Predicting Extreme Quantiles of Financial Returns: The Role and Information Content of Market Liquidity

Daniel Stašek[a]

[a] *Department of Finance, The Faculty of Economics and Administration, Masaryk University, Lipova 41a, Brno, 60200, Czech Republic*

Abstract

An accurate estimation of uncertainty related to the prices of financial assets is among the main interests of researchers and practitioners. Due to the ever changing nature of financial markets, it is still a challenge to find good explanatory variables of the market risks. Within these, we show that the liquidity measures bear useful information content related to the forecasts of extreme quantiles of price returns. In addition, we demonstrate the liquidity explanatory power to differ with the size of market capitalization on total sample of 190 companies. Lastly, we provide an evidence on the issue of mutual interchangeability of liquidity benchmarks and liquidity proxies.

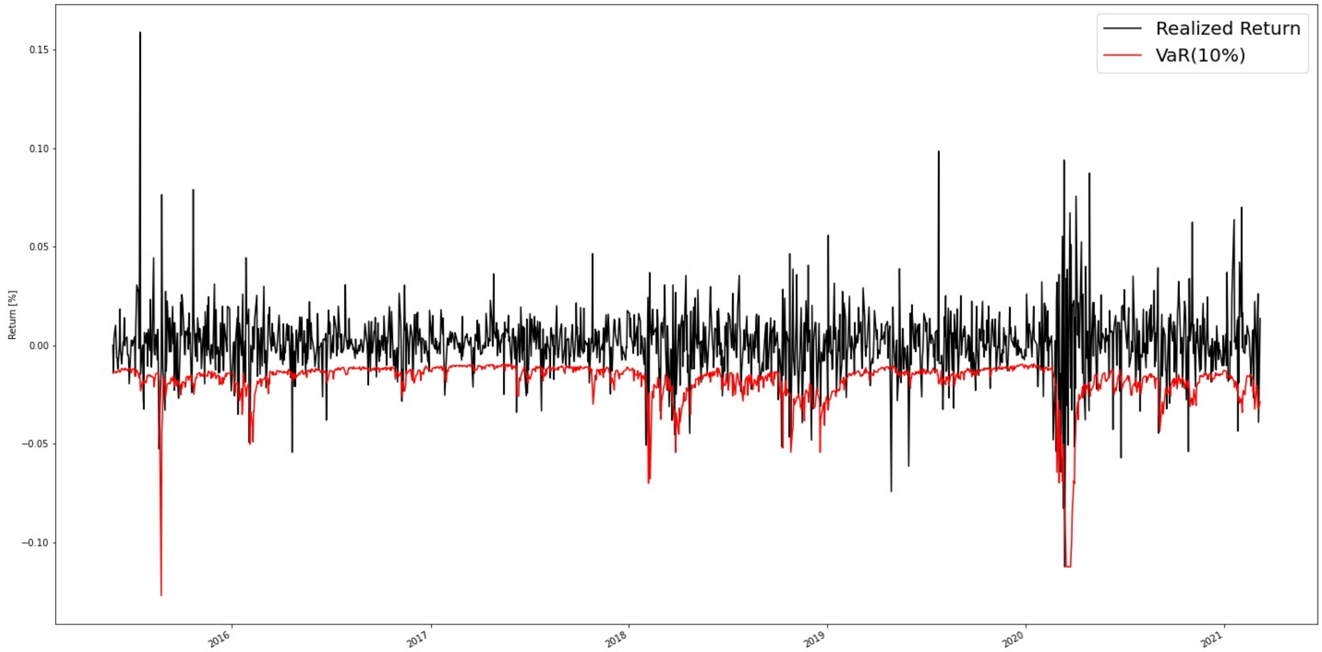
Keywords: Realized volatility, value-at-risk, liquidity measures, forecasting, quantile regression

1. Introduction

In the area of asset management, it is crucial to asses all relevant risks related to the current and future financial holdings. It is no wonder, that the more accurate the assessment of such risks, the better financial decisions can be taken and implemented. Therefore, the researchers aim is to come up, not only with methods, but also with relevant explanatory variables of financial risks. The literature usually differentiates the risks into groups of political, liquidity or interest rate risks to name a few. However, the most examined risk is the market risk related to the price changes of financial assets. The uncertainty related to the assets price changes is quantified by volatility, which is extensively studied over the past 60 years, since the seminal papers of Mandelbrot (1963) and Fama (1965), who were the first to describe the stylized facts of market price time series, and ARCH model proposed by Engle (1982). As the volatility is among the most used variables to appropriately address the market price risk, it cannot provide the full picture. The drawback is that if we assume the following data-generating process (DGP) $d\log(P_t) = \mu dt + \sigma_t dW_t$, for which the μ is a drift parameter, W_t denotes the Wiener process and σ_t defines the time dependent volatility, the price changes $\log(P_t)$ are normally independently distributed with $\log(P_t) \sim N(\mu, \sigma_t^2)$. This is, however, usually not the case of financial time series, which posses a serial dependence of its returns as well as the distribution of $\log(P_t)$ is in practice asymmetrical with fatter tails, than assumed by normal distribution. Because of this, the researchers proposed other measures, specifically related to left tails of an arbitrary distribution for $\log(P_t)$. The measure which can to a certain extent cope with the described drawbacks is the Value-at-risk (VaR). The VaR is defined as an anticipated minimal loss at a given probability level p (more rigorous definition in Section 4).

