LUND UNIVERSITY

School of Economics and Management

YAIR ELSNER

VALUE AT RISK – THE SQUARE ROOT RULE

MASTER THESIS IN FINANCIAL ECONOMICS, 2012-03-23, AT THE DEPARTMENT OF ECONOMICS OF THE SCHOOL OF ECONOMICS AND MANAGEMENT AT THE UNIVERSITY OF LUND, SWEDEN.

SUPERVISOR: BIRGER NILSSON

ABSTRACT

This paper tests the "Square Root Rule" (the SRR), a Basel sanctioned method of scaling 1-day Value At risk to higher time horizons. The SRR has come under serious assault from leading researchers focusing on its week theoretical basis: assuming i.i.d. asset returns. I performed an empirical test of the SRR on the 10-day horizon, the maximum horizon allowed by Basel, comparing SRR performance to directly estimating 10-day VaR by the same method 1-day was estimated, the "main-model". This test was performed for a number of well-known models. Performance was measured by means of Basel's own criteria for back-testing, the Christoffersen test and through general descriptive statistics. My results were that the SRR performs less well than its "main-model" alternative, but overall only slightly so. Based on this I claim that SRR usage as a thumb-rule is certainly defensible, and that even in circumstances calling for exact measures it's difficult to rule out SRR-scaling.

Prelude and acknowledgements

It is with great joy that I present this Master essay, the fruit of long and hard work. My studies at the University of Lund took place at a time in my life that has been all but tranquil, a far cry indeed from the dreamlike existence of the average student in these areas of the world. Without going into detail, I can say that having come to here, to the point of presenting this essay and to attaining my Master degree, despite the circumstances alluded to, fills me with a sense of accomplishment, content, and hope for the future. Having said this about myself, I want to also express my gratitude to those who have stood my side. My dear mother deserves all the credit imaginable, as does her husband. They know why. I also want to express my gratitude to the personnel at the School of Economics and Business Administration at the University of Lund, not the least because of their willingness to go to "logistical lengths" in helping me in throughout my path. My supervisor, Birger Nilsson, is worthy of praise, he has spent many evenings answering e-mails from me over the years! I also want to thank my late, beloved father, the thought of whom has accompanied and strengthened me at critical times. I only wish he was still here to witness this. Lastly, but not the least, I want to thank my small children, who never doubted their father's love and commitment, and never wavered in their own, during these years when I've been with less than I would have wanted in the ideal of worlds.

1 Introduction

1.1 Choice of subject

Value at risk is one of the most commonly used measurements of market risk (K. Dowd, Measuring Market risk, pages 9-11), and the so called “square root rule”, from henceforth “the SRR”, is a widely used (Christoffersen, Schuermann and Diebold (1998), page 2) thumb-rule for converting 1-day VaR to VaR of longer horizons by simply scaling the 1-day VaR estimate by the square root of time. The SRR gained important official support by being enshrined in Basel accords (Supervisory framework for the use of backtesting in with the internal models approach to market risk capital requirements, section III) as a legitimate method for banks to employ. The ease of use of the SRR certainly contributes to its attractiveness, not just in the plain sense of making life easy, but to enable decision makers to make quick estimates of risk in situations calling for swiftness or, alternatively, in more casual circumstances when exacter estimates, if they at all exist, are not of critical importance. Furthermore, considering the difficulty of risk-estimation, proof to which is the abundance of academic research and dispute concerning this, if it were the case that the SRR provides a reasonably good measurement of risk, such a fact could save resources that would otherwise be invested in research of other models and maintenance of these. After all, why invest developing expensive complicated models if there is a cheap and simple one at hands? Now, the SRR has come under heavy fire from academic circles criticizing it first and foremost for being completely unfit for longer horizons, but even for the short horizons of up until 10 days that Basel spoke about (K. Dowd, Long-term Value at risk). Some claimed that even Basel seemingly were showing second thoughts as to the appropriateness of the SRR (Provizionatou, Markose and Menkens (2002), page 2, quoting a 2002 Basel technical paper that didn’t mention the SRR but rather “an analytically appropriate method supported by empirical evidence”) but in the 2005 re-edition of the original Amendment there was again the SSR in all its glory. The academic critique has been mainly theoretical. Now, with all the credit due to theory, the real test, however, of any method, must be empirical, and there has been profoundly little of that concerning the SRR (an exception is ibid.). It is in light of this lack of empirical research, and with hope to provide a remedy to this, or at least a building block in such a remedy, that this paper has been written.

1.2 Purpose

This paper aims to empirically investigate the following:

- Is the SRR a valid way of obtaining estimates of h-day VaR? Does the SRR live up to the criteria recommended by Basel for evaluating risk measurements through backtesting? Does is pass other standard methods of backtesting?

- How does SRR-scaling compare to directly calculating the h-day VaR the same way 1-day VaR was calculated, the "main model"? Does the main model consistently over- or underestimate risk as compared to SRR-scaling? Does it result in significantly more or significantly less "exceptions", i.e. actual h-day losses greater than predicted? Are the sizes of these exceptions generally significantly greater for either model?

Based on the answers of these questions, about the SRR in general and about the SRR as compared to its main model alternative, I will attempt to answer the crucial questions: *Can the SRR serve as a sufficient replacement for more sophisticated models? Can the use of SRR be recommended, or at least defended, as a thumb-rule for VaR estimation*?

1.3 Summary

We find that for the main-models *not rejected* by backtesting using the Christoffersen test, the SSR scaling performed more or less equivalently to the each of these. It was even so that, for the main-model performing the best, the Student-t model, the SRR alternative outperformed it slightly, though by no means significantly so. Furthermore, we find that the SRR scaling throughout the models tested *underestimated risk* as compared to the main model. Again, the exception to this was the best-performing Student-t model, though the difference was negligible in that case. On average, we find for all models checked, that the SRR scaling results in exceptions of generally lesser magnitude than those of its main-model alternative (the only being the anyway soundly rejected Age-Weighted Simulation model). As a whole, my conclusion is that the SRR cannot be rejected as a model for converting 1-day VaR to h-day, certainly not as far as "thumb-rule usage" is concerned. Furthermore, concerning usage of the SRR in more "strict" circumstances, where as exact measurements as possible are called for, it is my opinion that the answer to this question is far from obvious. Research into the benefits of using complicated and expensive models for h-day VaR as opposed to the easier- and cheaper-to-be-employed SRR should be conducted.

