Forecast Comparison Tests Under Fat-Tails

Jihyun Kim Nour Meddahi Mamiko Yamashita

Toulouse School of Economics

May 6, 2021

Abstract

Forecast comparison tests are widely implemented to compare the performances of two or more competing forecasts. The critical value is often obtained by the classical central limit theorem (CLT) or by the stationary bootstrap (Politis and Romano, 1994) with regularity conditions, including the one where the second moment of the loss difference is bounded. However, the heavy-tailed nature of the financial variables can violate this moment condition. We show that if the moment condition is violated, the size of the test using the classical Normal asymptotics can be heavily distorted. The distortion is large especially when the tail of the marginal distribution of the loss differences is heavy. As an alternative approach, we propose to use a subsampling method (Politis, Romano, and Wolf, 1999) that is robust to fat tails. In the empirical study, we analyze several variance forecast tests by Hansen and Lunde (2006) and Bollerslev, Patton, and Quaedvlieg (2016). Examining several tail index estimators, we show that the second moment of the loss difference is likely to be unbounded especially when the popular squared error (SE) function is used as a loss function. We also find that the outcome of the tests may change if the subsampling is used.

1 Introduction

Forecast comparison tests are widely conducted in finance to evaluate performances of two or more competing forecasts of the financial variable. For example, an equal predictive ability (EPA) test suggested by Diebold and Mariano (1995) and West (1996) is a standard method used to compare two forecasts. White (2000) introduced the reality check (RC) for data snooping, which is a framework to compare multiple forecasts. Hansen (2005) then extended it to a superior predictive ability (SPA) test. Based on these tests, Hansen, Lunde, and Nason (2011) introduced a model confidence set, which is widely applied in the empirical studies. Whereas these tests deal with the null hypothesis defined with the unconditional mean, Giacomini and White (2006) study a setting in which the null hypothesis is defined with a conditional mean. Moreover, Giacomini and White (2006) suggest comparing the forecasting *methods* that include the model selection, the estimation method, the window of the data used for estimation, and so on.

These tests are conducted by computing the difference of losses from each forecast, and the common loss functions include the squared error (SE), absolute error (AE) and so on. In case of Diebold-Mariano-West type test, the null hypothesis is that the mean of the loss difference is zero, implying that the two forecasts perform equally on average. The p-value of the test statistic is computed either by the central limit theorem (CLT) or by the stationary bootstrap (Politis and Romano, 1994). In either case it is necessary to rely on the regularity conditions. Typically it is assumed that the loss difference process is strictly stationary and satisfies a suitable mixing condition with a sufficiently fast rate of mixing as well as existence of $(2+\delta)$th moment for $\delta > 0$.

However, the condition that the loss difference has a finite second moment may be violated in some situations, i.e., the second moment of the loss differences may not exist. Numerous studies indicate that the financial variables, such as financial returns and foreign exchange rates, have heavy-tailed distributions compared with the Normal distribution. Recent empirical studies find that the fourth moment of the financial returns are infinite (see amongst others Loretan and Phillips (1994), Gabaix (2009), Ibragimov (2009) and the references therein). If this is the case, the second moment of the realized variance (RV) can be unbounded.[1]

Noting that the second moment of the realized variance (RV) may be unbounded, we

[1] In a continuous model without jumps and without market microstructure noise, the fourth moment of returns is unbounded if and only if the second moment of the instantaneous variance infinite.

mainly focus our analysis on variance forecast tests. An equal predictive ability (EPA) test for the variance forecasts is typically conducted with the realized variance RV_t and two competing variance forecasts denoted by F_{1t} and F_{2t}, respectively. The loss difference with squared error function is given by

$$d_t = (RV_t - F_{1t})^2 - (RV_t - F_{2t})^2 = F_{1t}^2 - F_{2t}^2 - 2RV_t(F_{1t} - F_{2t})$$

Then, taking the square, we have

$$d_t^2 = (F_{1t}^2 - F_{2t}^2)^2 + 4RV_t^2(F_{1t} - F_{2t}) - 4RV_t(F_{1t} - F_{2t})(F_{1t}^2 - F_{2t}^2)$$

Then it is easy to invent an example where the second moment of the loss difference (d_t) is unbounded. If the third component is negligible and the second moment of the realized variance is unbounded, the second moment of d_t would also be unbounded.

We first review the generalized central limit theorem (CLT) with fat tails. With some assumptions, the asymptotic distribution is given by the Stable distribution (Feller, 1971). Using the results of Davis (1983) and Davis and Hsing (1995), we characterize the asymptotic distribution of the test statistic for the equal predictive ability (EPA) test and we show that it is a ratio of two correlated Stable random variable. We show that, when the left and the right tails are "well balanced", the asymptotic distribution of the test statistics is symmetric with tails similar to the Normal distribution, regardless of the tail index of the loss difference. However, when the tails are not "well balanced", the asymptotic distribution becomes asymmetric and skewed. Especially when the tail is heavier, the tail of the asymptotic distribution is more different from the Normal distribution. In these cases, the size property of a test using the Normal asymptotics is heavily distorted. We also provide an analysis of four important components that determine the tails of the loss difference, namely the tails of variable to be forecast and two competing forecasts and the choice of the loss function. We consider Bregman class and characterize the relation of the four components.

We then propose the subsampling method (Politis, Romano, and Wolf, 1999) as an alternative approach. It is well known that the subsampling is asymptotically valid in face of the fat tails unlike a bootstrap (Athreya (1987), Knight (1987)). We show in the simulation that the choice of the block size has a crucial impact on the validity of the subsampling in the finite sample, and that the appropriate block size depends on the tail index and the skewness parameter of the marginal distribution of the loss difference. We propose two data-driven selection methods that are combined with the minimum volatility method of Romano and Wolf (2001).

Finally, we conduct an empirical study. Motivated by the heavy tail in the realized variance (RV), we focus our analysis on tests to compare the variance forecasts. First, we inspect the estimators for the tail index parameter α and the tail balance parameter p of the loss difference. There are several estimators for α, such as the Hill estimator (Hill, 1975) and the log-log estimator (Gabaix and Ibragimov, 2011). We also obtain estimators under an assumption of the Stable distribution (Fama and Roll (1971) and McCulloch (1986)). Overall, the moment condition is more likely to be violated with the squared error (SE) function rather than the QLIKE function. When the squared error function is utilized, there are cases in which the tail index estimates for the loss difference are below one, indicating that even the first moment of the loss difference may be unbounded and the null hypothesis is not well defined. In addition, the outcome of the test may change when we implement the subsampling instead of the classical methods, such as the stationary bootstrap.

Throughout this paper, we denote by ℓ the slowly varying function at infinity, i.e., ℓ satisfies $\lim_{x\to\infty}\frac{\ell(tx)}{\ell(x)} = 1$ for all $t > 0$. They are not necessarily the same according to the context. Also we use the notation $f_1(x) \sim f_2(x)$ as $x \to \infty$ when the relationship $\frac{f_1(x)}{f_2(x)} \to c < \infty$.

1.1 Forecast comparison test

We consider a variable to be forecast y_t and two different forecasts, F_{1t} and F_{2t} where F_{1t} and F_{2t} are $\mathcal{F}_{t-1}-$measurable. Defining a loss function $L(\cdot)$, the null hypothesis of equal predictive ability test is given by

$$H_0 : \mathbb{E}\,[L(y_t, F_{1t})] = \mathbb{E}\,[L(y_t, F_{2t})].$$

In this paper, we mainly analyze two loss functions, SE and QLIKE, which are widely used in the literature, see Hansen and Lunde (2005), Patton (2011) among others. The two functions are defined as follows.

$$\begin{aligned} &\text{SE} &&: L(y_t, F_t) = (y_t - F_t)^2 \\ &\text{QLIKE} &&: L(y_t, F_t) = y_t/F_t - \log(y_t/F_t) \end{aligned}$$

Defining the loss difference sequence $d_t \equiv L(y_t, F_{1t}) - L(y_t, F_{2t})$, the null hypothesis is equivalent to $H_0 : \mathbb{E}[d_t] = 0$. The test statistic denoted by τ_{hac} is defined as

$$\tau_{hac} = \frac{1/T \sum_{t=1}^{T} d_t}{\sqrt{\hat{\Omega}_T / T}} \tag{1}$$

where $\hat{\Omega}_T$ is the heteroskedasticity and autocorrelation consistent (HAC) estimator of the long-run variance of d_t. With some regularity assumptions, the critical values can be obtained either by a central limit theorem with Normal asymptotics or by the stationary bootstrap (Politis and Romano, 1994). These methodologies are justified under the regularity conditions, and the existence of second moment of d_t is one of them. For the central limit theorem (CLT) with Normal asymptotics are shown in Ibragimov (1962), Oodaira and Yoshihara (1972) and Doukhan (1994), [2] and the conditions for the stationary bootstrap is shown in Hansen (2005).[3]

In this paper, we focus on the case where the second moment of the loss difference is unbounded, but its first moment is bounded. If the marginal distribution of the loss difference has even heavier tails, and the first moment of the loss difference does not exist, the null hypothesis $\mathbb{E}[d_t] = 0$ is not well-defined any more and it is beyond the scope of our analysis.

1.2 Related Literature

This paper is related to the papers which deal with the problems arising from fat tails and the works with the forecast comparison tests, especially the tests for variance forecasts.

There are several papers which consider the fat tails of the conditional variance and the high-frequency data such as the realized variance (RV). Kim and Meddahi (2020) analyze the behavior of the OLS estimator in the volatility regression. They find that the fat tail causes the estimator to converge to a random variable, rather than its true value. They propose an instrumental varriable (IV) estimation which is robust to the fat tails. Ibragimov (2007) studies the efficiency of linear estimators under the fat tails. Hill and Renault (2011)

[2] Let the stationary sequence X_t be centered at expectation and α-mixing. If $\sum_{n=0}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty, \mathbb{E}[X_1^{2+\delta}] < \infty$ for some $\delta > 0$ and $\sigma^2 = \mathbb{E}[X_1^2] + 2\sum_{k=1}^{\infty} \mathbb{E}[X_1 X_k] > 0$, then the sequence of processes $\{S_{[nt]}/\sigma_n : t \in [0,1]\}$ converges in the Skorohod topology to a standard Brownian Motion W on $[0,1]$ where $S_{[nt]} = \sum_{s=1}^{[nt]} X_s, \sigma_n^2 = \mathbb{E}[S_n^2]$.

[3] The vector of relative loss variables, $\{d_t\}$ is (strictly) stationary and α-mixing of size $-(2+\delta)(r+\delta)/(r-2)$ for some $r > 2$, $\delta > 0$ where $\mathbb{E}|d_t|^{r+\delta} < \infty$ and $Var(d_{k,t}) > 0$ for all $k = 1, \cdots, m$.

considers a tail trimming approach. Hill and Prokhorov (2016) considers a GARCH process with heavy tails and proposes a GEL estimator. Other estimation methods with Stable errors are proposed by Blattberg and Sargent (1971), Jurečková, Koenker, and Portnoy (2001), Samorodnitsky, Rachev, Kurz-Kim, and Stoyanov (2007), Andrews, Calder, and Davis (2009), Hallin, Swan, Verdebout, and Veredas (2013), Nolan and Ojeda-Revah (2013) and so on. Also, Mikosch and de Vries (2013) derives explicitly the finite sample expressions for the tail probabilities of the distribution of the OLS estimator, when the noise has heavy tails. Loretan and Phillips (1994) considers a stationarity test for a heavy-tailed time series processes. Ibragimov and Walden (2007) and Ibragimov (2009) analyze the impact of fat tails in portfolio selections or diversifications.

Forecast comparison tests are widely used in econometrics. In the variance forecasting, the problem of using a proxy of the latent variable has been studied in Patton (2011), Liu, Patton, and Sheppard (2015) and Li and Patton (2018). Patton (2020) considers a forecast comparison test with possibly misspecified forecasts. This paper is also related in spirit to Zhu and Timmermann (2020), who consider the Giacomini-White test and study the validity of the null hypothesis under a rolling window scheme. Our analysis with fat tails also concerns the validity of the null hypothesis when the first moment of the loss difference is unbounded.

2 Stable distribution

An α-stable distribution $\mathbb{S}_\alpha(\sigma, \beta, \mu)$ is characterized by four parameters: stable parameter $\alpha \in (0, 2]$, shit parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \geq 0$, and skewness parameter $\beta \in [-1, 1]$. If a random variable X has a Stable distribution, we write $X \sim \mathbb{S}_\alpha(\sigma, \beta, \mu)$ and its characteristic function is the following:

$$\mathbb{E}[e^{iuX}] = \begin{cases} \exp(i\mu u - \sigma^\alpha |u|^\alpha (1 - i\beta \operatorname{sign}(u) \tan(\pi\alpha/2))), & \alpha \neq 1 \\ \exp(i\mu u - \sigma |u| (1 + (2/\pi) i\beta \operatorname{sign}(u) \log |u|)), & \alpha = 1 \end{cases} \tag{2}$$

The special case arises when $\alpha = 2$, and a Stable distribution is a Normal distribution, i.e., $\mathbb{S}_2(\sigma, 0, \mu) = \mathcal{N}(\mu, 2\sigma^2)$. When $\alpha = 1$, it is a Cauchy distribution, i.e., $\mathbb{S}_1(\sigma, 0, \mu)$. When $\alpha < 2$, the second moment does not exist. For details about the Stable distribution, see Ibragimov and Linnik (1971), Brockwell and Davis (1991), Samorodnitsky and Taqqu (1994), Borodin and Ibragimov (1995) and Embrechts, Klüppelberg, and Mikosch (1997).

When β is positive (negative), the distribution is skewed to the right (left) respectively.

When $\beta = 0$, the distribution is symmetric. As α approaches to 2, β loses its effect and the distribution approaches to the Normal distribution regardless of value of β. Table 1 shows the quantile of a random variable $\mathbb{S}_\alpha(1, \beta, 0)$. Observe that when $\mu = 0$ and $\beta > 0$, the distribution is skewed to the right and it is centered in the negative region so that the mean is zero.

When simulating a Stable random variable with $\sigma = 1$ and $\mu = 0$, i.e., $\mathbb{S}_\alpha(1, \beta, 0)$, the following formula can be used from Theorem 3.1 of Weron (1996), or more explicitly from the equation (3.3) of Weron (1995):[4]:

$$\mathbb{S}_\alpha(1, \beta, 0) =^d \left(1 + \beta^2 \tan^2 \frac{\pi\alpha}{2}\right)^{1/(2\alpha)} \frac{\sin \alpha(\gamma - \gamma_0)}{(\cos \gamma)^{1/\alpha}} \left(\frac{\cos(\gamma - \alpha(\gamma - \gamma_0))}{W}\right)^{(1-\alpha)/\alpha} \tag{3}$$

for $\alpha \neq 1$. γ is uniform on $(-\pi/2, \pi/2)$, $\gamma_0 = -1/\alpha \arctan(\beta \tan(\pi\alpha/2))$. W is standard exponential, and γ and W are mutually independent. Note that when $\beta = 0$, the distribution is symmetric and reduced to the formula given by equation (1.7.3) of Samorodnitsky and Taqqu (1994). Then, a Stable random variable $\mathbb{S}_\alpha(\sigma, \beta, \mu)$ can be generated by $\sigma X + \mu$ where $X \sim \mathbb{S}_\alpha(1, \beta, 0)$.

3 Asymptotics under fat tails

This section studies the asymptotic distribution of the test statistics. The classical central limit theorem (CLT) to the Normal distribution is justified when the second moment of d_t is bounded. When the second moment of d_t is unbounded, it is still possible to derive the central limit theorem (CLT) under certain assumptions on the tail behavior. Let us define these assumptions with a notation of $\mathcal{R}(\alpha, p)$.

Definition 1. $\mathcal{R}(\alpha, p)$ Let $\{X_t\}$ be a strictly stationary sequence of random variables with marginal distribution function $F(x)$. Suppose that there exists $\alpha > 0$ such that

$$\mathbb{P}(|X_t| > z) = z^{-\alpha}\, \ell(z) \tag{4}$$

Moreover, the so-called tail-balance condition (Jessen and Mikosch, 2006)

$$\frac{\mathbb{P}(X_t > z)}{\mathbb{P}(|X_t| > z)} \to p, \qquad \frac{\mathbb{P}(X_t < -z)}{\mathbb{P}(|X_t| > z)} \to 1 - p \tag{5}$$

[4]This is related to the equation (2.3) of Chambers, Mallows, and Stuck (1976).

hold as $z \to \infty$ for $p \in [0,1]$. Then we write $X_t \sim \mathcal{R}(\alpha, p)$. We call α the tail index of X_t, and p the tail-balance parameter.

Example 1. Suppose that X is a standardized Stable random variable, $X \sim \mathbb{S}_\alpha(1, \beta, 0)$. When $\beta > -1$, as $x \to \infty$

$$\begin{aligned} \mathbb{P}(X > x) &\sim (1+\beta)C_\alpha \sigma^\alpha x^{-\alpha}, \quad \beta > -1 \\ \mathbb{P}(X < -x) = \mathbb{P}(-X > x) &\sim (1-\beta)C_\alpha \sigma^\alpha x^{-\alpha}, \quad \beta < 1 \end{aligned}$$

where

$$C_\alpha = \left(2\int_0^\infty x^{-\alpha} \sin x dx\right)^{-1} = \frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)$$

holds (Property 1.2.15, Samorodnitsky and Taqqu (1994)). When $\beta = -1$ ($\beta = 1$), the right (left) tail decays faster than a power. Therefore the tail balance parameter p is given as a function of β and it is

$$\frac{\mathbb{P}(X > x)}{\mathbb{P}(|X| > x)} = \frac{(1+\beta)C_\alpha x^{-\alpha}}{((1+\beta)C_\alpha x^{-\alpha}) + ((1-\beta)C_\alpha x^{-\alpha})} = \frac{1+\beta}{2} \tag{6}$$

For the details, see for example Fofack and Nolan (1999).

If $X_t \sim iid.\mathcal{R}(\alpha, p)$ with $\alpha \in (0, 2)$, X_t is said to belong to the domain of attraction of the α-Stable distribution (Davis and Resnick, 1985) and the central limit theorem (CLT) to the Stable distribution (Feller, 1971) holds:

$$\frac{1}{a_T}\sum_{t=1}^{T}(X_t - b_T) \xrightarrow[T\to\infty]{d} \mathbb{S}_\alpha(\sigma, \beta, 0) \tag{7}$$

where $a_T > 0$ and $b_T > 0$ satisfy the following conditions:

$$T\ \mathbb{P}(|X_t| > a_T x) \to x^{-\alpha} \text{ as } T \to \infty, \quad x > 0 \tag{8}$$

$$b_T = \int_{-a_T}^{a_T} x dF(x) \tag{9}$$

The problem of (7) is that the scaling factor a_T is difficult to obtain, since it depends on the tail index α. Motivated by this problem, the asymptotics of the self-normlized sum is often discussed, as it eliminates the scaling series a_T (Logan, Mallows, Rice, and Shepp, 1973). For some $r > \alpha$, the self-normalized sum denoted by $S_T(r)$ is defined as

$$S_T(r) = \frac{\sum_{t=1}^T X_t}{\left(\sum_{t=1}^T |X_t|^r\right)^{1/r}}.$$

Indeed, the denominator and the numerator have the same scaling factor and thus, the term a_T disappears. LePage, Woodroofe, and Zinn (1981) characterize the asymptotics of the self-normalized sum in an i.i.d. setting, and Davis (1983) and Davis and Hsing (1995) extended it in the time series context.

Now we modify our definition of our test statistics from τ_{hac} to τ_{var} so that the results of self-normalized sums are directly applied.

Definition 2. In testing the null hypothesis $\mathbb{E}[d_t] = 0$ against $\mathbb{E}[d_t] > 0$, our test statistic is denoted by τ_{var} and it is given by

$$\tau_{var} = \frac{1/T \sum_{t=1}^T d_t}{(\hat{\sigma}_T/T)^{1/2}}, \quad \hat{\sigma}_T = \frac{1}{T-1} \sum_{t=1}^{T} \left(d_t - \frac{1}{T} \sum_{t=1}^{T} d_t \right)^2 \tag{10}$$

While τ_{hac} considers the long-run variance, τ_{var} has the sample variance in the denominator. When $\alpha \in (0, 2)$, it is easily shown that τ_{var} converges weakly to the same distribution to the self-normalized sum with $r = 2$. The following proposition results from this observation and from Davis (1983).

Assumption 1. *Suppose that $\{d_t\}$ is a strictly stationary process. We assume that $\{d_t\}$ satisfies the following mixing condition. For any choice of integers,*

$$1 \le i_1 < \cdots < i_p < j_1 < \cdots j_q \le n, \quad j_1 - i_p > \ell,$$
$$|\mathbb{E}(U_{i_1 n} \cdots U_{i_p n} U_{j_1 n} \cdots U_{j_q n}) - \mathbb{E}(U_{i_1 n} \cdots U_{i_p n})\mathbb{E}(U_{j_1 n} \cdots U_{j_q n})| \le \alpha_{n,\ell}$$

where $\alpha_{n,\ell}$ is non-increasing in ℓ and $\alpha_{n,\ell(n)} \to 0$ as $n \to \infty$ for some sequence $\ell(n) \to \infty$ with $\ell(n) = o(n)$. This condition is weaker than the usual mixing conditions (Davis, 1983).

We also assume that a local dependence assumption: for all $x > 0$,

$$\limsup_{n\to\infty} S_{k,n}(x) = o(1) \text{ as } k \to \infty \text{where}$$

$$S_{k,n}(x) = n \sum_{j=2}^{n/k} \{\mathbb{P}(X_1 > a_n x, X_j > a_n x) + \mathbb{P}(X_1 > a_n x, X_j \leq -a_n x)$$
$$+\mathbb{P}(X_1 \leq -a_n x, X_j > a_n x) + \mathbb{P}(X_1 \leq -a_n x, X_j \leq -a_n x)\}.$$

Now we are ready to introduce the following proposition, stating the asymptotic distribution of the test statistics under the null hypothesis.

Proposition 1. *Suppose that the loss difference sequence is strictly stationary and $d_t \sim \mathcal{R}(\alpha, p)$ with $\alpha \in (1, 2)$ satisfying the Assumption 1. Then, under the null hypothesis that $\mathbb{E}[d_t] = 0$,*

$$\tau_{var} \quad \xrightarrow[T\to\infty]{d} \quad M(\alpha, p)$$

where $M(\alpha, p)$ is a ratio of two correlated Stable random variables and it is represented by

$$M(\alpha, p) = \frac{\sum_{j=1}^{\infty}(\delta_j Z_j - (2p-1)\mathbb{E}[Z_j \mathbb{I}_{Z_j \in (0,1]}]) - (2p-1)\alpha/(\alpha-1)}{(\sum_{j=1}^{\infty} Z_j^2)^{1/2}} \tag{11}$$

where $\mathbb{1}$ is an indicator function, $(\delta_j)_{j=1,2,\cdots}$ is an i.i.d. random sequence with $\mathbb{P}(\delta_j = 1) = p$ and $\mathbb{P}(\delta_j = -1) = 1 - p$ for all j. $(Z_j)_{j=1,2,\cdots}$ is independent of (δ_j) and for each j, $Z_j = (\sum_{k=1}^{j} E_k)^{-1/\alpha}$ where $(E_k)_{k=1,2,\cdots}$ is a sequence of independent random variable with exponential distribution with mean 1.

