



*RISK MEASURE ESTIMATES IN QUIET AND TURBULENT
TIMES: AN EMPIRICAL STUDY*

RESEARCH CENTER

ROSNAN CHOTARD, MICHEL DACOROGNA AND MARIE KRATZ

ESSEC WORKING PAPER 1618

NOVEMBER 2016

Risk Measure Estimates in Quiet and Turbulent Times: An Empirical Study

Rosnan Chotard[†], Michel Dacorogna* and Marie Kratz[†]

* DEAR Consulting, Switzerland

[†] ESSEC Business School, CREAR

November 24, 2016

Abstract

In this study we empirically explore the capacity of historical VaR to correctly predict the future risk of a financial institution. We observe that rolling samples are better able to capture the dynamics of future risks. We thus introduce another risk measure, the Sample Quantile Process, which is a generalization of the VaR calculated on a rolling sample, and study its behavior as a predictor by varying its parameters.

Moreover, we study the behavior of the future risk as a function of past volatility. We show that if the past volatility is low, the historical computation of the risk measure underestimates the future risk, while in period of high volatility, the risk measure overestimates the risk, confirming that the current way financial institutions measure their risk is highly procyclical.

2010 AMS classification: 60G70; 62M10; 62P05; 91B30; 91B70

JEL classification: C13; C22; C52; C53; G01; G32

Keywords: backtest; risk measure; sample quantile process; stochastic model; VaR; volatility

1 Introduction

The introduction of risk based solvency regulations has brought the need for financial institutions to evaluate their risk on the basis of probabilistic models. Since the riskmetrics attempt by JP Morgan, the capital needed by companies to cover their risk is identified to the quantile at a certain threshold α of the return distribution of the portfolio (see [17]). They called this risk measure Value-at-Risk (VaR), name that has been adopted by the financial community. The question to know if this is the appropriate risk measure to use for evaluating the risk of financial institutions has been heavily debated since the financial crisis of 2008/2009. For a review of the arguments on this subject, we refer e.g. to ([13]).

To estimate empirically a risk measure, besides its definition, there are at least two parameters to consider: one is the threshold at which this risk measure is estimated, the second is the historical sample used for estimating it. The aim of this study is to look at the dynamic behavior of risk measures when measured on various sample sizes at different thresholds. The first measure we examine is the VaR and the second one is the Sample Quantile Process (SQP) introduced by Muria ([15]), then studied by various authors for different stochastic processes (see Akahori ([1]), Akahori & Urata ([2]), Dassios ([8],[9]), Yor ([16]), Embrechts et al. ([12])). Since the SQP has a more dynamical view through the measure under which it is defined, we want to explore various choices of it to test its relevance in practice, especially in time of financial crisis, compared to the use of the historical VaR.

Indeed, ways of measuring risk are important, notably with the reinforcement of capital requirements regulation these past years (Basel II and III). The goal of this study is to find risk measures that provide, when computed on past data, an accurate vision of the future risk. In our paper, we analyze and optimize risk measures, using static and dynamic risk measures, conditional and unconditional measures, and allow the frequency of observations as well as the sample size to change, in order to find a suitable scheme for predicting future risks.

Many studies have looked at the performance of VaR estimation to predict future VaR (see, for instance [10] or [6, 4] or [11] and references therein). It has been observed that the performance varies very much according to various phases of the market, making it difficult to use in practice a pure historical and unconditional estimation of the risk measure. Our goal here is to go one step further and explore the behavior of these measures conditioned to past volatility. Our hypothesis is that during high-volatility periods, the VaR overestimates the risks for the following years and during low-volatility periods, the VaR underestimates these risks. As a consequence, should our assumption turn out to be verified, it would be possible to introduce anti-cyclical risk management rules that are at the opposite of what is currently required from financial institutions by the regulators.

In Section 2, we introduce and describe the S&P 500 index as well as its characteristics and define the variables of interest. In Section 3, we assess the historical VaR on this index allowing for changes in maturity, frequency of observations and thresholds. In Section 3.2, we present out-of-sample tests to find out how well VaR measures can represent risks. In Section 4, we introduce various examples of sample quantiles, playing on the choice of the measure. We discuss and compare it with VaR measures. In the last part (Section 5), inspired by a study presented in [7], we design an empirical study where the dynamic behavior of risk measures is conditioned on volatility and show that after a period of high volatility, the risk measure is overestimated. We conclude in Section 6.

2 Data

The data used in this study are the daily closing prices of the S&P 500 Index from Monday, January 2, 1987 to Friday, February 19, 2016. Since our goal is to look at the appropriateness of the VaR calculations from financial institutions, it seems to be natural to use a stock index and see how the various schemes perform on this data.

2.1 S&P 500 Price Time Series

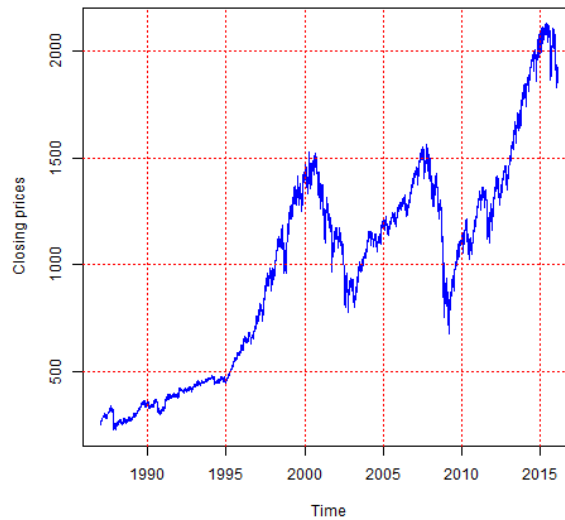


Figure 1: S&P 500 Index closing prices from January 1987 to February 2016

We see in Figure 1, that this period has been marked by a very strong increase of the index that quadruples its value from the beginning to the end, but also by marked draw-downs, from 2001 to 2004 (the internet bubble) and during the financial crisis of 2008/2009. This makes the sample particularly suitable to study various regimes and check the behavior of risk measures during these regimes.

2.2 S&P 500 Return Time Series

Let us denote the closing price at time t of the S&P 500, by $S(t)$ and by Δt the interval between two consecutive spot prices. We shall use daily or weekly intervals. We then define the log-return $X_{\Delta t}(t)$ given by:

$$X_{\Delta t}(t) = \ln \left(\frac{S(t)}{S(t - \Delta t)} \right) \quad (1)$$

We focus on the log-return to the extent that stock and index prices can easily be modelled by geometric Brownian Motions. The logarithmic transformation enables us to deal with a stationary time series. As a consequence, the log-returns are often modelled by normally distributed random variables (notably in the Black-Scholes-Merton or the Markovitz portfolio models).

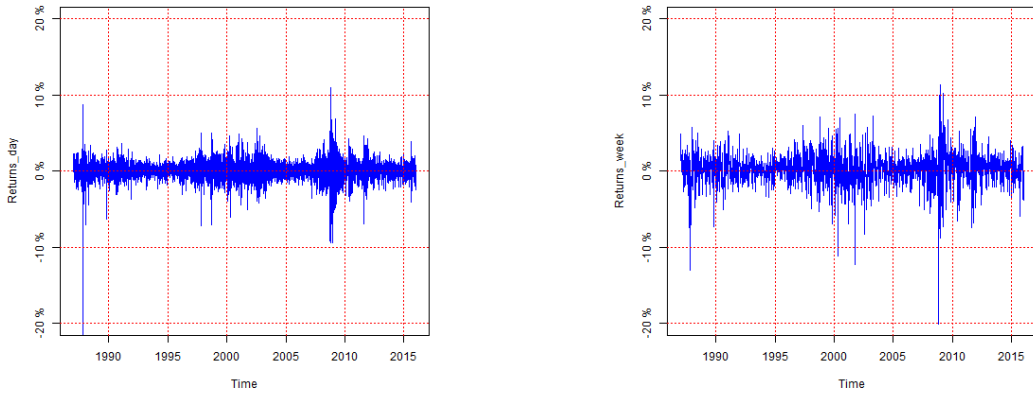


Figure 2: S&P 500 Index returns, with from left to right: daily returns and weekly returns

Note that the log-return is an approximation of the real return, for small values of the variable: $\ln(x) \sim x - 1$. Moreover, the log-return over a period of n days is the sum of the n daily log-returns. Indeed, denoting $X_{s,t}$ the log-return between the day s and the day t , we have:

$$X_{0,n} = \ln(S_n/S_0) = \sum_{i=1}^n \ln(S_i/S_{i-1}) := \sum_{i=1}^n X_i$$

As shown in Figure 2, both the daily and weekly returns time series look like noisy series. Augmented Dickey-Fuller tests, KPSS tests and Philips Perron tests do not reject the stationarity hypothesis of these series at the 1% level. These series are distinctly different from white noise since different volatility regimes (notably high volatility regimes) are clearly apparent in these time series. We note that the high volatility regimes correspond to periods of crisis on the US financial markets: the crash of 1987, the burst of the internet bubble 2001-2003 and finally the financial crisis of 2008/2009. We also see, when comparing daily and weekly returns, that the later have a higher volatility than the daily ones as expected from a pure diffusion process where the volatility grows like the square root of the aggregation factor.

However, the generally used assumption that returns are normally distributed is challenged by the numbers in Table 1 where we report descriptive statistics. Indeed, the skewness (negative) and excess kurtosis (highly positive) of these time series reveal that the left tail is more extended than the right tail (the distribution is not symmetric as the Gaussian distribution) and their distribution exhibit fat or heavy tails (which means that the probability of extreme events is higher than the Gaussian probability). Actually, the skewness of these series stems from the fact that the S&P 500 is an index, that is to say it is the weighted average of stocks, and these stocks have a stronger tendency to decrease at the same time than to increase at the same time. It should also be noticed that during periods of crisis, correlation between stocks tends to strongly increase, reinforcing the asymmetry.