2 Theory and previous research

2.1 Theory

2.1.1 Market Risk Management in general, Downside Risk and VaR

There are different types of financial risk, such as liquidity risk, credit risk and market risk. It is the latter type of risk that is subject of the measurement that is the focus of this paper, Value at Risk. By market risk I mean the risk of loss arising from (unexpected) changes in the market prices of assets (Dowd (2005), page 1). Individual investors, firms and even countries are all exposed to such risk, and the ability to describe and quantify this risk can be crucial to decision making. Let us illustrate this with a imagined example that will take us directly into the type risk that VaR measures: A supermarket about to buy a large amount of vegetables to be sold within a given period (presumably the time for which they're still fresh) would be interested in knowing how much money they are likely to lose if the vegetables aren't all purchased, certainly so if there are bills to be paid by the end of that same period. This is what's called "downside risk". The question the supermarket manager is asking himself is really, what's the worst case scenario? Of course, the worst case scenario is a total loss, for example due to a sudden complete customer unwillingness to purchase vegetables. This is not impossible; news of a terrible decease hitting vegetables could very well cause such a drop. However, it is indeed a very unlikely event, and our supermarket manager might be willing to accept that risk, which, even though great were it to occur, since such an occurrence is so unlikely he might feel it unnecessary to take into account and to guard himself against it. Perhaps then he will specify that he's interesting only in the worst case scenario that has a likelihood of at least, say, 1%. This brings us to the standard Value at Risk measurement. Formally, VaR is defined as:

$$-VaR = q_p$$

where p = 1-α and α is some chosen confidence level of interest, and q_p is the p-quantile of profit/loss over some holding period. The minus sign is so as to have the loss-estimate in positive numbers. Of course, we do not usually in advance, or even at time (!) know the exact distributions of future returns, so assumptions have to be made in order to regarding this. In addition to making assumptions about the *general* distribution of the returns, we usually, if not in fact always, make the assumption that passed returns contain information about future returns. To clarify these two points, let's again exemplify. We might first make the assumption that returns are normally distributed, and then make further assumptions regarding the mean and variance of these future returns based on passed history. We can now define VaR more exactly as:

$$-VaR_\alpha = max\ (\mathrm{r}^*|\Pr(\mathrm{r} < \mathrm{r}^*) < 1 - \alpha)$$

where r denotes the returns, alpha the confidence interval in question (Nilsson (2012), chapter 1).

2.1.2 Models for estimating VaR

The critical question is obviously how to estimate Value at Risk, which in effect is equivalent to asking what to assume about the random variable generating the returns of interest at the time horizon of interest. Kevin Dowd in "Measuring Market Risk" presents the most popular methods and divides them into two categories: parametric- and non-parametric approaches; these will be presented below in brief, and further expanded upon in the methods section:

- Normally- or Student-t-distributed returns: These are two common parametric approaches. Returns are assumed be either normally distributed or, if a distribution with fatter tails is wanted to increase the probability of great losses, t-distributed. The expected value and the variance of the random variable generating the returns is estimated based on historical data, and in the case for the Student-t model also the degrees of freedom based on the historical kurtosis. Based on this the distribution of the returns at the horizon of interested is determined, and the relevant quantile is easily calculated using standard critical values.
- Basic historical simulation (BHS): A non-parametric approach. No a-priori assumptions are made about the distribution of the returns. The assumption made, instead, is that passed history provides us with a sufficient description of the distribution. From a certain estimation window the relevant quantile is located and this is assumed as the estimate of VaR.
- Weighted historical simulation (AWS and VHS): Again a non-parametric approach. A critique against BHS is that, in giving equal weight to all observations, it assumes that returns at all points of the estimation window were generated by the same random variable. Empirical results point to that newer observations contain more relevant information about coming returns than old, and thus it makes sense to give higher weight to the former. This is called Age-weighted simulation (AWS). Another point is that financial market exhibit volatility clustering (as can be seen in tables 1a and 1b), where period of high or low volatility tend to cluster. In making an estimate based upon the unconditional volatility it makes sense to scale the observations to diminish the impact of periods with high volatility and to increase the impact of periods with low volatility.

It should be noted that regarding all methods mentioned above, an appropriate length of the estimation window has to be chosen. Also, if daily returns are used as basis, which is usually the case, a way to convert 1-day VaR to h-day VaR, where h>1, is called for. This brings us to the next section.

2.1.3 VaR beyond the 1-day horizon and the SRR

Calculating VaR beyond the 1-day horizon involves a not trivial problem concerning the amount of data required. If, again, we're basing our analysis on daily returns, which is what's common and also what's recommended by Basel (and what's done in this paper), and want to attain a 10-day VaR, then we require 10 years of data in order to be able to use an estimation window of 250 observations. This, of course, can be problematic, and thus the practice arose to rely upon scaling (Provizionatou, Markose and Menkens (2005), page 1). One of the most commonly known ways of calculating VaR beyond the 1-day horizon is to simply scale the 1-day VaR by the square-root of the number of days in the horizon. Formally:

$$VaR_{\alpha}(h) = \sqrt{h} * Var_{\alpha}(1)$$

Now, the theoretical foundations of the SRR is that *the variance of the sum of independently and identically (i.i.d.) distributed random variables* is indeed equal to the variance of one such variable times the number of variables, so that the standard deviation of that sum is equal to the standard deviation of one of the variables times the square root of the number of variables. Thus, assuming that the returns are indeed i.i.d., and denoting the sum of n days of (log) returns as r(n), and the standard deviation as s(r(n)), then

$$s\big(r(n)\big) = \sqrt{n} * s(r(1))$$

Were asset returns indeed to be i.i.d., no matter what distribution they belong to, the SRR would be impeccable (Christoffersen, Diebold and Schuermann (1998), page 3, and others). As we will see, however, researchers have attacked the usage of the SRR just on this point, for it is commonly accepted that high frequency asset returns are not i.i.d.