The development of VaR took a gradual steps over decades and therefore, its direct author is disputable. For this purpose, we mention an interesting review on the development of VaR, summarized by Holton (2002), in which the

Figure 1: Returns of Google and its predicted $VaR(10\%)$



Notes: Daily realized return in black, estimated VaR(10%) in red.

first authors are tried to be identified. Among the most popular (of its time), and quickly adopted by practitioners, was the methodology of $RiskMetrics^{TM}$, at foremost because it was made freely available to the industry. Followed by the extension of ARCH family models – the Asymmetric Power ARCH model (A-PARCH) of Ding et al. (1993) and Conditional Autoregressive Value-at-Risk (CAViaR) proposed by Engle and Manganelli (2004). In both cases these models try to address the deficiencies, such as the clustering feature of volatility, for which the original version of $RiskMetrics^{TM}$ did not account. Mainly due to its simplicity, currently applied models use the approach of Haugom et al. (2016), who proposed to estimate the corresponding VaR levels via quantile regression. We adopt this parsimonious approach in our research setup. The positive side of this estimation procedure, via quantile regression, is that we can use different VaR[1] estimates and aggregate them into an Expected shortfall (ES) measure of tail risk (as in Lyócsa et al. (2022)). This gives us an opportunity, to further extend our findings to the estimation of more rigour measure of tail risk – the ES. It was shown (e.g., by Acerbi and Tasche (2002)), that the VaR measure does not full-fill the axiom of sub-additivity and it cannot fully describe the tail risks, due it its limited information on what is happening below the estimated VaR level (Du and Escanciano (2017)).

In this manuscript, we predict one-day ahead VaR with different measures of market liquidity. The market liquidity is usually referred as: *"the ability to trade large quantities quickly without moving the price while maintaining the lowest possible cost"*. In contrast to volatility, which we precisely expressed in a simple mathematical formula, the definition of liquidity is rather loose and is perceived as an unobserved concept. In order to measure the liquidity, the researchers introduced different dimensions – among the ones of our interest is the price impact[2]

[1] Each of which is estimated on different probability level.

[2] With which we aim to measure the first part of the sentence defining liquidity: *"trade large quantities quickly without moving the price"*.

and trading cost measure[3]. We classify used liquidity measures into these categories in Section 3. Further, we follow the naming convention of Goyenko et al. (2009), who split the liquidity measures into groups of liquidity benchmarks and liquidity proxies. The liquidity benchmarks are derived from the trades-and-quotes (TAQ) data and its computation is problematic due to high requirements on cleaning and processing into a usable measure. Due to this, liquidity proxies were proposed, to overcome the drawbacks of liquidity benchmarks. However, this is not without a flaw, as the proxy does not allow us to accurately measure the price impact and trading costs. Due to this, several papers were introduced, which compares the informational content of liquidity proxies and benchmarks. For example, Bedowska-Sójka and Echaust (2020); Abdi and Ranaldo (2017) on equities, Karnaukh et al. (2015) on currencies and Marshall et al. (2012) on commodities. We see the defficiency of those papers in that they do measure the level of shared information via correlation or cross-entropy and solely on in-sample data. Therefore, we contribute to this literature in that we compare the information content of liquidity benchmarks and proxies in an application on VaR forecasting in an out-of-sample fashion.

To our extent knowledge, paper by Rubia and Sanchis-Marco (2013) is the only one closely related to us. Their aim was to improve the forecasting performance of VaR measure with liquidity related variables. In contrast to us, they used measures derived from limit-order book (LOB) and similarly employed the quoted spread. Additionally to them, we add a battery of liquidity measures – from liquidity proxies group and two additional liquidity benchmarks. Further, their sample ended in 2002, which is limiting in the practical implications of their findings.

With this paper, we bring several contributions to the current knowledge. Firstly, we uncover the predictive power of generally applied liquidity measures on extreme quantiles of financial returns. The set of measures, we implement for this purpose, has not been explored in the VaR forecasting literature. Further, we provide this evidence for 190 most capitalized U.S. stocks. Secondly, we examine whether, the researchers can use the liquidity proxies and liquidity benchmarks interchangeably. In other words, whether the liquidity proxies can bear the same amount of information as do the liquidity benchmarks in VaR forecasting. Lastly, we show that the explanatory power of liquidity can vary across the size of examined stocks.