Table 1: Descriptive statistics for the returns of the S&P 500 from January 1987 to February 2016

Statistics	Daily Returns	Weekly Returns
mean	0.03%	0.13%
standard deviation	1.17%	2.33%
skewness	-1.26	-0.84
excess kurtosis	27.48	6.49

Nevertheless, it can be noticed that the excess kurtosis strongly decreases with lower frequency of returns (weekly returns) whereas the skewness stays negative but also decreases. Looking at the aggregation properties: decrease of skewness and kurtosis, one may conclude that the center of the return distribution tends, under aggregation, to the Gaussian distribution (Central Limit Theorem).

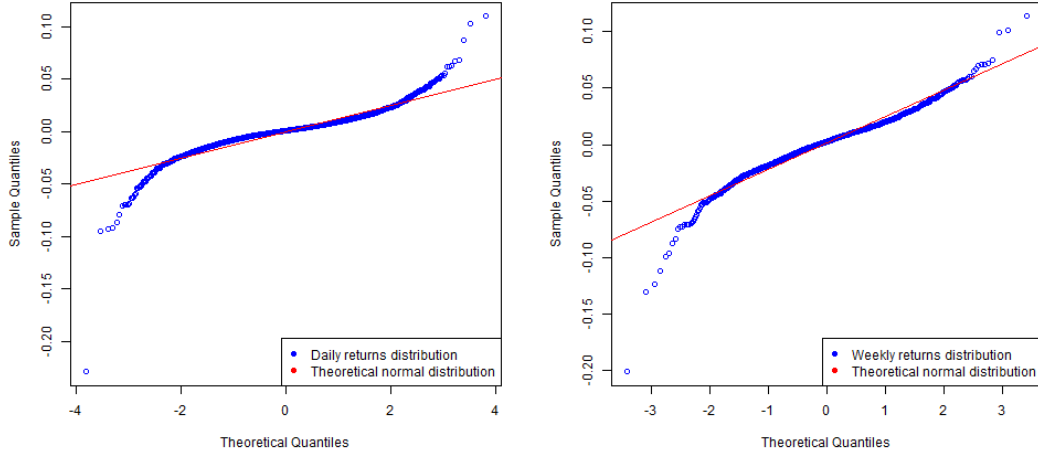


Figure 3: QQ plot returns, with from left to right: daily and weekly returns

In Figure 3, we show the Q-Q plots of both the daily and weekly returns series against the theoretical normal distribution (with mean and standard deviation calculated according to those of the respective samples). These plots are a useful tool for comparing two distributions between each other, as any deviation from the expected linear behavior means that the empirical distribution deviates from the theoretical one. Analyzing the above plots, it can be noticed that the middle of both the daily and weekly returns distributions fits with the red line (theoretical normal distribution). This illustrates nicely the CLT, when looking at the mean behavior.

However, the extreme values and the broad "S"-shape of the returns distributions show that these distributions are more fat-tailed than the normal distributions: with both heavier right and left tails than those of the normal distribution. Nevertheless, note that the weekly returns distribution is closer to the normal distribution than the daily returns distribution is, as pointed out in [7]. This is due to the fact that the second moment of the return distributions exists making the CLT applicable for the center of the distribution.

3 Evaluation of the Historical VaR on the S&P 500 Index

Recall that the Value-at-Risk of a random loss L , at level α , is simply the quantile of order α of L and is defined by

$$\mathbb{P}[L > \text{VaR}(\alpha)] \leq 1 - \alpha \quad \text{denoting} \quad \text{VaR}(\alpha) = \text{VaR}_\alpha(L).$$

The VaR is generally estimated on historical data, using the empirical quantile $F_n^{-1}(\alpha)$ associated to a n -loss sample (L_1, \dots, L_n) with $\alpha \in (0, 1)$, namely:

$$F_{n;L}^{-1}(\alpha) = \inf \left\{ x : \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(L_i \leq x)} \geq \alpha \right\}. \quad (2)$$

Here our observations are the log-returns $X = (X_i)_i$, so computing the VaR of the loss distribution at level α (for instance 95%) corresponds to computing the VaR of the log-returns X at level $1 - \alpha$ (5% respectively) and estimate it by $F_{n;X}^{-1}(1 - \alpha)$ (since $L = -X$).

Note also that, for a more precise estimation, we proceed by linear interpolation between the low and high nearest neighbours of the solution of (2).

In this study, we explore various ways of estimating this quantity. If the underlying process would be stationary and based on iid rv, the bigger the number of observations, the better the quantile estimation. So, it makes sense to first estimate a benchmark quantile on the whole sample. The second study is made on the usual samples used by financial institutions, namely 1, 3 or 5 years, and we update it every year. We call this here a fixed sample because it does not change on a whole year, even if it is not really a fixed sample as it changes after a year. The third study is based on rolling samples that are updated more frequently than one calendar year.

3.1 Computation of the historical VaR

3.1.1 On the full sample

In Table 2, we report the empirically estimated 95% and 99% VaRs', as well as the minimum and maximum returns of both daily and weekly returns time series over the whole sample. Moreover, we added the 95% and 99% VaRs' assuming that weekly or daily returns, denoted by X as in eq. (1), are normally distributed and computed as follows:

$$\text{VaR}_\alpha(X) = \mu + \sigma\phi^{-1}(1 - \alpha)$$

where μ is the mean of the series (daily or weekly returns), σ the standard deviation of the series, both estimated over the full sample, ϕ^{-1} the quantile of the standard Normal distribution, and $1 - \alpha$ being equal to 5% or 1% (respectively for the cases of 95% VaR or 99% VaR). These values constitute the benchmarks of the calculation of the risk measures on shorter samples.

Table 2: The risk measures computed on the whole sample for both daily and weekly returns

Model	Risk measures	Daily Returns	Weekly Returns
Empirical	95% VaR	-1.75%	-3.58%
Gaussian	95% VaR	-1.90%	-3.70%
Empirical	99% VaR	-3.15%	-7.01%
Gaussian	99% VaR	-2.70%	-5.29%
	min return	-22.90%	-20.08%
	max return	10.96%	11.36%

The minimum return and maximum return of both daily and weekly series reveal the negative skew of these series. Indeed, in both cases, the minimum return in absolute value is approximately twice greater than the maximum return. This provides evidence of the existence of a longer left tail for both time series.

Another remark is that in both cases, the returns losses are bigger in absolute value when looking at the 99% VaR than when looking at the 95% VaR, as expected.

Moreover, it can be noticed that weekly VaR (at 95% and 99%) are bigger in absolute value than daily VaR (about twice bigger). One may then assume that increasing the frequency of the returns leads to lower VaR values. It makes sense especially in crisis period when returns keep decreasing during several days in a row and notably because weekly returns volatility is twice as big as the daily returns volatility. This is also in line with the properties of normal distributions that aggregate following a square root scale of n where n is the aggregation factor.

Finally, it can be highlighted that the 95% VaRs' as given by the Gaussian model for daily and weekly returns time series are higher in absolute value than the empirical 95% VaRs', while it is the opposite for the 99% VaRs'. These results are one more hint to put into question the adequacy of the normal modelling of both the daily and weekly returns series. Indeed, should these series be normally distributed, all the empirically computed 95% and 99% VaRs' would be close to the results given by the Gaussian approximation. The overestimation at lower quantiles and the underestimation at higher quantile of the Gaussian model is typical for a process that is heavy-tailed distributed.

3.1.2 On a fixed sample

We first study the relevance of a VaR computed on a static sample of various size for both daily and weekly returns series. By relevance, we mean its capacity to predict the future VaR computed over one year ahead. We choose a sample of a certain size to estimate the VaR and then consider it as the predictor for the next year. Then we move the sample by one year and reiterate.

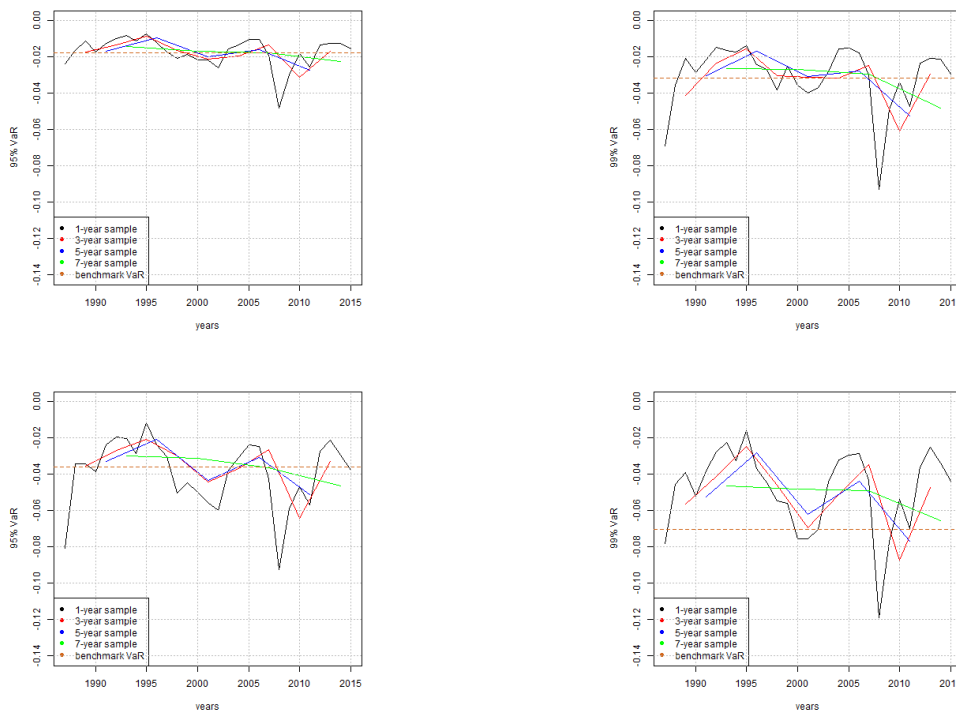


Figure 4: VaR on fixed samples, with from left to right and from top to bottom: 95% daily VaR, 99% daily VaR, 95% weekly VaR, 99% weekly VaR

We show in Figure 4 the 95% and 99% VaR for daily and weekly returns on fix samples of 1 year, 3 years, 5 years and 7 years and compare them to the benchmarks of Table 2, which are drawn as dotted lines.