2.2 Previous research

The SSR is quite popular among practitioners, even industry standard (Provizionatou, Markose and Menkens (2005), page 1), but more than, in 1996 the SRR received the very significant backing of the prime regulator, the Basel committee. In the "Amendment to the capital accord to incorporate Market Risk", section B.4, we find: "In calculating value-at-risk, an instantaneous price-shock equivalent to a 10-day movement in prices is to be used, i.e. the minimum "*holding period*" will be 10 trading days. Banks may use value-at-risk calculated at shorter holding periods scaled up to ten days by the square root of time …". It turns out, however that among researchers, the SRR, and the Basel embracement of this rule, has caused quite some discontent. Christoffersen, Diebold and Schuermann (from henceforth also C-D-S) (1998, page 3) points out that the SRR would indeed be correct if, as mentioned above, asset returns were i.i.d. If this were the case, then denoting the log-price at time t as p_t (p will throughout this paper mean ln(price)) and the return on that day as r_t, and further assuming a zero drift term, we have that

$$r_t(1) = p_t - p_{t-1} + \varepsilon_t$$

where the shock-term is i.i.d.(0,σ^2). Then we have for the h-day log-return that:

$$r_t(h) = p_{t+h} - p_{t-1} = \sum_{i=t}^{h} \varepsilon_i$$

with

$$r_t(h) \sim (0, h * \sigma^2)$$

and thus with a standard deviation of:

$$\sigma(1) * \sqrt{h}$$

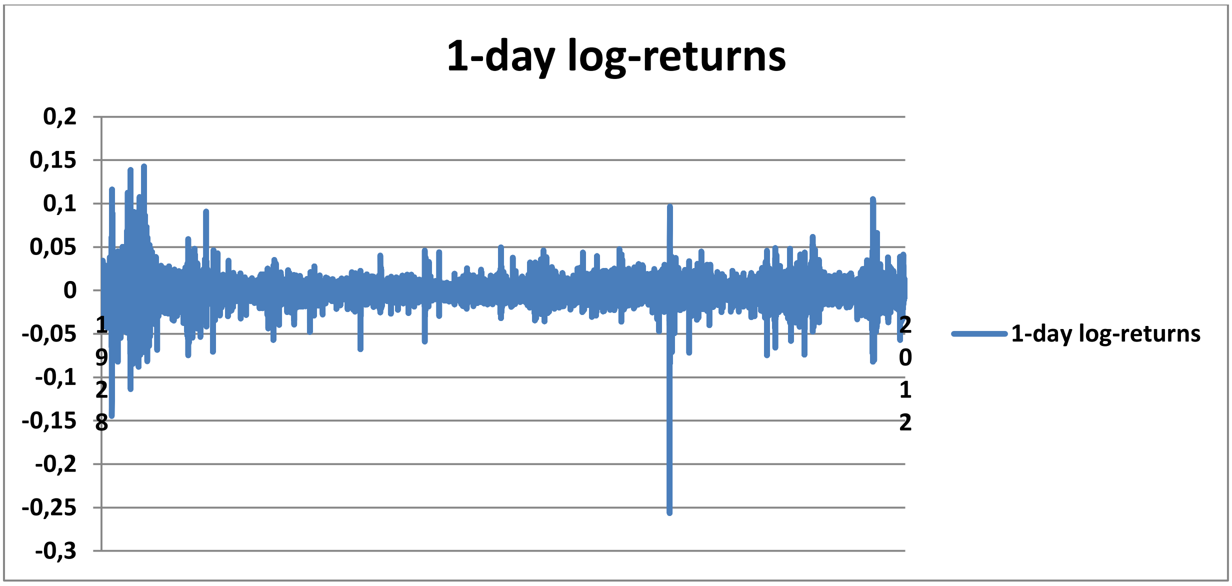
However, C-D-S (ibid) continue to claim that financial asset prices are all but i.i.d. Whether high-frequency portfolio returns are conditional *mean* independent has and is a bone on contentment among researchers but it has been soundly demonstrated by hundreds of papers that they certainly are *not* conditional *variance* independent. C-D-S (ibid, page 5) proceed to quote Drost and Nijman (1993) who derived formulas conversion of 1-day volatility for h-day volatility if the volatility component of the DGP follows a GARCH(1,1) and note that these formulas look nothing like the SRR scaling; furthermore, as Dorst and Nijman (ibid.) show, as h goes to infinity volatility fluctuations disappear, which is the opposite of the SRR where volatility fluctuations are magnified with time. The last point, I must however note, deviates from what is directly relevant to this paper, as Basel allows scaling up to 10 days only, and this is our focus here. An almost to the exact word identical paper is the one by Diebold, Hickman, Inoue and Schuermann (1998) (Diebold and Schuermann seem to be the connecting factor here). Danielsson and Zigrand (2003, page 2) also take aim on the i.i.d. assumption and quotes Engle (1982) that returns are not i.d.d. due to the presence of volatility clusters but add to this that when including a jump-process, in their opinion more apt to describe the commonly observed extreme-tail phenomena, the SRR *underestimates* risk at an increasing pace, as the probability of a jump increases with time. Dowd (2003, page 3), on the other hand, claims that the SRR *overestimates* risk at an increasing pace. Dowd focuses on the fact that in the long-run, the compounding of the drift-term will dominate (ibid., page 8), as it grows proportionally to h, while volatility grows proportionally only to the square-root of h. Thus, while "in fact" volatility fluctuation die out in the long run, SRR rises indefinitely. For a positive drift term, "actual" VaR will flatten out with time, reaching a limit which it will not pass with probability one, while SRR VaR will eventually pass this limit. For a zero drift term, actual VaR is at least bounded from above by the value of the original investment. SRR VaR, however, is not, and will eventually break through even that boundary. Dowd's arguments are indeed convincing. The problem with them, however, is that it is unclear against whom they are aimed. Dowd speaks about boundary-breaking horizons of decades! Whoever suggested using SRR in such a case? Now, it's true that the title of Dowd's paper indicates that his focus is on long-term VaR in which case we might have assumed he's taking aim mostly at mere laymen; however, he clearly says (ibid, page 4) that: "We suggest that the estimation of VaR should *not* involve the square-root-rule, which can be miss-leading, even for relatively short horizons…", and he does mention Basel explicitly (ibid. page 3) so it is unclear to me what his intention was. In all fairness, it should be noted that the choir has not been all negative; Allen, Boudoukh and Saunders (2008, page 10) wrote in defense of the SSR even at longer horizons of up to one year!