Proof. See Theorem 3.1 and Remark 3.1 of Davis and Hsing (1995). Regarding the convergence of types, consult Theorem A1.5 (page 554) of Embrechts, Klüppelberg, and Mikosch (1997). □

Proposition 1 argues that when $\alpha \in (1, 2)$, the limiting distribution of τ_{var} is not a standard Normal distribution under the null hypothesis. It is represented as a ratio of two correlated Stable random variables. The shape of the distribution depends on the tail index α and the tail balance parameter p. Therefore, if $\alpha < 2$ but a researcher wrongly assumes that it is greater than 2, then the test result may suffer from size distortions. If a researcher takes into account the fact, obtaining the critical value is still challenging since the value of (α, p) is not known and thus must be estimated.

Figure 1 shows the approximated density function of $M_{\alpha,p}$ for $\alpha = 1.1, 1.5, 1.9$ and for $p = 0.0, 0.5, 1.0$. When $p = 0$ (the left tail is heavier), the asymptotic distribution is heavily skewed to the right and most of them are positive. On the other hand, when $p = 1.0$ (the right tail is heavier), the asymptotic distribution is heavily skewed to the left and most of them are negative. When $p = 0.5$, they are symmetric around zero.

Table 5 shows some tail characteristics of $M_{\alpha,p}$. The left columns are the quantiles of 1, 5, 95 and 99%. The right columns indicates the p-values of ± 2.32 and ± 1.64. In the classical case where d_t has a finite second moment, the p-values are 1 and 5%. When $p = 0.5$, the p-values are somewhat similar to 1% and 5%, implying that there are mild size distortions. When $p = 0$ and $p = 1$ however, the size is heavily distorted. Suppose that we test the null hypothesis $H_0 : \mathbb{E}[d_t] = 0$ against an alternative hypothesis $H_1 : \mathbb{E}[d_t] > 0$ using a critical value 1.64. When $p = 0$, we reject the null hypothesis for over 75% of time. On the other hand, if the alternative hypothesis is given by $H_1 : \mathbb{E}[d_t] < 0$, then we never reject the null hypothesis.

Proposition 2 introduces the asymptotics under the alternative hypothesis. It shows that the test statistic τ_{var} diverges under the alternative hypothesis, and therefore the test is consistent regardless of the value (α, p). One remark is that the divergence rate can be slower when α is smaller, i.e., when the marginal distribution of d_t has a heavier tail.

Proposition 2. *Suppose that* $d_t \sim \mathcal{R}(\alpha, p)$*, and the test statistic* τ_{var} *is given by* (10)*. Under the alternative hypothesis* $\mathbb{E}[d_t] > 0$*,* $\tau_{var} \xrightarrow{p} \infty$ *and thus the test is consistent. The divergence rate is* $T^{1-1/\alpha}\ell(T)$ *if* $1 < \alpha < 2$ *and* $T^{1/2}$ *if* $\alpha > 2$*, where* $\ell(\cdot)$ *is given by* (4)*. In a special case where* $\ell(T) = 1$*,* τ *diverges more slowly if the value of* α *is smaller.*

4 Tail relations

We have shown possible size distortions when the marginal distribution of the loss difference d_t has a fat tail. In this section, we analyze the relationship between the tails of the loss difference and those of the variable to be forecast (y_t), two competing forecasts (F_{1t} and F_{2t}) as well as the choice of a loss function (L). It is obvious that the choice of a loss function is important and forecast evaluation depends heavily on the choice.

Among many loss functions, we consider a homogeneous Bregman class. A function that belongs to the Bregman class has the form

$$L(y, f) = \phi(y) - \phi(f) - \phi'(f)(y - f),$$

where ϕ is a convex function with subgradient ϕ' (Savage, 1971). Functions of Bregman class are known to be consistent for the conditional mean, i.e., no other quantity leads to a lower expected loss than the conditional mean (Geiting (2011), Patton (2020)), and the squared error (SE), QLIKE functions are examples of Bregman class. Efron (1991) and Patton (2011) then argue that homogeneity or scale invariance is a desirable property of a loss function, i.e., $L(cy, cf) = |c|^b L(y, f)$ holds for all $c \in \mathbb{R}$ and $y, f \in \mathcal{D}$, where $\mathcal{D}$ is the domain of y, f. For example, if $\mathcal{D} = \mathbb{R}$ and $\phi(x) = |x|^a$ with $a > 1$, then the Bregman representation yields the loss function

$$L_a(y, f) = |y|^a - |f|^a - a \times sign(f)|f|^{a-1}(y - f)$$

which is homogeneous of order a and it nests the squared error that arises when $a = 2$. Patton (2011) introduced a rich and flexible family of homogeneous Bregman function on $\mathcal{D} = (0, \infty)$,[5] namely

$$L_a(y, f) = \begin{cases} |y|^a - |f|^a - a|f|^{a-1}(y - f) & \text{if } a \in \mathbb{R} \setminus \{0, 1\} \\ y/f - \log(y/f) - 1 & \text{if } a = 0 \\ y \log(y/f) - y + f & \text{if } a = 1 \end{cases}$$

Clearly, $L_2(y, f)$ is the squared error loss and $L_0(y, f)$ is the QLIKE.

4.1 Homogeneous Bregman with $a > 1$

Suppose that the target variable y_t and two forecasts have the following relation.

$$y_t = F_{1t} + F_{2t} + \epsilon_t$$

where

$$\epsilon_t \sim iid.\mathbb{S}_\alpha, \quad F_{1t} \sim iid.\mathbb{S}_{\alpha_1}, \quad F_{2t} \sim iid.\mathbb{S}_{\alpha_2}$$

and ϵ_t, F_{1t} and F_{2t} are independent each other. In this example, each forecast indicates a part of the conditional mean of y_t, and ϵ_t stands for the unforecastable component. Also, the tail index of y_t is given by $\min\{\alpha, \alpha_1, \alpha_2\}$.

[5] Patton (2011) studies the variance forecast tests, and thus the support of the target variable (conditional variance) and its forecast is given by $(0, \infty)$.

4.1.1 Squared error ($a = 2$)

Suppose that we select the squared error (SE) as the loss function.

$$L(y, F) = (y - F)^2.$$

Then the loss difference d_t is given by

$$\begin{aligned} d_t &= (y_t - F_{1t})^2 - (y_t - F_{2t})^2 = (F_{2t} + \epsilon_t)^2 - (F_{1t} + \epsilon_t)^2 \\ &= F_{2t}^2 - F_{1t}^2 + 2\epsilon_t(F_{2t} - F_{1t}) \end{aligned}$$

Under the independence assumption, we have $\mathbb{E}[d_t^2] < \infty$ if

$$\alpha_1, \alpha_2 > 4, \quad \alpha > 2.$$

In particular, if $\alpha_1 < 4$, $\alpha_2 < 4$ or $\alpha < 2$, then $\mathbb{E}[d_t^2] = \infty$ (see Cline and Samorodnitsky (1994) for example). Note also that $\mathbb{E}|d_t| < \infty$ if

$$\alpha_1, \alpha_2 > 2, \quad \alpha > 1,$$

which is required to construct a meaningful null hypothesis $\mathbb{E}[d_t] = 0$. We show in this very simple example that, if the squared error loss function is utilized, the necessary condition for the second moment of the loss difference to be existent is that the tail index of the target variable y_t is larger than 2 (recall that the tail index of y_t is given by the minimum among $\alpha, \alpha_1, \alpha_2$). Suppose that the tail index of the target variable y_t is below 2, then the second moment of the loss difference is infinite.

4.1.2 Homogeneous Bregman Class with $a > 1$

Now consider a Bregman loss function with $a > 1$

$$L_a(y, f) = |y|^a - |f|^a - a \times sign(f)|f|^{a-1}(y - f).$$

Then the loss difference d_t is given by

$$d_t = L_a(y_t, F_{1t}) - L_a(y_t, F_{2t})$$
$$= |F_{2t}|^a - |F_{1t}|^a + a \times sign(F_{2t})|F_{2t}|^{a-1}(F_{1t} + \epsilon_t) - a \times sign(F_{1t})|F_{1t}|^{a-1}(F_{2t} + \epsilon_t)$$

We can show that $\mathbb{E}[d_t^2] < \infty$ if

$$\alpha_1, \alpha_2 > 2a, \quad \alpha > 2.$$

In particular, we can show that if $\alpha_1 < 2a$, $\alpha_2 < 2a$ or $\alpha < 2$, then $\mathbb{E}[d_t^2] = \infty$. Note also that $\mathbb{E}|d_t| < \infty$ holds if

$$\alpha_1, \alpha_2 > a, \quad \alpha > 1.$$

Note that for any value of a, the necessary condition for the second moment of the loss difference to be existent is that the tail index of the target variable y_t is larger than 2, as in the case of the squared error (SE) function. The moment condition becomes more strict as a increases.

4.2 QLIKE loss function ($a = 0$)

It is more convenient to consider a different relationship between the target variable y_t and the two competing forecasts. Suppose that y_t is given by a product of two forecasts:

$$y_t = F_{1t}F_{2t}$$

and both F_{1t} and F_{2t} are defined on $(0, \infty)$, and they are independent each other. We further assume that $\mathbb{E}[F_{1t}] = \mathbb{E}[F_{2t}] = 1$ for simplicity. We now assume that $F_{it} \sim iid\mathbb{S} + (\alpha)$ with $\alpha > 1$, where $\mathbb{S}+(\alpha)$ is a stable distribution (or similar concept (see, e.g., Davis and Mikosch, 1998)) defined on $(0, \infty)$ with a tail index α. It then follows immediately that

$$\mathbb{E}[d_t] = \mathbb{E}[F_{2t} - F_{1t} - (\log F_{2t} - \log F_{1t})] = 0$$

since F_{1t} and F_{2t} are identically distributed. Moreover, it is easy to see that if $\alpha < 2$, then $\mathbb{E}[d_t^2] = \infty$.

5 Differenciation and cancellation of tails

In this section, we examine if taking the difference of the losses cancels out the heavy tails. Suppose that the target variable y_t is given by

$$y_t = F_t + \varepsilon_t$$

where F_t has a fat tail. Suppose that there are two forecasts, $F_{jt}, j \in \{1,2\}$ that captures the component F_t:

$$F_{jt} = \lambda_j F_t + u_{jt}, \quad \text{for } j \in \{1,2\}$$

Then, under MSE,

$$L(y_t, F_{jt}) = (1-\lambda_j)^2 F_t^2 + (\varepsilon_t - u_{jt})^2 + 2(1-\lambda_j)(\varepsilon_t - u_{jt})F_t$$

and thus the loss difference is given by

$$\begin{aligned} d_t &= (\lambda_1^2 - \lambda_2^2 - 2(\lambda_1 - \lambda_2))F_t^2 + (u_{1t}^2 - u_{2t}^2 - 2(u_{1t} - u_{2t})) \\ &\quad +2\left[(1-\lambda_1)(\varepsilon_t - u_{1t}) - (1-\lambda_2)(\varepsilon_t - u_{2t})\right]F_t \end{aligned}$$

and thus, the fat-tail cancels out when $\lambda_1 = \lambda_2$. However, as long as $\lambda_1 \neq \lambda_2$, the tail of d_t is given by the tail of F_t^2. If the fourth moment of F_t is unbounded, then the second moment of d_t is also unbounded.

6 Subsampling

6.1 Asymptotic theory of subsampling

In the previous section, we show that the use of the central limit theorem (CLT) with the Normal asymptotics can lead to a size distortion if the marginal distribution of $\{d_t\}$ has fat tails. We also show that the degree of distortions depends on the tail index and the tail balance parameter.

Then it is natural to ask if there is a robust alternative way to obtain the critical value, when one thinks that the second moment of the loss difference may be unbounded. One possible way is to use Proposition 1. Once the estimators for (α, p) is obtained, we can obtain the quantiles of $M_{\hat{\alpha},\hat{p}}$ by simulation. However, estimating (α, p) precisely is difficult in the finite sample, and it is more reasonable not to rely on the central limit theorem (CLT) since the quantiles of $M_{\alpha,p}$ are sensitive to the values of (α, p). From this reason, we investigate the subsampling method as an alternative. Subsampling is known to be robust to the fat tails (Politis, Romano, and Wolf, 1999) unlike the bootstrap (Athreya (1987), Knight (1989)). The following assumption and theorem are from Kokoszka and Wolf (2004) which is an extension of Politis, Romano, and Wolf (1999).

Assumption 2. *Suppose that we have observed a sample* $X_1, \cdots, X_n$ *and* $\hat{\theta}_n$ *is the estimator of* θ *and* J_n *is the sampling distribution of* $\tau_n(\hat{\theta} - \theta)/\hat{\sigma}_n$ *where* $\hat{\sigma}_n > 0$*. Set also*

$$J_n(x) = \mathbb{P}\left[\frac{\tau_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n} \le x\right]$$

There are nondegenerate distributions J*,* V*,* W *such that* W *has no mass at the origin, and positive sequences* $\{t_n\}$ *and* $\{u_n\}$ *such that* $\tau_n = t_n/u_n$ *and*

$$J_n \xrightarrow[T\to\infty]{d} J, \quad t_n(\hat{\theta}_n - \theta) \xrightarrow[T\to\infty]{d} V, \quad u_n\hat{\sigma}_n \xrightarrow[T\to\infty]{d} W$$

Theorem 1. *Suppose that the process* $\{X_t\}$ *is strictly stationary and strong mixing, and Assumption 2 holds. For some integer* $b < n$*, define*

$$L_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} I\left\{\frac{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n)}{\hat{\sigma}_{n,b,t}} \le x\right\}$$

where $\hat{\theta}_{n,b,t}$ *is the statistic* $\hat{\theta}$ *evaluated at the data set* $(X_t, X_{t+1}, \cdots, X_{t+b-1})$*, and* $\hat{\sigma}^2_{n,b,t}$ *is the sample variance of the same data set. Also, assume that*

$$b \to \infty, \quad \frac{b}{n} \to 0, \;\; \frac{\tau_b}{\tau_n} \to 0, \;\; \frac{t_b}{t_n} \to 0 \quad \textit{hold.}$$

Then, the following conclusions hold:

(i) If x *is a continuity point of* $J(\cdot)$*, then* $L_{n,b}(x) \xrightarrow{p} J(x)$

(ii) If $J(\cdot)$ *is continuous, then* $\sup_x |L_{n,b}(x) - J(x)| \xrightarrow{p} 0$

(iii) Denote

$$c_{n,b}(1-\alpha) = \inf\{x : L_{n,b}(x) \ge 1-\alpha\} \textit{ and } c(1-\alpha) = \inf\{x : J(x) \ge 1-\alpha\}$$

If $J(\cdot)$ is continuous at $c(1-\alpha)$, then

$$\mathbb{P}\left[\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \leq c_{n,b}(1-\alpha)\right] \to 1-\alpha$$

i.e., the subsampling confidence intervals yield asymptotically correct coverage probability.

Using Theorem 1, we introduce two corollaries which make clear the procedure to obtain the critical values with subsampling. In practice, it is a challenge to verify that the loss difference d_t satisfies Assumption 2. The case of independent observations are studied in paper 11 of Politis, Romano, and Wolf (1999). McElroy and Politis (2002) study this assumption in the case of serial correlation, and Kokoszka and Wolf (2004) extend it to the GARCH-type processes.

Corollary 1. *Suppose that Assumption 2 and other assumptions in Theorem 1 are satisfied for the process $\{d_t\}$. With a data set $(d_1, \cdots, d_n)$, define*

$$\hat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} d_t, \quad \tau_n = \sqrt{n}$$

and $\hat{\sigma}_n$ is the sample standard deviation of d_t. Then the critical value associated with a block size b for a test with

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) > 0$$

of level $q \in [0,1]$ can be obtained by the $1-q$ sample quantile of $\left\{\frac{\tau_b(\hat{\theta}_{n,b,t})}{\hat{\sigma}_{n,b,t}}\right\}$ with $t = 1, \cdots, (n-b+1)$.

Corollary 2. *Suppose that Assumption 2 and other assumptions in Theorem 1 are satisfied for the process $\{d_t\}$ with*

$$\hat{\theta}_n = \frac{1}{n}\sum_{t=1}^{n} d_t, \quad \tau_n = \sqrt{n}$$

and $\hat{\sigma}_n$ is the sample standard deviation of d_t. Then the critical value associated with a block size b for a test with

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) \neq 0$$

of level $q \in [0,1]$ can be obtained by equal-tailed confidence interval and symmetric confidence interval (paper 3, page 72 of Politis, Romano, and Wolf (1999)). The rejection rule given *the equal-tailed confidence interval is given by*

$$\textit{Reject the null if } \hat{\theta}_n < c_{n,b}\left(\frac{\alpha}{2}\right) \textit{ or } \hat{\theta}_n > c_{n,b}\left(1-\frac{\alpha}{2}\right)$$

and the rejection rule given the symmetric confidence interval is given by

$$\textit{Reject the null if } |\hat{\theta}_n| > c_{n,b,|\cdot|}(1-\alpha)$$

where

$$c_{n,b,|\cdot|} = \inf\{x : L_{n,b,|\cdot|}(x) \geq 1-\alpha\}$$

and

$$L_{n,b,|\cdot|}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbb{1}\{\frac{\tau_b(|\hat{\theta}_{n,b,t} - \hat{\theta}_n|)}{\hat{\sigma}_{n,b,t}} \leq x\}$$

6.2 Finite sample properties of subsampling

Suppose that we have a sample of loss difference $d_1, d_2, \cdots, d_T$. Theorem 1 states that the subsampling-based critical values are valid asymptotically if the block size b is chosen such that $b \to \infty$ and $b/T \to 0$ as $T \to \infty$. However, it does not provide the finite sample performance with a choice of the block size. In this section, we analyze the finite sample properties of the subsampling by the Monte Carlo simulations. We consider a simple setting where $d_t \sim iid.\mathbb{S}_\alpha(\sigma, \beta, \mu)$, and find that the choice of a block size is crucial on the performance in the finite sample, and the appropriate block size depends on the tail index α and the skewness parameter β.

We then propose two ways of selecting the block size which are combined with the minimum volatility method of Romano and Wolf (2001). These algorithms work in our simulations. Regarding the power property, we find that the rejection rate is high under the alternative hypothesis when α is larger, but it remains low when α is small, which confirms the slow rate of divergence in Proposition 2.

The simulation procedure is the following. We generate 2000 replications of d_t independently from the Stable distribution using the formula (3) with the sample size $T \in \{250, 500, 1000, 2500\}$. We vary the parameter values such that $\alpha \in \{1.1, 1.3, 1.6, 1.9\}$, $\beta \in \{-1, -0.5, 0, 0.5, 1\}$. We fix the value of σ to be 1 and we set $\mu = 0$ when we impose the null hypothesis. For each replication, we decide whether we reject the null hypothesis or not with a level of 5%. Most of the case, we consider the one-sided test with

$$H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] > 0$$

The alternative hypothesis is imposed by setting $\mu = 100/T$ and $\mu = 500/T$ in order to study the local power.

6.2.1 Benchmark case: classical Normal asymptotics

Consider a test with the null and the alternative hypothesis given as follows:

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) > 0$$

and the test statistics τ_{var} is given by equation (10). As a benchmark, we compute the rejection rate when the classical Normal asymptotics is used, i.e., to conduct a 5% level test, we reject the null hypothesis when $\tau_{var} > 1.64$. Table 6 shows the rejection rate from the simulation. It confirms our theoretical findings shown in Table 5. When α is small, the rejection frequency is far from the targeted value of 5%, and the distortion is larger when β is larger in the absolute term. When β is negative, the null hypothesis is rejected too often, and when β is positive the rejection frequency is too low.

Table 7 indicates the rejection rate in case of two-sided test, i.e., the alternative hypothesis is given by $\mathbb{E}(d_t) \neq 0$. In this case, the rejection rate is symmetric for the value of β. We find similar values of rejection rate for the same value of $|\beta|$, and we tend to over reject the null hypothesis.

6.2.2 Subsampling with fixed blocksize

We still consider a test with the following null and the alternative hypotheses:

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) > 0$$

We then conduct the subsampling with fixed values of block sizes. When $T = 250$, the block sizes are from 20 to 200 with an increment of 5. When $T = 500$, the block sizes are from 20 to 380 with an increment of 10. When $T = 1000$, they are from 20 to 740 with an increment of 20. When $T = 2500$, they are from 20 to 2180 with an increment of 60.

For each replication $\{d_t^{(s)}; t = 1, \cdots, T\}$ and for each block size b, we conduct a 5%-

level test by following the procedure of Corollary 1. The results are shown in Figure 2. It shows the rejection frequencies as a function of the block size. In each case, the rejection frequencies are stable near the region of 5%, and then it increase as the block size increases. When β is negative, i.e., the distribution of d_t is skewed to the left, the rejection rate increases rapidly as the block size increases, and it grows even faster with smaller values of α. We also find that, when $\beta \geq 0$, the rejection rate is still increasing in the block size, but it is less sensitive than the case of $\beta < 0$. In either case, the size property of the test is heavily dependent on the choice of the block size.

6.2.3 Appropriate block sizes

In the previous section, we observed that the size of the test is heavily dependent on the choice of the block size in the finite sample. It is therefore important to carefully select the appropriate block size. For this purpose, we assess the bounds of block size which leads a desirable rejection rate. We run a simulation with 5,000 replication where we conducted a one-sided test of level 5%:

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) > 0$$

We consider $\alpha \in \{1.1, 1.3, 1.5, 1.7, 1.9\}$ and $\beta \in \{-1, -0.5, -0.25, 0, 0.25, 0.5, 1\}$.[6]

We define the *appropriate* block sizes as the one with which the rejection rate under the null hypothesis falls between 0.04 and 0.06. With this definition, we derive the bounds of the appropriate block sizes from the simulation for each (α, β). Table 8 summarizes the lower bound and the upper bound.

Figure 3 depicts the relation of the appropriate block size bound and the value of β. First observation is that, when α is small, the block size is dependent on the value of β. As α approaches to 2, the block size becomes more independent of the value of β. It is quite intuitive since the skewness parameter β loses its meaning as α approaches to 2. More interestingly, we find that the appropriate block size is increasing in β for any value of α.