The rest of this manuscript is organized as follows. In Section 2 we introduce the data-set from which all relevant variables were computed. Next, in Section 3 we define all variables used in the VaR forecasting tasks. This is followed by Section 4, in which we define the VaR models, their estimation procedures as well as the methodology necessary to evaluate the resulting forecasts. Finally, we present the Results and conclude.

2. Financial data

We used a unique data set[4] containing trades and bid/ask quotes for the most capitalized companies traded on the US exchanges. Those data are collected from the Security Information Processor (SIP). The SIP consolidates all bid and ask quotes from all trading venues into a single data feed (also known as Consolidated Tape or Consolidated Feed). Algoseek collects the data via ultra-low latency connection, therefore, in our data set, there may occur certain time discrepancies if compared to direct feeds from the exchanges, however, working with the daily data, we can

[3] Related to the remaining part of the definition: *"lowest possible cost"*.

[4] Obtained via Algoseek company, which we are very grateful to for providing us with the data and their willingness and support.

neglect this issue[5]. The advantage of using SIP data has an advantage of the ability to analyse the US stock market as a whole. The stocks are traded at 18 different market centers[6] at the same time. If we had data from direct feeds, we could observe the price changes on a single exchange only. With SIP, we can see the price changes on all of them including the national best bid offer (NBBO).

The data are provided on as–is basis, which could result in bad trades, prices or crossing quotes. Therefore, the data needed an appropriate cleaning. We follow the cleaning procedure of Barndorff-Nielsen et al. (2009), out of which we used the following rules:

R1. Delete entries with a timestamp outside of 09:30:00 − 16:00:00 (EST time) time window. This left us with records during the official market hours.

R2. Keep entries originating from New York Stock Exchange (NYSE) only. Delete the rest.

R3. Delete records with prices (bid or ask) equal to[7] 0.

R4. We replace quotes with the same timestamp for a single entry with the median bid and median ask price and median volume.

R5. Delete entries in which the bid price is higher than ask price at any time t.

R6. Delete entries for which the spread is more that 50 times the median spread on that day.

Further, we need to clean the trades data which are also used in liquidity measures computation. For this purpose, we applied the same rule R4 on trades data. Further, we followed the rule of Barndorff-Nielsen et al. (2009):

R7. Delete entries with prices that are above the 'ask' plus the bid–ask spread. Similar for entries with prices below the 'bid' minus the bid–ask spread.

As the trades from the consolidated feed do not contain the aggressor tag, meaning we do not certainly know whether the record of a trade is buyer or seller initiated, we used the simple Lee-Ready algorithm to classify our trades. We resorted to this type of classificator due to its simplicity, which is reflected in low computational requirements, that was crucial in our setup.

We omit the rules that delete entries with different conditions. In our data-set, the conditions are coded. To decode the the conditions would take serious amount of computational time, thus, we decided to keep the trades in as-is for those rules. On the other hand, the conditions to be deleted may include extended hours trades, market center official close/open (which we already dealt with) among others[8].

Data for pre- and post-market hours are also available, however we decided to use the official market hours only starting at 09:30:00 to 16:00:00 (EST time). Further, a day before US holidays, some of our measures indicated high iliquidity, we decided to smooth them by the average last 5 days liquidity.

The span of our data starts at 2010 until 25th May 2021, with focus on most capitalized companies traded on the US stock market[9].

---

[5]It likely to be in mili- or micro-seconds, which is not our concern at all. That is why because according to law, the quote must be send to the SIP first, before it is sent to direct feed.

[6]However, the stocks can be traded off-exchange, one of the example are dark pools or orders executed on internal brokerage books.

[7]Some of the exchanges send quotes equal to 0, to indicate the end of the regular market hours.

[8]Official docs of NYSE $https : //www.nyse.com/publicdocs/nyse/data/Daily_TAQ_Client_Spec_v3.0.pdf$, page 17. Or algoseek $https : //us - equity - market - data - docs.s3.amazonaws.com/algoseek.US.Equity.TAQ.pdf$ page 12.

[9]For full reference on used stocks, see Appendix A

3. Liquidity measures

We follow the naming convention of Goyenko et al. (2009) and refer to the measures derived from ultra-high frequency data as liquidity benchmarks, and those derived from lower-frequency as liquidity proxies. We compute the spread measures in dollars and weight them appropriately.