3 Method

3.1 Choice of asset, source of data and length of time series

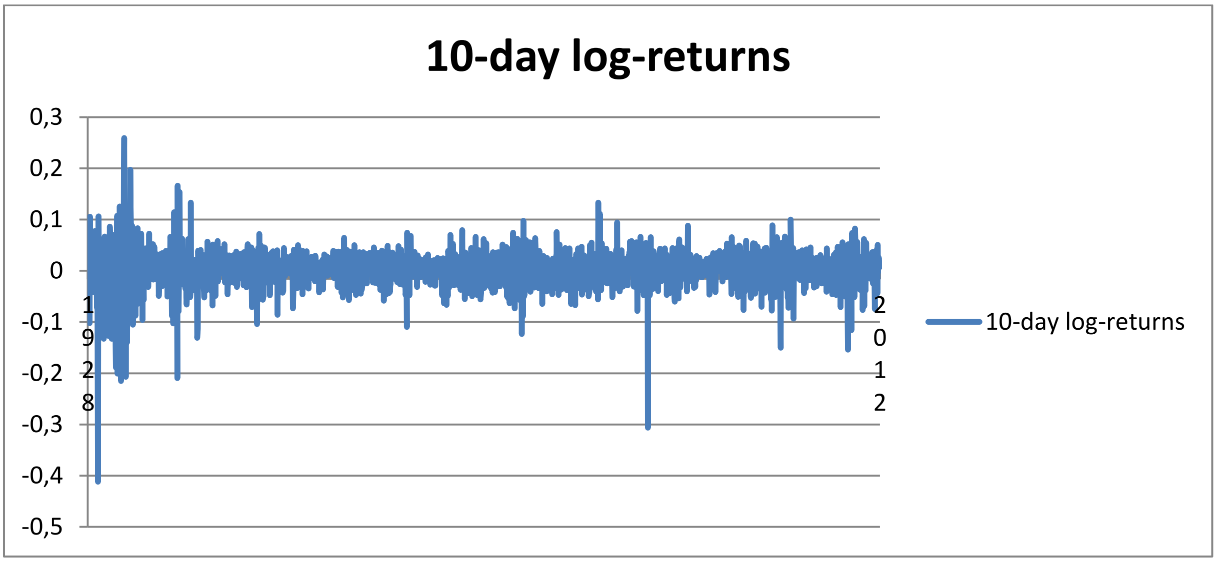
The data upon this research has been conducted was the Dow Jones Industrial index between the years 1928 and 2012; the time-series of prices was downloaded from Yahoo Finance. This rather long period was chosen for two reasons, first of all so as to generally decrease the impact of statistical peculiarities but more specifically also 10-day VaR was calculated based on aggregates of 10 consecutive days of 1-day return *without overlapping*, as is practice (Provizionatou, Markose and Menkens (2005), page 1), thus reducing the amount of observations to 10% of the amount of daily observations, necessitating a larger sample than had only 1-day VaR been the object. The DJI index was chosen as it is quite a standard benchmark workhorse of financial research, and considered representative of the American Industrial economy as a whole, and log prices are illustrated in table below in table 1a:

Table 1a



10-day log-returns was calculated as the sum of 10 consecutive 1-day log returns, each such sum consisting of non-overlapping, of 10-day returns. The 10-day log returns are illustrated below in table 1b.

Table 1b



3.2 On the calculation of VaR

There are a couple of issues that need to be decided upon for each model of VaR estimation:

- The confidence level needs to be decided upon, obviously being the same for all models so as to make comparison possible.
- The estimation length of the window has to be determined.

3.2.1 On the confidence level

Throughout this paper the level of 99% was used to calculate VaR. This is in accordance not only with standard practice but also with the Basel accords, which states (Amendment to the capital accord to incorporate Market Risk, section B.4) that: "In calculating value-at-risk, a 99^{th} percentile, one-tailed *confidence interval* is to be used".

3.2.2 On the estimation window

Seeing as there will always be an element of arbitrariness to any choice estimation window length I set out to choose a standard one. This also fits well with one of the objectives of this paper, namely focusing on the usage of thumb-rule estimation of VaR. Now, seeing as the length of 250 observations is mentioned in several papers (Provizionatou, Markose and Menkens (2005, page 7), Dowd and Cotter (2007), page 6) and considering that this is the length of the evaluation window for backtesting suggested by Basel (Supervisory Framework etc., section II), this was the length chosen.

3.2.3 On the calculation of the SRR

SRR, it should be noted, is in of itself not a model for calculating VaR, but a model for converting 1-day VaR as calculated by whatever model, to h-day VaR, rather than using the "main-model" also for this, directly. Thus calculating h-day SRR VaR involves the following two steps:

- Estimating 1-day VaR according to the main-model.
- Multiplying the 1-day VaR estimate by the square root of h.

3.3 Models to compare with

The models to I chose to compare the SRR to are all models listed as popular benchmark models by K. Dowd (2005, chapters 4 and 6), mentioned in brief above in section 1.2.2. presented in detail below:

- The Parametric Normal: Returns are assumed to normally distributed, i.e.

  $$r(h)_t \sim N(\mu_t, \sigma^2{}_t),$$

  where μ_t is the expected return at time t and $\sigma^2{}_t$ is the variance of the returns at time t. (Obviously if r(1) is normally distributed, so is r(10).) These two parameters were determined for each observation based on the last 250 observations preceding, i.e. on $r_{t-250}(h),\ldots r_{t-1}(h)$, by Maximum Likelihood, i.e.

  $$\hat{\mu}_t(h) = \frac{1}{250}\sum_{i=t-250}^{t-1} r(h)_i$$

  and

  $$\widehat{\sigma^2}_t(h) = \frac{1}{249}\sum_{i=t-250}^{t-1} (r(h)_i - \hat{\mu}_t(h))^2$$

  (B. Nilsson, Lecture notes, chapter 3, section 4a). Once an estimate of $\mu_t(h)$ and $\sigma^2{}_t(h)$ were obtained, an estimate of $VaR_t(h)$, the h-day VaR at time t, was easily calculated according to:

  $$VaR_t(h)= \mu_t(h) - k_{critical} * \sigma_t(h)$$

  where $k_{critical}$ is the critical value for the requested first percentile attained from the standard normal, in this case approximately -2.33. A rolling window was employed so that the next estimate of VaR, i.e. $VaR_{t+1}(h)$, occurring h days later, would be based on estimated of μ_{t+1} and $\sigma^2{}_{t+1}$ them in turn based on the h-day returns $r_{t+h-249}(h),\ldots r_{t+h}(h)$. The relevant comparison for backtesting would then be $r_{t+1}(h)$, the time t+1 h-day return realization. The attraction of this model lies partly in its simplicity; only two parameters have to be determined; also due to the possibility of appealing to the central limit theorem when aggregating sums of independent variables, which might be close to the truth for certain portfolios (Dowd, 2005, page 154). Furthermore the historically colossally wide usage of the normal to describe asset returns makes it an obvious candidate, even though research shows that financial returns to have excess kurtosis, something which can lead to underestimation of losses using the normal (ibid., page 157). Another problem with this approach is that the normal allows infinitely large losses, whilst losses normally are bounded from above due to limited liability. This might lead to an overestimation of losses (ibid.). Another obvious problem with this approach is that volatility is assumed to be constant and does not account for volatility clustering.

