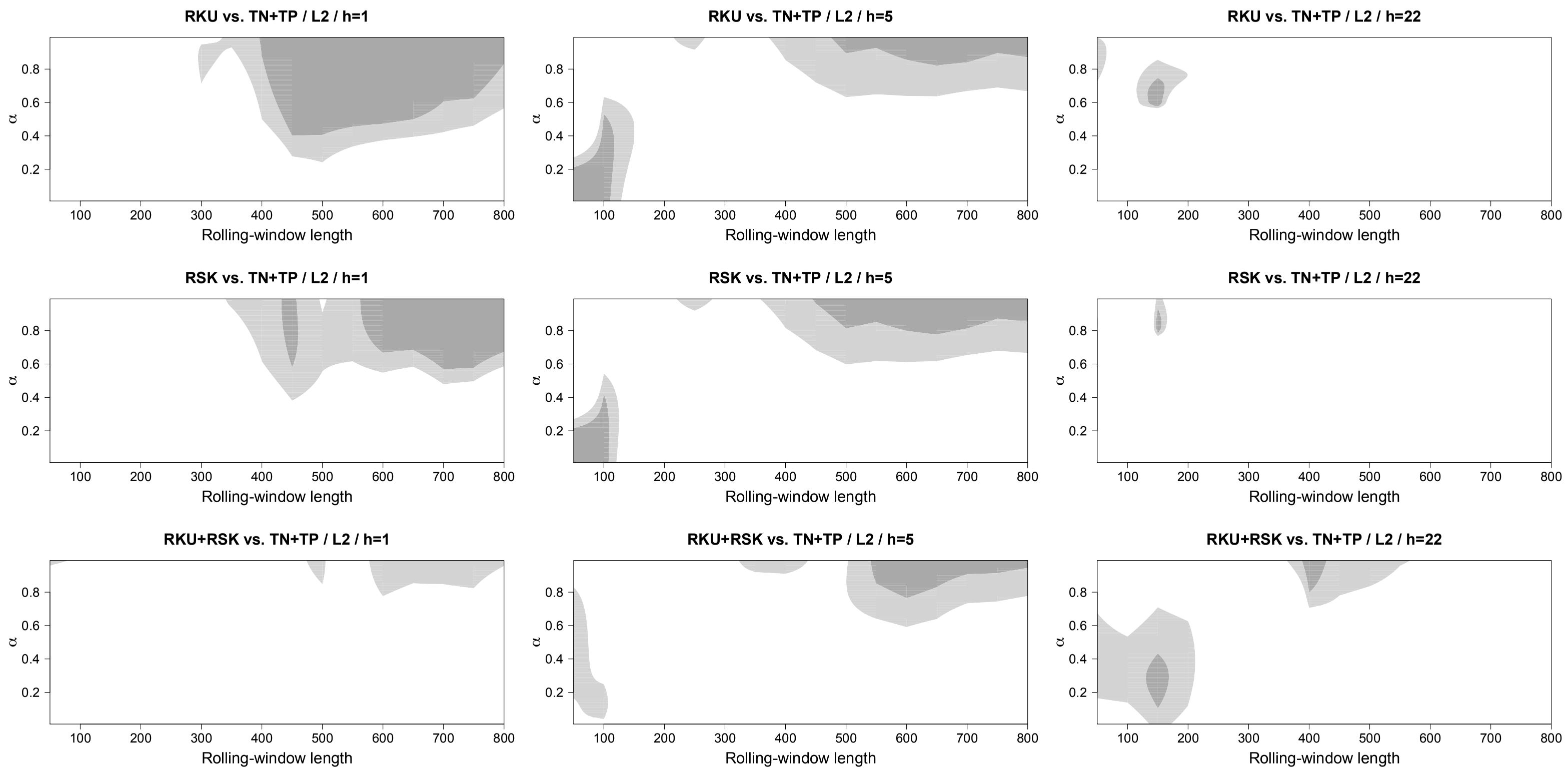


Figure 2: Baseline Results



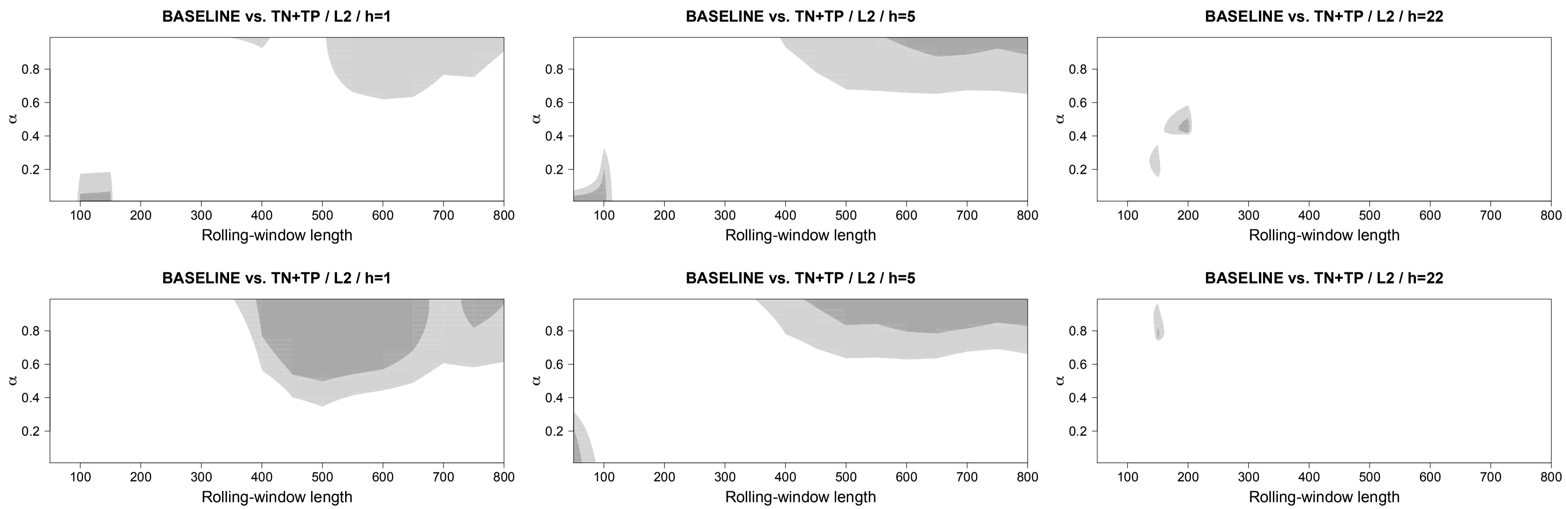
Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include the tail-index estimator 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.