In order to understand the reason why the appropriate block size is increasing in β, let us suppose that $\beta = -1$. Then the left tail of the distribution of d_t is heavier than the right tail, and its median is positive. The sample path $\{d_t\}$ of size T is then likely to include a few

[6] The considered block sizes are from 20 to 240 with an increment of 5, and $b = 241, 242, 243$ in case of $T = 250$. When $T = 500$, they are from 20 to 480 with an increment of 10, and $b = 490$. When $T = 1000$, they are from 20 to 960 with an increment of 20. When $T = 2500$, they are from 20 to 2352 with an increment of 53 and $b = 2410, 2420, 2430$.

observations of large negative values and many observations of small values in the absolute term. When we take a subsample of size $b < T$, the test statistics from each subsample can be heavily influenced by whether at least one of the few large negative observations is included or not. If they are included in one subsample, it is likely that the test statistic of this subsample is pulled to the negative value. Also note that as the block size increases, the number of subsamples including the large negative observations increases as well.

Now, the test statistics located at the neighborhood of 95% quantile of its asymptotic distribution is likely to be driven by many small and positive values of d_t.[7] It is because the test statistics have the sample mean of d_t in the numerator, and its denominator is proportional to the sample standard deviation. Given one sample path $\{d_t\}$ of size T, these test statsitics cannot be replicated if the subsample includes the large negative observations. Therefore, we need a small number of block size so that majority of the subsampled test statistics are computed without the observations of large negative values.

Figure 4 depicts the relation of the appropriate block size bound and the value of α. We notice that the appropriate block size is increasing in α when β is negative or zero. On the other hand when β is positive, i.e., the distribution is skewed to the right, the optimal block size should decreasing in α.

6.2.4 Block size setting rule

We have conducted the subsampling with fixed block sizes to test the null hypothesis:

$$H_0 : \mathbb{E}(d_t) = 0, \quad H_1 : \mathbb{E}(d_t) > 0$$

and show that the optimal block size depends on the tail index parameter α and the skewness parameter β. Regarding the choice of the block size, Romano and Wolf (2001) propose a so-called "minimum volatility method" (Politis, Romano, and Wolf, 1999). This method is conducted by the algorithm below.

1. For each $b \in [b_{min}, b_{max}]$, compute the critical values, c_b.
2. For each b, compute the volatility index VI_b as the standard deviation f the critical values in a neighborhood of b. More specifically, for a small integer k, VI_b = sample standard deviation of $\{c_{b-k}, c_{b-k+1}, \cdots, c_{b+k-1}, c_{b+k}\}$.
3. Pick the value b^* such that $b^* = \arg\min_b VI_b$

[7] Recall that we are conducting an one-sided test.

In this paper, we adopt this methodology, and the question becomes how to select the values of b_{min} and b_{max}. We consider the following two methods, for which the estimated values, $\hat{\alpha}$ and $\hat{\beta}$ are used to extract the information of (α, β). In the simulation, we adopt the Quantile-based method (QM) of McCulloch (1986).

The first method is to find the values of b_{min} and b_{max} from Table 8 which indicates the appropriate block size bounds from the simulation. Once $\hat{\alpha}$ and $\hat{\beta}$ are obtained, we can find the optimal block size from this table by a linear interpolation. The second approach is to derive a formula to obtain b_{min} and b_{max} as a function of α and β. Again, the estimated parameter values can be plugged in in reality. Based on the findings in the previous section, we propose the following formula.

$$b_{min} = C_{min} T^{0.33}, \quad C_{min} = (\beta + 2)\alpha \tag{12}$$

$$b_{max} = C_{max} T^{0.66}, \quad C_{max} = 0.5(\beta(2 - \alpha) + 2)\alpha^2 \tag{13}$$

In this way, b_{min} and b_{max} increases with β but its effect decreases as α approaches to 2.

Table 9 indicates the rejection frequencies when we follow the first method of block selection. We find that the rejection frequencies are overall larger than 5% but we have obtained much improvements compared to the Normal asymptotics as shown in Table 6. Then we look at the power property. We impose the alternative hypothesis by setting the mean as $100/T$ and $500/T$. When $\mathbb{E}(d_t) = 100/T$ as shown in Table 10, the rejection frequencies are larger as α increases. When $\alpha = 1.1$ however, the rejection frequency is almost the same as under the null hypothesis. This indicates that the divergence rate of the test statistics is slow when the tail is very heavy. Table 11 shows the rejection rate when the mean of d_t is given by $500/T$. In this table, the rejection rates are larger than Table 10, and they reach over 90% when $\alpha = 1.9$. However when $\alpha = 1.1$, the rates are still low, indicating a poor power.

Table 12 and Table 13 demonstrates the rejection frequencies when we follow the second method of block size selection. Table 12 is the one when the true values of (α, β) are used, whereas Table 13 corresponds to the case where estimated values of (α, β) are used. Overall, it has a good size property.

Regarding the power, we find a similar pattern as we witnessed with the previous analysis. As shown in Table 14 where the mean of d_t is given by $100/T$ and in Table 15 where $\mathbb{E}(d_t) = 500/T$, the rejection rates are higher when α is larger. However when $\alpha = 1.1$, the rejection rate remains low, indicating the difficulty in correctly rejecting the null hypothesis.

7 Empirical studies

In the previous sections, we find that the classical testing procedures using the central limit theorem with the Normal asymptotics may lead to a severe size distortion when the second moment of the loss difference is unbounded, and that the subsampling method is robust to the fat tails. By simulations we showed that the subsampling approach is valid in the finite sample as long as the block size is properly chosen, and we proposed two data-driven ways to select the block size. In this section, we apply these findings to the data.

We focus our analysis on the variance forecast tests. In the financial econometrics, the conditional variance is one of the main interest since it "is empirically the dominant time-varying characteristic of the distribution" (Andersen, Bollerslev, Diebold, and Labys, 2003). There exist several modeling approach to the conditional variance, such as ARCH (Engle, 1982) or GARCH (Bollerslev, 1986) type models, the stochastic volatiltiy models (Nelson (1990), Drost and Werker (1996)) and so on. Different models produce different conditional variance forecasts and thus, the evaluation of the variance forecasts is important. Evaluating variance forecasts has been much improved by the emergence of the high-frequency data such as the realized variance (RV) which is defined as the sum of intra-daily squared returns (Andersen and Bollerslev, 1998). However, the validity of the moment condition in the variance forecast comparison tests is not discussed in the literature. As is discussed in the introduction, if the fourth moment of the returns is unbounded, then the second moment of the daily realized variance is also unbounded, which may make the second moment of the loss difference unbounded.

First we review several estimators for the tail index. Then, we study the variance forecast tests by Hansen and Lunde (2005) and Bollerslev, Patton, and Quaedvlieg (2016)

7.1 Review of tail index estimators

Suppose that X_t is a non-negative variable that satisfies $\mathbb{P}[|X_t| > x] \sim Cx^{-\alpha}$ as $x \to \infty$, and the data set $\{X_t; t = 1, \cdots, T\}$ is available. Also denote by $\{X_{(i)}\}_{i=1}^T$ the order statistics, with $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(T)}$.

7.1.1 Hill estimator

The Hill estimator arises in the i.i.d. context as a conditional maximum likelihood estimator (MLE) and it is defined as

$$\hat{\alpha}(k_n) = \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{X_{(i)}}{X_{(k_n)}} \right)^{-1} \tag{14}$$

$k_n \leq T$ is chosen so that $k_n \to \infty$ and $k_n/T \to 0$ as $T \to \infty$. The results by Hsing (1991) and Resnick and Stărică (1995) indicate that the Hill estimator is asymptotically robust with respect to the deviations from independence; Resnick and Stărică (1998) prove consistency under ARCH-type dependence. See also Hill (2010) for some other processes including ARFIMA, FIGARCH, explosive GARCH, nonlinear ARCH-GARCH and so on.

Valid standard errors of the Hill estimator are available only for some specific models with serial correlation. Therefore we do not provide any of them and we follow the literature by providing Hill's plots, that is by varying the integer k_n. A flat area is viewed as a good estimator of the tail index.

7.1.2 Log-log estimator

Another estimator of the tail index is introduced by Gabaix and Ibragimov (2011). They showed the consistency of the log-log rank-size regression with a parameter $\gamma > 0$:

$$\log(t - \gamma) = a - b \log X_{(t)}, \quad t = 1, \cdots, n$$

and the tail index estimator is given by $\hat{b}_n$. They also show that the minimum bias is obtained when $\gamma = 1/2$. In an i.i.d. setting the asymptotic expansion hold $\hat{b}_n = \alpha + \alpha\sqrt{2/n}\,\mathcal{N}(0,1) + O_p(\log^2 n/n)$. However they do not show it for dependent series. They also do not discuss on how to select the truncation size n.

7.1.3 Moment estimator

Another estimator of the tail index is introduced by Dekkers, Einmahl, and De Haan (1989), denoted by $\hat{\alpha}_M = \hat{\gamma}_M^{-1}$ where

$$\hat{\gamma}_M = M_{k_n}^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_{k_n}^{(1)})^2}{M_{k_n}^{(2)}} \right)$$

where $M_k^{(j)} = \frac{1}{k}\sum_{j=1}^{k}(\log X_{(j)} - \log X_{(k+1)})^j$. Under an i.i.d. setting and $\gamma \in \mathbb{R}$, $\hat{\alpha}_M$ is a consistent estimator if $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

7.1.4 Bias reduced estimator

Another estimator is introduced by Feuerverger and Hall (1999), and then simplified by Gomes and Martins (2004) denoted by $\hat{\alpha}_{br} = \hat{\gamma}_{br}^{-1}$ where

$$\hat{\gamma}_{br} = \frac{1}{k_n}\sum_{j=1}^{k_n} U_j + \left(\frac{1}{k_n}\sum_{j=1}^{k_n} jU_j\right)\frac{\sum_{j=1}^{k_n}(2j - k_n - 1)U_j}{\sum_{j=1}^{k_n} j(2j - k_n - 1)U_j}$$

7.1.5 Stable assumption

Another approach to estimate the tail index is to assume that the marginal distribution of X_t is given by the Stable distribution, i.e., $X_t \sim \mathbb{S}_\alpha(\sigma, \beta, \mu)$. The useful feature of this approach is that the estimator for the skewness parameter β provides the estimates of the tail balance parameter p as in (5) by $\hat{p} = \frac{1+\hat{\beta}}{2}$. Since the size distortion of the Normal asymptotics is larger when p is different from 0.5, examining the estimates for the tail balance parameter is useful. In the empirical study, we apply the quantile-based method (QM) proposed by McCulloch (1986).

Estimation of the four parameters of the Stable distribution is a challenge because the density function is not known except for some special cases, and thus methods such as the maximum likelihood (ML) are difficult to employ. However there are several estimation methods, using the sample quantile (Fama and Roll (1971), McCulloch (1986)), numerical approximation of the likelihood function (DuMouchel (1973), Brorsen and Yang (1990)), or the characteristic function (Koutrouvelis, 1980). In this paper, we mainly consider Quantile Method (QM). Weron (1995) finds that the quantile based method by McCulloch (1986) performs well and can be computed fast as long as the tail index is greater than or equal to 0.6.

7.1.6 Quantile Method

Fama and Roll (1971) provided estimators of parameters of symmetric Stable distribution with $1 < \alpha \leq 2$, $\beta = 0$ and $\mu = 0$. Denoting by the fth population quantile by q_f and its sample counterpart by $\hat{q}_f$, the estimate for σ is given by

$$\hat{\sigma} = \frac{\hat{q}_{0.72} - \hat{q}_{0.28}}{1.654}$$

and $\hat{\alpha}$ is obtained such that $\frac{\hat{q}_f - \hat{q}_{1-f}}{2\hat{\sigma}}$ is the fth quantile of $\mathbb{S}_{\hat{\alpha}}(1, 0, 0)$. They find that $f = 0.95, 0.96, 0.97$ works.

Since the estimators of Fama and Roll (1971) has asymptotic bias in $(\hat{\alpha}, \hat{\sigma})$ and it has restricted to the case where $\beta = 0, \mu = 0$, McCulloch (1986) improved their method. He provided consistent estimators for the four parameters with $0.6 \leq \alpha \leq 2$. Define

$$v_1 = \frac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}}, \quad v_2 = \frac{q_{0.95} + q_{0.05} - 2q_{0.50}}{q_{0.95} - q_{0.05}}$$

and let $\hat{v}_1$ and $\hat{v}_2$ be the sample analogue. He provides tables of ϕ_α and ϕ_β such that

$$\hat{\alpha} = \phi_\alpha(\hat{v}_1, |\hat{v}_2|), \quad \hat{\beta} = sign(v_2)\phi_\beta(\hat{v}_1, |\hat{v}_2|)$$

and we use the linear interpolation to apply the table. He also provides a table of $\phi_\sigma(\alpha, \beta) = (q_{0.75} - q_{0.25})/\sigma$, and the estimates for σ is obtained by

$$\hat{\sigma} = \frac{q_{0.75} - q_{0.25}}{\phi_\sigma(\hat{\alpha}, |\hat{\beta}|)}$$

Again, the linear interpolation is used to apply the table. Tables of ϕ_α, ϕ_β and ϕ_σ are demonstrated in Table 2, Table 3 and Table 4.

7.1.7 Characteristic function

Garcia, Renault, and Veredas (2011), Carrasco and Florens (2002). Also consider Pickands estimators. The simpler version of Carrasco and Florens (2002) is Feuerverger and McDunnough (1981).

Suppose that we have a iid. data $\{X_1, \cdots, X_n\}$. The empirical characteristic function is given by

$$\psi_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}$$

Carrasco and Florens (2002) considers a continuum of moment conditions

$$h(t, X_j; \theta) = e^{itX_j} - \psi_\theta(t)$$

where ψ_θ is the true characteristic function of X_j, and the moment condition is given by $\mathbb{E}^\theta[h(t, X_j; \theta)] = 0$ for all t.

The simpler version is introduced by Feuerverger and McDunnough (1981). Define a random variable $y_{jt} = \exp(itX_j)$, we can characterize their moments b

$$\mathbb{E}(y_{jt}) = \psi(t), \quad \mathbb{E}(y_{jt}y_{js}) = \psi(t+s)$$

Applying CLT with $t = (t_1, \cdots, t_k)$,

$$\sqrt{n}\left(\underbrace{\frac{1}{n}\sum_{j=1}^{n} y_{jt}}_{=\psi_n(t)} - \psi(t)\right) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, \Omega)$$

where Ω is characterized with

$$cov(y_{js}, y_{jt}) = \mathbb{E}(y_{js}\overline{y_{jt}}) - \mathbb{E}(y_{js})\mathbb{E}(\overline{y_{jt}}) = c(s-t) - c(s)c(-t)$$

Defining

$$z_n = \begin{bmatrix} Re\ \psi_n(t_1) \\ \vdots \\ Re\ \psi_n(t_k) \\ Im\ \psi_n(t_1) \\ \vdots \\ Im\ \psi_n(t_k) \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum \cos(t_1 X_j) \\ \vdots \\ \frac{1}{n}\sum \cos(t_k X_j) \\ \frac{1}{n}\sum \sin(t_1 X_j) \\ \vdots \\ \frac{1}{n}\sum \sin(t_k X_j) \end{bmatrix}$$

and

$$z(\theta) = \mathbb{E}\begin{bmatrix} \cos(t_1 X_j) \\ \vdots \\ \cos(t_k X_j) \\ \sin(t_1 X_j) \\ \vdots \\ \sin(t_k X_j) \end{bmatrix} = \begin{bmatrix} Re\ \psi(t_1) \\ \vdots \\ Re\ \psi(t_k) \\ Im\ \psi(t_1) \\ \vdots \\ Im\ \psi(t_k) \end{bmatrix} = \begin{bmatrix} z_1(t_1;\theta) \\ \vdots \\ z_1(t_k;\theta) \\ z_2(t_1;\theta) \\ \vdots \\ z_2(t_k;\theta) \end{bmatrix}$$

whee z_1 and z_2 are defined by

$$z_1(t;\theta) = \exp(-\sigma^\alpha |t|^\alpha)\cos(\Phi), \quad z_2(t;\theta) = \exp(-\sigma^\alpha |t|^\alpha)\sin(\Phi)$$

with

$$\Phi = \begin{cases} t\mu + \sigma^\alpha |t|^\alpha \beta sign(t) \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ t\mu - \sigma^\alpha |t|^\alpha \beta sign(t) \frac{2}{\pi} \log|t| & \alpha = 1 \end{cases}$$

Then, we have the CLT

$$\sqrt{n}(z_n - z(\theta)) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, \Omega_z)$$

where Ω_z is the variance covariance matrix of $[\cos(t_1 X_j), \cdots, \cos(t_k X_j), \sin(t_1 X_j), \cdots, \sin(t_k X_j)]$. Specifically,

$$\begin{aligned} cov(\cos(sX_j), \cos(tX_j)) &= \frac{z_1(s+t) + z_1(s-t)}{2} - z_1(s)z_1(t) \\ cov(\cos(sX_j), \sin(tX_j)) &= \frac{z_2(s+t) - z_2(s-t)}{2} - z_1(s)z_2(t) \\ cov(\sin(sX_j), \sin(tX_j)) &= \frac{z_1(s-t) - z_1(s+t)}{2} - z_2(s)z_2(t) \end{aligned}$$

Using this, the estimator for θ can be obtained by

$$\hat{\theta} = \arg\max_\theta \left\{ -\frac{1}{2}\log(\det\Omega) - \frac{n}{2}(z_n - z_\theta)'\Omega^{-1}(z_n - z_\theta) \right\}$$

7.2 Variance forecasts

We take the data from two papers, Hansen and Lunde (2005) and Bollerslev, Patton, and Quaedvlieg (2016). In both papers, they conduct the Superior Predictive Ability (SPA) test (Hansen, 2005) for the variance forecasts. They obtain the p-value by the stationary bootstrap (Politis and Romano, 1994) with an assumption that the second moment of the loss difference is bounded. The loss functions we study in this paper are SE and QLIKE,

$$\begin{aligned} &\text{SE} &&: L(y, F) = (y - F)^2 \\ &\text{QLIKE} &&: L(y, F) = y/F - \log(y/F) \end{aligned}$$

These are commonly used in the variance forecasting literature, since they are "robust" to the measurement error (Patton, 2011). The evaluation of the variance forecast is done with a proxy of the conditional variance, such as the squared daily returns or the realized variance (RV). The "robust" loss functions ensure that using the proxy leads to the correct ranking of the variance forecasts as long as the proxy is unbiased.

QLIKE is preferred by many researchers since "the average QLIKE loss will be less affected (generally) by the most extreme observations in the sample (Patton, 2011)." However QLIKE loss function involves the ratio of the target variable and the forecast, and thus the existence of the second moment is not evident. For example the student-t distribution can be represented as a ratio of two independent random variables, both of which have thin tails.

Then, we compare the p-values for equal predictive ability (EPA) tests, using (i) the Normal asymptotics, (ii) the stationary bootstrap and (iii) the subsampling. Since the superior predictive ability (SPA) test is more complicated, we consider the pair-wise equal predictive ability test for several pairs of forecasts.

7.3 Hansen and Lunde (2005)

Hansen and Lunde (2005) compare 330 ARCH-type models and test if their benchmark model, GARCH(1,1) is outperformed by other models. They conclude that GARCH(1,1) is outperformed by models which take into account the leverage effect, and the best performing model is A-PARCH(2,2) of Ding, Granger, and Engle (1993).

We take their data on IBM stock returns from June 1, 1999 through May 31, 2000 (T=254). Figure 5 shows the time series of the realized variance (RV) and two forecasts based on GARCH(1,1) and A-PARCH(2,2). The realized variance fluctuates more than the forecasts especially in the period of April 2000. When the realived variance is high, the forecast based on A-PARCH(2,2) tends to be higher than that of GARCH(1,1).

In this paper, we select 8 variance models out of 330 models they consider in the paper, as summarized in Table 16. We include the GARCH(1,1) model with a zero mean and the Normal innovation as the benchmark. The "student t", "Constant mean" and "GARCH-M" specifications are based on the GARCH(1,1) model but each of them is different from the benchmark model in that the innovation follows the student t distribution ("student t"), that the non-zero constant conditional mean is included ("Constant mean") and that the conditional mean is proportional to the conditional variance ("GARCH-M"). We also consider four other variance specifications, ARCH(1), GJR-GARCH(2,2) of Glosten, Jagannathan, and Runkle (1993), NGARCH(2,2) defined as the A-PARCH model without the leverage effect, and A-PARCH(2,2) model. These four models are specified with a zero mean and a Normal innvation.

7.3.1 Tail index estimators

Figure 6 shows the Hill and the log-log estimators for the realized variance (RV), with the truncation parameter $k_n \in [20, 120]$. The log-log estimator is more stable than the Hill estimator, and its value is around 2.5. The Hill estimator is stable around the value of 2. Noting that the second moment exists if the tail index being greater than 2, the existence of the second moment of the realized variance is not obvious.

Figure 7 and Figure 8 depict the Hill and the log-log estimators for the loss sequence $L(RV_t, F_t)$ where the left panels show the estimates with the squared error (SE), and the right panels with the QLIKE function. When the squares error (SE) function is utilized, the estimates are not sensitive to the value of the threshold k_n, and they are below two. This pattern is observed for all eight different forecasts. This phenomenon is consistent with our analysis with the squared error loss function. When a variable to be forecast has unbounded second moment, then using the squared error (SE) can lead to the violation of the moment condition. When the QLIKE function is used, the estimates are more sensitive to the value of the threshold k_n, and the estimates are around the value of 2. From these plots, we find that the tail index estimates tend to be larger with the QLIKE function rather than the squared error (SE) function. Yet the existence of the second moment is not obvious even if the QLIKE function is used.

Figure 9 and Figure 10 depict the Hill and the log-log estimators with the loss difference $d_t = L(RV_t, F_{1t}) - L(RV_t, F_{jt})$ where F_{1t} is the forecast from the benchmark model given by GARCH(1,1). The other forecasts F_{jt} are the froecasts based on the other models specified in Tbale 16. We find the similar pattern as we found in Figure 7 and Figure 8. When the squared error is utilized, the tail index estimates of the loss difference tend to be smaller than two, indicating that the existence of the second moment is questionable. On the other hand, with the QLIKE function, the violation of the moment condition is less obvious.

Finally, Table 17 summarizes the tail index and the skewness parameter estimates based on an assumption that each sequence is i.i.d. and the Stable distribution. Focusing on the loss difference sequence, the tail index estimators ranges from 1.00 to 1.24 (with squared error) and from 1.03 to 1.27 (with QLIKE), indicating unbounded second moments. Moreover, the estimates for the skewness parameter β range from -0.52 to 0.29 (with squared error) and from -0.70 to 0.52 (with QLIKE), indicating unbalanced tails.

7.3.2 Pair-wise Equal Predictive Ability (EPA) test.

We conduct the equal predictive ability (EPA) test for the selected pairs of forecast and study the difference of the p-value when (i) Normal asymptotics, (ii) the stationary bootstrap and (iii) the subsampling are used. When the subsampling is used, the block size is selected either by the formula (12) and (13) or by referring to Table 8. We compute the p-values for the following two one-sided tests:

$$\begin{aligned} &\text{EPA-1:} \quad H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] > 0 \\ &\text{EPA-2:} \quad H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] < 0 \end{aligned}$$

where $d_t = L(RV_t, F_{1t}) - L(RV_t, F_{jt})$, $j = 2, 3, \cdots, 8$. F_{1t} refers to the forecast based on the benchmark GARCH(1,1) model. Other forecasts F_{jt} are the forecasts based on the other models specified in Table 16.