- The Parametric Student-t: Completely similar to what was described above about the Parametric Normal regarding all aspects, just that instead $k_{ckritical}$ is taken from the Student-t distribution. As for the attractiveness of the Student-t is that it allows us to adjust for the problem of excess kurtosis. We can in fact choose our kurtosis by choosing the degrees of freedom, since

  kurtosis=3(df-2)/(df-4)

  for df>=5 where df are the degrees of freedom, and similarly choose our degrees of freedom to suite our sample kurtosis. What was done here is to have a rolling window estimation of the kurtosis given by the standardized fourth moment as

$$kurtosis = \frac{1}{248}\sum_{t-250}^{t}\left(\frac{r(h)_i - \mu_t}{\sigma_t}\right)^4$$

  updating this for each observation and calculating

$$df_t = \frac{(4 * kurtosis - 3)}{(kurtosis - 3)}$$

  rounded to the nearest integer. To use this formula when the sample kurtosis equals 3, i.e. when there is no excess kurtosis, the degrees of freedom have been set equal to 100 000, giving us a distribution very close to the normal, with $k_{critical}$=-2.32.
- Basic Historical Simulation (BHS): On the basis of the 250 latest observations, VaR is calculated as the observation just higher than the first percentile in the estimation window. In other words, to calculate $VaR_t(h)$, we order the observations $r(h)_{t-250},\ldots,r(h)_{t-1}$ in increasing order, denoting them $r(h)^*_1,\ldots r(h)^*_{250}$ and VaR is calculated as

$$VaR_t(h) = \frac{r(h)_2^* + (h)r_3^*}{2}$$

  the average between the values just above and just below the relevant percentile (Dowd, ibid., page 84). The attractiveness of this approach lies in the fact that no a priori assumptions are made regarding the distribution of the returns, instead letting the returns, so to say, speak for themselves. Of course, the assumption that is made is just that, that passed returns contain information about future returns, but this is an assumption that's almost always made, and that seems sound empirically (though here it's taken to its extreme in the sense of regarding the last-period realizations to the only possible). What is less convincing, perhaps, is that equal weight are given to relatively old observations as to new, something that rhymes not so well a couple of stylized facts, first of all the observable existence of volatility clusters and secondly the fact that even obvious (to the representative contemporary observer) changes in risk occurring only late in the sample, say due to some event, would fail to show up in the VaR estimate because of the overwhelming weight given to those observations that occurred before the event. Also, we might, in the case last described, have a "ghost effect" of the consequences of the event showing up in VaR estimates quite a while after the impact already has diminished due to the observations effected by the event being numerous enough to influence (ibid. page 92)!

- Age-Weighted Simulation (AWS): A way to deal with the problems mentioned regarding BHS is by weighing observations according to their age. The idea is that newer observations are more relevant than older, should therefore each observations is weighted so that more weight is given no new later ones according to:

$$r_{weighted_{t-i}}(h) = w_i r_{t-i}(h)$$

where w_i is the weight given to the observation i days old according to

$$w_i = \frac{\lambda^{i-1}(1-\lambda)}{1-\lambda^{250}}$$

where w_1 is set to $\frac{1-\lambda}{1-\lambda^{250}}$ so that the weights sum to 1 and λ is a factor of decay, in this research set to 0.999 making for a rather slow decay. The weighted observations are the sorted in increasing order, similarly to BHS, as:$r^*_{weighted_1}, \dots, r^*_{weighted_{250}}$, while the un-weighted returns as $r^*_{un-weighted_1}, \dots, r^*_{un-weighted_{250}}$ so that $r^*_{weighted_1}$ is $r^*_{un-weighted_i}$ weighted, i.e. each un-weighted return is sorted so that it's has the same index in the vector of un-weighted, sorted returns as has, weighted, in the vector of weighted, sorted returns. The weights are similarly sorted in a vector of weighted, so that w_i will be the weight given to $r^*_{un-weighted_1}$ to obtain $r^*_{weighted_1}$. The weights, which seeing as they sum to 1 might be taken as probabilities, are then summed up going from the weight corresponding to the greates weighted loss, and once the relevant percentile has been reached, the next un-weighted return is taken as the measure of VaR.
- Volatility-Weighted Simulation modeling volatility by a GARCH model: Here the idea is similar to AWS, just that observations are weighted according to their volatility rather than their age. Next period's volatility is estimated and previous observations are scaled by multiplying by the next period estimated standard deviation and dividing by their own period's standard deviation. The idea is to scale observations to what they "would have been" had volatility in their period been like the one estimated for the next period. Of course, this approach entails estimating time-varying volatility. GARCH models have proven themselves astoundingly accurate for this purpose, despite their simplicity, and there is in fact research showing that complicated, structural models used at banks are not superior to a simple GARCH(1,1) (Berkowitz and O'Brian (2002, page 1094)). Therefore, I decided to include the GARCH as a proxy for the more advanced models used by quant teams at banks, and thereby attempting to answer the second main question of this paper, namely whether SRR scaling can serve as cheaper and more convenient substitute even for these more complicated models. In the GARCH(1,1) volatility is time-varying as given by:

$$\sigma_t^2 = \beta_1 + \beta_2 \varepsilon_{t-1}^2 + \beta_3 \sigma_{t-1}^2$$

Usually the Beta-parameters are estimated through Maximum Likelihood methods from the returns, something that should ideally be done at each holding period. (Nilsson (2012, chapter 3, section 3c)). An alternative to this, that sidesteps the issue of continuous re-estimation of the beta-parameters, is the Exponentially Weighted Moving Average model (EWMA): Here, the volatility is given by

$$\sigma_{t+1}^2 = \frac{1-\lambda}{1-\lambda^t} \sum_{i=t-250}^{t} \lambda^{t-i} \varepsilon_t^2$$

This is the model employed by RiskMetrics with λ =0.94. It can be shown that for sample sizes reasonably large, the EWMA is a special case of the GARCH(1,1) with β_1=0, β_1=1-λ β_2=λ (ibid.) For matters of computational simplicity the EWMA was chosen instead of a "regular" GARCH(1,1). An initial value of ε_0=0 was used, and σ_0^2 equal to the the variance of the entire sample, this being quite standard procedure (ibid.). As usual for each time t VaR_t was calculated based on 250 observations, each observation t-250<=i<=t-1 weighted by

$$r(h)_{i_{weighted}} = \frac{\sigma_{t+1}}{\sigma_i} r(h)_i$$

and VaR_t is taken as weighted return just among these weighted returns that's just above the first percentile.