After analysing the 4 graphs, few remarks can be made:

- The first interesting feature is the fact that the VaR is not constant over time. The values oscillate around the benchmark value.
- As noticed when discussing the numbers of Table 2, the 99% VaR is always strictly smaller than the 95% VaR for both daily and weekly returns time series.
- As observed in the full sample, the weekly VaR is smaller than the daily VaR, notably due to the higher weekly returns volatility compared to the daily returns volatility.
- For the four graphs, it can be observed that 1-year VaR and 3-year VaR have a high volatility regime compared to the 5-year VaR and the 7-year VaR which have a low volatility regime. This can be explained by the VaR measure itself to the extent that only 2.5 points are below the 99% threshold for a 1-year sample, compared to 7.5 for 3-year samples, 12.5 for 5-year samples and 17.5 for 7-year samples. The higher the number of points, the less the volatility (tendency of averaging).
- As expected, the VaR estimated on a longer sample reacts slower than those estimated on a shorter sample, introducing a delay that could be costly for risk management.
- The last remark concerns the correlation of the different VaRs'. Indeed, the 1-year and the 3-year VaRs' are highly correlated. On the other hand, the 5-year and the 7-year VaRs' are also very correlated between them. However, there is a smaller correlation between less-to-3-year VaR and greater-to-3-year VaR.

Our conclusions

Following this graph, it appears that the fix 1-year and 3-year sample VaRs' fluctuate a lot around the benchmark VaR as it would be expected since the sample size seems quite short. Using the result on such a short sample, without confidence intervals, makes risk management either under-evaluate or over-evaluate the risk capital.

The 5-year and 7-year sample sizes seem more appropriate because the calculated VaRs' on these sample sizes are closer to the benchmark VaR. Indeed, in the case of daily VaR, the fix 5-year and 7-year sample VaR lie slightly below the losses of daily returns for a threshold of 95% and slightly above the losses for a threshold of 99%. However, the 2007-2008 crisis made these VaRs' collapse, thus overestimating the losses for the following years. It is a behavior corresponding to what is often seen in times of crisis for market agents; they tend to overestimate the risk in the middle of the turmoil.

In the case of weekly returns, the fix sample 99% VaRs' lie clearly above the benchmark, whereas, in the 95% case, they hover around it. It may be explained by the fact that the 99% VaRs' have been evaluated using a Gaussian model, because of a lack of enough data to compute them empirically. It is why the computed VaRs' oscillate around 5.53%, which is the Gaussian value, instead of oscillating around 7.01% that corresponds to the empirical value computed over the whole sample (see Table 2 for the benchmark values).

To conclude, due to the small size of sample, the fixed 1-year and 3-year sample 95% VaR are too volatile and tend to misrepresent potential losses for following years. The best sample size seems to be 5 and 7-year, at the loss of certain dynamics and a delayed response. That is why, we try to remedy this problem by using rolling samples, as developed in the next subsection.

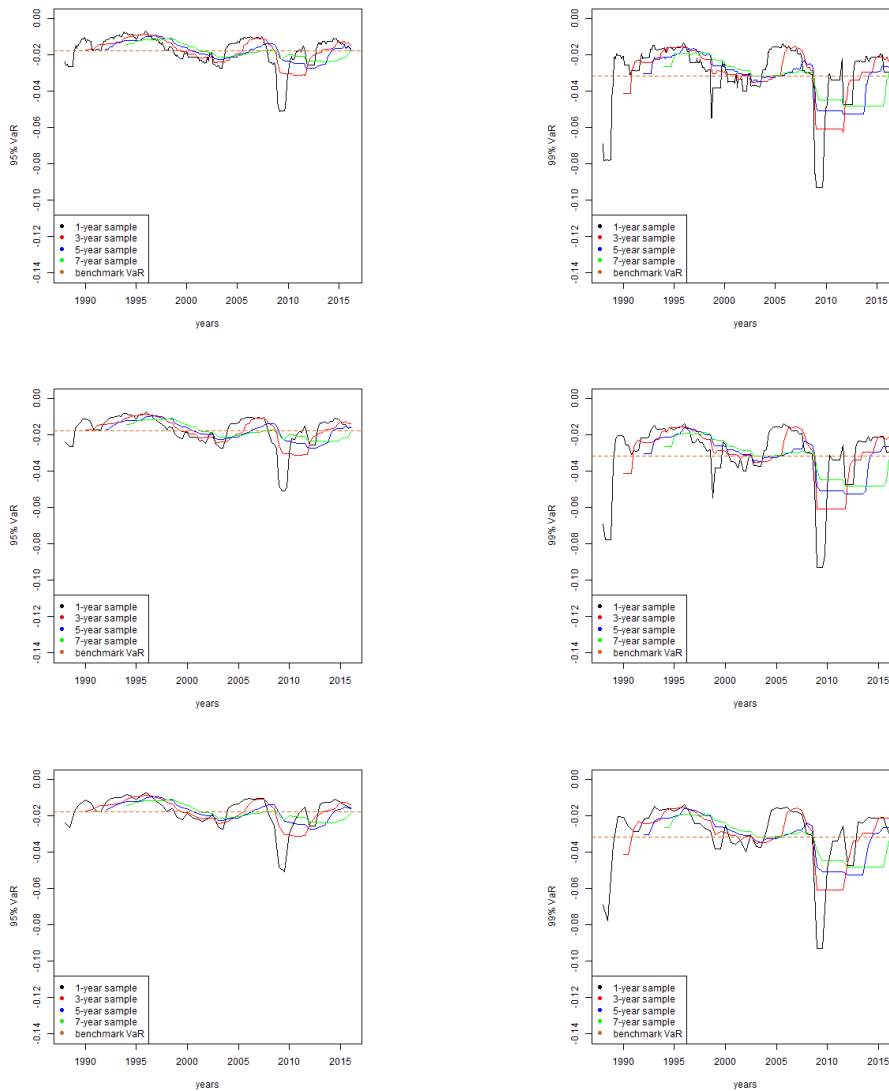


Figure 5: Daily VaR on rolling samples, with from left to right: 95% VaR and 99% VaR, and from top to bottom: samples rolling every month, every 3 months and every 6 months

3.1.3 On rolling samples

Using a rolling sample means that we do not wait for a full testing year to update the VaR computation. Instead, we update it at a shorter sample. In Figures 5 and 6, we display the daily and weekly returns, respectively, considering both 95% and 99% VaRs' for 1-year, 3-year, 5-year and 7-year samples rolling every month, every 3 months, and every 6 months. At this periodicity, we include new data and drop old ones. Note that, as previously, the weekly returns for one year sample 99% VaRs' are evaluated using a Gaussian approximation.

If we compare the VaR obtained with fixed samples and rolling samples using the remarks made on the previous section, we have:

- The VaR computed with rolling samples is still not constant over time. However, rolling samples offer a greater continuity to VaR. Indeed, there is a smaller difference between consecutive VaR with rolling samples because of the overlapping samples and the fact that the values adapt faster,

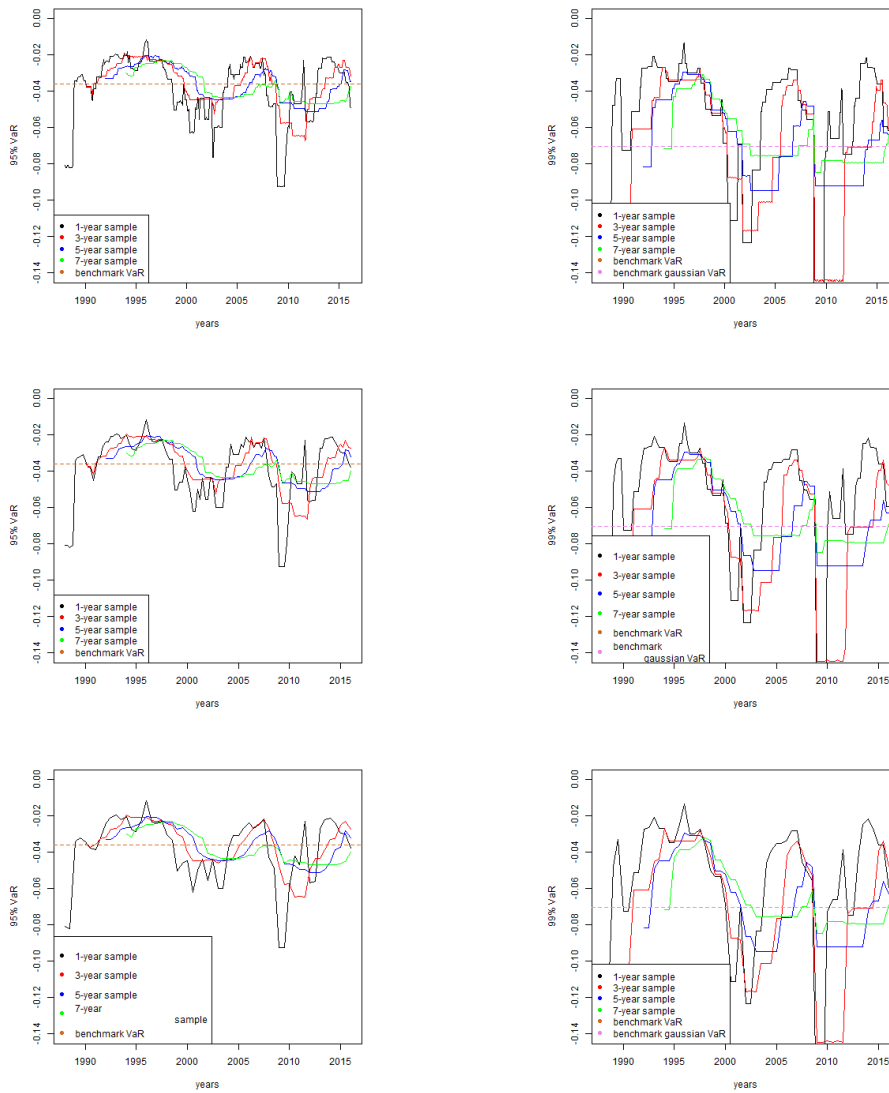


Figure 6: Weekly VaR on rolling samples, with from left to right: 95% VaR and 99% VaR, and from top to bottom: samples rolling every month, every 3 months and every 6 months

as expected (the new data being integrated at a higher frequency).