When the Normal asymptotics are used, we compute the test statistic using the long-run variance with the Bartlett Kernel of Newey and West (1987). Likewise, we conduct the stationary bootstrap using the test statistic with the long-run variance. The bootstrap resampling is 5000 with an average block length of 5.

The p-values for the pair-wise equal predictive ability (EPA) tests are demonstrated in Table 18 in which the squared error (SE) function is used as the loss function. The p-values are different across the methodology of the test. Especially when we compare the GARCH(1,1) model and the constant mean model, the Normal asymptotics and the stationary bootstrap suggest to reject the null hypothesis at level 10%, whereas the subsampling-based approach suggests not to reject the null hypothesis. A similar disagreement is observed in the EPA-2 test with student t specification.

Table 19 shows the p-values when the QLIKE is used as the loss function. In this test, the p-values are sensitive to the selection of the block size. In the EPA-1 test with the constant mean model, the p-value based on the block-size selection with the formula leads to the p-value of 24.3%, whereas it is 2.9% when the block size is selected according to Table 8. It is because the selected block sizes are different across the block size selection.

7.4 Bollerslev, Patton, and Quaedvlieg (2016)

Bollerslev, Patton, and Quaedvlieg (2016) propose an extension of HAR model (Corsi, 2009), called a HARQ model which incorporates the realized quarticity (RQ). Taking the 5-

minute realized variance as the proxy of the conditional variance, they empirically evaluate its forecast performance. They use the data of S&P500 index from April 9, 2001 through August 30, 2013, with 4,096 samples using 1000 rolling window for estimation and the remaining 3096 observations for forecast evaluation. They conduct a test whether their HARQ model significantly outperforms other 7 models they consider, i.e., AR, HAR, HAR-with-Jumps (HAR-J),[8] Continuous-HAR (CHAR),[9] Semivariance-HAR (SHAR),[10] ARQ and HARQ-F.[11] Their null and alternative hypothesis are given by

$$H_o: \quad \min_{k=1,\cdots,7} \mathbb{E}(L_t(y_t, F_{k,t}) - L_t(y_t, F_{0,t})) \leq 0$$
$$H_1: \quad \min_{k=1,\cdots,7} \mathbb{E}(L_t(y_t, F_{k,t}) - L_t(y_t, F_{0,t})) > 0$$

where $F_{0,t}$ denotes the forecast with the benchmark HARQ model. Rejection of H_0 implies the loss of the HARQ model is significantly lower than all the other models. The critical values are obtained by the stationary bootstrap of Politis and Romano (1994) with 999 re-samplings with an average block length of 5. The P-value is 0.063 with SE and 0.871 with QLIKE.

7.4.1 Tail index estimators

Figure 13 shows the Hill and the log-log estimators for the realized variance (RV) used in Bollerslev, Patton, and Quaedvlieg (2016), with the truncation parameter $k_n \in [20, 3000]$. Both the Hill and the log-log estimators are decreasing in the threshold parameter k_n, and it is stable around the value below two. It indicates that the existence of the second moment is questionable for the realized variance.

Figure 14 and Figure 15 depict the Hill and the log-log estimators for the loss sequence $L(RV_t, F_t)$, where the left panels show the estimates with the squared error (SE) and the right panels with the QLIKE function. When the squared error (SE) function is utilized, the estimates are not sensitive to the value of the threshold k_n, and they are below two. This pattern is observed for all eight different forecasts. This phenomenon is again consistent

[8]See Andersen, Bollerslev, and Diebold (2007). The model is $RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t$. J_t is computed using the Bi-Power variation (BPV) measure. $J_t = \max(RV_t - BPV_t, 0)$

[9]The model is $RV_t = \beta_0 + \beta_1 BPV_{t-1} + \beta_2 BPV_{t-1|t-5} + \beta_3 BPV_{t-1|t-22} + _t$.

[10]See Patton and Sheppard (2015). They decompose the RV due to the negative and positive intraday returns

[11]ARQ is an extension of AR(1) model with RQ component. HARQ-F model is the model to put RQ component to every component in the HAR model.

with our analysis with the squared error loss function. When a variable to be forecast has unbounded second moment, then using the squared error (SE) can lead to the violation of the moment condition. When the QLIKE function is used, the estimates are more sensitive to the value of the threshold k_n, and the estimates are around the value of 2.

Figure 16 and Figure 17 depict the Hill and the log-log estimators with the loss difference $d_t = L(RV_t, F_{2t}) - L(RV_t, F_{jt})$ where F_{2t} is the forecast from the benchmark model given by HAR. We find again the similar pattern as before. In most of the cases, the estimated tail index is below the value of two, indicating that the existence of the second moment is questionable.

Finally, Table 20 summarizes the tail index and the skewness parameter estimates based on the assumption that each sequence is strictly stationary and its unconditional distribution is given by the Stable distribution. Looking at the loss difference sequence, the tail index estimates ranges from 0.63 to 1.00 (with squared error) and from 0.97 to 1.32 (with QLIKE). When the squared error is used, the tail estimates are below one, indicating that the null hypothesis is not well defined. When QLIKE is used the tail index estimates are larger but all of them are still blow two, indicating unbounded second moments. The estimates for the skewness parameter range from -0.23 to 0.13 (with the squared error) and from -0.52 to 0.10 (with QLIKE), indicating unbalanced tails.

7.4.2 Pair-wise equal predictive ability (EPA) test

In ths exercise, we conduct the same analysis as with the data set of Hansen and Lunde (2005). We compute the p-values for the following two one-sided tests:

$$\begin{aligned} &\text{EPA-1:} \qquad H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] > 0 \\ &\text{EPA-2:} \qquad H_0 : \mathbb{E}[d_t] = 0, \quad H_1 : \mathbb{E}[d_t] < 0 \end{aligned}$$

where $d_t = L(RV_t, F_{2t}) - L(RV_t, F_{jt}), j = 1, 3, \cdots, 8$. F_{2t} refers to the forecast based on the benchmark HAR model. Other forecasts F_{jt} are the forecasts. When the Normal asymptotics are used, we compute the test statistic using the long-run variance with the Bartlett Kernel of Newey and West (1987). Likewise, we conduct the stationary bootstrap using the test statistic with the long-run variance. The bootstrap resampling is 5000 with an average block length of 5.

Table 21 demonstrates the p-values for the equal predictive ability tests when the squared error is used as a loss function. Table 22 demonstrates the p-values with the QLIKE. These values should be examined with care since the tail index estimates are sometimes below one, indicating that the null hypothesis is not well defined. For the EPA-2 test with CHAR model, the p-values based on the Normal asymptotics and the stationary bootstrap are below 1%, whereas the p-values based on the subsampling is above 25%. This example shows that the outcome of a test may change according to the assumptions we made on the loss differences.

8 Conclusion

In this paper, we analyze forecast comparison tests under fat tails. We show that the heavy-tailed nature of the financial variables can violate the moment condition, which is necessary to apply the classical central limit theorem or the stationary bootstrap. We characterize the asymptotic distribution of the test statistic for the equal predictive ability (EPA) test, when the second moment of the loss difference is unbounded, and its distribution has a regularly varying tails. It is a ratio of two correlated Stable random variable given by

$$M(\alpha,p) = \frac{\sum_{j=1}^{\infty}(\delta_j Z_j - (2p-1)\mathbb{E}[Z_j \mathbb{I}_{Z_j\in(0,1]}]) - (2p-1)\alpha/(\alpha-1)}{(\sum_{j=1}^{\infty} Z_j^2)^{1/2}}$$

We show that, when the tails are well balanced (i.e., $p = 0.5$), the asymptotic distribution of the test statistics is symmetric with tails similar to the Normal distribution. However, as p increases or decreases, the asymptotic distribution becomes asymmetric and skewed. As a result, the size property of a test using the Normal asymptotics is heavily distorted especially when α is smaller and $|p - 0.5|$ is larger. We also provide an analysis of four important components which determines the tails of the loss difference, namely the tails of variable to be forecast and two competing forecasts and the choice of the loss function. We consider a homogeneous Bregman class and characterize the relation of the four components.

As a method which is robust to the fat-tailedness, we consider subsampling method (Politis, Romano, and Wolf, 1999), since it is well known that subsampling is robust to the fat-tails. We show in the simulation that infinite sample, the choice of the block size has an impact on whether the subsampling method lieads to the correct size under the null hypothesis. In the finite sample, the appropriate block sizes depends on the tail index α and the skewness. We propose two methods of block size selection which are combined with the minimum volatility methods, which is known as Romano-Wolf method (Romano and Wolf, 2001).

Finally we conduct an empirical study to inspect the estimated α and p of the loss difference process. There are several estimators of α, such as Hill estimator (Hill, 1975) and log-log estimator (Gabaix and Ibragimov, 2011). We also obtain the estimator under an assumption of Stable distribution (Fama and Roll (1971) and McCulloch (1986)). Over all, the moment condition is more likely to be violated with the squared error (SE) function rather than the QLIKE function. Also the outcome of the test may change when we implement subsampling instead of classical methods such as stationary bootstrasp.

Appendix

Appendix A: proof of proposition 1

The result in case of $\alpha > 2$ is standard and thus its proof is omitted. When $0 < \alpha < 1$, and when $1 < \alpha < 2$ under the null hypothesis, results from Davis (1983) is directly applied. when $1 < \alpha < 2$ under the alternative hypothesis,

$$\tau = \frac{T}{a_T} \times \frac{1/T \sum_{t=1}^{T} d_t}{\left(1/a_T^2 \sum_{t=1}^{T} d_t^2 - 1/T(1/a_T \sum_{t=1}^{T} d_t)^2\right)^{1/2}}$$

Since $a_T = T^{1/\alpha}\ell(T)$, using the fact that $1/\ell(T)$ is slowly varying at infinity and that $T^{\epsilon}\ell(T) \to \infty$ if $\epsilon > 0$, $T/a_T \to \infty$. The second term converges to a positive random variable. Therefore $\tau \xrightarrow{p} \infty$.

References

ANDERSEN, T. G., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39(4), 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *The Review of Economics and Statistics*, 89(4), 701–720.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

ANDREWS, B., M. CALDER, AND R. A. DAVIS (2009): "Maximum Likelihood Estimation for α-Stable Autoregressive Processes," *Ann. Statist.*, 37(4), 1946–1982.

ATHREYA, K. B. (1987): "Bootstrap of the Mean in the Infinite Variance Case," *Ann. Statist.*, 15(2), 724–731.

BLATTBERG, R., AND T. SARGENT (1971): "Regression with Non-Gaussian Stable Disturbances: Some Sampling Results," *Econometrica*, 39(3), 501–510.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31(3), 307 – 327.

BOLLERSLEV, T., A. J. PATTON, AND R. QUAEDVLIEG (2016): "Exploiting the Errors: a Simple Approach for Improved Volatility Forecasting," *Journal of Econometrics*, 192(1), 1 – 18.

BORODIN, A. N., AND I. A. IBRAGIMOV (1995): *Limit Theorems for Functionals of Random Walks*. American Mathematical Soc.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*. Springer Series in Statistics.

BRORSEN, B. W., AND S. R. YANG (1990): "Maximum Likelihood Estimates of Symmetric Stable Distribution Parameters," *Communications in Statistics - Simulation and Computation*, 19(4), 1459–1464.

Chambers, J. M., C. L. Mallows, and B. W. Stuck (1976): "A Method for Simulating Stable Random Variables," *Journal of the American Statistical Association*, 71(354), 340–344.

Cline, D., and G. Samorodnitsky (1994): "Subexponentiality of the Product of Independent Random Variables," *Stochastic Processes and their Applications*, 49(1), 75 – 98.

Corsi, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7(2), 174–196.

Davis, R., and S. Resnick (1985): "More Limit Theory for the Sample Correlation Function of Moving Averages," *Stochastic Processes and their Applications*, 20(2), 257 – 279.

Davis, R. A. (1983): "Stable Limits for Partial Sums of Dependent Random Variables," *Ann. Probab.*, 11(2), 262–269.

Davis, R. A., and T. Hsing (1995): "Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance," *Ann. Probab.*, 23(2), 879–917.

Dekkers, A. L. M., J. H. J. Einmahl, and L. De Haan (1989): "A Moment Estimator for the Index of an Extreme-Value Distribution," *Ann. Statist.*, 17(4), 1833–1855.

Diebold, F. X., and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), 253–263.

Ding, Z., C. W. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1(1), 83 – 106.

Doukhan, P. (1994): *Mixing: Properties and Examples*, Lecture Notes in Statistics 85. Springer, New York.

Drost, F. C., and B. J. Werker (1996): "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74(1), 31 – 57.

DuMouchel, W. H. (1973): "On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution," *Ann. Statist.*, 1(5), 948–957.

Efron, B. (1991): "Regression Percentiles Using Asymmetric Squared Error Loss," *Statistica Sinica*, 1(1), 93–125.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling extremal events : for insurance and finance*, Applications of mathematics ; 33. 1st ed. 1997.. edn.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

Fama, E. F., and R. Roll (1971): "Parameter Estimates for Symmetric Stable Distributions," *Journal of the American Statistical Association*, 66(334), 331–338.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications*. Wiley.

Feuerverger, A., and P. Hall (1999): "Estimating a Tail Exponent by Modelling Departure From a Pareto Distribution," *Ann. Statist.*, 27(2), 760–781.

Feuerverger, A., and P. McDunnough (1981): "On Some Fourier Methods for Inference," *Journal of the American Statistical Association*, 76(374), 379–387.

Fofack, H., and J. Nolan (1999): "Tail Behavior, Modes and other Characteristics of Stable Distributions," *Extremes*, 2, 39–58.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1(1), 255–294.

Gabaix, X., and R. Ibragimov (2011): "Rank 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, 29(1), 24–39.

Garcia, R., E. Renault, and D. Veredas (2011): "Estimation of Stable Distributions by Indirect Inference," *Journal of Econometrics*, 161(2), 325–337.

Geiting, T. (2011): "Making and Evaluating Point Forecasts," *Journal of the American Statistical Association*, 106(494), 746–762.

Geweke, J., and S. Porter-Hudak (1983): "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4(4), 221–238.

Giacomini, R., and H. White (2006): "Tests of Conditional Predictive Ability," *Econometrica*, 74(6), 1545–1578.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *The Journal of Finance*, 48(5), 1779–1801.

GOMES, M., AND M. MARTINS (2004): "Bias Reduction and Explicit Semi-parametric Estimation of the Tail Index," *Journal of Statistical Planning and Inference*, 124(2), 361 – 378.

HALLIN, M., Y. SWAN, T. VERDEBOUT, AND D. VEREDAS (2013): "One-Step R-estimation in Linear Models With Stable Errors," *Journal of Econometrics*, 172(2), 195 – 204, Latest Developments on Heavy-Tailed Distributions.

HANSEN, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23(4), 365–380.

HANSEN, P. R., AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?," *Journal of Applied Econometrics*, 20(7), 873–889.

HANSEN, P. R., AND A. LUNDE (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131(1), 97 – 121.

HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): "The Model Confidence Set," *Econometrica*, 79(2), 453–497.

HILL, B. (1975): "A Simple General Approach to Inference About the Tail of a Distribution," *The Annals of Statistics*, 3, 1163–1174.

HILL, J. B. (2010): "On Tail Index Estimation for Dependent, Heterogeneous Data," *Econometric Theory*, 26(5), 1398–1436.

HILL, J. B., AND A. PROKHOROV (2016): "GEL Estimation for Heavy-tailed GARCH Models With Robust Empirical Likelihood Inference," *Journal of Econometrics*, 190(1), 18 – 45.

HILL, J. B., AND E. RENAULT (2011): "Generalized Method of Moments with Tail Trimming," *Mimeo*.

HSING, T. (1991): "On Tail Index Estimation Using Dependent Data," *The Annals of Statistics*, 19(3), 1547–1569.

Hurvich, C. M., R. Deo, and J. Brodsky (1998): "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series," *Journal of Time Series Analysis*, 19(1), 19–46.

Ibragimov, I. A. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and Its Applications*, 7, 349–382.

Ibragimov, I. A., and Y. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Ibragimov, R. (2007): "Efficiency of Linear Estimators Under Heavy-tailedness: Convolutions of α-symmetric Distributions," *Econometric Theory*, 23(3), 501517.

——— (2009): "Portfolio Diversification and Value at Risk Under Thick-tailedness," *Quantitative Finance*, 9(5), 565–580.

Ibragimov, R., and J. Walden (2007): "The Limits of Diversification When Losses May Be Large," *Journal of Banking & Finance*, 31(8), 2551 – 2569.

Jessen, A., and T. Mikosch (2006): "Regularly Varying Functions," *Publications de lInstitut Mathématique. Nouvelle Série*, 80.

Jurečková, J., R. Koenker, and S. Portnoy (2001): "Tail Behavior of the Least-squares Estimator," *Statistics & Probability Letters*, 55(4), 377 – 384.

Kim, J., and N. Meddahi (2020): "Volatility Regressions with Fat Tails," *Journal of Econometrics*, 218(2), 690 – 713.

Knight, K. (1987): "Rate of Convergence of Centered Estimates of Autoregressive Parameters for Infinite Variance Autoregressions," *Journal of Time Series Analysis*, 8(1), 51–60.

——— (1989): "On the Bootstrap of the Sample Mean in the Infinite Variance Case," *Ann. Statist.*, 17(3), 1168–1175.

Kokoszka, P., and M. Wolf (2004): "Subsampling the Mean of Heavy-tailed Dependent Observations," *Journal of Time Series Analysis*, 25(2), 217–234.

Koutrouvelis, I. A. (1980): "Regression-Type Estimation of the Parameters of Stable Laws," *Journal of the American Statistical Association*, 75(372), 918–928.

Kruse, R., C. Leschinski, and M. Will (2018): "Comparing Predictive Accuracy under Long Memory, With an Application to Volatility Forecasting*," *Journal of Financial Econometrics*, 17(2), 180–228.

LePage, R., M. Woodroofe, and J. Zinn (1981): "Convergence to a Stable Distribution Via Order Statistics," *Ann. Probab.*, 9(4), 624–632.

Li, J., and A. J. Patton (2018): "Asymptotic Inference About Predictive Accuracy Using High Frequency Data," *Journal of Econometrics*, 203(2), 223 – 240.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015): "Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes," *Journal of Econometrics*, 187(1), 293 – 311.

Logan, B. F., C. L. Mallows, S. O. Rice, and L. A. Shepp (1973): "Limit Distributions of Self-normalized Sums," *Ann. Probab.*, 1(5), 788–809.

Loretan, M., and P. C. Phillips (1994): "Testing the Covariance Stationarity of Heavy-tailed Time Series: an Overview of the Theory With Applications to Several Financial Datasets," *Journal of Empirical Finance*, 1(2), 211 – 248.

McCulloch, J. H. (1986): "Simple Consistent Estimators of Stable Distribution Parameters," *Communications in Statistics - Simulation and Computation*, 15(4), 1109–1136.

McElroy, T., and D. N. Politis (2002): "Robust Inference for the Mean in the Presence of Serial Correlation and Heavy-Tailed Distributions," *Econometric Theory*, 18(5), 1019–1039.

Mikosch, T., and C. G. de Vries (2013): "Heavy Tails of OLS," *Journal of Econometrics*, 172(2), 205 – 221, Latest Developments on Heavy-Tailed Distributions.

Nelson, D. B. (1990): "ARCH models as diffusion approximations," *Journal of Econometrics*, 45(1), 7 – 38.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Nolan, J. P., and D. Ojeda-Revah (2013): "Linear and Nonlinear Regression With Stable Errors," *Journal of Econometrics*, 172(2), 186 – 194, Latest Developments on Heavy-Tailed Distributions.

Oodaira, H., and K. Yoshihara (1972): "Functional Central Limit Theorems for Strictly Stationary Processes Satisfying the Strong Mixing Condition," *Kodai Math. Sem. Rep.*, 24(3), 259–269.

Patton, A. J. (2011): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics*, 160(1), 246–256.

——— (2020): "Comparing Possibly Misspecified Forecasts," *Journal of Business & Economic Statistics*, 38(4), 796–809.

Patton, A. J., and K. Sheppard (2015): "Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility," *The Review of Economics and Statistics*, 97(3), 683–697.

Politis, D. N., and J. P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89(428), 1303–1313.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*. Springer, New York.

Resnick, S., and C. Stărică (1995): "Consistency of Hill's Estimator for Dependent Data," *Journal of Applied Probability*, 32(1), 139–167.

——— (1998): "Tail Index Estimation for Dependent Data," *The Annals of Applied Probability*, 8(4), 1156–1183.

Robinson, P. M. (1995): "Log-Periodogram Regression of Time Series with Long Range Dependence," *Ann. Statist.*, 23(3), 1048–1072.

Romano, J. P., and M. Wolf (2001): "Subsampling Intervals in Autoregressive Models with Linear Time Trend," *Econometrica*, 69(5), 1283–1314.

Samorodnitsky, G., S. T. Rachev, J.-R. Kurz-Kim, and S. V. Stoyanov (2007): "Asymptotic Distribution of Unbiased Linear Estimators in the Presence of Heavy-tailed Stochastic Regressors and Residuals," *Probability and Mathematical Statistics*, 27(2), 275–302.

Samorodnitsky, G., and M. Taqqu (1994): *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. CRC Press.

Savage, L. J. (1971): "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association*, 66(336), 783–801.

Weron, R. (1995): "Performance of the Estimators of Stable Law Parameters," Discussion paper.

Weron, R. (1996): "On the Chambers-Mallows-Stuck method for simulating skewed stable random variables," *Statistics & Probability Letters*, 28(2), 165 – 171.

West, K. D. (1996): "Asymptotic Inference about Predictive Ability," *Econometrica*, 64(5), 1067–1084.

White, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68(5), 1097–1126.

Zhu, Y., and A. Timmermann (2020): "Can Two Forecasts Have the Same Conditional Expected Accuracy?," .