3.4 The choice of h

Since Basel (Amendment etc., section B.4) permitted the usage of SRR scaling up to 10 days, and since there has been a case made that the inaccuracy of the SRR increases with time (K. Dowd, Long-term etc., page 11) the choice was made to compare the SRR to the its main-model alternatives at the 10-day horizon.

3.5 On the treatment of the data

The data was in daily prices. To make the research independent of nominal money-levels, I decided, as is also standard, to work with returns rather than prices. Furthermore, since I needed to calculate VaR on the 10-day horizon, I chose to deal with log-returns that have the comfortable characteristic of additivity. Denoting the log-price at day t as p_t, we therefore have that the h-day return going from time t_1, $r_{t1}(h)$, is given by

$$r_{t_1}(h) = \sum_{t=t_1}^{t_1+h} (\ln(p_t) - \ln(p_{t-1})) = \ln(p_{t_1+n}) - \ln(p_{t_1-1})$$

3.6 Empirical tests

The empirical tests were conducted in accordance with the two main questions raised in this paper, namely to compare SRR-scaling to direct main-model estimation, and to generally evaluation the performance of the SRR. More specifically:

- General descriptive statistics: I performed backtesting, calculating the amount of losses exceeding the VaR estimate, from henceforth call "exceptions", which in turn was calculated at the standard 1% level. For each model tested, this was done at the 10-day horizon. The percentage of exceptions for the SRR model was then compared to that of the main-model. Furthermore, the average magnitude of the exceptions was compared, as well as the instance of the largest exception.

- To evaluate the performance of SRR to general criteria, I compared the percentage of excess losses to the standard guidelines set up by Basel for evaluating VaR models according to backtesting. This was done, again, on all models, thus giving a relative measurement of the SRR performance to Basel standards as compared to these other models.
- Furthermore, the Christoffersen test was employed to evaluate each of the models, and their results were compared.

The above will be explained below in greater detail.

3.6.1 Backtesting

Backtesting in general terms amounts to comparing the risk measure predicted by the model in question to the actual realizations of profits/losses. If we're to claim our model to be "reasonably accurate" then surely our predictions should not deviate "too much" from the actual realizations. What is "too much" is obviously a delicate question, and different methods of backtesting exist. Since the major underpinning of this paper is the fact that Basel committee permits the SSR, I have used the Basel criteria for backtesting. Furthermore, I've used the so called Christoffersen-test. This will be elaborated upon below.

3.6.2 The Basel criteria

Basel (Amendment to the capital accord to incorporate Market Risk, section B.2) lists a number of requirements that bank have to comply with in to be eligible for application of the minimum multiplication factor, among them: "(a) The bank should have an independent risk control unit… (b) The unit should conduct a regular back-testing program, i.e. an ex-post comparison of the risk measure generated by the model against actual daily changes in portfolio over longer periods of time… (b) … A review of the overall risk management process should take place at regular intervals (ideally not less than once a year) and should specifically address, at the minimum: … the verification of the models accuracy through frequent back-testing as described in (b) above and in the accompanying document: *Supervisory framework for the use of backtesting in with the internal models approach to market risk capital requirements*", and in this referred-to document (section III) three "zones" are defined: "The green zone corresponds to backtesting results that do not themselves suggest a problem with the quality or accuracy of the bank's model. The yellow zone encompassed results that do raise questions in this regard, but where such a conclusion is not definite. The red zone indicates a backtesting result that almost certainly indicates a problem with the bank's risk model". The document then continuous to define the zones according to the number of losses, or "exceptions", observed larger than the VaR estimated by the model. Assuming 250 trading days up and including 4 exceptions would land us in the green zone, between 4 and 9 exceptions in the yellow zone, and more than 9 exceptions in the red zone. In table 2 below this is illustrated with accommodating percentages.

Table 2

	Green	Yellow	Red
Number of exceptions per 250 trading days	0-4	5-9	>=10
Percentage of exceptions	0%-1.6%	2%-3.6%	>=4%

What we can immediately see is that even the green zone allows for a slight margin above the 99% coverage level required by Basel themselves. Obviously even if the model is “correct” there is a positive probability of obtaining a sample with more than 1% exceptions, it would be undesirable to reject a correct model based on such an observation unless the number of excess exceptions was “sufficiently large”. Similarly there is a concern of not rejecting an incorrect model based on inflating the margin of excess exceptions allowed “too much”. In other word, the margin had to be chosen in order to balance reasonable between type 1 and type 2 errors. The regulators based their choice of cut off points on the following probabilities copied from the original document (ibid.), with the original explanations in table 3 below:

Table 3 (copied from the Basel committee (1996))