Figure 3: Accounting for Realized Skewness and Realized Kurtosis



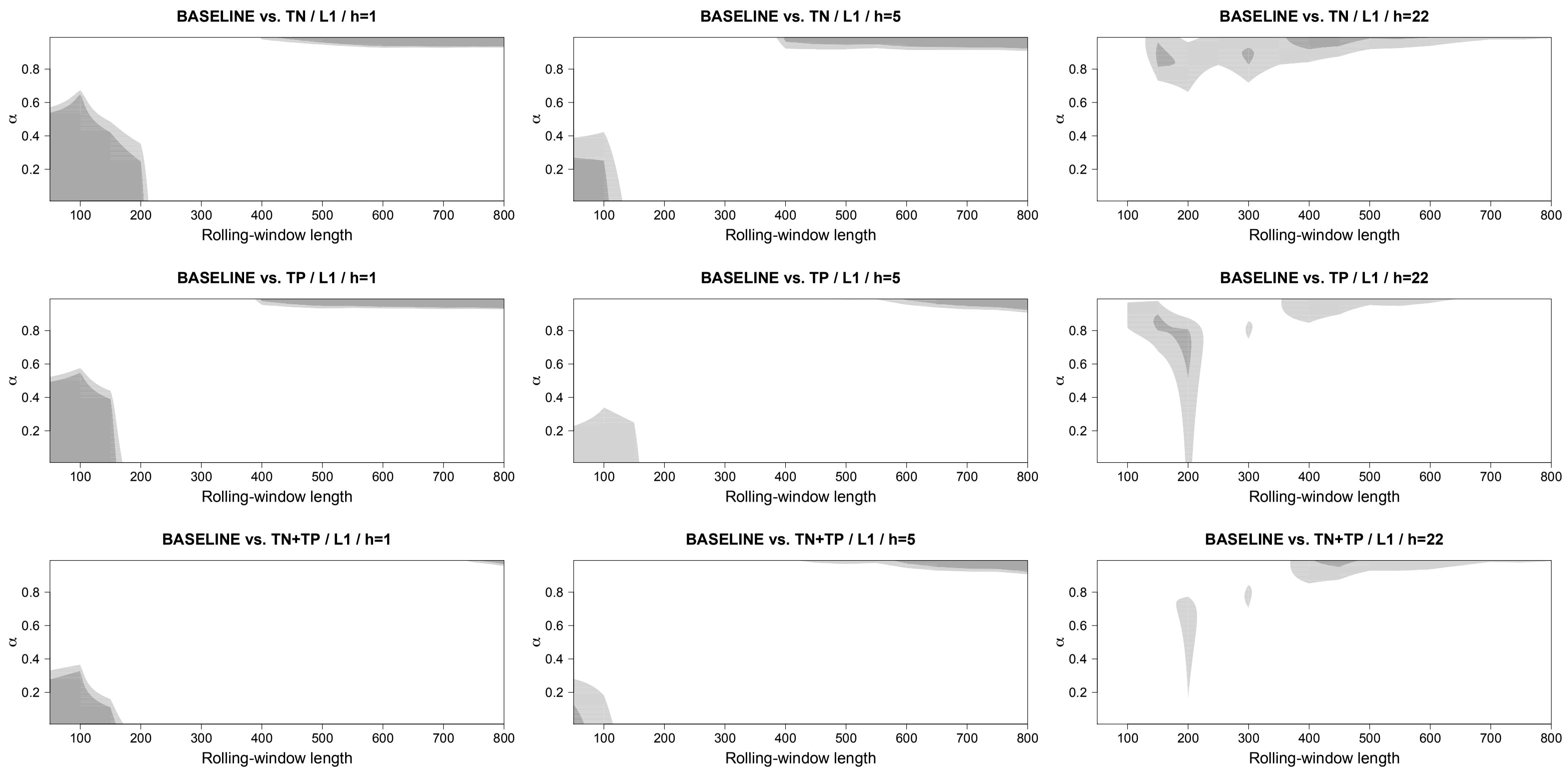
Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include the tail-index estimator 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.