9 Figures

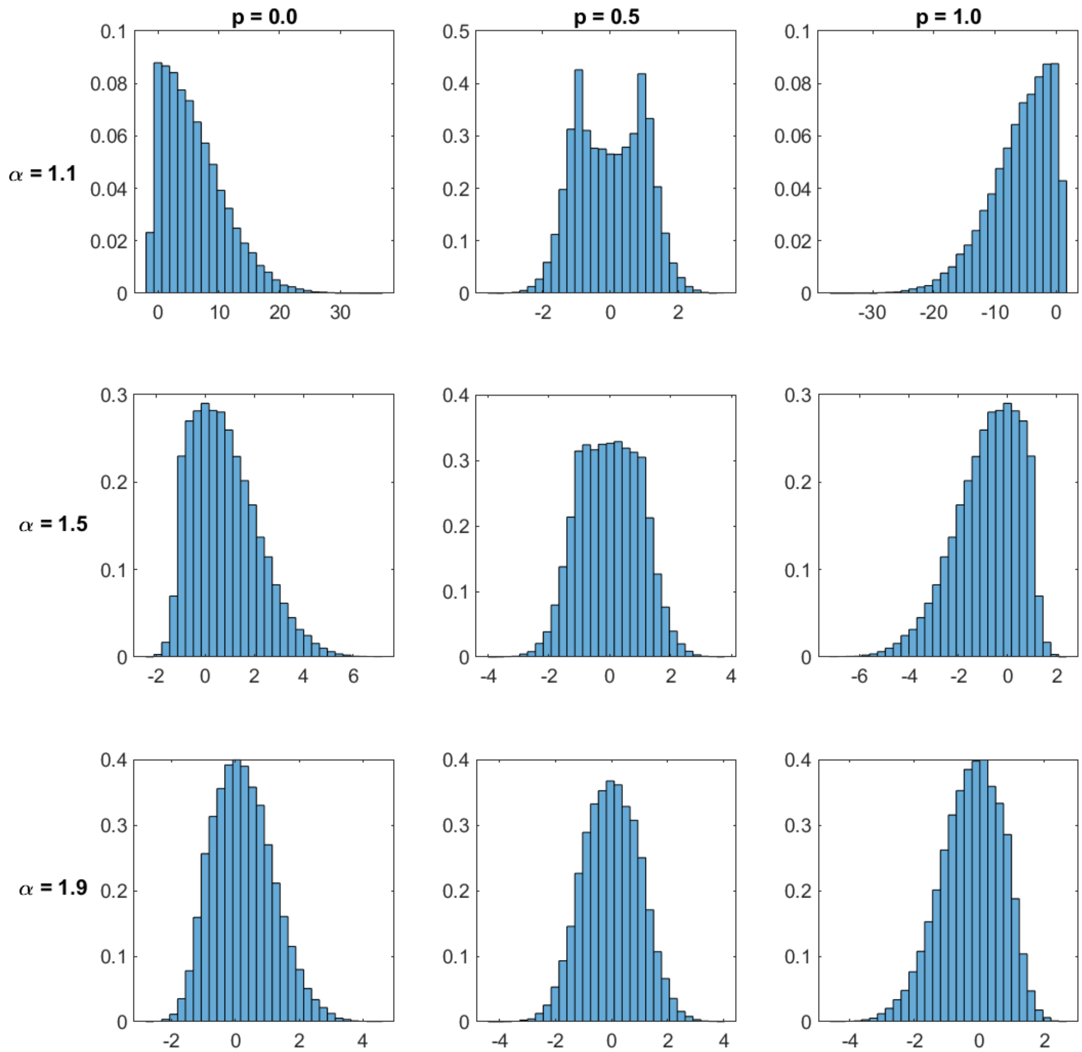

Figure 1: **Distribution of $M_{\alpha,p}$.** This is a histogram for processes generated according to equation (11) where the infinite sum in the numerator is truncated at 10000. The number of replications is 50,000.

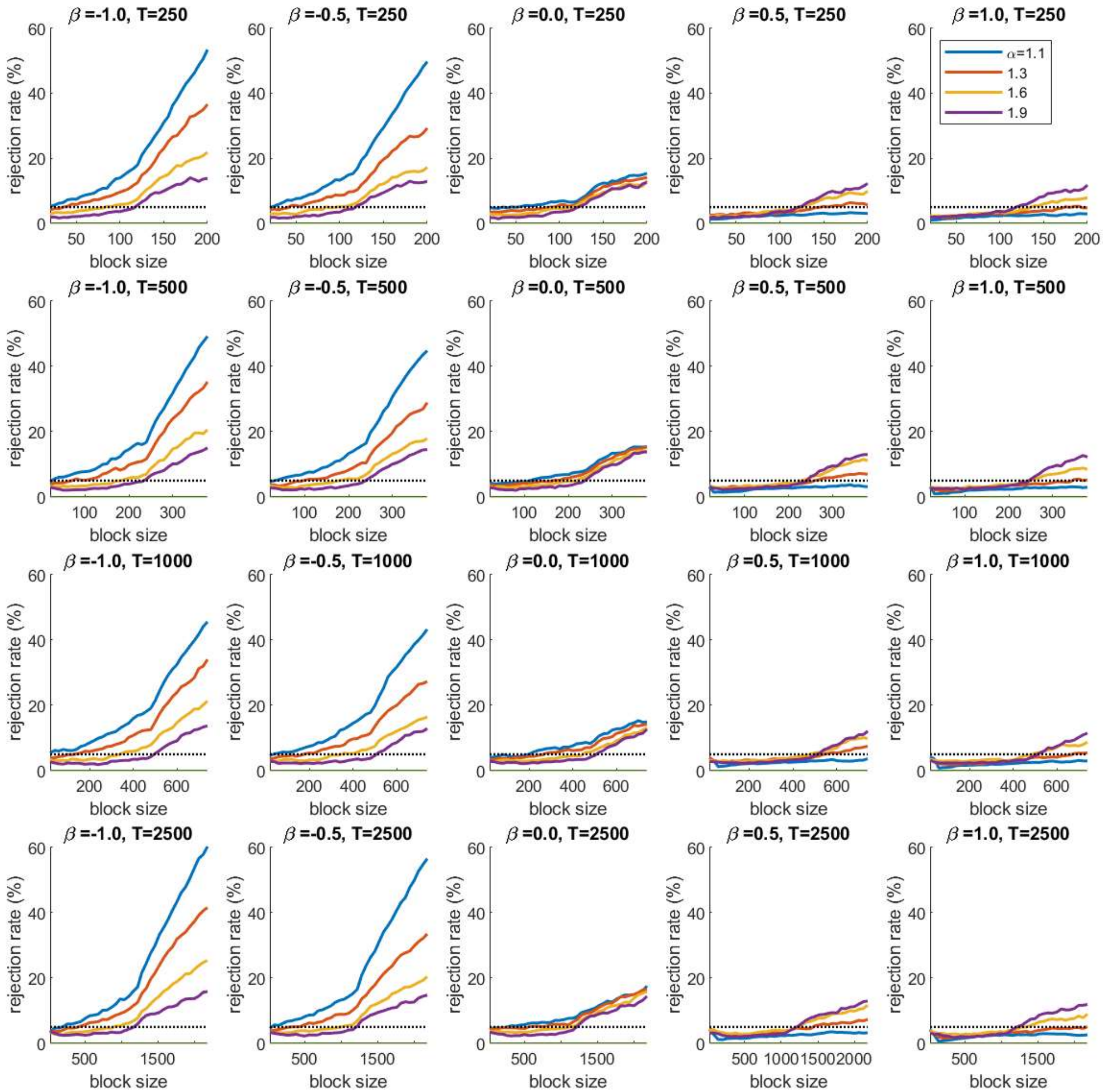

Figure 2: **Relation of the rejection rate under the null hypothesis and the block size**. We replicate 2000 series of $d_t \sim iid.\mathbb{S}_\alpha(1, \beta, 0)$ of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we conduct a one-sided test of level 5%. We compute τ_{var} and reject the null hypothesis according to the procedure of Corollary 2 when the block size is fixed.

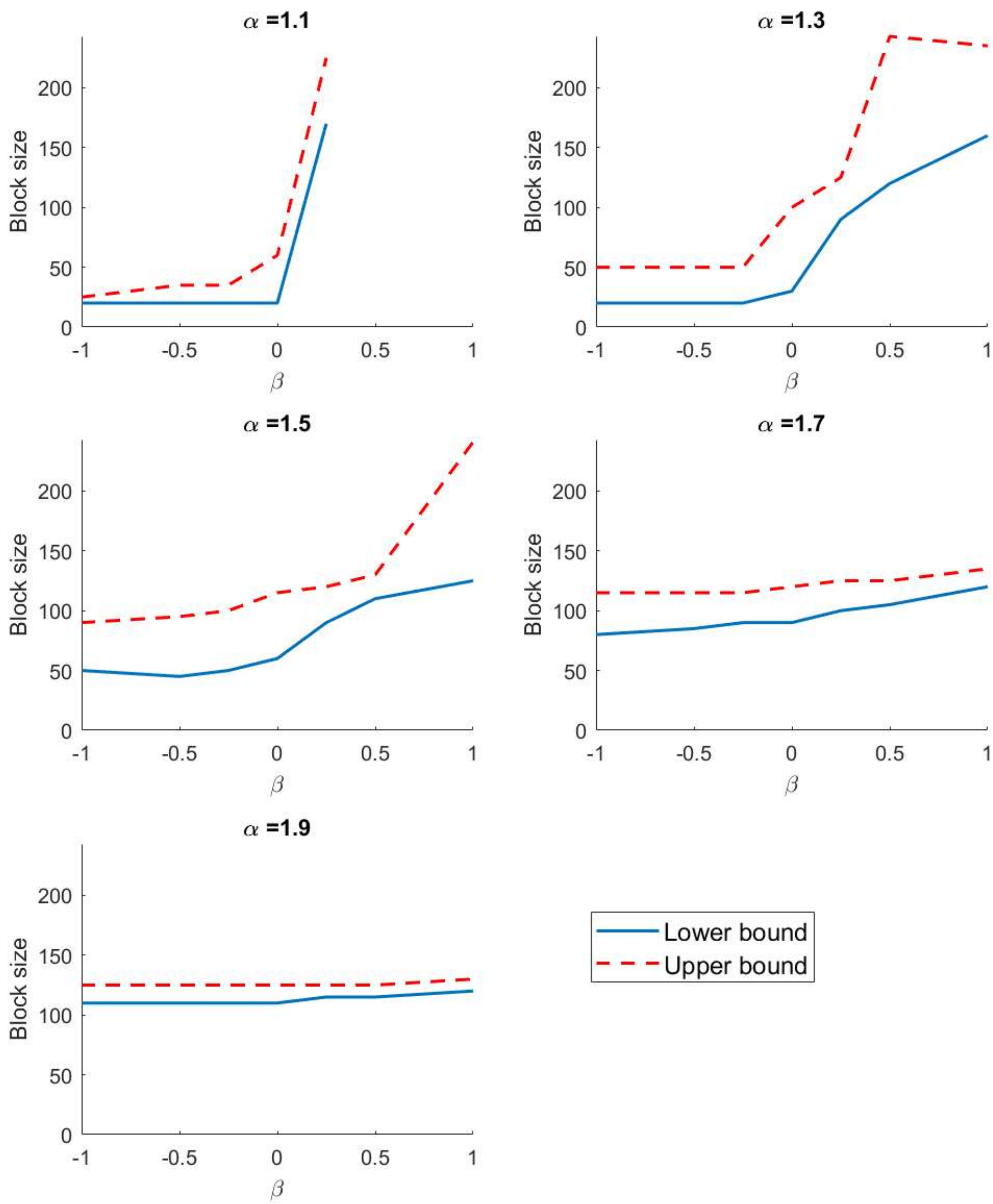

Figure 3: **Appropriate block sizes and β.** We define the *appropriate* block sizes as the one with which the rejection rate under the null hypothesis falls between 0.04 and 0.06. With this definition, we derive the bounds of the appropriate block sizes from the simulation for each (α, β). This figure shows the relation with β.

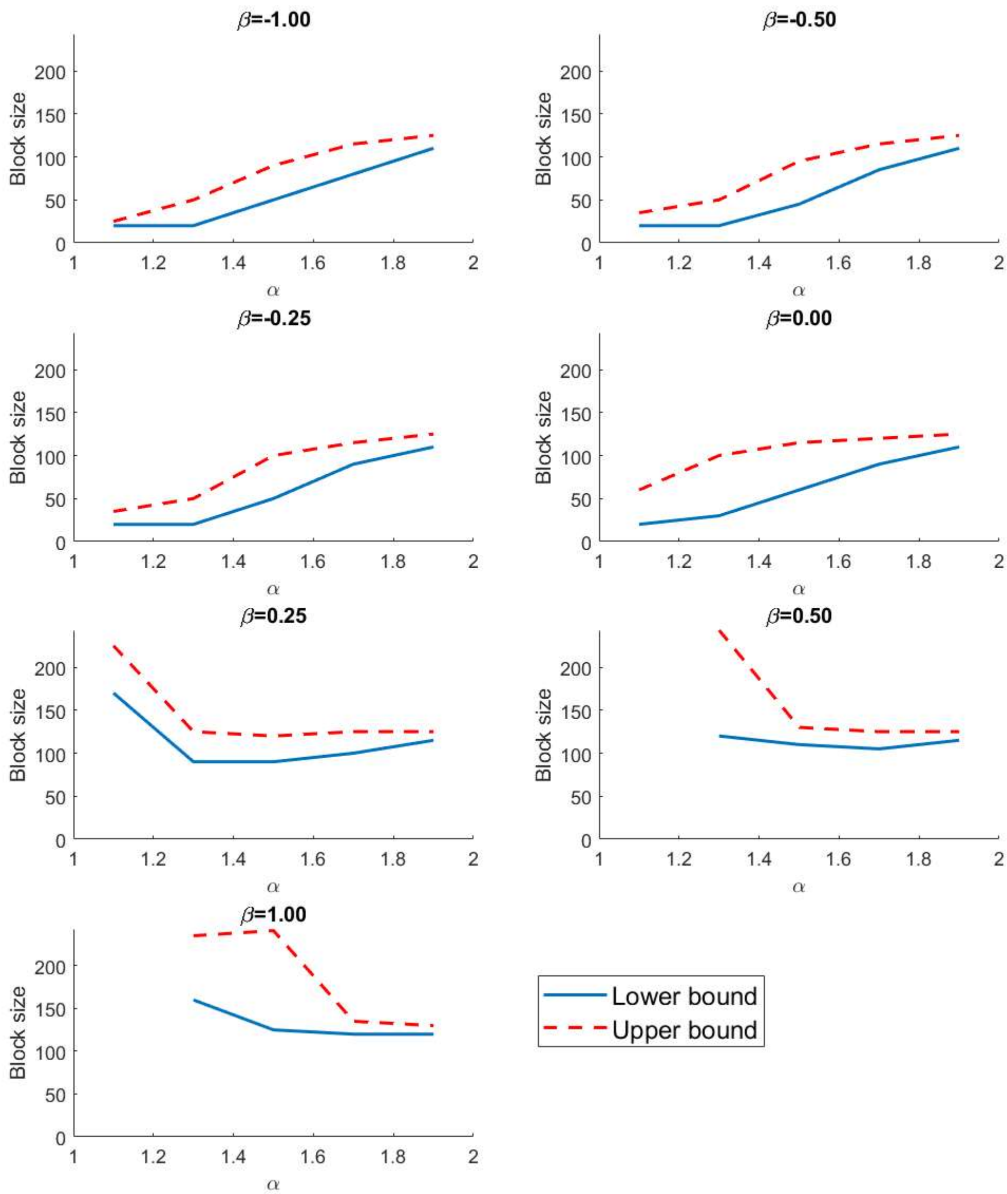

Figure 4: **Appropriate block sizes and α.** We define the *appropriate* block sizes as the one with which the rejection rate under the null hypothesis falls between 0.04 and 0.06. With this definition, we derive the bounds of the appropriate block sizes from the simulation for each (α, β). This figure shows the relation with α.

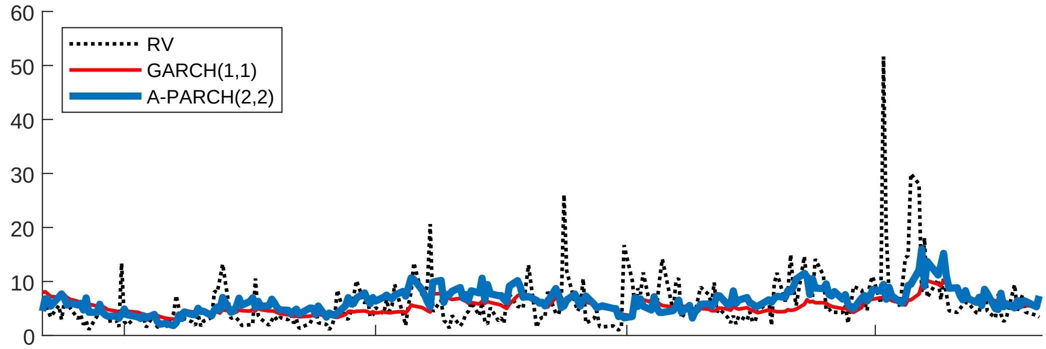

Figure 5: Time series of the realized variance (RV) and two forecasts based on GARCH(1,1) and A-PARCH(2,2)

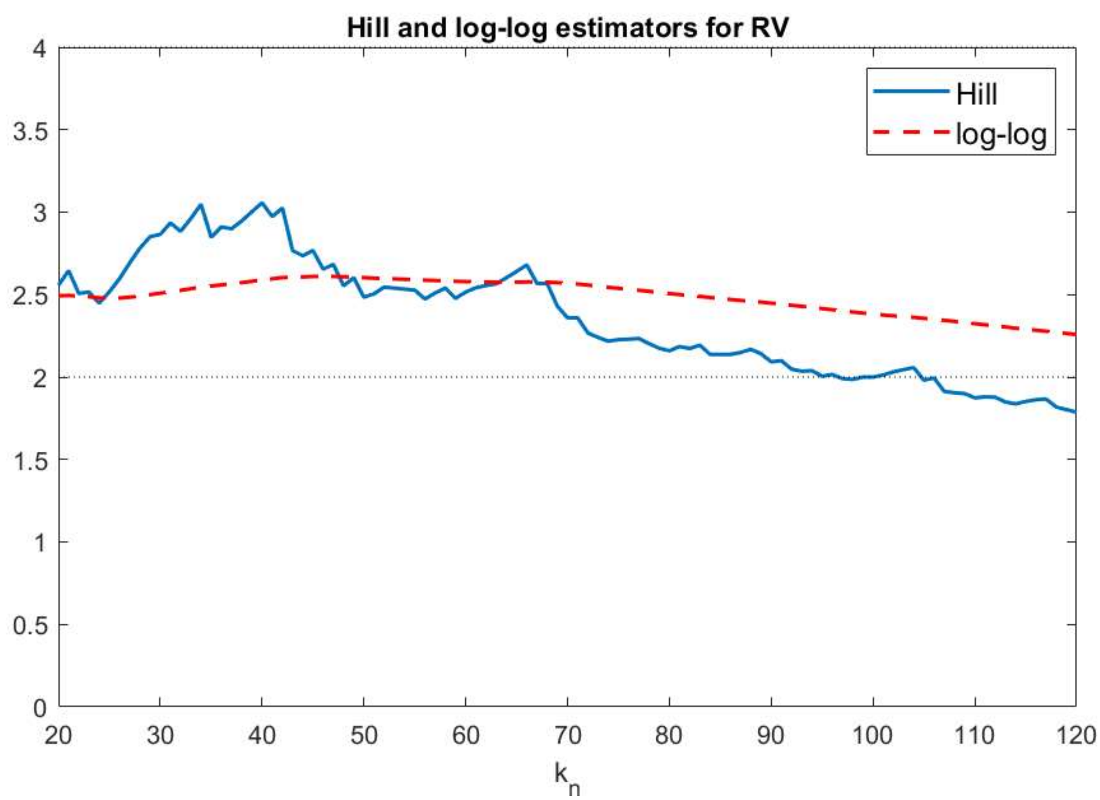

Figure 6: Tail index estimates for RV (Hansen and Lunde, 2005)

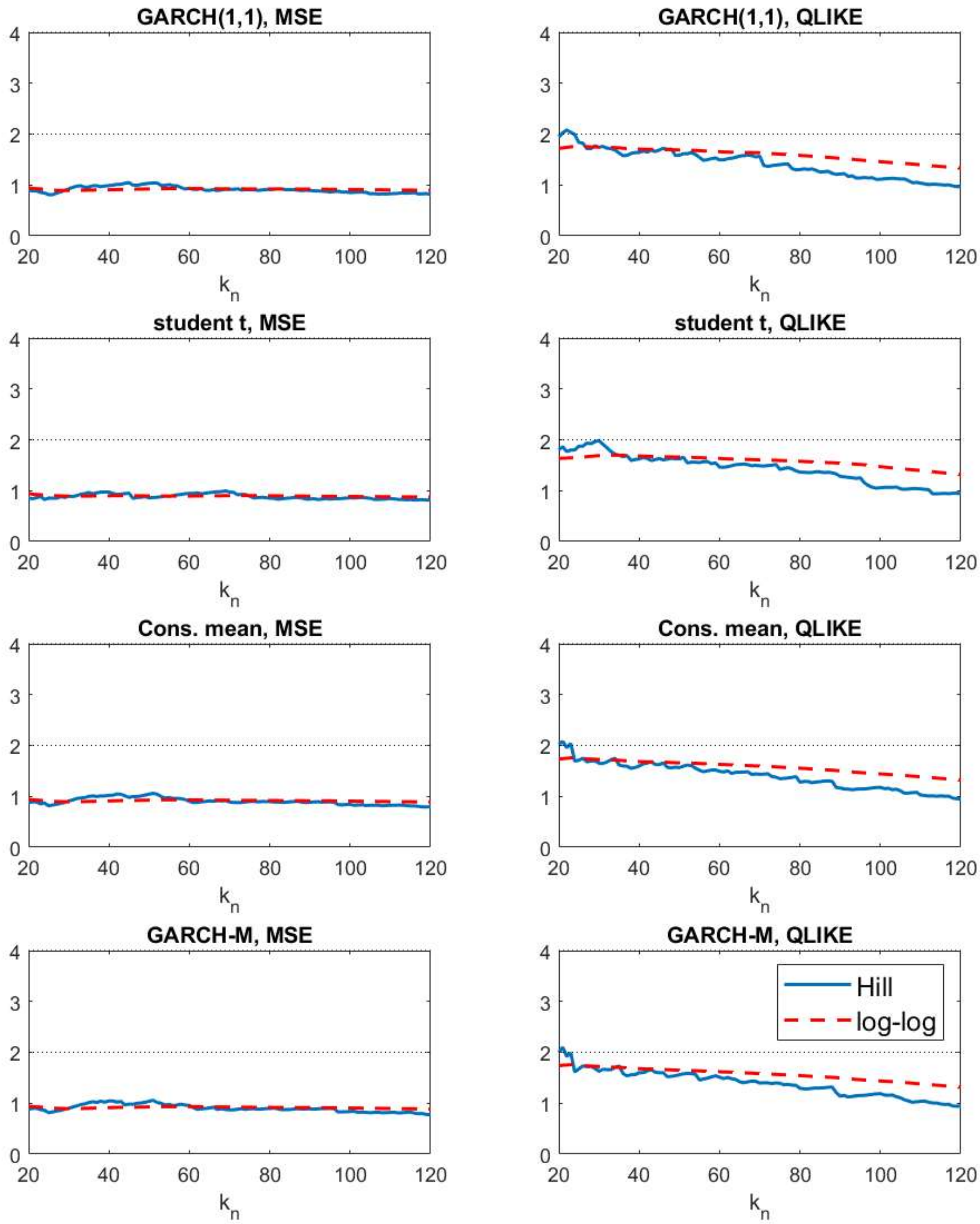

Figure 7: Tail index estimates, SE and QLIKE losses (Hansen and Lunde, 2005)