3.1. Liquidity benchmarks

The process of the measurement of quoted spread follows each re-quotation on given trading venue. In other words, for every order submission or cancellation that takes place at the best bid or ask position, the quoted spread is calculated as:

$$\$qs_{i,t} = ask_{i,t} - bid_{i,t} \tag{1}$$

where i denotes the ith re-quotation at day t. As we aim to predict daily value-at-risk, we have to aggregate the quoted spreads as recorded at each re-quotation to its daily observations. In this manner we follow two aggregation strategies, first a simple mean:

$$\$QS_t = \frac{1}{I}\sum_{i=1}^{I} \$qs_{i,t} \tag{2}$$

and then volume weighted mean:

$$\$wQS_t = \frac{1}{I}\sum_{i=1}^{I} \omega_i \$qs_{i,t}, \tag{3}$$

where ω_i is the weight for given re-quotation, calculated as $\omega_i = \frac{volume_{i,t}}{volume_t}$, in which the $volume_{i,t}$ denotes the total volume at best bid and ask prices and $volume_t$ the sum of $volume_{i,t}$ over all i. Similar process is followed in the construction of the two remaining spread metrics. The effective spread:

$$\$es_{k,t} = 2 \cdot \left| p_{k,t} - \frac{(bid_{k,t} - ask_{k,t})}{2} \right| \tag{4}$$

which in contrast to the quoted spread uses each trade k, to record its value. Therefore, the p_k denotes the execution price of kth trade at day t. and the valid best $bid_{k,t}$ and $ask_{k,t}$ prices at the time of the trade k. The aggregation procedure follows Equation 2, for Equation 3 it is weighted by the size of trade k. This is same as in case of realized spread, which we calculate as follows:

$$\$es_{k,t} = 2 \cdot \left| p_{k,t} - p_{k+5,t} \right| \tag{5}$$

where $p_{k+5,t}$ denotes the price of a trade approximately 5 minutes after trade k. The 5 minute is approximate because at some less traded stocks, the trade may or even may not occur during the 5 minutes. In this case we take the price of trade, that is closest to 5 minutes treshold. If no trade occurred in interval if $(4min; 6min)$ after the trade, the computation of realized spread for kth trade is omitted.

3.2. Liquidity proxies

As a counterpart to the benchmark measures, we use 2 cost measures and 2 price impact measures in order to better compare, whether the benchmark contain more information needed for better VaR predictions. Those measure are usually used on frequencies higher than one day, therefore instead of using daily data to estimate these proxies, we use the intraday data to estimate the daily liquidity proxies. In order to not avoid possible miscalculation due to market frictions, which are of a main interest in intraday-prices estimators, we decided to use several frequencies for our computations. The following proxies are computed using 5 minute, 15 minute, 30 minute frequencies (more information in Section 4).

First spread proxy is Roll (1984) estimator of the effective spread:

$$Roll_t = \begin{cases} 2 \cdot \sqrt{-cov(r_\tau, r_{\tau-1})} & \text{if } cov(r_\tau, r_{\tau-1}) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where r_τ is vector of mid-price returns of day t with frequency τ. In order to measure daily $Roll_t$ we resorted to different return frequencies $\tau \in \{5Min, 15Min, 30Min\}$ frequencies. Next, we utilize Zeros – liquidity proxy measure introduced by Lesmond et al. (1999):

$$Zeros_t = zeros_\tau / \mathcal{T} \tag{7}$$

where $\mathcal{T}$ is the total number of price observations at day t (which depends on the data frequency) and $zeros_\tau$ is the total number of zero return price buckets at frequency τ. First price impact proxy is Amihud (2002) – illiquidity ratio:

$$Amihud_t = average(\frac{r_\tau}{volume_\tau}) \tag{8}$$

Second, price impact measure is the Roll's impact. Following the approach of Goyenko et al. (2009), we divide the $Roll_t$ by a total dollar volume[10]:

$$Impact_t = \frac{Roll_t}{Volume_t} \tag{9}$$

4. Value-at-Risk Models

There are several, well established and widely applied by practitioners, methods such as historical simulation or $RiskMetrics^{TM}$ method. In order to overcome its deficiencies[11] further models were proposed, such as Assymetric Power ARCH model (A-PARCH) of Ding et al. (1993) or Conditional Autoregressive Value-at-Risk (CAViaR) proposed by Engle and Manganelli (2004) among others. Generally, the ARCH family models are difficult to estimate and therefore we decided to use, a rather parsimonious, quantile regression model of Koenker and Bassett Jr (1978), which was recently applied for VaR predictions by Haugom et al. (2016).

For a known and continuous cumulative distribution function F of returns R_t, we can define the VaR at given probability level p as:

$$VaR_p(R_t) = \inf\{var : F(var) \geq p\} \tag{10}$$

However, as the true cumulative distribution F of returns is unknown and discrete, we can define the VaR for

[10]The numerator can be exchange by any liquidity proxy measure.
[11]Most importantly the inability to capture most of the stylized facts – time varying volatility and asymmetrical returns.

our sample data R_t more conveniently as:

$$\Pr(R_t < VaR_t|\Omega_{t-1}) = p \tag{11}$$

which is conditioned by the past available information Ω_{t-1} and p is the probability for which we want estimate the VaR_t.