Figure 4: Results for Bad and Good Realized Volatility



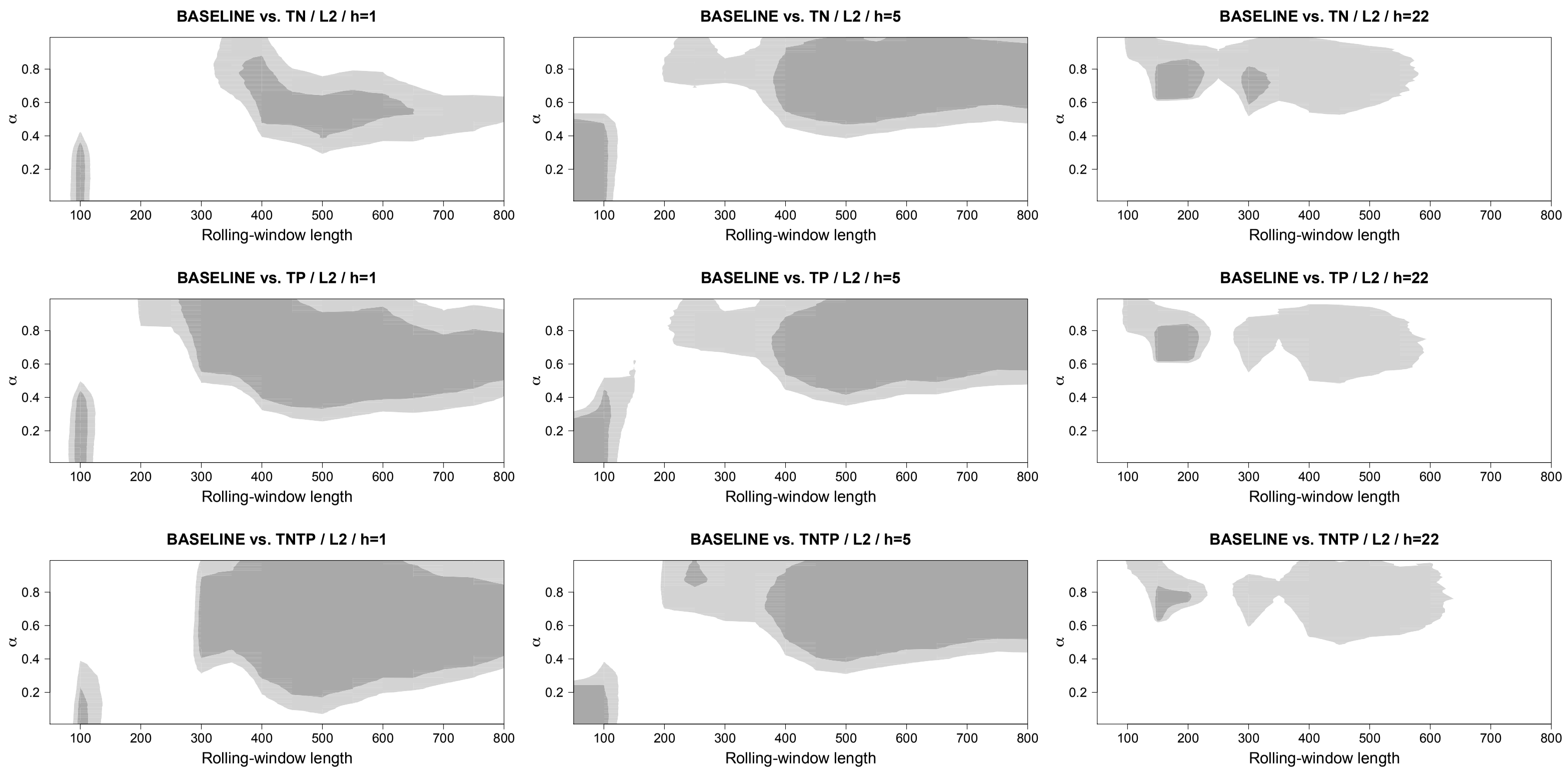
Note: Upper row: bad realized volatility. Lower row: Good realized volatility. Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include the tail-index estimator 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.

Figure 5: Results for L1 Loss



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 the tail-index estimator 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.