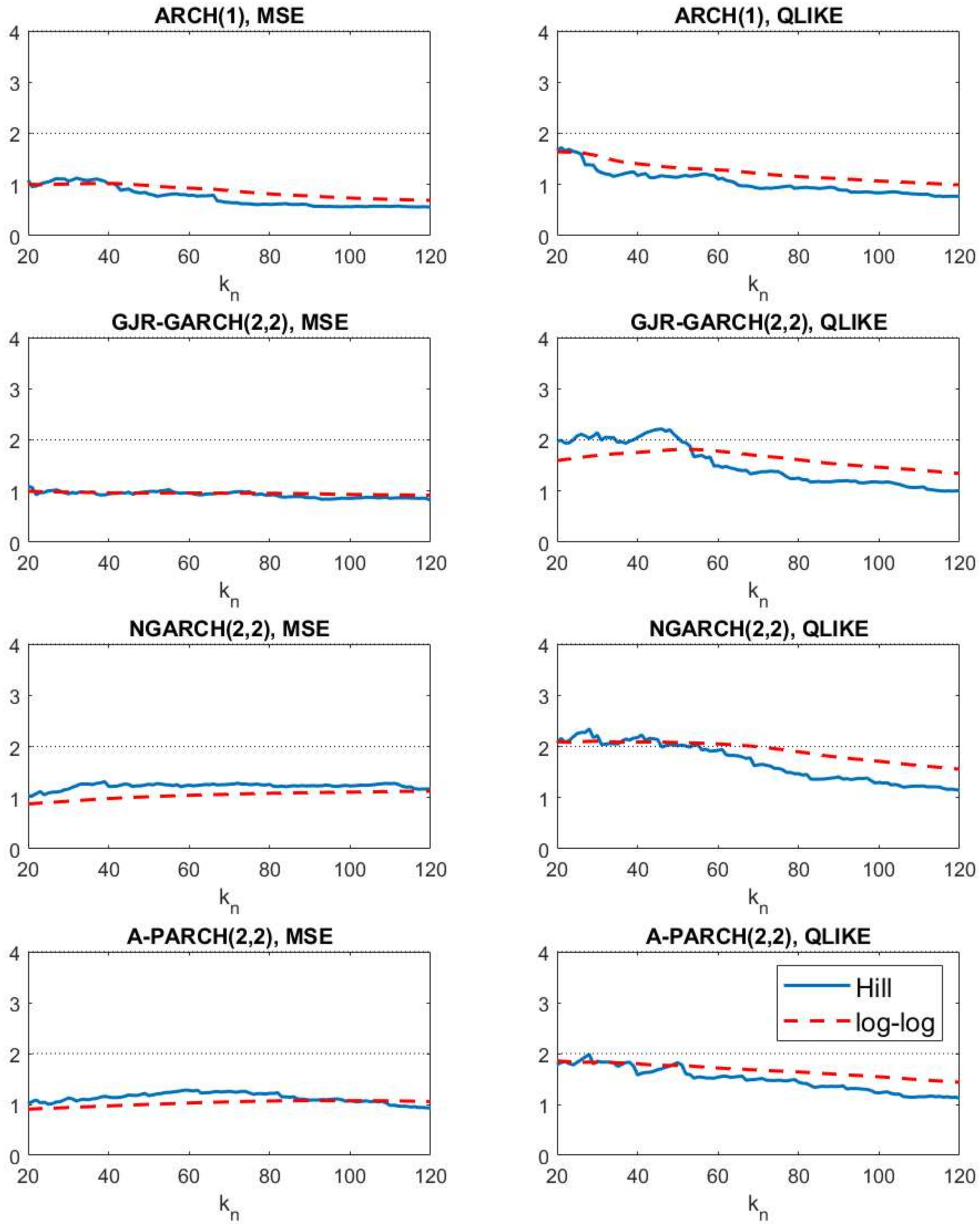

Figure 8: Tail index estimates, SE and QLIKE losses (Hansen and Lunde, 2005)-2

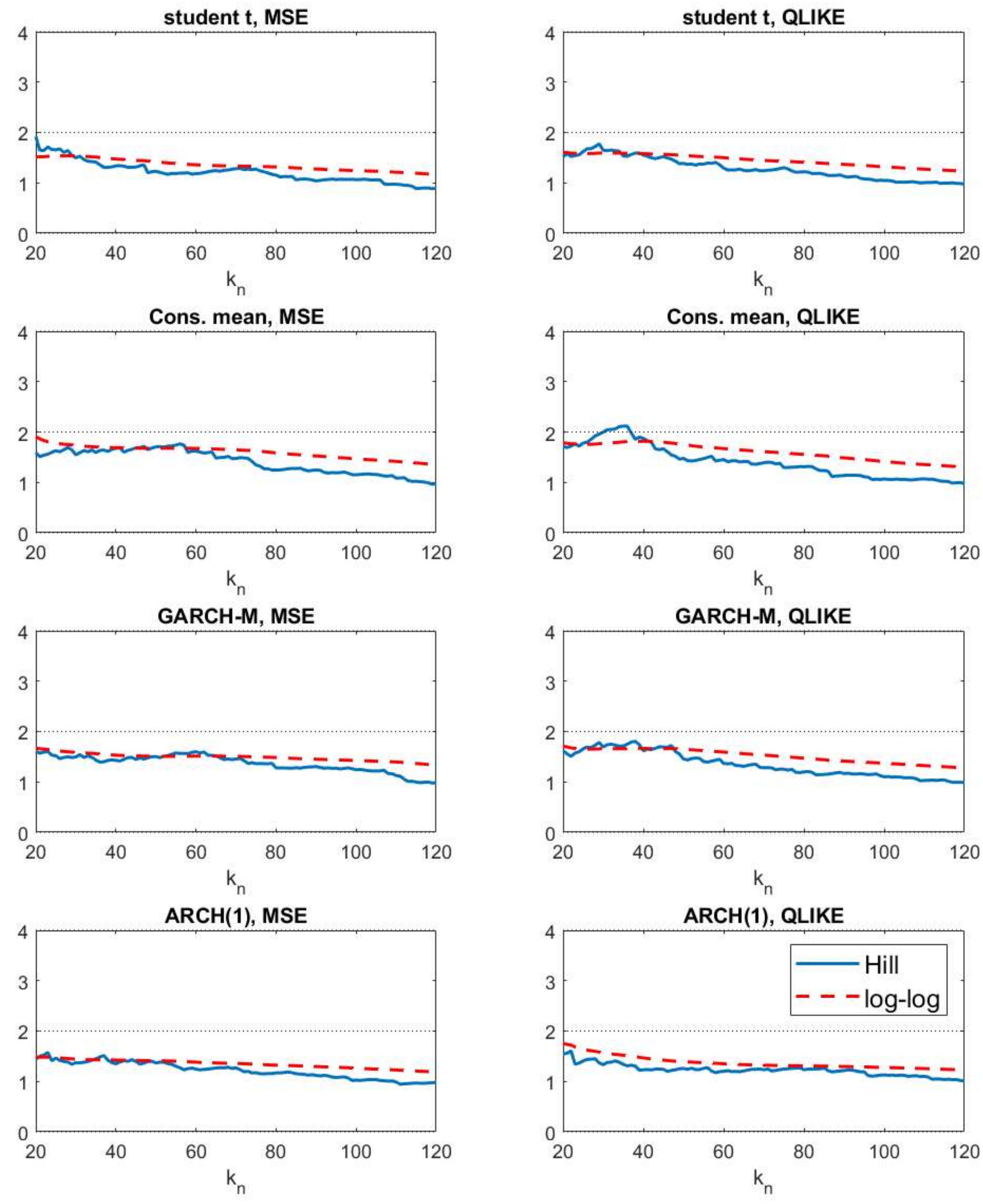

Figure 9: **Tail index estimates, loss difference (Hansen and Lunde, 2005)**. The panels show the tail index estimators for the loss difference, $L(RV_t, F_{jt}) - L(RV_t, F_{1t})$ where F_{1t} is the forecast based on the GARCH(1,1) model with zero mean and Normal innovation.

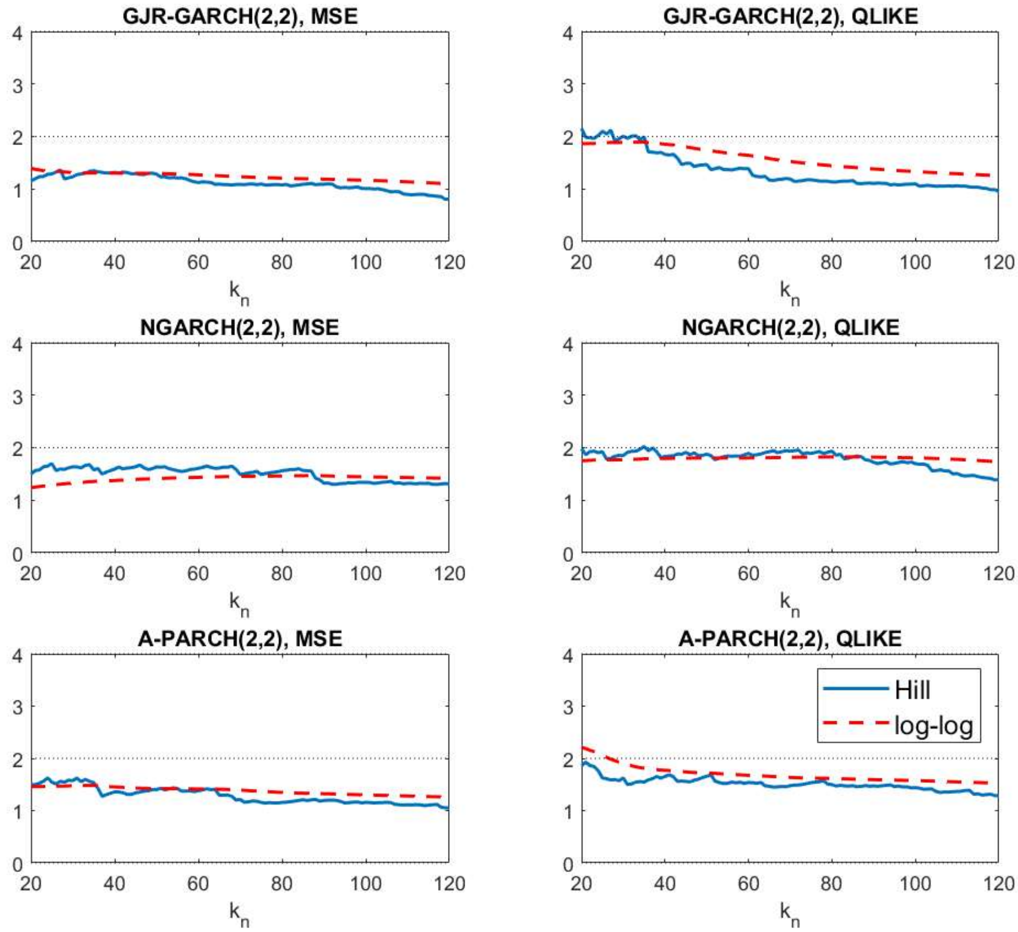

Figure 10: **Tail index estimates, loss difference (Hansen and Lunde, 2005) continued**. The panels show the tail index estimators for the loss difference, $L(RV_t, F_{jt}) - L(RV_t, F_{1t})$ where F_{1t} is the forecast based on the GARCH(1,1) model with zero mean and Normal innovation.

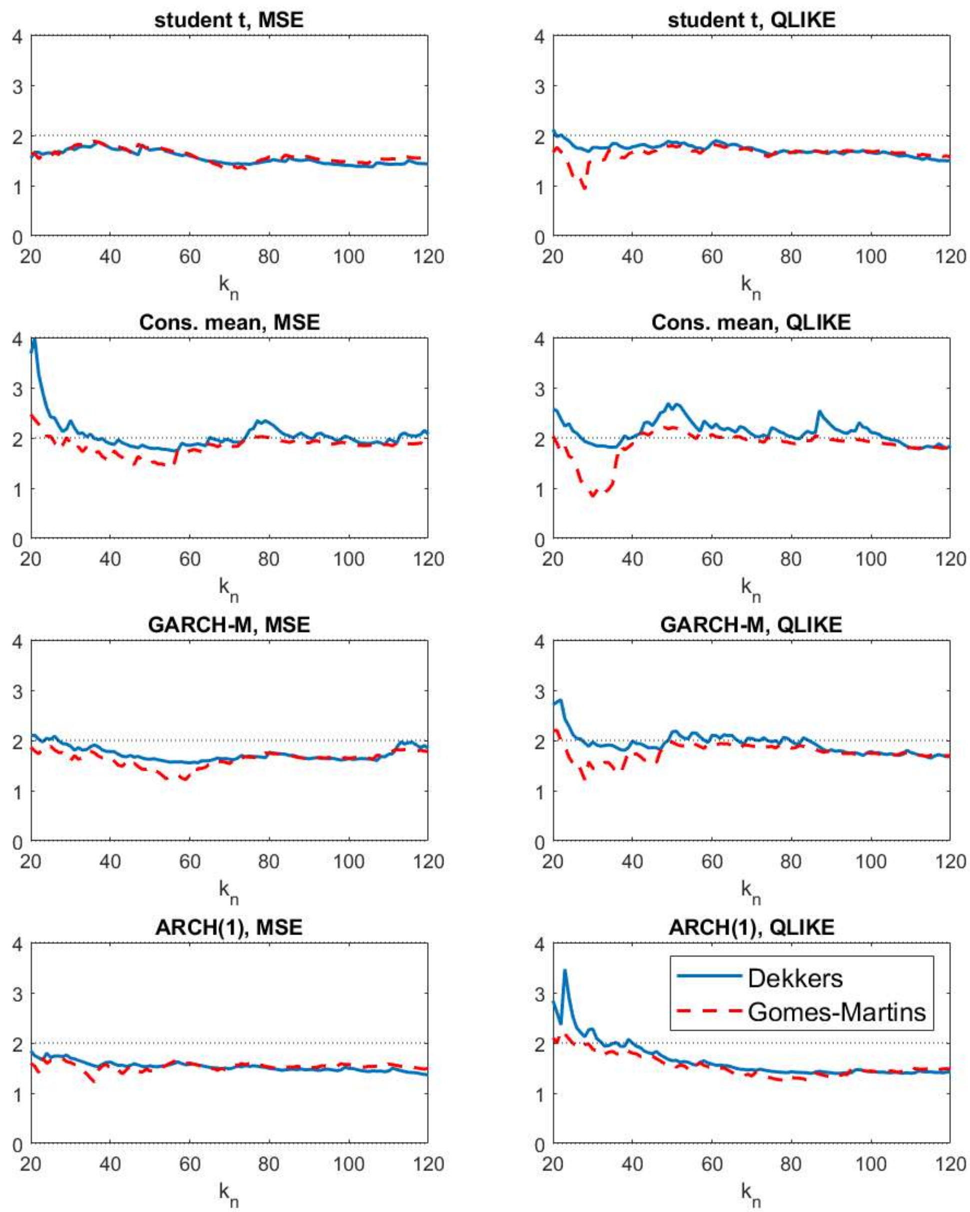

Figure 11: **Tail index estimates, loss difference (Hansen and Lunde, 2005)**. The panels show the tail index estimators for the loss difference, $L(RV_t, F_{jt}) - L(RV_t, F_{1t})$ where F_{1t} is the forecast based on the GARCH(1,1) model with zero mean and Normal innovation.

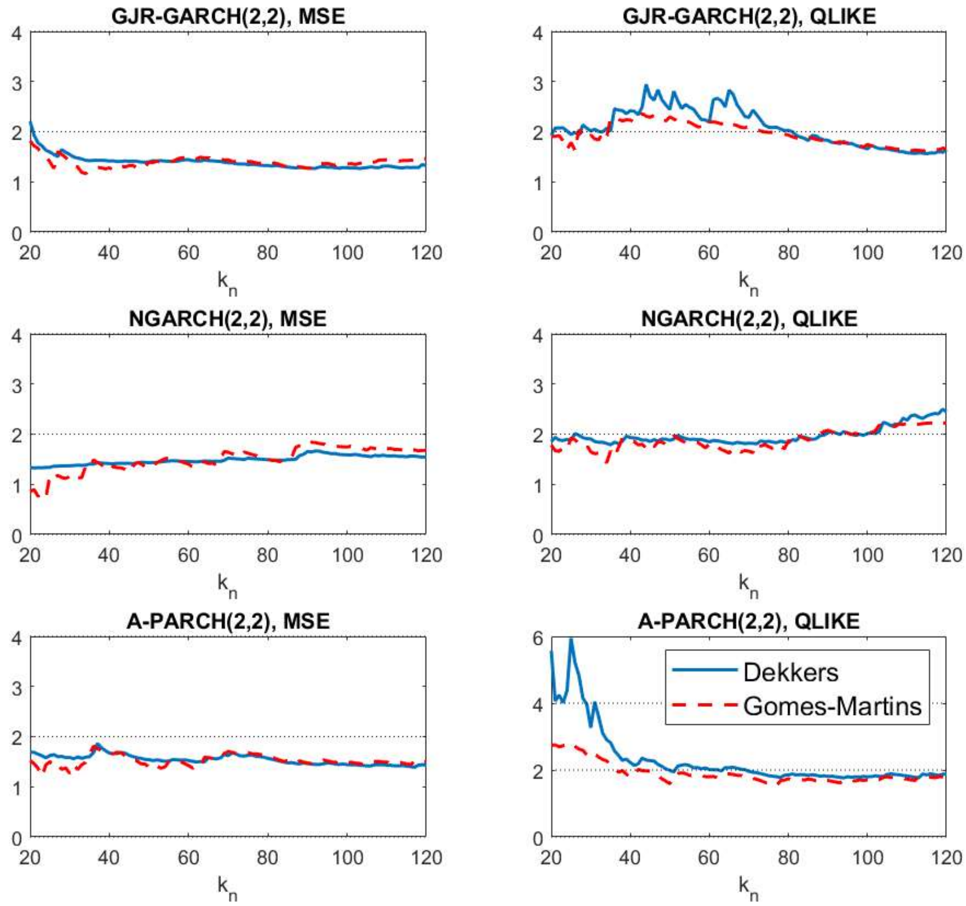

Figure 12: **Tail index estimates, loss difference (Hansen and Lunde, 2005) continued**. The panels show the tail index estimators for the loss difference, $L(RV_t, F_{jt}) - L(RV_t, F_{1t})$ where F_{1t} is the forecast based on the GARCH(1,1) model with zero mean and Normal innovation.

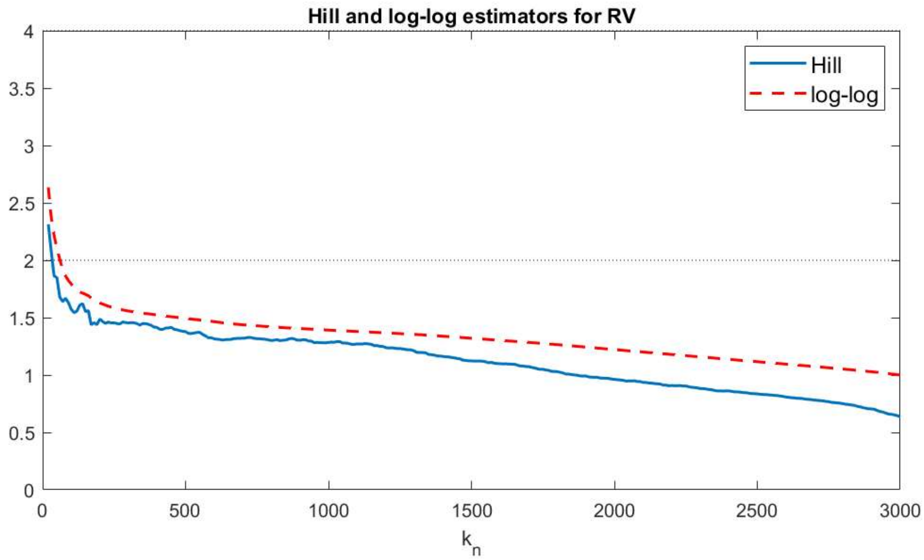

Figure 13: Tail index estimates for RV (BPQ, 2016)

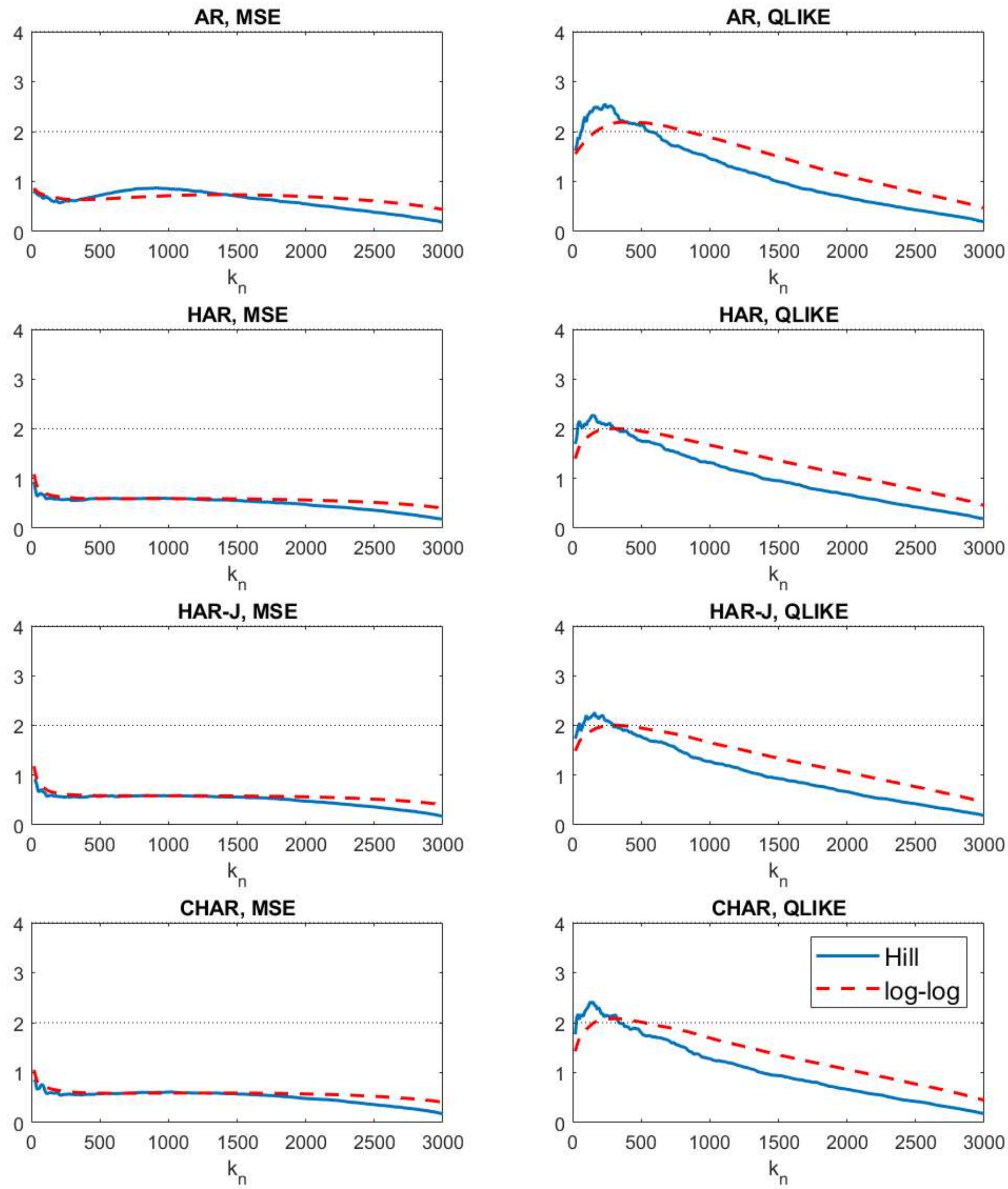

Figure 14: Tail index estimates, SE and QLIKE loss (BPQ, 2016)