- As expected, the 99% VaR is always strictly smaller than the 95% VaR for both daily and weekly returns time series with rolling samples.
- We still find the expected behavior that the weekly VaR is smaller than the daily VaR for both 95% and 99% levels.
- With rolling samples, it can be noticed that all VaRs' with different samples intersect each other in numerous points. However, the volatility of the VaR decreases with the sample size as noticed in fix samples.
- As for the case of fixed samples, a lag is observed between the different VaRs'. Indeed, there is a lag of the 3-year rolling samples VaRs' compared to 1-year rolling samples VaR, and so on and so forth for longer rolling samples. As for the fixed sample, this is due to the sample size. The longer the sample, the more time it will take for VaR to change due to a new event on the market. Moreover, in the case of rolling samples, the overlap is also longer, which increases the effect.

- Finally, the 1-year and 3-year VaRs' are still very correlated, so are the 5-year and 7-year VaRs' between them. However, the 4 curves in each graph seems more correlated between each other than in graphs with fix samples.

To conclude, we see that the behaviors of the VaRs', computed on rolling samples, are very similar to those of the VaRs' computed on fixed samples, except that they are smoother because of the big overlap between consecutive samples. Thus, it is advisable to use rolling samples in order to update more frequently the VaR estimates.

3.2 Out-of-sample backtest of the historical VaR

Here we want to check whether realized losses, observed ex post, are in line with the VaR forecasts. The statistical procedure designed to compare realizations with forecasts is known as *backtesting*. It is based, for the VaR, on the observation that when VaR at level α is consistently well estimated, the occasions on which realized losses exceed VaR forecasts (called *VaR exceptions* or *VaR violations*), should form a sequence of iid Bernoulli variables with probability $(1 - \alpha)$. Hence the use of a binomial test based on the so-called violation process counting how many times the estimated VaR has been violated during the following year. The literature on backtesting VaR estimates is large and we refer e.g. to pioneers in the domain, e.g. Kupiec (1995, [14]) and Davé & Stahl (1998, [10]). Christoffersen examined in [5] the assumption of independence of violations and proposed then a test said of *conditional coverage* (see e.g. [4] for an overview of tests of conditional coverage).

Here we apply the binomial test on our data set, then test the independence assumption.

We calculate ex-ante VaR using 1 year rolling window and daily returns on a fix sample of 1, 3 and 5 years. Let denote $\text{VaR}_{t-1}(\alpha)$ the VaR calculated during a period whose ending year is $t - 1$ and for a threshold α . For example, if we calculate a VaR using daily data on the period 1987- 1989 (3 years), we note it $\text{VaR}_{1989}(\alpha)$.

First we intend to assess the quality of the ex-ante $\text{VaR}_{t-1}(\alpha)$ using daily returns $X_{t,i}$ of the following year t . It is possible to introduce the daily loss function $L_{t,i}$ measured during a given year t , such that $L_{t,i} = -X_{t,i}$. The number of observable returns for the following year is noted by n_t .

Introducing the daily violation process $(V_{t,i})_i$ of VaR measured during a year t , $V_{t,i} = \mathbf{1}_{(L_{t,i} > \text{VaR}_{t-1}(\alpha))}$, and assuming for the moment that $(V_{t,i})_{i,t}$ are mutually independent for all i and t , then the total number N_t of daily violations of $\text{VaR}_{t-1}(\alpha)$ over year t follows, for any t , a Binomial distribution:

$$N_t := \sum_{i=1}^{n_t} V_{t,i} \sim B(n_t, 1 - \alpha).$$

We also compute the VaR_{t-1} taking longer samples of 3 and 5 years. In Figure 7, we plot the ratios of VaR violations, N_t/n_t , that is to say the number of days that $\text{VaR}_{t-1}(\alpha)$ has been violated during the year t over the number of observations during the same year. We also draw the benchmark value $(1 - \alpha)$ and the 95% confidence intervals for the proportion of VaR violations around it.

We observe the following.

Using 1-year VaR with daily data. In this case, it can be noticed that the ratios often exceed the calculated 1-year 95% VaR the following years. Thus, this is not a stable and appropriate measure that enables risk managers to protect them from large losses. Nevertheless, the ratios of violation are smaller in the case of 99% VaR (only 6 ratios are greater than the upper bound of the confidence interval).

Using 3-year VaR with daily data. In this case, the ratios of violation for the 95% VaR are often outside

the confidence interval. However, ratios of violation have only been greater 6 times than the upper bound of the confidence interval for the 99% VaR.

Using 5-year VaR with daily data. In this case where VaRs' have been calculated over 5-year samples, we find approximately the same results as with 1-year and 3-year samples VaR. The 95% VaRs' have often been breached. However, the ratios of violation for 99% VaR are close to the 1% benchmark in most cases.

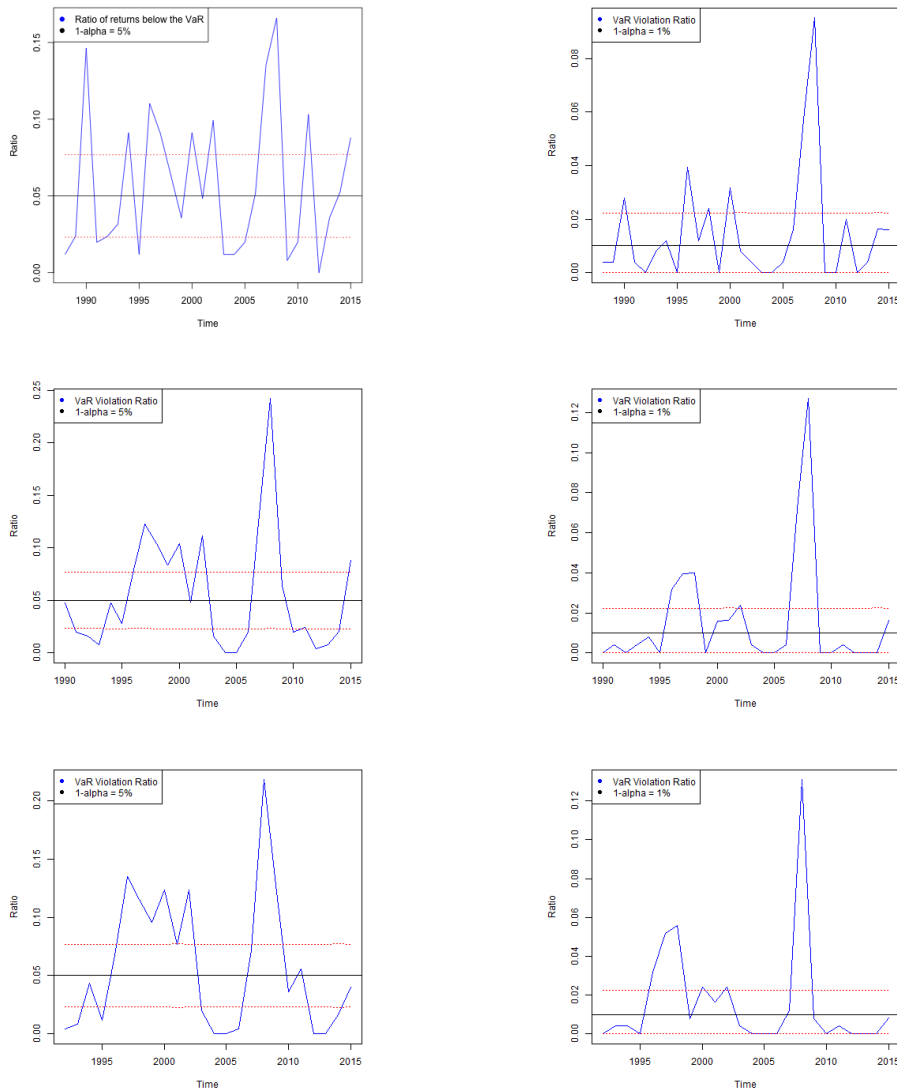


Figure 7: Out-of-sample test using daily data, with from left to right and from top to bottom: of the 1-year 95% VaR, of the 1-year 99% VaR, of the 3-year 95% VaR, of the 3-year 99% VaR, of the 5-year 95% VaR and of the 5-year 99% VaR

3.2.1 Binomial test under the independence assumption

We now apply the binomial test under the assumption of independence of the $V_{t,i}$, testing the proportion parameter p_t of violations, according to:

$$\left| \begin{array}{l} H_0 : p_t = 1 - \alpha \\ H_1 : p_t > 1 - \alpha \end{array} \right. \quad (\text{underestimation of VaR})$$

and introducing the square of the binomial *score* test statistic (for year t):

$$Z_t = \sqrt{n_t} \frac{N_t/n_t - (1 - \alpha)}{\sqrt{\alpha(1 - \alpha)}}$$

which is compared with the standard normal distribution (since $Z_t \underset{H_0}{\sim} N(0, 1)$).