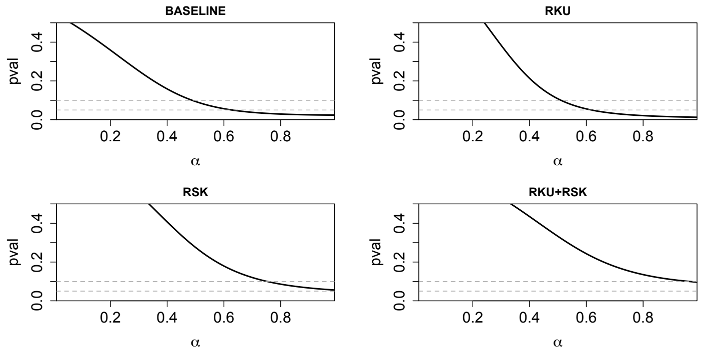
Figure 6: Simulation Results



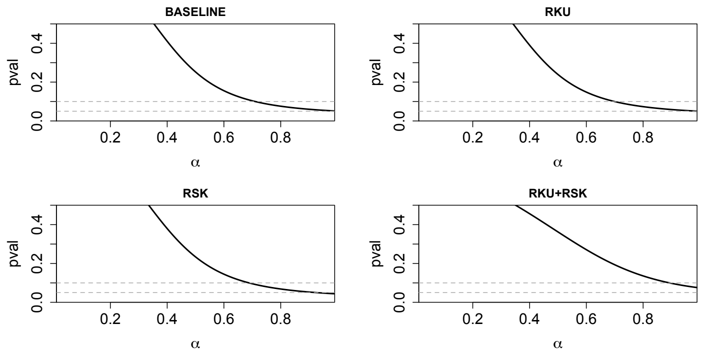
Note: Black area: Diebold-Mariano test is significant at the 5% level. Gray area: Diebold-Mariano test is significant at the 10% level. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include the tail-index estimator 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. Results are based on 1000 simulation runs.