To further estimate the VaR_t we regress the daily returns R_t on lagged volatilities following the approach of Haugom et al. (2016) or Lyócsa and Todorova (2020) we use the Heterogenous Autoregressive Quantile Regression model (HAR-QREG) as:

$$R_{p,t+1} = \hat{\beta}_0 + \hat{\beta}_1 RV_t^d + \hat{\beta}_2 RV_t^w + \hat{\beta}_3 RV_t^m + \epsilon_{p,t} \tag{12}$$

in which the RV_t^d, RV_t^w, RV_t^m stands for daily, weekly and monthly realized volatility. ϵ_t need not to be normally, identically and independently distributed (NIID) as is the case of classical OLS estimator. To express the Equation 13 in Value-at-risk like framework, we can write:

$$\widehat{VaR}_{p,t+1}|RV_t^d, RV_t^w, RV_t^m = \hat{\beta}_0 + \hat{\beta}_1 RV_t^d + \hat{\beta}_2 RV_t^w + \hat{\beta}_3 RV_t^m + \epsilon_{p,t} \tag{13}$$

The HAR-QREG model will be used as a benchmark to models extended by liquidity measures. We implement several groups of competing models. First competing model is a set of all liquidity measures defined in lower. Second and third competing model extend the benchmark with liquidity benchmarks and liquidity proxies respectively. All of those competing models are extended by a single measure $X_{i,t}$ at a time:

$$\widehat{VaR}(X_{i,t})_{p,t+1} = \hat{\beta}_0 + \hat{\beta}_1 RV_t^d + \hat{\beta}_2 RV_t^w + \hat{\beta}_3 RV_t^m + \hat{\beta}_4 X_{i,t} + \epsilon_{p,t} \tag{14}$$

where i denotes the ith measure from the subset of measures chosen to serve as a competing model. We define three different sets of liquidity measures that are used, in order to explore the information content of chosen liquidity measures:

all = Bench $\cup$ prox

bench = $\{\omega ES^{1D}, ES^{1D}, \omega QS^{1D}, QS^{1D}, \omega RS^{1D}, RS^{1D}, \omega ES^{5m}, ES^{5m}, \omega QS^{5m}, QS^{5m}, \omega RS^{5m}, RS^{5m}\}$

Prox = $\{Zeros^{5m}, Zeros^{15m}, Zeros^{30m}, Impact^{5m}, Impact^{15m}, Impact^{30m}, Roll^{5m}, Roll^{15m}, Roll^{30m},$
$Amihud^{5m}, Amihud^{15m}, Amihud^{30m}, Amihud^{1D}\}$

, in which *all* defines the set of all liquidity measures, *bench* comprises all liquidity benchmarks and analogically *prox* aggregates the liquidity proxies.

Then, we resort to adopt the complete subset regression approach of Elliott et al. (2013), which resulted in better volatility predictions (e.g. in Lyócsa and Stašek (2021)). For the final prediction, we use a simple mean over the all i predictions:

$$\widehat{VaR}_{p,t+1} = \frac{1}{N}\sum_{i=1}^{N} \widehat{VaR}(X_{i,t})_{p,t+1} \tag{15}$$

where N is the total number of liquidity variables, used to beat the benchmark model.

4.1. Evaluation metrics and tests

For the evaluation of models defined in Section 4, we use several back-testing methods. First one is the test of Unconditional coverage of Kupiec et al. (1995), which measure the proportion of failures during the examined period. Put differently, if we estimate $\widehat{VaR}_{0.1}$ (VaR at 0.1% probability level), we should expect the market return to exceed the level of $\widehat{VaR}_{0.1}$ (also called a violation) in exactly 10.0% of cases. If the proportion of violation are significantly different from the level p, we can conclude that the model is not appropriate. When the proportion of violation is greater than p, then the risk managers will incurr higher losses than expected. On the other hand, when the ratio is smaller than p, the managers would tend to overestimate the risks and employ less capital in their decision, which would consequently lead into an "under-investment", leading to lower profits.

Second evaluation criterion we employ was proposed by Christoffersen (1998) and its main aim is to examine the independence of the observed violation. In other words, if at day t we observe a violation, it is undesirable to observe violation at any upcoming day $t+1$ or $t+i$, where i is reasonably small number. The null hypothesis states that the duration between the violations has no memory.