Note that we could also have carried other two- or one-sided score tests, for instance

$$\left| \begin{array}{l} H_0 : p_t = 1 - \alpha \\ H_1 : p_t \neq 1 - \alpha \end{array} \right. \quad \text{or} \quad \left| \begin{array}{l} H_0 : p_t \leq 1 - \alpha \\ H_1 : p_t > 1 - \alpha \end{array} \right. ,$$

testing N_t against a binomial distribution, as done at the 99% level for the Basel backtesting regime for the latter one.

The event (rejection of H_0) at size γ is then given by ($Z_t > \phi^{-1}(1 - \gamma)$) where $\phi^{-1}(1 - \gamma)$ is the quantile of order $1 - \gamma$ of the standard normal distribution, for instance 1.65 if we consider $\gamma = 5\%$.

Results of the binomial test under independence assumption

For each year, we perform the binomial test. The testing samples are 28, 26 and 24 years (from 1988 to 2015) for VaR estimated on a 1-year, 3-year and 5-year sample, respectively. We qualify the year as a "success" when H_0 cannot be rejected at the 5% significance level (*i.e.* when the p-value is above 5%), and as a "failure" otherwise. The results are displayed in Table 3.

Table 3: Binomial tests with the 1, 3 and 5-year sample VaR using daily data

Sample	1 year		3 years		5 years	
	95% VaR	99% VaR	95% VaR	99% VaR	95% VaR	99% VaR
number of successes	18	22	17	20	16	18
number of failures	10	6	9	6	8	6

We observe that the 95% VaR is more rejected than the 99% one (about a third for 95%, and a fourth for 99%); this is more or less consistent whatever the estimation sample size (1, 3 or 5 years). The better performance of the 99% VaR (than the 95% one) is due to the nature of our test. Rejecting H_0 means that we underestimate the true VaR. Nevertheless, a closer look at the rejected years shows that those are always the years of crisis, e.g. 2007 (Bear Stearns' bankruptcy) is underestimated for all samples, as well as 2001 with the internet bubble crash.

3.2.2 Test of independence

In order to test the condition of independence of the $V_{t,i}$, we use the duration-based test of independence detailed in [6].

The rationale behind this VaR backtesting is to assess the clustering in VaR violations (or hit). This issue is of importance to the extent that from a Risk Management’s point of view, big losses in a short period of time would decrease the creditworthiness of a financial institution and may lead this entity to bankruptcy.

In [6], the authors test the spacings (or durations) between VaR violations using the fact that a discrete geometric distribution can be approximated by a continuous exponential distribution. The null hypothesis of exponential spacings (constant hazard model; the duration has no memory) is tested against a Weibull alternative (in which the hazard function may be increasing or decreasing), with shape parameter β . When $\beta = 1$, the duration follows an exponential distribution and we find again constant hazard function. As a result, the duration-based test can be expressed as

$$\left. \begin{array}{l} H_0 : \beta = 1 \quad (\text{independence}) \\ H_1 : \beta \neq 1 \end{array} \right\} .$$

Results of the independence test

In Table 4, we present the duration-based test every year considering the same six cases as for the binomial test (see Table 3). Note that when the VaR is not violated or only once a year, it is not possible to perform the duration-based test, that is why the item ”number of years with insufficient data” is added in the table. Note that all the reported tests lead to the rejection (failure) or the acceptance (success) of H_0 at a significance level of 5%.

Table 4: Independence tests with the 1, 3 and 5-year sample VaR using daily data

Sample	1 year		3 years		5 years	
	95% VaR	99% VaR	95% VaR	99% VaR	95% VaR	99% VaR
risk measures						
number of years with insufficient data	1	14	3	16	6	13
number of successes	27	12	21	10	15	11
number of failures	0	2	2	0	3	0

In Table 4, it can be noticed that the number of rejections of H_0 is extremely small compared with the acceptance number for both 95% and 99% VaR. Nevertheless, the results given for the 99% VaR have to be taken with more caution due to a smaller number of years for which the test can be performed than when the test cannot be performed. Our test gives rejection when there is an underestimation of the VaR. Moreover, we just saw that, for 99%, these rejections happen at the dawn of the crisis. It is thus not surprising that the size of the test shows independence of the times of occurrence of those rejections. Indeed it is a well known fact that financial crises in one market are difficult to predict, which means independence in a first approximation.

4 Estimation of Sample Quantile Process on the S&P 500 Index

4.1 Definitions

In the previous sections, we have considered a quantile-based measure through the VaR, introducing already some dynamics via rolling windows. We formalize this, turning to the notion of sample quantile process, as follows.

Let $L = (L_t)_t$ denote the loss stochastic process. The Sample Quantile Process (SQP) $(Q_{T,\alpha,t}(L))_{t \geq 0}$

of L at threshold α with respect to a random measure μ defined on \mathbb{R}^+ , is defined by (see e.g. [12]):

$$Q_{T,\alpha,t}(L) = \inf \left\{ x : \frac{1}{\int_t^{t+T} \mu(s) ds} \int_t^{t+T} \mathbf{1}_{(L_s \leq x)} \mu(s) ds \geq \alpha \right\}. \quad (3)$$

Note that choosing μ as the Lebesgue measure in (3), corresponds to the rolling window VaR. Indeed we have

$$Q_{T,\alpha,t}(L) = \inf \left\{ x : \frac{1}{T} \int_t^{t+T} \mathbf{1}_{(L_s \leq x)} ds \geq \alpha \right\}, \quad (4)$$

which corresponds to the continuous version of (2) on the interval $[t, t + T)$. Nevertheless, being evaluated on the observations, it means to take its discrete version

$$\widehat{Q}_{T,\alpha,t}(L) = \inf \left\{ x : \frac{1}{T} \sum_{i \in [t, t+T)} \mathbf{1}_{(L_i \leq x)} \geq \alpha \right\}. \quad (5)$$

In this study, we explore the SQP as a predictor of the risk, when considering μ as a random measure. We focus on the particular case where μ is defined by $\mu(s) = |L_s|^p$, with $p \in \mathbb{N}^+$ a positive integer. It includes the case of Lebesgue measure, taking $p = 0$.

In the next section, we observe sample paths of SQP over a stock index, fixing the interval length T to 1 year, and varying p of the risk measure μ . We represent those samples paths on a monthly frequency, evaluating them via (5), or equivalently via $\widehat{Q}_{T,1-\alpha,t}(X)$ using the log returns observations. Note that here, we do not refine the computation taking a linear interpolation between the low and high nearest neighbours of the solution of (5), as we did for the VaR.

4.2 Evaluation of the Sample Quantile on the S&P 500 Index

In Figure 8, we plot the sample paths of SQP at level α at monthly frequency, as well as the monthly rolling VaR and the VaR benchmark on the overall sample, both VaR at the same level α . The sample, T , is taken to be one year and we consider thresholds $\alpha = 95\%$ and 99% , respectively. One should note that the monthly rolling VaR corresponds to observations/points of the sample path SQP with $p = 0$ taken at monthly frequency. However, we can observe that this latter seems to consistently underestimate the risk represented by the rolling VaR; it can be explained by the fact that the rolling VaR has been computed via a linear interpolation, while the SQP's are not.

When $p = 0.5$, the values of the SQP are more volatile. Indeed, it can be noticed that the SQP range is wider when $p = 0.5$ than when it is equal to 0. We also notice that the SQP has smaller values than the VaR for both thresholds. Hence, the SQP, with $p > 0$, is a measure that enables a bigger discrimination between the losses. We can also highlight the fact that the differences between the SQP's for 95% and 99% have diminished.

For $p = 1$, the properties found for $p = 0.5$ are also present and accentuated. The volatility of the values of the SQP has once again increased. The range of the SQP is also bigger when $p = 1$ than when equal to 0 and 0.5. As for $p = 0.5$, the SQP is a measure that enables a bigger discrimination between the losses. The differences between the SQP has again diminished. To conclude, the SQP for $p = 1$ is extremely close to the SQP for $p = 0.5$ and continues the trend seen for the later.

In addition, the SQP with $p = 2$ is very close to the previous SQP with $p = 1$. However, the differences between the SQP with the different thresholds are extremely small and almost vanished. Hence, it is difficult, if not impossible, to distinguish between the 95% and 99% thresholds. The 95% threshold

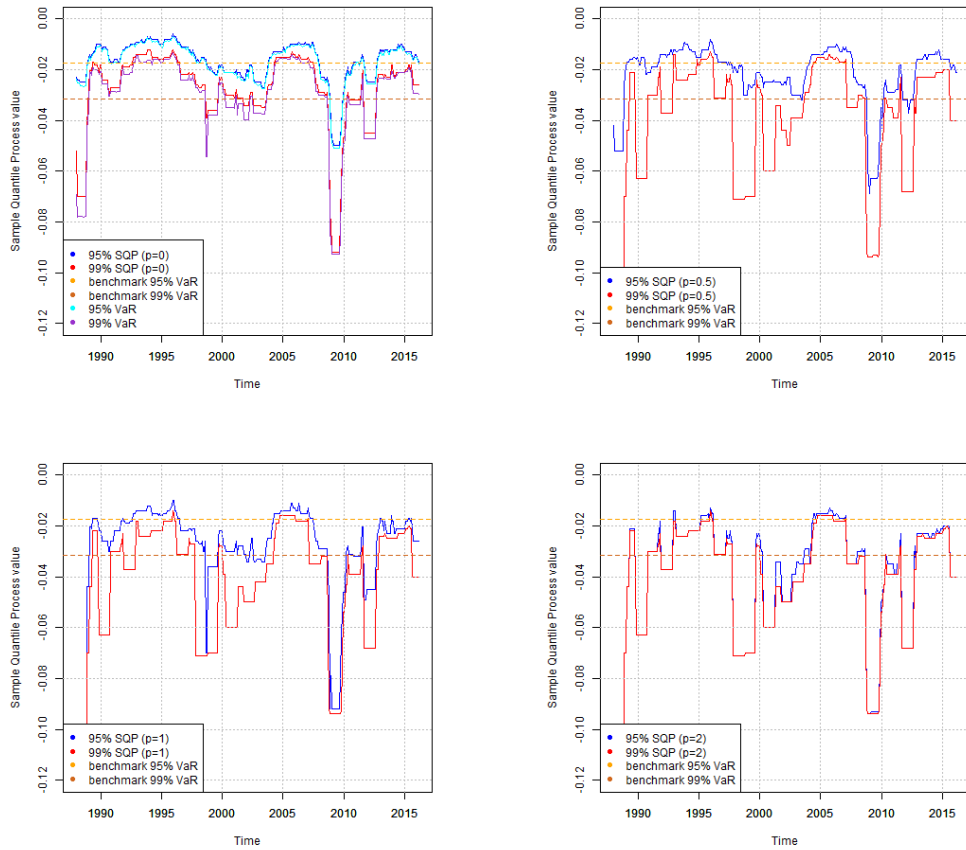


Figure 8: Sample Quantile Processes rolling every month with $T = 1\text{year}$. From left to right and from top to bottom: $p = 0$, $p = 0.5$, $p = 1$ and $p = 2$. The y-scale is the same in the four graphs to allow comparison between the different measures.