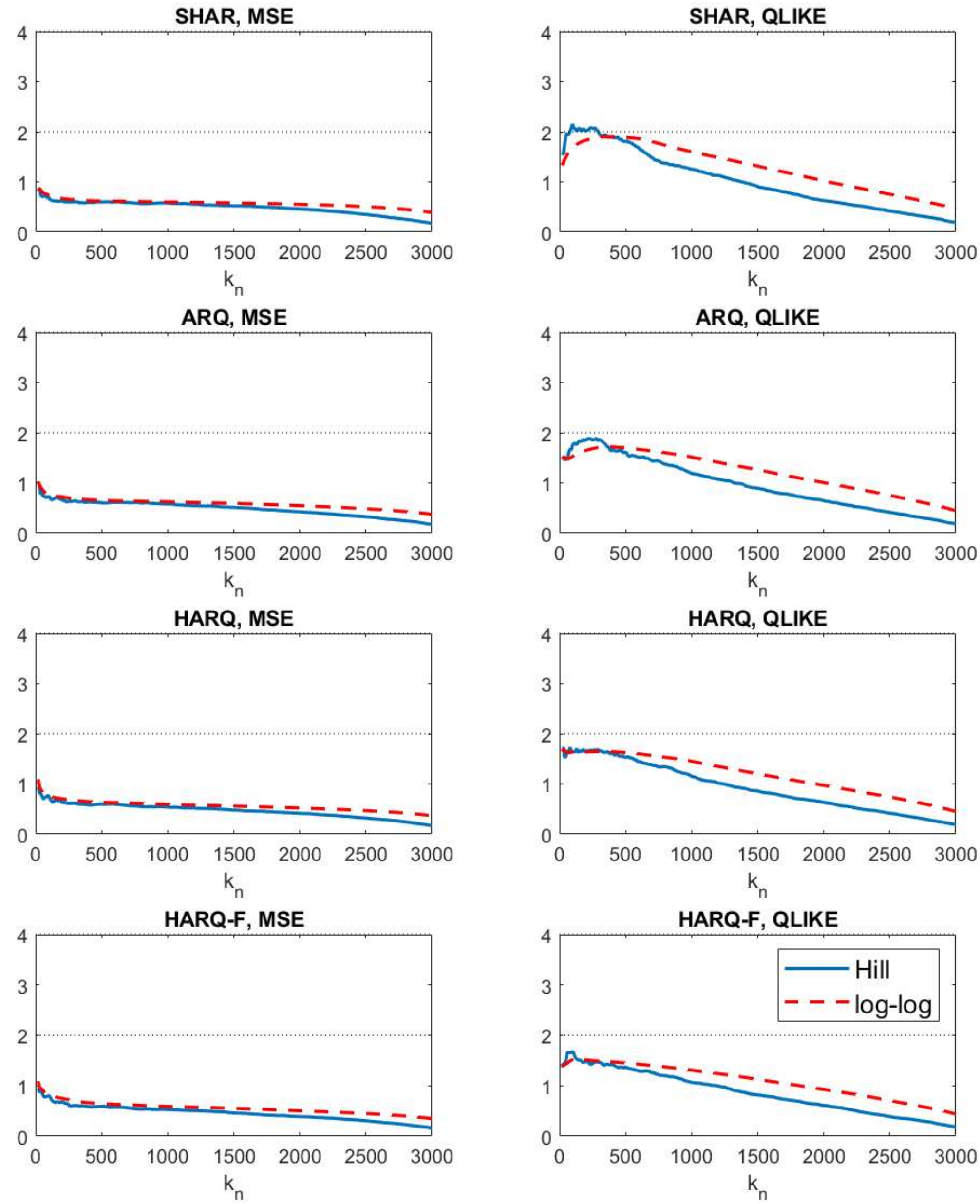

Figure 15: Tail index estimates, SE and QLIKE loss (BPQ, 2016)-2

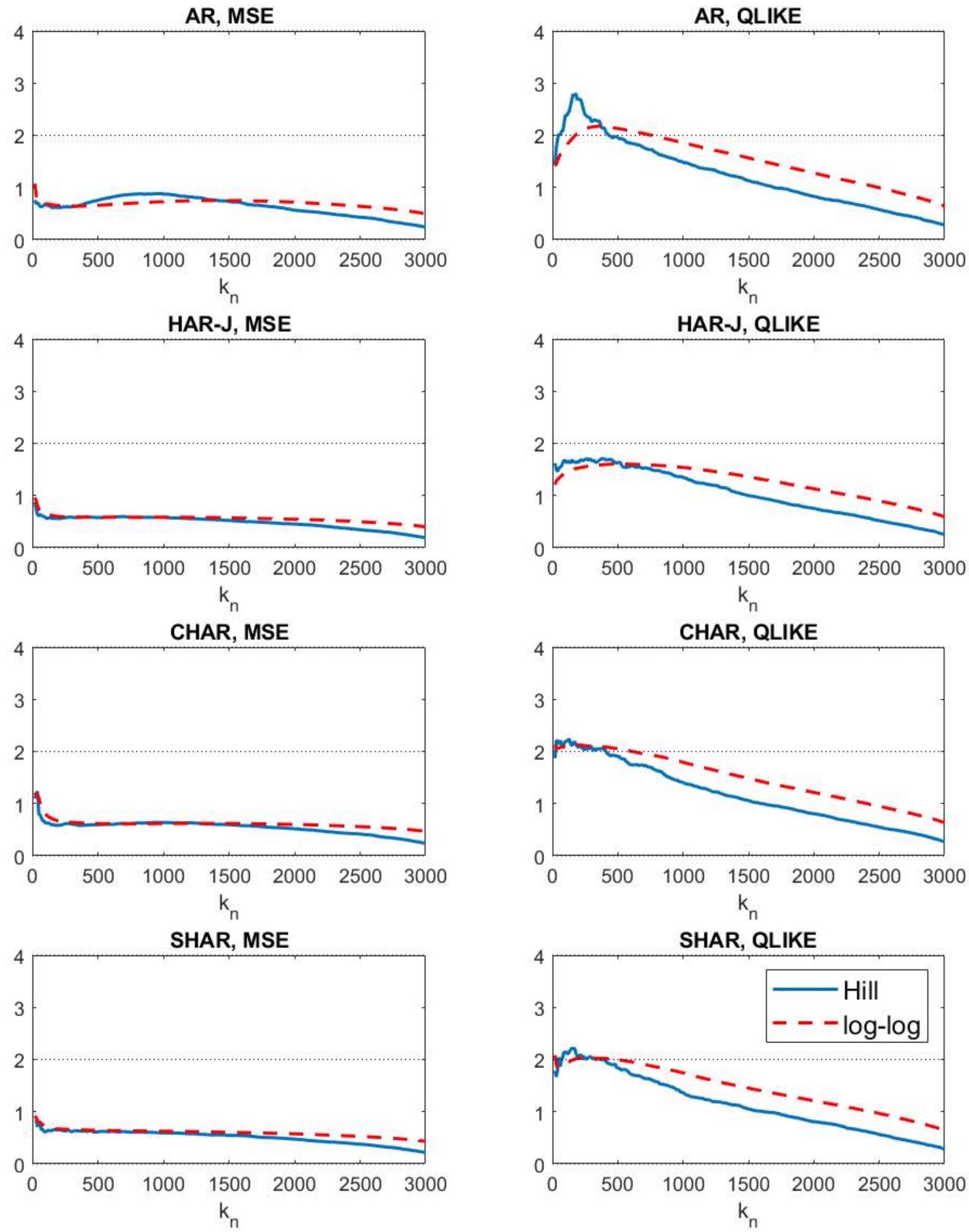

Figure 16: **Tail index estimates, loss difference (BPQ, 2016).** The panels show the tail index estimators for the loss difference that is computed with the benchmark forecast based on the HAR model

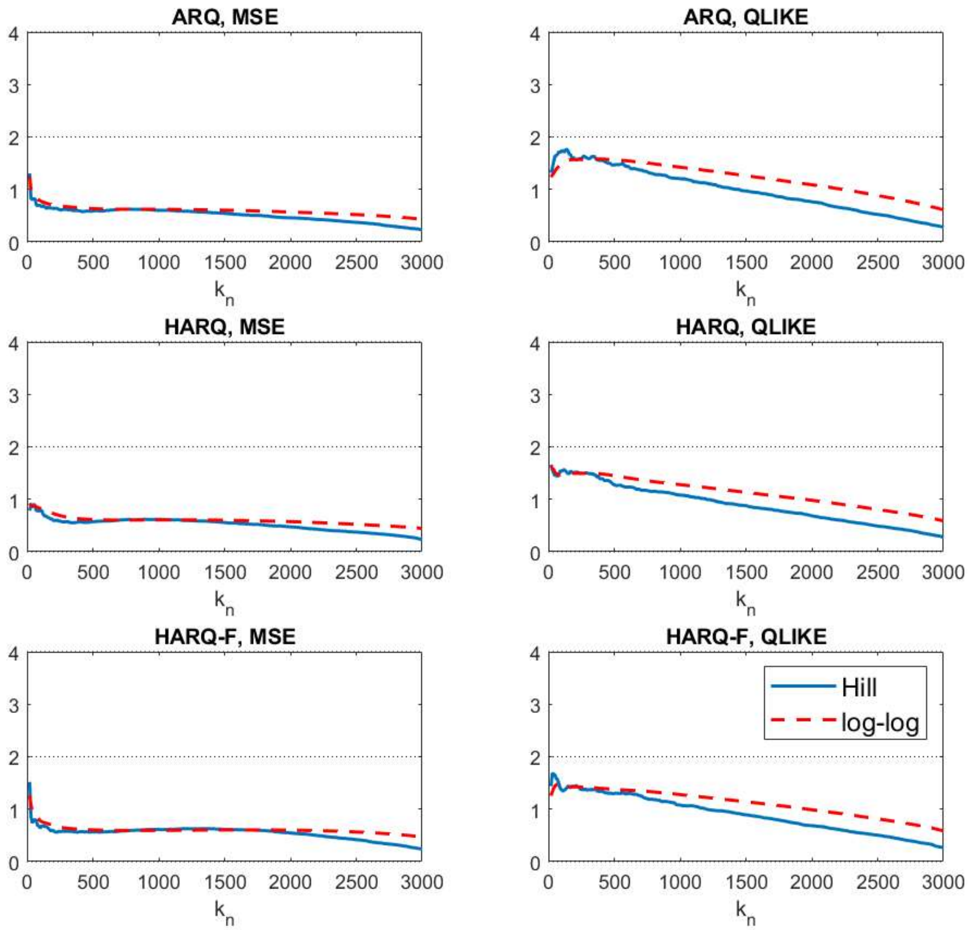

Figure 17: **Tail index estimates, loss difference (BPQ, 2016) continued.** The panels show the tail index estimators for the loss difference that is computed with the benchmark forecast based on the HAR model

10 Tables

Table 1: Quantiles of simulated Stable distribution $\mathbb{S}_\alpha(1, \beta, 0)$

		Quantiles in percentage								
β	α	0.01	0.1	1	5	50	95	99	99.9	99.99
0	1.1	-1118.83	-167.99	-21.19	-5.12	0.01	5.11	21.65	174.04	1809.97
	1.3	-394.00	-67.29	-11.95	-3.79	0.01	3.84	12.37	60.10	463.16
	1.6	-110.90	-25.00	-6.13	-2.83	0.01	2.88	6.32	22.71	121.47
	1.9	-29.25	-8.34	-3.63	-2.40	0.01	2.44	3.77	8.07	30.37
0.3	1.1	-810.15	-122.77	-16.96	-5.50	-1.77	4.83	26.00	219.64	2296.50
	1.3	-299.65	-51.40	-9.53	-3.49	-0.48	4.18	14.76	73.31	566.61
	1.6	-88.79	-20.12	-5.18	-2.70	-0.14	3.05	7.23	26.71	143.08
	1.9	-24.30	-7.07	-3.49	-2.37	-0.02	2.47	3.93	9.14	34.84
0.5	1.1	-597.79	-91.69	-14.11	-5.81	-2.94	4.65	28.89	249.50	2615.06
	1.3	-231.51	-39.94	-7.78	-3.34	-0.80	4.41	16.30	81.72	632.48
	1.6	-72.01	-16.49	-4.52	-2.62	-0.24	3.16	7.78	29.19	156.45
	1.9	-20.41	-6.15	-3.41	-2.36	-0.03	2.49	4.03	9.79	37.55
0.7	1.1	-377.39	-59.50	-11.02	-6.29	-4.09	4.49	31.73	279.01	2929.76
	1.3	-156.58	-27.39	-5.92	-3.30	-1.11	4.63	17.82	89.89	696.35
	1.6	-52.43	-12.28	-3.89	-2.56	-0.34	3.29	8.32	31.55	169.16
	1.9	-15.70	-5.10	-3.35	-2.35	-0.05	2.50	4.14	10.40	40.10
1	1.1	-8.77	-8.50	-8.09	-7.64	-5.80	4.26	35.94	322.71	3395.67
	1.3	-4.94	-4.59	-4.03	-3.48	-1.56	4.97	20.03	101.74	789.02
	1.6	-4.54	-4.05	-3.27	-2.54	-0.48	3.46	9.13	34.91	187.23
	1.9	-4.95	-4.28	-3.25	-2.34	-0.08	2.53	4.29	11.28	43.66

Note. The quantiles are based on 50,000 simulations.

Table 2: $\phi_\alpha(v_1, |v_2|)$ from McCulloch (1986)

	$|v_2|$						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
$v_1 = 2.44$	2.000	2.000	2.000	2.000	2.000	2.000	2.000
2.50	1.916	1.924	1.924	1.924	1.924	1.924	1.924
2.60	1.808	1.813	1.829	1.829	1.829	1.829	1.829
2.70	1.729	1.730	1.737	1.745	1.745	1.745	1.745
2.80	1.664	1.663	1.663	1.668	1.676	1.676	1.676
3.00	1.563	1.560	1.553	1.548	1.547	1.547	1.547
3.20	1.484	1.480	1.471	1.460	1.448	1.438	1.438
3.50	1.391	1.386	1.378	1.364	1.337	1.318	1.318
4.00	1.279	1.273	1.266	1.250	1.210	1.184	1.150
5.00	1.128	1.121	1.114	1.101	1.067	1.027	0.973
6.00	1.029	1.021	1.014	1.004	0.974	0.935	0.874
8.00	0.896	0.892	0.887	0.883	0.855	0.823	0.769
10.00	0.818	0.812	0.806	0.801	0.780	0.756	0.691
15.00	0.698	0.695	0.692	0.689	0.676	0.656	0.595
25.00	0.593	0.590	0.588	0.586	0.579	0.563	0.513

Table 3: $\phi_\beta(v_1, |v_2|)$ from McCulloch (1986)

	$\|v_2\|$						
	0.0	0.1	0.2	0.3	0.5	0.7	1.0
$v_1 = 2.44$	0.000	2.160	1.000	1.000	1.000	1.000	1.000
2.50	0.000	1.592	3.390	1.000	1.000	1.000	1.000
2.60	0.000	0.759	1.800	1.000	1.000	1.000	1.000
2.70	0.000	0.482	1.048	1.694	1.000	1.000	1.000
2.80	0.000	0.360	0.760	1.232	2.229	1.000	1.000
3.00	0.000	0.253	0.518	0.823	1.575	1.000	1.000
3.20	0.000	0.203	0.410	0.632	1.244	1.906	1.000
3.50	0.000	0.165	0.332	0.499	0.943	1.560	1.000
4.00	0.000	0.136	0.271	0.404	0.689	1.230	2.195
5.00	0.000	0.109	0.216	0.323	0.539	0.827	1.917
6.00	0.000	0.096	0.190	0.284	0.472	0.693	1.759
8.00	0.000	0.082	0.163	0.243	0.412	0.601	1.596
10.00	0.000	0.074	0.147	0.220	0.377	0.546	1.482
15.00	0.000	0.064	0.128	0.191	0.330	0.478	1.362
25.00	0.000	0.056	0.112	0.167	0.285	0.428	1.274

Table 4: $\phi_\sigma(\alpha, |\beta|)$ from McCulloch (1986)

	$\lvert\beta\rvert$				
	0.0	0.25	0.5	0.75	1.0
$\alpha = 2.0$	1.908	1.908	1.908	1.908	1.908
1.9	1.914	1.915	1.916	1.918	1.921
1.8	1.921	1.922	1.927	1.936	1.947
1.7	1.927	1.930	1.943	1.961	1.987
1.6	1.933	1.940	1.962	1.997	2.043
1.5	1.939	1.952	1.988	2.045	2.116
1.4	1.946	1.967	2.022	2.106	2.211
1.3	1.955	1.984	2.067	2.188	2.333
1.2	1.965	2.007	2.125	2.294	2.491
1.1	1.980	2.040	2.205	2.435	2.696
1.0	2.000	2.085	2.311	2.624	2.973
0.9	2.040	2.149	2.461	2.886	3.356
0.8	2.098	2.244	2.676	3.265	3.912
0.7	2.189	2.392	3.004	3.844	4.775
0.6	2.337	2.635	3.542	4.808	6.247
0.5	2.588	3.073	4.534	6.636	9.144

Table 5: Quantiles and p-values on $M_{\alpha,p}$, truncation 15000

		Quantile				P-Value (%)			
p	α	Q1%	Q5%	Q95%	Q99%	-2.32	-1.64	1.64	2.32
0.0	1.1	-0.91	-0.45	15.81	20.99	0.00	0.00	76.67	70.82
	1.5	-1.36	-1.00	3.27	4.47	0.00	0.18	24.91	13.61
	1.9	-1.64	-1.20	1.90	2.64	0.02	0.97	7.89	2.07
0.5	1.1	-2.02	-1.52	1.52	2.02	0.29	3.43	3.48	0.30
	1.5	-2.17	-1.58	1.58	2.16	0.62	4.27	4.30	0.56
	1.9	-2.27	-1.61	1.63	2.25	0.85	4.70	4.88	0.76
1.0	1.1	-20.99	-15.81	0.45	0.91	70.82	76.67	0.00	0.00
	1.5	-4.47	-3.27	1.00	1.36	13.61	24.91	0.18	0.00
	1.9	-2.64	-1.90	1.20	1.64	2.07	7.89	0.97	0.02
Classical		-2.32	-1.64	1.64	2.32	1.00	5.00	5.00	1.00

Note. The left columns show the quantiles of 1,5,95 and 99% of simulated $M_{\alpha,p}$. The right columns are the p-value of classical critical values from the standard Normal distribution. These values are obtained based on the representation (11) with 50000 replications and infinite sum truncated at 15,000.

Table 6: Size property: Normal asymptotics (One-sided 5% level test)

	T	β =-1.0	-0.5	0	0.5	1.0
		$p = 0$	0.25	0.5	0.75	1
$\alpha = 1.1$	250	76.30	62.95	2.85	0.00	0.00
	500	76.55	63.00	3.05	0.00	0.00
	1000	77.15	64.10	3.00	0.00	0.00
	2500	77.55	63.35	3.20	0.00	0.00
$\alpha = 1.3$	250	43.75	25.70	3.40	0.20	0.10
	500	43.10	25.20	3.90	0.10	0.00
	1000	46.40	25.70	3.40	0.20	0.05
	2500	44.30	25.35	4.05	0.20	0.00
$\alpha = 1.6$	250	18.50	9.60	3.85	1.45	0.60
	500	17.90	11.05	4.95	1.60	0.40
	1000	17.85	10.35	4.20	1.65	0.75
	2500	17.90	10.15	4.75	1.90	0.70
$\alpha = 1.9$	250	7.10	5.85	4.80	4.30	3.60
	500	8.50	7.25	6.00	5.10	4.20
	1000	6.45	5.55	4.70	3.80	3.25
	2500	6.65	5.70	4.80	3.85	3.25

Unit: Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we compute τ_{var} and reject the null hypothesis when $\tau_{var} > 1.64$.

Table 7: Size property: Normal asymptotics (Two-sided 5% level test)

	T	β =-1.0	-0.5	0	0.5	1.0
		$p = 0$	0.25	0.5	0.75	1
$\alpha = 1.1$	250	73.60	57.75	1.85	58.85	73.70
	500	74.00	57.45	2.45	57.85	72.65
	1000	74.40	58.85	2.35	58.25	72.20
	2500	74.40	57.85	2.55	57.70	73.15
$\alpha = 1.3$	250	37.90	19.65	2.35	20.15	38.95
	500	37.60	18.30	2.75	19.90	39.70
	1000	39.60	19.45	3.25	19.45	38.80
	2500	37.95	19.50	3.25	18.90	36.80
$\alpha = 1.6$	250	13.30	6.55	3.55	6.30	13.30
	500	12.65	6.75	4.10	6.75	12.65
	1000	12.60	6.35	4.00	5.90	13.35
	2500	12.40	6.10	3.95	6.30	12.20
$\alpha = 1.9$	250	5.35	5.05	4.90	4.45	5.05
	500	6.60	6.20	5.75	5.60	5.95
	1000	5.35	4.75	4.50	4.65	5.50
	2500	5.15	4.60	4.35	4.60	5.25

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we compute τ_{var} and reject the null hypothesis when $|\tau_{var}| > 1.96$.