Figure 7: Results for Forecast Combination

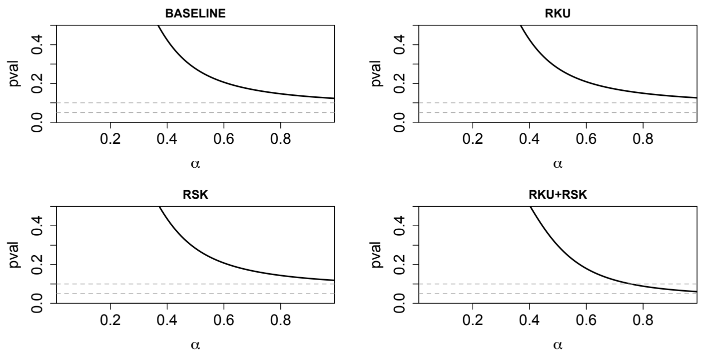
Panel A: $h = 1$



Panel B: $h = 5$

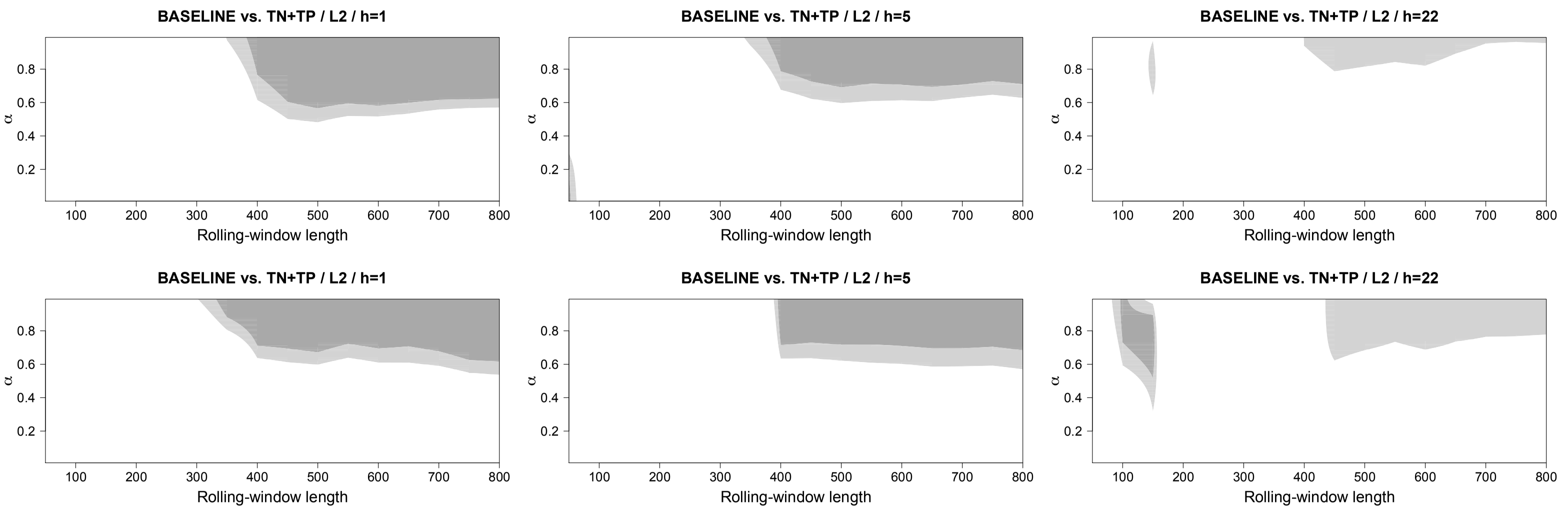


Panel c: $h = 22$



Results are based on rolling-window estimates. The horizontal axis displays the asymmetry parameter. The vertical axis displays the p-value of the Diebold-Mariano test. Null hypothesis: the two series of forecasts are equally accurate. Alternative hypothesis: the forecasts from the model extended to include the tail-index estimators (TN and TP) is more accurate. Forecasts Combination is achieved by choosing in every out-of-sample period the median forecast as computed across rolling-estimation windows, where the length of the rolling-estimation windows varies from 50, 100, ..., to 800 observations.

Figure 8: Results for the Nonlinear Transformations



Note: Upper row: Square-root model. Lower row: Log model. 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 the tail-index estimator 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.

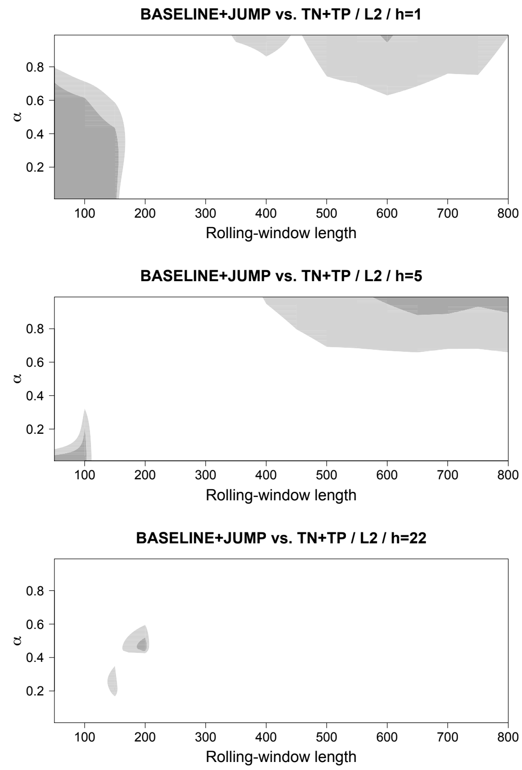
Table 1: Summary Statistics

Statistic	RV	RSK	RKU	TN	TP
Min.	0.001	-35.766	2.396	-4.533	-4.185
1st Qu.	0.040	-1.534	5.364	-1.847	-1.981
Median	0.089	-0.935	8.116	-1.38 0	-1.570
Mean	0.385	-1.416	19.726	-1.147	-1.362
3rd Qu.	0.249	-0.543	14.637	-0.694	-0.952
Max.	46.876	0.302	1294.250	4.610	3.783

Note: The column for realized volatility shows the statistics for annualized realized volatility.

Appendix

Figure A1: Results for a Model Featuring Jumps



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 the tail-index estimator 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.