collapses in fact on to the 99% threshold. As can be seen on the last graph, both SQP's hover around the $\text{VaR}(99\%)$. Moreover, for the 99% threshold, the SQP with $p = 2$ is equal the SQP with $p = 1$, which makes it not interesting. We then conclude that this choice of random measure to define the SQP may not be the best one to assess the intensity of losses, but might be good to model the $\text{VaR}(99\%)$ of the S&P 500 index.

To illustrate all these points, we report and draw in Table 5 and its figure, the average values achieved by the various SQP's over the full sample as a function of the parameter p and the threshold α . Moreover, in the table, we compare them to the VaR benchmarks already reported in Table 2. We see that the SQP with $p = 0$ under-evaluates on average the VaR for both α , while, for all other p 's, it over-evaluates VaR. We note also that the higher the exponent p of the measure μ , the closer the averages between the 95% and the 99% get together. The ratio between the two thresholds goes down. Last, we see that the SQP 99% has the same average for $p = 1$ and $p = 2$. In fact, we noted already when discussing Figure 8, that both measures give the same results for all cases. In the figure of Table 5, we report more thresholds and we see that for a threshold far in the tail, 99.5%, we already reach saturation at $p = 0.5$. Taking a power $p > 0.5$ reduces the distinction between thresholds and the SQP's seem to all converge to a given value with increasing p 's. The value $p > 0.5$ gives the wider differences between thresholds. It seems that it is the value of p for which the description of the tail by the SQP is the most discriminating.

Risk measures	$\alpha=95\%$	$\alpha=99\%$	Ratio ($\frac{99\%}{95\%}$)
SQP (p=0)	-1.66%	-2.80%	1.69
VaR	-1.75%	-3.15%	1.80
SQP (p=0.5)	-2.18%	-4.32%	1.98
SQP (p=1.0)	-3.24%	-4.47%	1.38
SQP (p=2.0)	-4.33%	-4.47%	1.03

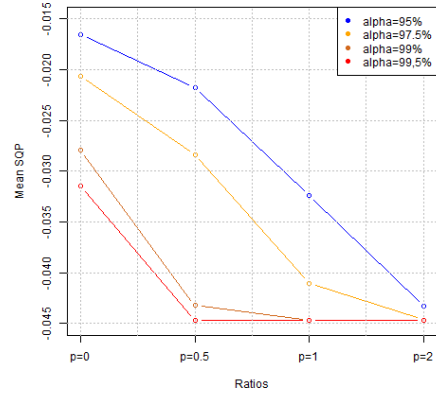


Table 5: Average Sample Quantile Processes over the whole sample as a function of the power p for various threshold α .

5 Volatility impact

5.1 Introduction

Inspired by the empirical study on Foreign Exchange done by M. Dacorogna *et al.* in [7], we explore the behavior of the Sample Quantile Processes and Value-at-Risk as a function of past volatility. According to the results of chap.7 in [7], after a period of high volatility, the returns tend to be negatively correlated, which means that the volatility should diminish and thus also the quantiles in the future. To be able to condition on volatility, we define an empirically measured annual volatility, also called in the literature realized volatility, of the S&P 500 returns, taken over a rolling sample of $T = 1$ year at a time t , as:

$$v_t = \hat{\sigma}(t) := \sqrt{\frac{1}{N-1} \sum_{i=t-N+1}^t \left(X_i - \frac{1}{N} \sum_{i=t-N+1}^t X_i \right)^2} \quad (6)$$

where N is the number of days in a year, and the returns X_i are defined as in eq. (1). We create so a time series of annual volatility $(v(t))_t$, which can be used as a benchmark of the current market volatility. Note that v_t defined in (6) is the annual daily volatility, but we annualize it by considering $v_t \times \sqrt{252}$; we keep the same notation in the following.

In Figure 9, we present this time series $(v(t))_t$ of annual volatility of the S&P 500 Index returns between 1987 and 2015. We also indicate the average annual volatility of the index calculated over the whole sample (represented by the red line). Its value is 18.63%. We can see that the volatilities above this benchmark mostly stand for periods of high instability or crisis (and not only in the USA). The high volatility of the period 1987-1989 is for instance explained by the New York Stock Exchange crash in October 1987. In 1997, Asia was hit by a crisis, as well as Russia in 1998 and Argentina in 1999-2000. In 2001, the United States experienced the bursting of the internet bubble. Following the Lehman Brother's bankruptcy, the period of 2008-2009 was a period of high volatility. And finally, the sovereign debt crisis in Europe also impacted the S&P 500 Index in 2011-2012. This is an illustration of the increased dependence between the markets during times of crisis.

We conclude from this behavior that our measure of volatility is a good measure that reacts well at various market states. It is thus a reasonable proxy to qualify the times of high risks and to use it to discriminate between quiet periods and periods of crisis. We use it to condition our statistics and see if

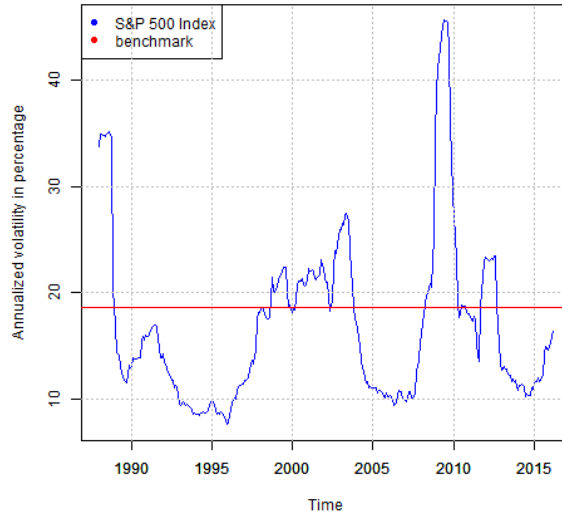


Figure 9: Annual volatility of the S&P 500 returns. The red line is for the annualized volatility computed over the whole sample

we can detect different behaviors of the price process during these periods in comparison to quiet times. It was the basis of the study in [7] that concluded to two different behaviors as a function of volatility: in times of low volatility, consecutive returns tend to be positively auto-correlated, while, when the past volatility is high, future returns tend to be negatively auto-correlated with current returns. We use it again here to look at the ability of the risk measures to correctly predict the future risks.

5.2 Dependence between the annualized volatility and the Sample Quantile Processes

For our study, we introduce a new random variable, the ratio of SQP's, that qualifies the difference between the current and future risk. It is defined as :

$$R_{t,\alpha} = \frac{Q_{T,\alpha,t+1}}{Q_{T,\alpha,t}} \quad (7)$$

where $t + 1$ means the time $t +$ one year later. If the observed value of the ratio is close to one, it means that the historically estimated risk measure has correctly predicted the future one for this sample path. If the realized ratio is above or below one, the observed future risk is higher (lower, respectively) than what was calculated, hence would lead to an under-evaluation (over-evaluation, respectively) of the needed capital.

Now we look at this random ratio statistically and see how far it is from 1. To do so, as a distance, we choose the Root Mean Square (RMS) distance and compute the associated RMSE:

$$RMSE := \sqrt{\frac{1}{N} \sum_{t=1}^N (R_{t,\alpha} - 1)^2} \quad (8)$$

where N is the number of years on which the ratio is tested. This RMSE measures the "error" of the forecasted risk for this sample path.

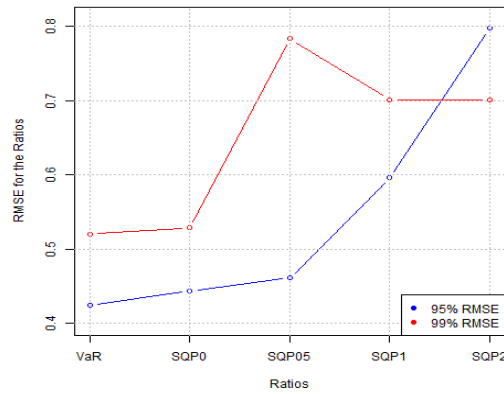


Figure 10: RMSE of the Ratios of the various measures, VaR_{α} , $SQP_{p\alpha}$ with $p = 0, 0.5, 1, 2$ with $\alpha = 95\%$ (in blue) and $\alpha = 99\%$ (in red)

We plot in Figure 10 the value of this quantity for the various measures we use. We see that the error increases with the threshold. It is smaller for 95% than for 99% except for the last measure of SQP with $p=2$, where it is inverted because the error levels off after $p=1$. We notice also that the error increases with the power p of the SQP and is the smallest for the rolling VaR. The values are clearly above 0, suggesting that the various measures do not forecast well the future risk. To validate statistically this statement, we could test other sample paths by testing other stock indices.