Table 8: Appropriate block sizes

		$\beta=-1$		$\beta=-0.5$		$\beta=-0.25$		$\beta=0$		$\beta=0.25$		$\beta=0.5$		$\beta=1$	
α	T	b_{min}	b_{max}	b_{min}	b_{max}	b_{min}	b_{max}	b_{min}	b_{max}	b_{min}	b_{max}	b_{min}	b_{max}	b_{min}	b_{max}
1.1	250	20	25	20	35	20	35	20	60	170	225				
	500	20	50	30	60	30	60	20	160	340	480				
	1000	20	80	40	100	40	100	20	260	660	920				
	2500	20	126	20	126	20	126	20	709	20	2410	20	20	20	20
1.3	250	20	50	20	50	20	50	30	100	90	125	120	243	160	235
	500	40	90	30	110	20	130	30	210	210	280	270	490	300	480
	1000	40	180	40	220	40	260	40	440	20	520	520	660	560	960
	2500	20	391	73	497	73	603	20	1080	20	1345	20	1663	1557	2430
1.5	250	50	90	45	95	50	100	60	115	90	120	110	130	125	241
	500	100	160	50	190	100	210	150	230	190	240	220	270	250	490
	1000	180	360	200	440	220	460	260	480	420	500	480	540	500	600
	2500	285	815	285	921	232	1027	73	1186	20	1239	20	1292	1239	1504
1.7	250	80	115	85	115	90	115	90	120	100	125	105	125	120	135
	500	160	220	180	230	180	230	190	240	200	240	220	250	240	280
	1000	300	480	340	480	340	500	380	500	420	500	460	520	480	540
	2500	762	1133	868	1186	868	1186	815	1186	921	1186	974	1239	1133	1292
1.9	250	110	125	110	125	110	125	110	125	115	125	115	125	120	130
	500	210	250	210	250	210	250	220	250	230	250	240	260	240	260
	1000	440	500	460	500	460	500	460	500	460	500	460	500	460	500
	2500	1080	1239	1080	1239	1133	1239	1133	1239	1133	1239	1080	1239	1080	1239

We define the *appropriate* block sizes as the one with which the rejection rate under the null hypothesis falls between 0.04 and 0.06. With this definition, we derive the bounds of the appropriate block sizes from the simulation for each (α, β). b_{min} refers to the lower bound and b_{max} the upper bound.

Table 9: Size property w/block size with table, QM-estimated (α, β)

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	7.05	7.25	6.05	1.80	1.90
	500	5.30	5.85	7.00	1.85	1.50
	1000	5.50	5.15	7.65	2.95	2.50
	2500	5.60	5.40	5.30	2.65	2.80
1.3	250	6.65	6.60	5.70	5.60	3.90
	500	5.90	6.65	6.95	6.10	3.95
	1000	6.55	7.00	7.55	6.50	4.30
	2500	5.85	6.15	6.30	4.25	4.20
1.6	250	7.40	6.65	5.85	6.80	6.05
	500	7.00	6.20	5.95	6.55	6.40
	1000	7.25	6.70	6.15	6.85	6.10
	2500	5.80	5.15	4.80	4.05	4.65
1.9	250	8.15	7.25	7.25	7.45	6.85
	500	8.30	8.55	8.30	8.15	6.95
	1000	7.85	7.80	7.80	6.70	7.40
	2500	8.30	8.35	7.30	6.35	6.15

Unit:Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we estimate (α, β) by QM methods, choose b_{min} and b_{max} according to Table 8, then apply the Romano-Wolf method (Romano and Wolf, 2001).

Table 10: Power property w/block size with table, QM-estimated (α, β), $\mathbb{E}(d_t) = 100/T$

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	7.40	7.55	10.30	2.15	1.90
	500	5.55	6.55	9.05	1.85	1.60
	1000	5.60	5.35	8.05	2.75	2.55
	2500	5.60	5.60	5.45	2.70	2.80
1.3	250	10.20	13.15	19.05	16.80	8.05
	500	8.15	10.30	14.55	10.45	5.55
	1000	8.40	9.40	11.45	8.60	5.15
	2500	6.60	7.90	7.50	4.60	4.45
1.6	250	27.65	32.30	37.40	43.55	42.80
	500	21.40	23.40	25.95	28.40	26.05
	1000	15.30	16.60	19.25	18.65	15.30
	2500	9.65	10.30	10.35	8.45	9.40
1.9	250	60.35	63.05	64.80	68.10	71.55
	500	47.30	49.05	51.35	51.95	53.80
	1000	32.35	32.90	33.75	34.30	33.65
	2500	22.50	22.15	21.25	21.15	19.85

Unit:Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. The mean of d_t is given by $100/T$. For each replication, we estimate (α, β) by QM methods, choose b_{min} and b_{max} according to Table 8, then apply the Romano-Wolf method (Romano and Wolf, 2001).

Table 11: Power property w/block size with table, QM-estimated (α, β), $\mathbb{E}(d_t) = 500/T$

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	8.45	9.00	15.85	13.35	3.05
	500	6.30	8.20	15.30	3.70	1.90
	1000	6.10	6.60	12.85	3.55	2.60
	2500	5.95	6.05	7.80	2.75	2.90
1.3	250	21.00	25.20	34.10	58.55	74.45
	500	16.70	23.00	33.85	59.15	50.80
	1000	14.60	18.95	28.40	34.10	17.70
	2500	9.05	12.35	16.55	12.35	7.80
1.6	250	56.90	61.50	65.55	68.55	75.00
	500	53.20	59.95	65.05	71.40	81.05
	1000	46.10	52.95	59.35	68.30	80.65
	2500	32.05	37.95	45.75	53.30	58.10
1.9	250	90.85	92.10	92.55	93.55	93.95
	500	90.20	91.95	92.70	93.95	95.55
	1000	86.30	88.15	90.50	91.75	94.55
	2500	77.25	80.00	83.00	85.55	88.60

Unit:Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. The mean of d_t is given by $500/T$. For each replication, we estimate (α, β) by QM methods, choose b_{min} and b_{max} according to Table 8, then apply the Romano-Wolf method (Romano and Wolf, 2001).

Table 12: Size property w/block size with formula, true (α, β)

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	4.55	5.75	5.40	1.40	1.15
	500	3.90	4.20	4.45	1.00	1.10
	1000	4.20	4.35	4.70	1.35	0.65
	2500	4.20	4.60	4.70	3.95	3.40
1.3	250	4.80	5.85	4.75	2.35	1.60
	500	4.35	4.35	4.05	2.65	1.95
	1000	4.55	4.45	4.50	3.20	2.75
	2500	4.30	4.55	4.10	3.45	3.25
1.6	250	4.25	5.00	4.35	3.90	3.35
	500	3.90	3.80	3.95	3.20	2.70
	1000	3.65	3.45	3.60	3.35	2.80
	2500	3.65	4.25	4.20	3.50	3.50
1.9	250	4.30	4.40	4.35	4.35	4.20
	500	2.90	3.00	3.10	3.40	2.95
	1000	2.75	2.80	2.75	2.45	2.60
	2500	3.10	2.70	2.65	2.75	2.45

Unit: Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we choose b_{min} and b_{max} according to equations (12) and (13) and then apply the Romano-Wolf method (Romano and Wolf, 2001). In choosing b_{min} and b_{max}, we apply the true vaue of α and β

Table 13: Size property w/block size with formula, QM-estimated(α, β)

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	5.45	5.30	5.55	1.45	1.60
	500	5.85	5.60	6.25	1.30	1.00
	1000	5.85	5.25	6.00	1.15	1.20
	2500	5.65	6.60	5.50	0.70	0.50
1.3	250	5.30	5.40	5.50	3.75	2.50
	500	5.30	6.40	6.05	3.40	2.25
	1000	5.50	4.95	5.05	2.75	1.95
	2500	5.05	5.35	4.45	2.95	1.60
1.6	250	5.20	4.85	4.90	4.50	4.00
	500	5.10	5.40	4.95	3.55	3.15
	1000	4.10	3.60	3.25	3.00	2.55
	2500	3.90	3.75	3.40	2.95	2.90
1.9	250	5.45	4.85	6.05	5.45	5.55
	500	3.95	3.40	3.35	3.55	2.95
	1000	2.65	2.90	2.50	2.70	2.55
	2500	2.75	2.40	2.35	2.20	2.25

Unit: Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we choose b_{min} and b_{max} according to equations (12) and (13) and then apply the Romano-Wolf method. In choosing b_{min} and b_{max}, we apply the estimated values of α and β

Table 14: Power property w/block size with formula, QM-estimated(α, β), $\mathbb{E}(d_t) = 100/T$

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	5.70	6.45	8.20	1.85	1.75
	500	6.00	6.10	7.60	1.45	1.05
	1000	5.90	5.15	6.70	1.25	1.25
	2500	5.80	6.65	5.90	0.70	0.55
1.3	250	9.10	11.75	16.55	14.10	7.75
	500	8.55	9.65	11.85	7.35	4.30
	1000	6.85	7.95	8.15	5.00	2.95
	2500	6.15	6.35	5.95	3.65	2.15
1.6	250	27.65	32.95	40.55	46.90	50.40
	500	19.25	23.35	26.45	28.05	25.90
	1000	13.05	15.55	15.80	16.40	13.35
	2500	8.50	8.85	8.70	7.65	6.30
1.9	250	64.10	67.55	70.70	75.65	78.30
	500	48.55	51.45	54.10	57.75	59.15
	1000	32.10	34.10	35.65	36.40	38.00
	2500	16.45	16.55	17.05	17.00	16.30

Unit: Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we choose b_{min} and b_{max} according to equations (12) and (13) and then apply the Romano-Wolf method. In choosing b_{min} and b_{max}, we apply the estimated values of α and β

Table 15: Power property w/block size with formula, QM-estimated(α, β), $\mathbb{E}(d_t) = 500/T$

		β				
α	T	-1	-0.5	0	0.5	1
1.1	250	6.45	8.00	14.10	13.25	2.65
	500	6.65	7.55	11.55	3.55	1.25
	1000	6.15	6.25	9.85	1.70	1.25
	2500	6.00	7.35	6.95	0.85	0.60
1.3	250	20.40	26.25	35.45	49.10	70.45
	500	17.65	24.10	34.55	53.10	67.70
	1000	13.80	18.70	27.95	35.15	19.20
	2500	9.85	12.50	16.20	10.65	5.65
1.6	250	62.20	66.30	69.85	74.20	79.55
	500	58.60	67.55	72.80	78.80	85.10
	1000	52.40	61.00	69.35	79.70	89.45
	2500	38.05	45.45	53.75	63.80	72.95
1.9	250	92.95	93.65	94.35	94.50	95.65
	500	94.20	95.10	96.00	96.50	97.35
	1000	94.05	95.20	96.00	97.30	97.90
	2500	88.85	90.75	93.60	96.20	98.40

Unit: Percent

Note. We replicate 2000 time series of size $T \in \{250, 500, 1000, 2500\}$. For each replication, we choose b_{min} and b_{max} according to equations (12) and (13) and then apply the Romano-Wolf method. In choosing b_{min} and b_{max}, we apply the estimated values of α and β

Table 16: Specification of the 8 selected models (Hansen and Lunde, 2005)

GARCH(1,1)

$$y_t = \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t = \omega + \alpha y_t^2 + \beta v_{t-1}$$

Student t

$$y_t = \sqrt{v_{t-1}} \sqrt{(\nu - 2)/\nu}\; u_t, \quad u_t \sim iid.t(\nu)$$
$$v_t = \omega + \alpha y_t^2 + \beta v_{t-1}$$

Constant mean

$$y_t = c + \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t = \omega + \alpha \varepsilon_t^2 + \beta v_{t-1}, \quad \varepsilon_t = \sqrt{v_{t-1}} u_t$$

GARCH-M

$$y_t = c + \lambda v_{t-1} + \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t = \omega + \alpha \varepsilon_t^2 + \beta v_{t-1}, \quad \varepsilon_t = \sqrt{v_{t-1}} u_t$$

ARCH(1)

$$y_t = \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t = \omega + \alpha y_t^2$$

GJR-GARCH(2,2)

$$y_t = \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t = \omega + \sum_{j=0}^{1} (\alpha_j + \gamma_j \mathbb{1}_{y_{t-j}>0}) y_{t-j}^2 + \sum_{j=1}^{2} \beta_j v_{t-j}$$

NGARCH(2,2)

$$y_t = \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t^\delta = \omega + \sum_{j=0}^{1} \alpha_j y_{t-j}^{2\delta} + \sum_{j=1}^{2} \beta_j v_{t-j}^\delta$$

A-PARCH(2,2)

$$y_t = \sqrt{v_{t-1}} u_t, \quad u_t \sim iid.\mathcal{N}(0,1)$$
$$v_t^\delta = \omega + \sum_{j=0}^{1} \alpha_j (|y_{t-j}| - \gamma_j y_{t-j})^{2\delta} + \sum_{j=1}^{2} \beta_j v_{t-j}^\delta$$

Table 17: Estimation results for the Stable distribution, Hansen and Lunde (2005)

Process		Tail index $\hat{\alpha}$		Skewness $\hat{\beta}$		Tail balance $\hat{p}$	
		SE	QLIKE	SE	QLIKE	SE	QLIKE
RV		1.67		1.00		1.00	
Loss sequence, $L(RV_t, F_{jt})$							
$j = 1$	GARCH	0.83	1.52	1.00	1.00	1.00	1.00
2	student	0.85	1.57	1.00	1.00	1.00	1.00
3	cons.mean	0.83	1.46	1.00	1.00	1.00	1.00
4	GARCH-M	0.82	1.44	1.00	1.00	1.00	1.00
5	ARCH	0.80	0.98	1.00	1.00	1.00	1.00
6	GJR-GARCH	0.88	1.70	1.00	1.00	1.00	1.00
7	NGARCH	1.49	1.63	1.00	1.00	1.00	1.00
8	A-PARCH	1.26	1.38	1.00	1.00	1.00	1.00
Loss difference sequence, $L(RV_t, F_{jt}) - L(RV_t, F_{1t})$							
$j = 2$	student	1.03	1.19	-0.12	-0.32	0.44	0.34
3	cons.mean	1.20	1.16	-0.03	0.16	0.48	0.58
4	GARCH-M	1.12	1.17	-0.13	0.17	0.44	0.59
5	ARCH	1.03	1.07	-0.52	-0.70	0.24	0.15
6	GJR-GARCH	1.00	1.03	0.03	0.11	0.51	0.56
7	NGARCH	1.24	1.24	0.29	0.52	0.64	0.76
8	A-PARCH	1.15	1.27	0.10	0.31	0.55	0.65

Note. F_{jt} corresponds to the forecasts based on 8 models defined in Table 16. Parameters are estimated using the quantile-based method (QM) by McCulloch (1986) applicable to a sequence with mean ero. For RV and L, the parameters are estimated for a demeaned sequence. For the loss differences, we do not demean it. The tail balance parameter is computed by $\hat{p} = (1 + \hat{\beta})/2$ based on equation (6).

Table 18: P-Values for pair-wise EPA tests with SE, Hansen and Lunde (2005)

	Normal asymptotics	Stationary bootstrap	Subsampling Formula	Subsampling Table
EPA-1				
Student t	91.65	92.92	75.56	75.56
Cons.mean	8.35	8.54	17.65	28.11
GARCH-M	11.87	10.82	27.03	27.03
ARCH	99.64	99.82	97.42	100.00
GJR-GARCH	43.00	42.42	38.67	38.67
NGARCH	17.72	15.50	21.92	22.67
A-PARCH	27.61	27.60	36.84	25.16
EPA-2				
Student t	8.35	7.08	24.44	24.44
Cons.mean	91.65	91.46	82.35	71.89
GARCH-M	88.13	89.18	72.97	72.97
ARCH	0.36	0.18	2.58	0.00
GJR-GARCH	57.00	57.58	61.33	61.33
NGARCH	82.28	84.50	78.08	77.33
A-PARCH	72.39	72.40	63.16	74.84

Unit: Percent

Note. We conduct the equal predictive ability (EPA) test with $H_0 : \mathbb{E}[d_t] = 0$ and $H_1 : \mathbb{E}[d_t] > 0$ (EPA-1) and $H_1 : \mathbb{E}[d_t] < 0$ (EPA-2). Each row represents the p-value of the test statistics when $d_t = L(RV_t, F_{1t}) - L(y_t, F_{jt})$ where F_{1t} is the forecast based on the benchmark GARCH(1,1) model, and F_{jt} is the forecast based on the model indicated in the first column in the table. The loss function is the squared error, i.e., $L(y, F) = (y - F)^2$. When the subsampling is used, the block size is selected either by the formula (12) and (13) or by referring to Table 8.

Table 19: P-Values for pair-wise EPA tests with QLIKE, Hansen and Lunde (2005)

	Normal asymptotics	Stationary bootstrap	Subsampling Formula	Subsampling Table
EPA-1				
Student t	93.66	94.36	83.54	74.09
Cons.mean	7.83	9.04	24.32	2.86
GARCH-M	9.80	10.66	24.34	0.00
ARCH	99.96	99.94	100.00	100.00
GJR-GARCH	35.89	37.30	42.08	38.20
NGARCH	17.59	17.84	22.87	24.54
A-PARCH	27.90	29.76	20.73	0.00
EPA-2				
Student t	6.34	5.64	16.46	25.91
Cons.mean	92.17	90.96	75.68	97.14
GARCH-M	90.20	89.34	75.66	100.00
ARCH	0.04	0.06	0.00	0.00
GJR-GARCH	64.11	62.70	57.92	61.80
NGARCH	82.41	82.16	77.13	75.46
A-PARCH	72.10	70.24	79.27	100.00

Unit: Percent

Note. We conduct the equal predictive ability (EPA) test with $H_0 : \mathbb{E}[d_t] = 0$ and $H_1 : \mathbb{E}[d_t] > 0$ (EPA-1) and $H_1 : \mathbb{E}[d_t] < 0$ (EPA-2). Each row represents the p-value of the test statistics when $d_t = L(RV_t, F_{1t}) - L(y_t, F_{jt})$ where F_{1t} is the forecast based on the benchmark GARCH(1,1) model, and F_{jt} is the forecast based on the model indicated in the first column in the table. The loss function is QLIKE, i.e., $L(y, F) = y/F - \log(y/F)$. When the subsampling is used, the block size is selected either by the formula (12) and (13) or by referring to Table 8.

Table 20: Estimation results for the Stable distribution, BPQ (2016)

	Process	Tail index $\hat{\alpha}$		Skewness $\hat{\beta}$		Tail balance $\hat{p}$	
		SE	QLIKE	SE	QLIKE	SE	QLIKE
	RV_t	1.07		1.00		1.00	
Loss sequences, $L(RV_t, F_{jt})$							
j =1	AR	0.81	1.75	1.00	1.00	1.00	1.00
2	HAR	0.59	1.46	1.00	1.00	1.00	1.00
3	HAR-J	0.58	1.46	1.00	1.00	1.00	1.00
4	CHAR	0.58	1.47	1.00	1.00	1.00	1.00
5	SHAR	0.58	1.40	1.00	1.00	1.00	1.00
6	ARQ	0.59	1.39	1.00	1.00	1.00	1.00
7	HARQ	0.56	1.33	1.00	1.00	1.00	1.00
8	HARQ-F	0.55	1.17	1.00	1.00	1.00	1.00
Loss difference sequences, $L(RV_t, F_{2t}) - L(RV_t, F_{jt})$							
$j = 1$	AR	1.00	1.32	-0.14	-0.52	0.43	0.24
3	HAR-J	0.63	1.17	0.13	0.03	0.57	0.51
4	CHAR	0.69	1.25	0.12	-0.11	0.56	0.45
5	SHAR	0.64	1.21	0.13	0.10	0.57	0.55
6	ARQ	0.66	1.09	-0.10	-0.05	0.45	0.47
7	HARQ	0.67	0.99	-0.08	0.05	0.46	0.53
8	HARQ-F	0.70	0.97	-0.23	-0.24	0.39	0.38

Note. The loss difference is computed with the benchmark forecast F_{2t}, based on the HAR model. The parameters are estimated for a demeaned sequence such that $\mu = 0$ and then applied the quantile-based method (QM) by McCulloch (1986). The tail balance parameter is computed by $\hat{p} = (1 + \hat{\beta})/2$ based on equation (6).

Table 21: P-Values for pair-wise EPA tests with SE, BPQ(2016)

	Normal asymptotics	Stationary bootstrap	Subsampling Formula	Subsampling Table
		EPA-1		
AR	26.02	18.94	18.16	7.52
HAR-J	26.20	17.24	22.60	24.83
CHAR	11.11	5.28	8.73	10.04
SHAR	17.01	9.56	37.60	35.93
ARQ	9.95	3.46	5.13	5.91
HARQ	9.76	2.68	22.83	24.65
HARQ-F	14.21	7.54	18.08	15.84
		EPA-2		
AR	73.98	81.06	81.84	92.48
HAR-J	73.80	82.76	77.40	75.17
CHAR	88.89	94.72	91.27	89.96
SHAR	82.99	90.44	62.40	64.07
ARQ	90.05	96.54	94.87	94.09
HARQ	90.24	97.32	77.17	75.35
HARQ-F	85.79	92.46	81.92	84.16

Unit: Percent

Note. We conduct the equal predictive ability (EPA) test with $H_0 : \mathbb{E}[d_t] = 0$ and $H_1 : \mathbb{E}[d_t] > 0$ (EPA-1) and $H_1 : \mathbb{E}[d_t] < 0$ (EPA-2). Each row represents the p-value of the test statistics when $d_t = L(RV_t, F_{2t}) - L(y_t, F_{jt})$ where F_{2t} is the forecast based on the benchmark HAR model, and F_{jt} is the forecast based on the model indicated in the first column in the table. The loss function is the squared error, i.e., $L(y, F) = (y - F)^2$. When the subsampling is used, the block size is selected either by the formula (12) and (13) or by referring to Table 8.

Table 22: P-Values for pair-wise EPA tests with QLIKE, BPQ(2016)

	Normal asymptotics	Stationary bootstrap	Subsampling Formula	Subsampling Table
		EPA-1		
AR	100.00	100.00	98.94	78.40
HAR-J	83.06	87.50	59.53	61.26
CHAR	99.73	99.40	71.64	65.48
SHAR	0.01	0.00	1.06	17.64
ARQ	99.83	100.00	70.66	93.20
HARQ	68.14	69.60	72.83	69.68
HARQ-F	99.88	100.00	100.00	100.00
		EPA-2		
AR	0.00	0.00	1.06	21.60
HAR-J	16.94	12.50	40.47	38.74
CHAR	0.27	0.60	28.36	34.52
SHAR	99.99	100.00	98.94	82.36
ARQ	0.17	0.00	29.34	6.80
HARQ	31.86	30.40	27.17	30.32
HARQ-F	0.12	0.00	0.00	0.00

Unit: Percent

Note. We conduct the equal predictive ability (EPA) test with $H_0 : \mathbb{E}[d_t] = 0$ and $H_1 : \mathbb{E}[d_t] > 0$ (EPA-1) and $H_1 : \mathbb{E}[d_t] < 0$ (EPA-2). Each row represents the p-value of the test statistics when $d_t = L(RV_t, F_{2t}) - L(y_t, F_{jt})$ where F_{2t} is the forecast based on the benchmark HAR model, and F_{jt} is the forecast based on the model indicated in the first column in the table. The loss function is QLIKE, i.e., $L(y, F) = y/F - \log(y/F)$. When the subsampling is used, the block size is selected either by the formula (12) and (13) or by referring to Table 8.