Third criterion has a quasi-form of classical loss function and was proposed by Lopez (1999). The difference is that we calculate the total loss only in cases of violations:

$$L_t(R_t, \widehat{VaR}_{p,t}) = \begin{cases} 1 + (R_t - \widehat{VaR}_{p,t})^2, \text{if } R_t < \widehat{VaR}_{p,t} \\ 0, \text{otherwise} \end{cases} \tag{16}$$

, where the loss $L_t(.)$ is aggregated by a simple mean for final comparison. Furthermore, we apply a test of Diebold and Mariano (1995) (DM) test on the losses $L_t(.)$.

5. Results

The first question of interest is to answer, whether the widely applied liquidity measures, defined in Section 3, can improve the forecasting performance of VaR. As was shown in Section 4.1, we employ a battery of statistical tests and loss functions, to thoroughly describe different aspects of the forecasts. It can be the case, that one of the test is improved, but the other decrease in evaluated accuracy. Therefore, the final choice of relevant test or loss function depends on the real-world application. From this point of view, our tests and losses should provide enough information necessary, to decide the relevance of forecasting improvement.

We decided to test the VaR at three different levels of 10%, 5% and 1% (as depicted in e.g. Table1), due to the fact, that each of the level can be explained by different explanatory variables, as its behavior may be different. On the 10% level we can observe an improvement in case of Kupiec test, in which the extension by liquidity measures improved the equality to the 10% in case of 6 stocks. Further improvement is shown in the Lopez loss, which average over all stocks decreased by 4%. In contrast to that, the statistically significant improvement was observed only in case of 12 stocks. Another metric that has improved is the Lopez ratio, which in our case point to the fact that the violation ratio of the augmented QHAR-RV-all (tightly connected to the test of Kupiec) was closer to the expected ratio of 10% in case of 6 more companies. In contrast to the improvements, the addition of liquidity measure lead to a decrease in time independent violations as measured by the Christoffersen test. Despite some

Table 1: Comparison of baseline model against the augmented version

	10% VaR		5% VaR		1% VaR	
	HAR-RV	HAR-RV-all	HAR-RV	HAR-RV-all	HAR-RV	HAR-RV-all
Christof.	7.10%	10.33%	26.20%	8.60%	57.89%	32.80%
Kupiec	16.94%	13.59%	30.48%	23.66%	76.84%	64.55%
Lopez ratio	0.484	0.516	0.463	0.537	0.342	0.658
Lopez Loss	1.297	1.248	1.459	1.374	2.551	2.093
DM test	–	6.84%	–	4.21%	–	10.00%

Notes: The abbreviation Christoff. stands for the duration test of Christoffersen and Pelletier (2004). Values in the first two rows shows the ratio of rejections (either of duration test or kupiec) to the total number of examined assets. Lopez ratio is the ratio of cases in which the Violation ratio was closer to the VaR level. DM test is the Diebold-Mariano test and the value depicts the number of cases in which the losses were statistically better, than the baseline model. Each such result is presented at three different violation probabilities – 10%, 5% or 1%.

minor improvements, our results do not improve the $VaR_{10\%}$ as we expected based on previous literature (e.g., Rubia and Sanchis-Marco (2013)).

Further results in Table 1, shows better improvements in case of $VaR_{5\%}$ for all evaluation metrics. In case of the Christoffersen test we achieved to increase the time independency of individual violation in more than 33 stocks in comparison to the baseline model QHAR-RV. For Kupiec the baseline model extension lead to an improvement in case of 13 stock and in Lopez ratio for 14 stocks. Compared to $VaR_{10\%}$, we report decreased cross-sectional average Lopez loss, in this case of more than 6%. However, the reason for these improvements can be contradicted by a generally worse performance of the baseline model, thus, it remains an open question, if we used different benchmark model, whether the improvements were of the reported size.

The third column of Table 1 compares the predictions for $VaR_{1\%}$. The test in case of the baseline models reject the nulls of Christoffersen and Kupiec in more than 57.89% cases. Despite the high improvements demonstrated by their extension with liquidity measures, we decided to omit the $VaR_{1\%}$ from the results presented in further sections. Simply, because our baseline model cannot predict $VaR_{1\%}$ without passing the basic tests, and thus, any variable addition into such model, will likely lead to an improved forecasting performance.

Table 2: Comparison of basic models split across company size

	10% VaR, high-cap		10% VaR, low-cap		5% VaR, high-cap		5% VaR, low-cap	
	HAR-RV	HAR-RV-all	HAR-RV	HAR-RV-all	HAR-RV	HAR-RV-all	HAR-RV	HAR-RV-all
Christof.	6.52%	13.04%	6.82%	13.64%	28.26%	13.04%	26.67%	4.44%
Kupiec	4.35%	6.52%	22.73%	20.45%	34.78%	23.91%	28.89%	26.67%
Lopez ratio	0.522	0.478	0.413	0.587	0.435	0.565	0.565	0.435
Lopez Loss	1.228	1.177	1.433	1.360	1.379	1.293	1.614	1.544
DM test	–	4.35%	–	6.52%	–	2.17%	–	2.17%

With the improved results on 5% level and slightly on the 10% level, it is naturally leading us to try to explore, why the improvements took place in the given assets. In this manner, we followed a simple proposition, that the liquidity could be higher in case of companies with higher market capitalization. In other words, it could be the case, that the our liquidity measure explain the respective VaR levels differently, as the assumed intrinsic liquidity should be different.