In the following, since RMSE does not give us any indication on the sign of $(R_{t,\alpha} - 1)$, we study the dynamics of the ratios in order to understand where the difference of RMSE from 0 comes from. We show in Figure 11 both the annualized volatility and the realized ratios of SQP's for both thresholds 95% and 99% and for various powers: $p = 0, 0.5, 1$ and 2.

The first remark is that these ratios become even higher when p increases for both thresholds. Secondly, the realized ratios can differ from 1 either on the upper side or on the lower side. Thirdly, it can be noticed that for all the graphs, the realized ratios of SQP's are negatively correlated with the annualized volatility. This particular feature is of importance in our study because it means that when the volatility in year t is high in the market, the realized ratio of the SQP in year $t + 1$ on the SQP in t is quite low and, as we said before, this situation means that the risks have been over-evaluated by the SQP calculated in year t . Conversely, when the volatility is low in year t , the realized ratio of the SQP is often much higher than 1 the next year, which means that the risks for the year $t + 1$ have been under-evaluated by the calculation of the SQP in year t . One further remark, the SQP does not vary between $p=1$ and $p=2$ for the 99% threshold; the values are identical. It is also why we see, on Figure 10, that the two last values are the same for this threshold.

By comparison, we present in Figure 12 the realized ratios of the VaR for both thresholds and the annualized volatility in order to verify whether we find the same feature as we do for the ratios of SQP. Indeed, the Value-at-Risk measure seems also negatively correlated with the annualized volatility. Hence, if the volatility is low in year t , the risks for the next year will probably be underestimated by the calculation of Value-at-Risk measure, no matter the threshold taken into account, while it will be the reverse if the current volatility is high. This is confirming the behavior we have seen with the RMSE, which is similar for VaR and for SQP even if a little smaller in the case of VaR. Overall, the dynamics is the same: higher than 1 ratio for low volatility and lower than 1 ratio for high volatility.

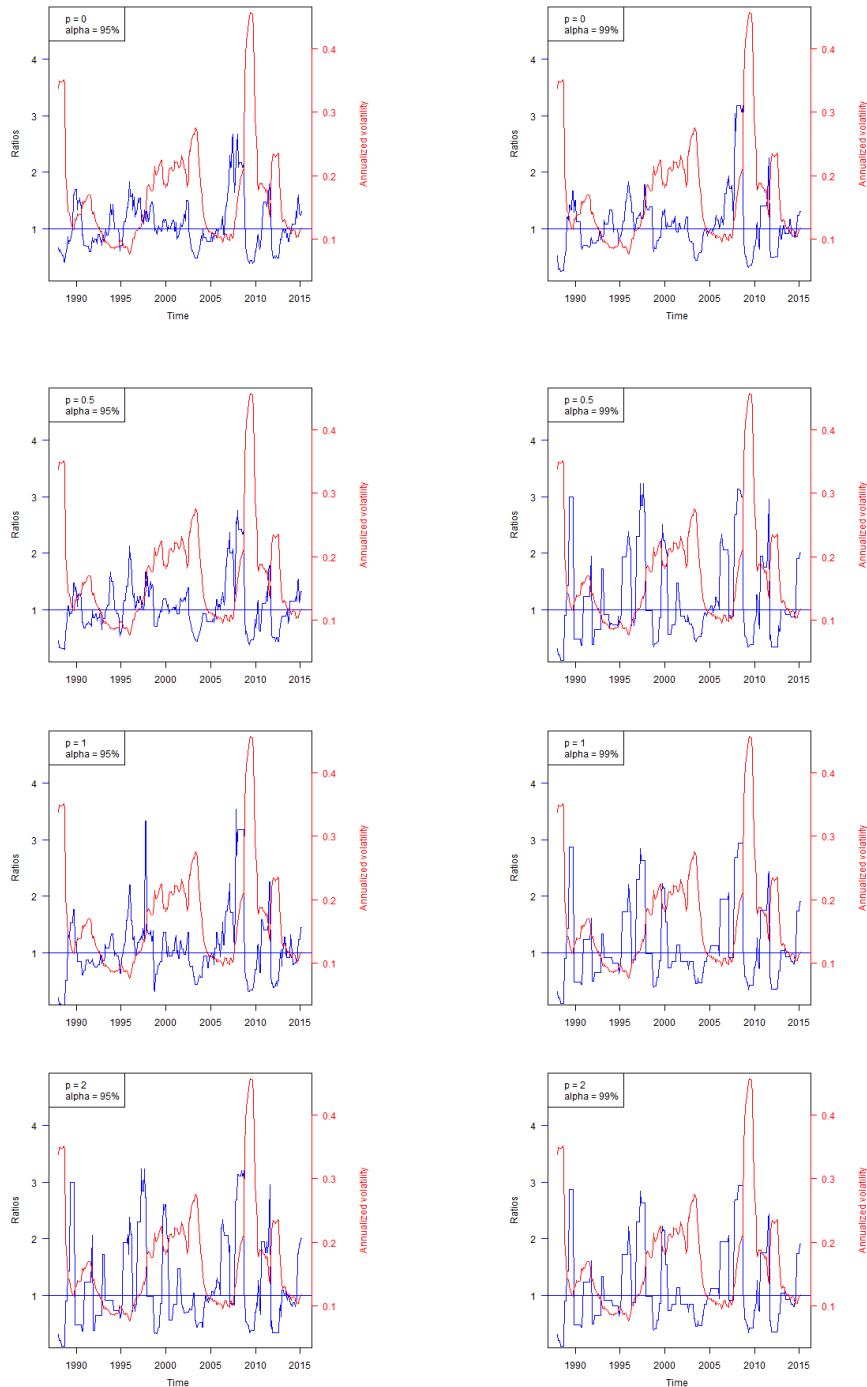


Figure 11: Ratios of Sample Quantile Processes and annualized volatility, with, from top to bottom : $p=0$, $p=0.5$, $p=1$, $p=2$, and from left to right: $\alpha = 95\%$ and $\alpha = 99\%$

In order to better understand this phenomenon, we present, in Figure 13, the various realized ratios of SQP's as a function of the annual volatility. Such a function highlights better the existing negative correlation between these two quantities. This negative correlation between the ratios of SQP's (and the VaR) with the annual volatility is clearly noticeable (and notably with the simple linear regression line that is decreasing). The more volatility there is in year t , the lower are the ratios $R_{t,\alpha}$, which means that losses in year $t + 1$ have been overestimated with the measures calculated in year t . We see that this underestimation can be very big since we see that the realized ratios can be as large as 3; in other words, the risk next year is 3 times the risk measured during the current year for this sample. The overestimation

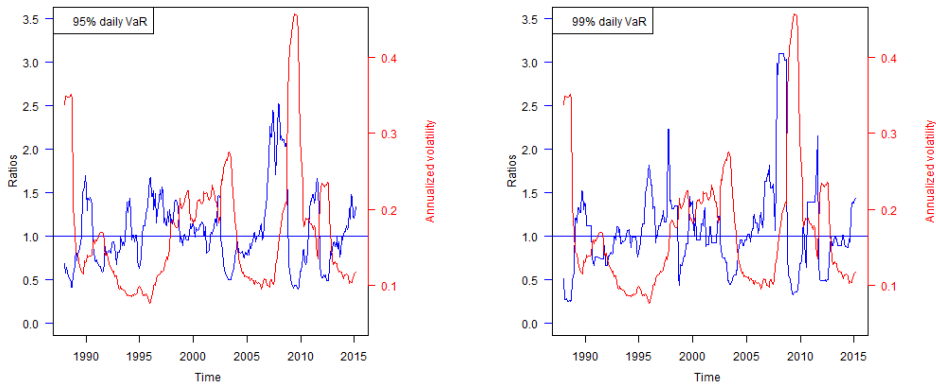


Figure 12: Ratios of Value-at-Risk and annualized volatility, with from left to right: $\alpha = 95\%$ and $\alpha = 99\%$

is not as high but still we observe ratios much below 0.5, *i.e.* the risk computed at the height of the crisis is twice the size of the risk measured a year later. We also notice that the underestimation is very systematic for high volatility, whereas it is not for low volatility. Indeed, in this latter case, we see also values below 1, while we do not see any value above one when the volatility is very high (above 35%). As noticed with the RMSE, the values for the VaRs' are slightly less dispersed than those of the SQP's with $p = 0$ in the two top graphs of Figure 13.

The question is: Which of these various measures does show the strongest counter-cyclical behavior? To answer it, we compute the empirical linear correlation between the annual volatility and the different ratios, and report the results in Table 6. The values are relatively high all above 30%, as illustrated in the slopes of Figure 13. We note in this table that the sample quantile with $p = 0.5$ presents the strongest negative correlation for $\alpha = 95\%$, while the highest for $\alpha = 99\%$ is with $p = 0$. The correlation is more pronounced for the lower threshold of 95%, except for $p = 2$. We notice that VaR and SQP for $p = 0$ are quite close as expected since they only differ due to the linear interpolation of the VaR. We can then conclude that, for those observations, SQP with $p = 0$, or VaR, are the most suited measure for designing counter-cyclical models. However, more exploration of this behavior with other stock indices is needed to better understand the best parametrization.