Indeed, in Table 2 we show these improvements in the case of $VaR_{5\%}$. Interestingly, the best highest increase is for the low-cap companies and in the case of the Christoffersen test, for which we decreased the number of time-independent violations for 10 cases, with also improved average loss by approximately 5%. Similar improvements hold for the high-cap companies, additionally an increase in the Lopez ratio is by 7%. This is an interesting finding, in counter intuitive results, that the liquidity measure do play a role in case of both, the high-cap and low-cap stocks. The improvements would not be observed in the case of the mid-caps.

Table 3: Comparison of baseline model to the competing extension of liquidity proxies and benchmarks

	10% VaR			5% VaR		
	HAR-RV	HAR-RV-prox	HAR-RV-bench	HAR-RV	HAR-RV-prox	HAR-RV-bench
Christof.	7.69%	8.20%	8.74%	22.58%	6.99%	10.22%
Kupiec	14.29%	14.75%	14.75%	27.96%	26.88%	26.88%
Lopez ratio	0.519	0.481	0.519	0.547	0.453	0.542
Lopez Loss	1.289	1.253	1.250	1.448	1.374	1.374
DM test	-	4.23%	4.23%	-	3.68%	4.21%

Notes: The abbreviation Christoff. stands for the duration test of Christoffersen and Pelletier (2004). Values in the first two rows shows the ratio of rejections (either of duration test or kupiec) to the total number of examined assets. Lopez ratio is the ratio of cases in which the Violation ratio was closer to the VaR level. DM test is the Diebold-Mariano test and the value depicts the number of cases in which the losses were statistically better, than the baseline model. Each such result is presented at three different violation probabilities – 10%, 5% or 1%.

So far, we showed that the improvements in case of $VaR_{10\%}$ are rather mixed, but in case of $VaR_{5\%}$ they can stand out in most of the measures used. Therefore, to answer our second question, we split the all liquidity measures into groups of liquidity proxies and benchmarks, and answer, whether the proxies or benchmarks provide more relevant information content for our main task of VaR forecasting. For this purpose with present show the results in Table 3

Interestingly, in case of $VaR_{10\%}$ the values in both QHAR-RV-prox and QHAR-RV-bench are almost equal, but, they in both cases decreased the accuracy in cases of Christoffersen and Kupiec. Increased, and roughly same improvement is observer in the case of Lopez loss – by 3%. In case of the second column in which we present the results for $VaR_{5\%}$, both extensions improve most of the evaluation criterions, most notably in case of the test of Christoffersen. Thus, the conclusions we can draw are similar to those based on Table 1. In case of $VaR_{5\%}$ the improvements are present over most of the evaluation criterion's. With claim that, the proxies do perform equally well as the liquidity benchmarks – answering the third research question of this paper. In case of VaR forecasting, we can use the liquidity benchmarks and proxies interchangeably, but this finding is limited to the $VaR_{5\%}$.

6. Conclusion

In this paper we explored the relationship between the VaR and liquidity measures. We showed, that under certain circumstances, the liquidity measures do improve the forecasts of VaR, specifically in case of $VaR_{5\%}$. Furthermore, we demonstrated that these improvements led to an similar performance in cases of models extended either by liquidity benchmarks or liquidity proxies. Based on this observations we concluded, that the liquidity proxies can be used in order to mimic the information content of liquidity benchmarks in case of VaR forecasting framework. Lastly, we demonstrated that the information content of liquidity measures differ across the size of the examined companies. Thus, the size of company matters, but in contrast to our prior hypothesis, we showed that the lower the size, the better the results were achieved.

7. Acknowledgments

This work was supported from Operational Programme Research, Development and Education-project *"Internal Grant Agency of Masaryk University"* (No.CZ.02.2.69/0.0/0.0/19_073/0016943).

References

Abdi, F., Ranaldo, A., 2017. A simple estimation of bid-ask spreads from daily close, high, and low prices. The Review of Financial Studies 30, 4437–4480.

Acerbi, C., Tasche, D., 2002. Expected shortfall: a natural coherent alternative to value at risk. Economic notes 31, 379–388.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of financial markets 5, 31–56.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes.