Table 6: Linear correlation between the various ratios and the annual volatility (in bold the largest values)

Ratios for	Correlation	
	$\alpha = 95\%$	$\alpha = 99\%$
VaR	-42.48%	-38.82%
SQP (p=0)	-41.86%	-38.93%
SQP (p=0.5)	-43.89%	-36.89%
SQP (p=1)	-39.59%	-37.29%
SQP (p=2)	-35.49%	-37.29%

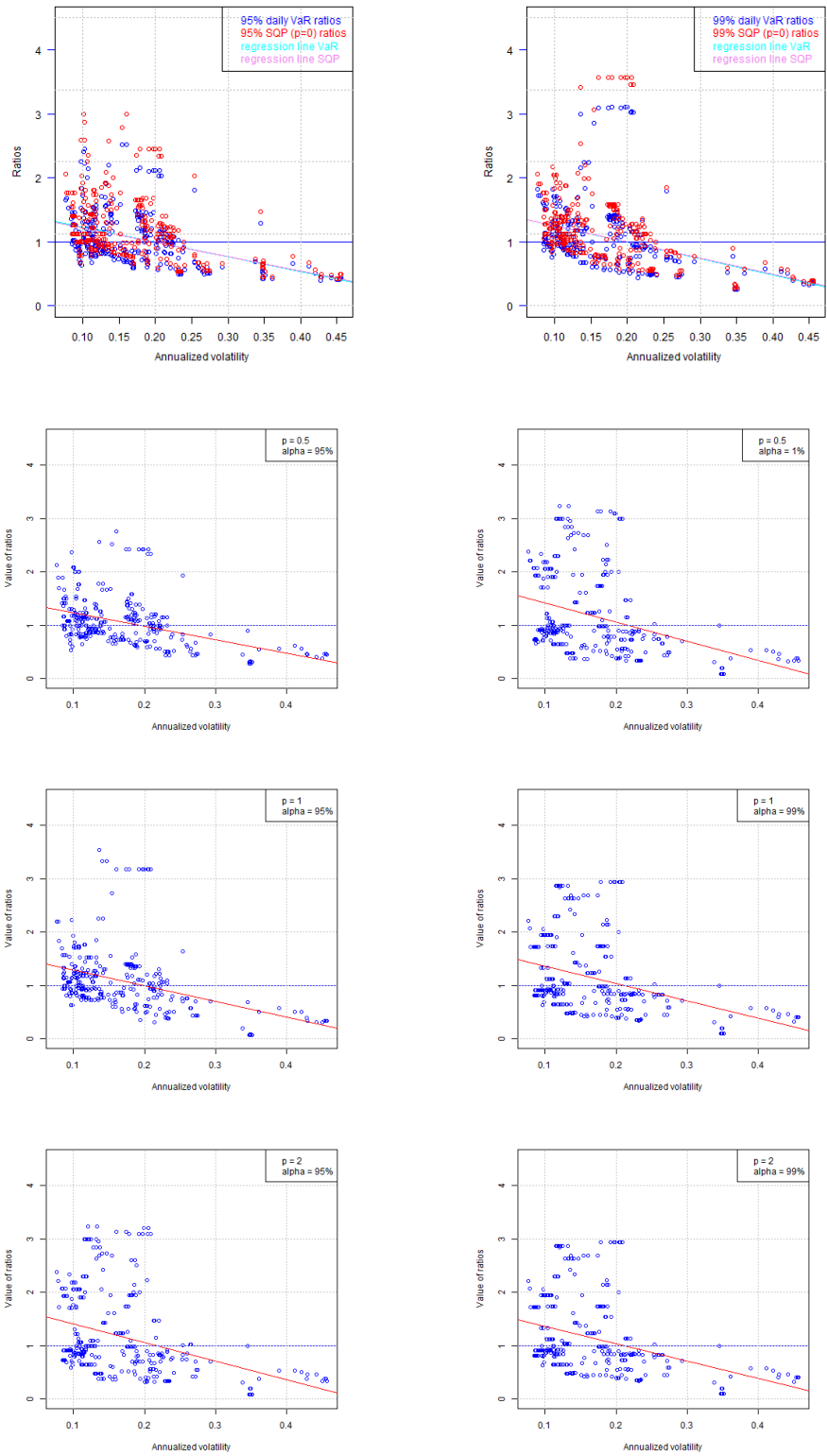


Figure 13: Ratios of Sample Quantile Processes (and Value-at-Risk for the first two graphs) as a function of annualized volatility, with, from top to bottom : $p = 0$, $p = 0.5$, $p = 1$, $p = 2$ and from left to right: $\alpha = 95\%$ and $\alpha = 99\%$

6 Conclusion

Popularized by JP Morgan in 1993 [17], the Value-at-Risk measure is still, nowadays, mostly used by banks and some insurance companies for risk management purposes. This use has been reinforced by regulators even if it has been shown that Value-at-Risk was not a coherent measure contrary to Expected Shortfall in [3]. However, as a lesson learned from the 2008 financial crisis, the Basel Committee is thinking of moving towards Expected Shortfall. Nonetheless, this study remains timely since VaR is the measure used by both banks and insurances and could be generalized for Expected Shortfall at a later stage.

To summarize this study, we observe that estimation of Value-at-Risk over past data is best done over monthly rolling samples to allow for quick adaptation to new market conditions. The sample period must be long enough to contain tail observations. Hence the introduction of Sample Quantile Process (SQP) that can be defined in many ways depending on the choice of measure μ w.r.t. which it is defined. Here this SQP is studied considering $\mu(s) = |X_s|^p ds$ for various positive values of p . We observe that it gives better results for a low power of p as 0 or 0.5; for larger values, it blurs the difference between the thresholds α . Further studies, playing on the measure μ , are needed to design a SQP process that is best suited to describe the risk.

A subsequent study of the dynamics of risk measures (VaR or SQP) as a function of realized volatility has been conducted. It confirms the assumption we presented in the introduction of this study, that is to say that during high-volatility periods, those risk measures overestimate the risks for the following years, whereas during low-volatility periods, they underestimate them. Therefore, would this observation be made on other indices, it would mean that the risk management imposed by regulators favors procyclical behaviors of the market actors. Indeed, it does not reflect the genuine potential risks and losses. Following a pure historical estimation of risk measures, banks would be required to keep more capital than needed during a crisis (period of high volatility), while the requirements during quiet period grossly underestimate the risk, leaving the banks, or financial institutions in general, unprepared in case of crisis.

To conclude, in view of our results, the introduction of anti-cyclical risk management measures, which are at the opposite of what it is currently required for financial institutions by regulators, would better prepare the financial system to cope with future crises by enhancing the capital requirements in quiet times and relaxing them during the crisis. This is what we are currently developing. Studies are under way to design SQP with the right dynamical behavior that would be a good basis for anti-cyclical regulation.

Acknowledgments

This project has received the support from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 318984 - RARE. Marie Kratz also acknowledges with gratitude the hospitality, when working on this project, of ETH Zurich, RiskLab, as a FIM visitor during Spring 2016, and of SCOR Switzerland in July 2016. Part of the empirical study has been performed by Rosnan Chostard, ESSEC student in M.Sc. in Management (Grande École) as a research project under the supervision of Prof. Kratz. We wish to warmly thank Marcel Bräutigam, PhD student of M. Kratz, for reviewing the code and helping improving it. We are also grateful to Prof. Jiro Akahori for fruitful discussions on the SQP and with whom the next developments are conducted.

References

- [1] J. Akahori. Some formulae for a new type of path-dependent option. *Ann. Appl. Probab.*, 5(2):383–388, 1995.
- [2] J. Akahori and M. Urata. A class of multi-period coherent risk measures defined in terms of sample quantiles. *Preprint Ritsumeikan University*, 2011.
- [3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risks. *Mathematical Finance*, 9:203–228, 1999.
- [4] J. Berkowitz, P. F. Christoffersen, and D. Pelletier. Evaluating value-at-risk models with desk-level data. *Management Science*, 57:2213–2227, 2011.
- [5] P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1998.
- [6] P. F. Christoffersen and D. Pelletier. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, 2:84–108, 2004.
- [7] M. Dacorogna, R. Gencay, U.A. Muller, O. Pictet, and R. Olsen. *An introduction to High-Frequency Finance*. Academic Press, 2001.
- [8] A. Dassios. The distribution of the quantile of a brownian motion with drift and the pricing of related path-dependent options. *Ann. Appl. Probab.*, 5(2):389–398, 1995.
- [9] A. Dassios. Sample quantiles of stochastic processes with stationary and independent increments. *Ann. Appl. Probab.*, 6(3):1041–1043, 1996.
- [10] R. Davé and G. Stahl. On the accuracy of VaR estimates based on the variance-covariance approach. *Risk Measurement, Econometrics and Neural Networks*, pages 198–232, 1998.
- [11] P. Embrechts, R. Frey, and A. McNeil. *Quantitative Risk Management: Concepts, Techniques and Tools, Revised Edition*. Princeton Series in Finance, 2015.
- [12] P. Embrechts and G. Samorodnitsky. Sample quantiles of heavy tailed stochastic processes. *Stochastic processes and their applications*, 59(2):217–233, 1995.
- [13] S. Emmer, M. Kratz, and D. Tasche. What is the best risk measure in practice? a comparison of standard risk measures. *Journal of Risk*, 18(2):31–60, 2015.
- [14] P. H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3:73–84, 1995.
- [15] R. Muria. A note on look-back options based on order statistics. *Hitotsubashi J. Commerce Management*, 27:15–28, 1992.
- [16] M. Yor. The distribution of brownian quantiles. *J. Appl. Probab.*, 32:405–416, 1995.
- [17] P. Zangari. *RiskMetrics Technical Document*. available on <https://www.ms-ci.com/www/research-paper/>, 1996.

PARIS

ESSEC Business School

3, avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tél. +33 (0)1 34 43 30 00
www.essec.edu

ESSEC Executive Education

CNIT BP 230
92053 Paris-La Défense
France
Tél. +33 (0)1 46 92 49 00
www.executive-education.essec.fr

ESSEC Asia-Pacific

5 Nepal Park
Singapore 139408
Tél. +65 6884 9780
www.essec.edu/asia

SINGAPOUR

Contact :

Centre de Recherche
+33 (0)1 34 43 30 91
research.center@essec.fr

ISSN 1291-9616



affilié à la