Bedowska-Sójka, B., Echaust, K., 2020. Do liquidity proxies based on daily prices and quotes really measure liquidity? Entropy 22, 783.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics 2, 84–108.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International economic review , 841–862.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. journal of business and economics statistics, v. 13 .

Ding, Z., Granger, C.W., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of empirical finance 1, 83–106.

Du, Z., Escanciano, J.C., 2017. Backtesting expected shortfall: accounting for tail risk. Management Science 63, 940–958.

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. Journal of Econometrics 177, 357–373.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the econometric society , 987–1007.

Engle, R.F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of business & economic statistics 22, 367–381.

Fama, E.F., 1965. The behavior of stock-market prices. The journal of Business 38, 34–105.

Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? Journal of financial Economics 92, 153–181.

Haugom, E., Ray, R., Ullrich, C.J., Veka, S., Westgaard, S., 2016. A parsimonious quantile regression model to forecast day-ahead value-at-risk. Finance Research Letters 16, 196–207.

Holton, G.A., 2002. History of value-at-risk.

Karnaukh, N., Ranaldo, A., Söderlind, P., 2015. Understanding fx liquidity. The Review of Financial Studies 28, 3073–3108.

Koenker, R., Bassett Jr, G., 1978. Regression quantiles. Econometrica: journal of the Econometric Society , 33–50.

Kupiec, P.H., et al., 1995. Techniques for verifying the accuracy of risk measurement models. volume 95. Division of Research and Statistics, Division of Monetary Affairs, Federal . . . .

Lesmond, D.A., Ogden, J.P., Trzcinka, C.A., 1999. A new estimate of transaction costs. The review of financial studies 12, 1113–1141.

Lopez, J.A., 1999. Methods for evaluating value-at-risk estimates. Economic Review-Federal Reserve Bank of San Francisco , 3.

Lyócsa, Š., Plíhal, T., Vỳrost, T., 2022. Forecasting day-ahead expected shortfall on the eur/usd exchange rate: The (i)relevance of implied volatility. Working paper .

Lyócsa, Š., Stašek, D., 2021. Improving stock market volatility forecasts with complete subset linear and quantile har models. Expert Systems with Applications 183, 115416.

Lyócsa, Š., Todorova, N., 2020. Trading and non-trading period realized market volatility: Does it matter for forecasting the volatility of us stocks? International Journal of Forecasting 36, 628–645.

Mandelbrot, B., 1963. The variation of certain speculative prices. The Journal of Business 36, 394–419.

Marshall, B.R., Nguyen, N.H., Visaltanachoti, N., 2012. Commodity liquidity measurement and transaction costs. The Review of Financial Studies 25, 599–638.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of finance 39, 1127–1139.

Rubia, A., Sanchis-Marco, L., 2013. On downside risk predictability through liquidity and trading activity: A dynamic quantile approach. International Journal of Forecasting 29, 202–219.

Appendix A. Symbols of used assets

AAPL, AMGN, AXP, BA, CAT, CPB, CRM, CSCO, CSX, CTAS, CTSH, CTXS, CVS, CVX, D, DE, DFS, DGX, DHI, DHR, DIS, DISCA, DOV, DRI, DTE, DUK, DVA, DVN, EBAY, ECL, ED, EFX, EIX, EL, EMN, EMR, EOG, EQR, ETN, ETR, EW, EXC, EXPD, EXPE, F, FAST, FCX, FDX, FE, FFIV, FIS, FISV, FITB, FMC, GD, GE, GILD, GIS, GLW, GNW, GOOG, GPC, GPS, GS, GWW, HAL, HAS, HBAN, HD, HES, HIG, HON, HPQ, HRL, HST, HSY, HUM, IBM, ICE, IFF, INTC, INTU, IP, IPG, IR, IRM, ISRG, ITW, IVZ, JNJ, JNPR, JPM, K, KEY, KIM, KLAC, KMB, KMX, KO, KR, L, LEG, LEN, LH, LLY, LMT, LNC, LOW, LUV, MCD, MCHP, MCK, MCO, MRK, MSFT, NKE, NOC, NOV, NRG, NSC, NTAP, NTRS, NUE, NVDA, NWL, NWSA, OKE, OMC, ORCL, ORLY, OXY, PAYX, PBCT, PCAR, PEG, PEP, PFE, PFG, PG, PGR, PH, PHM, PKI, PLD, PM, PNC, PNW, PPG, PPL, PRU, PSA, PWR, PXD, QCOM, RF, RHI, RL, ROK, ROP, ROST, RSG, SBUX, SCHW, SEE, SHW, SJM, SLB, SNA, SO, SPG, SRE, STT, STZ, SWK, SYK, SYY, T, TAP, TGT, TJX, TMO, TROW, TRV, TSN, TXN, TXT, UNH, V, VZ