Model is accurate			Model is inaccurate: Possible alternative levels of coverage								
Exceptions	**Coverage = 99%**		**Exceptions**	**Coverage = 98%**		**Coverage = 97%**		**Coverage = 96%**		**Coverage = 95%**	
(our of 250)	**exact**	**type 1**	**(our of 250)**	**exact**	**type 2**	**exact**	**type 2**	**exact**	**type 2**	**exact**	**type 2**
0	8,1 %	100,0 %	0	0,6 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
1	20,5 %	91,9 %	1	3,3 %	0,6 %	0,4 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
2	25,7 %	71,4 %	2	8,3 %	3,9 %	1,5 %	0,4 %	0,2 %	0,0 %	0,0 %	0,0 %
3	21,5 %	45,7 %	3	14,0 %	12,2 %	3,8 %	1,9 %	0,7 %	0,2 %	0,1 %	0,0 %
4	13,4 %	**24,2 %**	4	17,7 %	26,2 %	7,2 %	5,7 %	1,8 %	0,9 %	0,3 %	0,1 %
5	6,7 %	10,8 %	5	17,7 %	43,9 %	10,9 %	12,8 %	3,6 %	2,7 %	0,9 %	0,5 %
6	2,7 %	4,1 %	6	14,8 %	61,6 %	13,8 %	23,7 %	6,2 %	6,3 %	1,8 %	1,3 %
7	1,0 %	1,4 %	7	10,5 %	76,4 %	14,9 %	37,5 %	9,0 %	12,5 %	3,4 %	3,1 %
8	0,3 %	0,4 %	8	6,5 %	86,9 %	14,0 %	52,4 %	11,3 %	21,5 %	5,4 %	6,5 %
9	0,1 %	**0,1 %**	9	3,6 %	93,4 %	11,6 %	66,3 %	12,7 %	32,8 %	7,6 %	11,9 %
10	0,0 %	0,0 %	10	1,8 %	97,0 %	8,6 %	77,9 %	12,8 %	45,5 %	9,6 %	19,5 %
11	0,0 %	0,0 %	11	0,8 %	98,7 %	5,8 %	86,6 %	11,6 %	58,3 %	11,1 %	29,1 %
12	0,0 %	0,0 %	12	0,3 %	99,5 %	3,6 %	92,4 %	9,6 %	69,9 %	11,6 %	40,2 %
13	0,0 %	0,0 %	13	0,1 %	99,8 %	2,0 %	96,0 %	7,3 %	79,5 %	11,2 %	51,8 %
14	0,0 %	0,0 %	14	0,0 %	99,9 %	1,1 %	98,0 %	5,2 %	86,9 %	10,0 %	62,9 %
15	0,0 %	0,0 %	15	0,0 %	100,0 %	0,5 %	99,1 %	3,4 %	92,1 %	8,2 %	72,9 %

Notes: The table reports both exact probabilities of obtaining a certain number of exceptions from a sample of 250 independent observations under several assumptions about the true level of coverage, as well as type 1 or type 2 error probabilities derived from these exact probabilities.

The left-hand portion of the table pertains to the case where the model is accurate and its true level of coverage is 99%. Thus, the probability of any given observation being an exception is 1% (100% - 99% = 1%). The column labelled "exact" reports the probability of obtaining exactly the number of exceptions shown under this assumption in a sample of 250 independent observations. The column labeled "type 1" reports the probability that using a given number of exceptions as the cut-off for rejecting a model will imply erroneous rejection of an accurate model using a sample of 250 independent observations. For example, if the cut-off level is set at five or more exceptions, the type 1 column reports the probability of falsely rejecting an accurate model with 250 independent observations is 10.8%.

The right-hand portion of the table pertains to models that are inaccurate. In particular, the table concentrates of four specific inaccurate models, namely models whose true levels of coverage are 98%, 97%, 96% and 95% respectively. For each inaccurate model, the "exact" column reports the probability of obtaining exactly the number of exceptions shown under this assumption in a sample of 250 independent observations. The columns labelled "type 2" report the probability that using a given number of exceptions as the cut-off for rejecting a model will imply erroneous acceptance of an inaccurate model with the assumed level of coverage using a sample of 250 independent observations. For example, if the cut-off level is set at five or more exceptions, the type 2 column for an assumed coverage level of 97% reports the probability of falsely accepting a model with only 97% coverage with 250 independent observations is 12.8%.

We might therefore complete our above summary of the zones by noting that the upper limits chosen for the green and yellow zone respectively correspond to a likelihood of a type 1 error, i.e. of falsely rejecting a correct model, of 24.2% and 0.1%. Thus, a correct model falling into the yellow zone would have been falsely rejected based on 4 exceptions cut-off point with likelihood 24.2%. It should be noted that models are not punished, according to this system, for being overly conservative. A model grossly inflating the risk, resulting in zero exception would pass this test with honors. Of course, this is understandable as the aim of the regulator is to limit risk-taking. The opposite objective, to stimulate the economy by encouraging reducing capital requirements as much as possible, can safely be assumed to be taken care of by the banks themselves. However, for the more theoretical purpose of evaluating how "true" the models in fact are, and not just how well they fulfill the purpose of the regulator, over-conservative models would desirably be rejected as-well. This is where the Christoffersen- test dealt with in the next section comes in.

3.6.2.1 The Christoffersen-Test

This test, or the version of the test used here, is a likelihood ratio test, measuring the whether the observed frequency of exceptions is sufficiently close to the predicted one. If x is the number of exceptions, n is the number of observations and the predicted probability of exceptions is p, then the Christoffersen test can be written formally as:

$$LR_C = 2ln\left[\left(1-\frac{x}{n}\right)^{n-x}\frac{x^x}{n}\right] - 2ln[(1-p)^{n-x}p^x]$$

where

$$LR_C \sim \chi^2(1)$$

(Dowd (2005, page 328)).

As it can be seen, this test punishes also a deviance to the downside, i.e. when the number of actual exceptions is lower than that predicted by the model.

4 Results and analysis

Below, in table 4a to 4f, the VaR estimates according to each main-model are illustrated together with its corresponding SRR estimate of VaR:

Table 4a

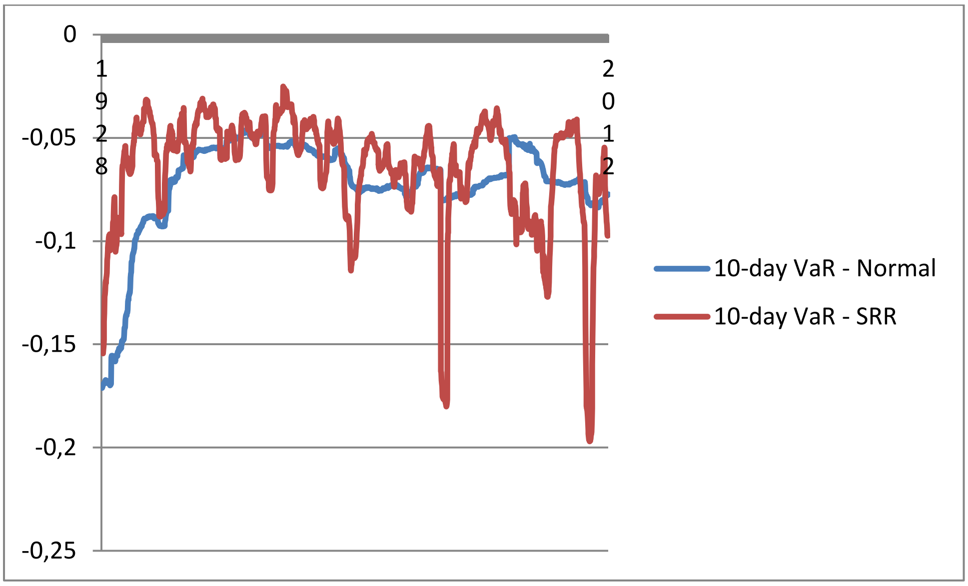


Table 4b

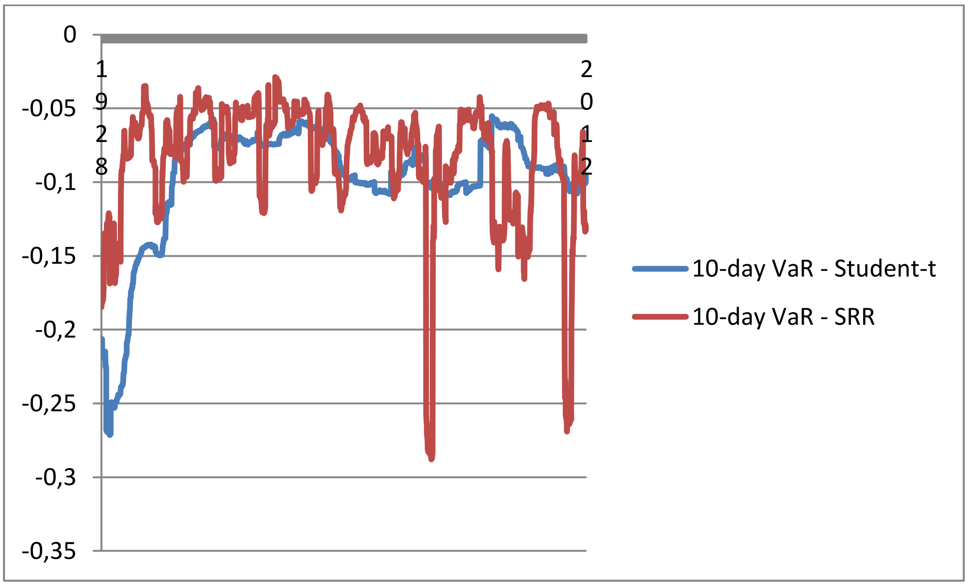


Table 4c

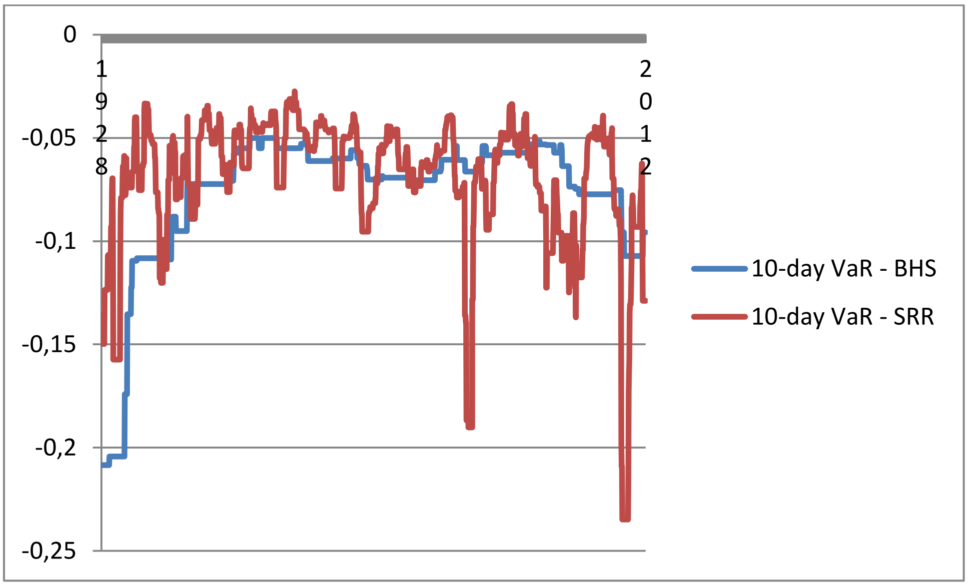


Table 4d

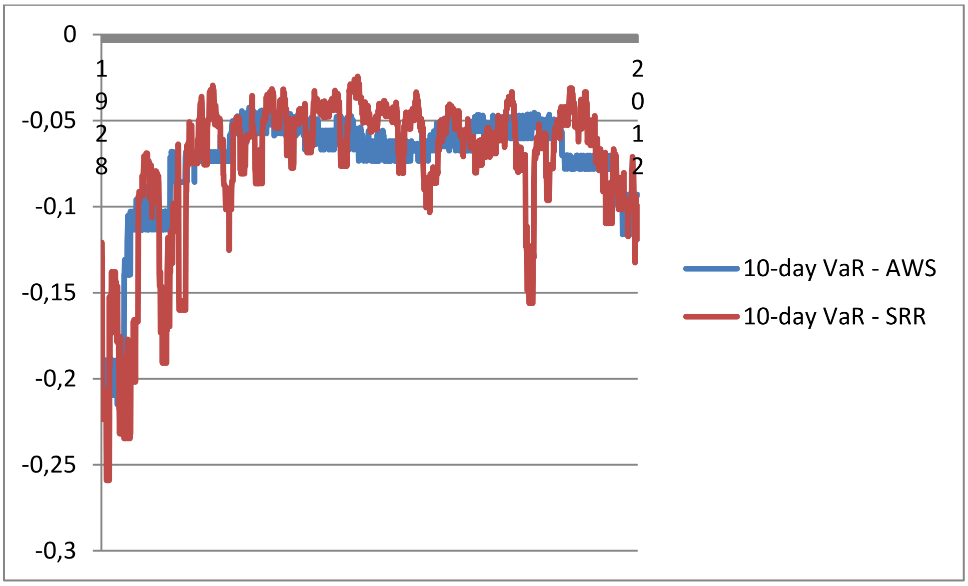
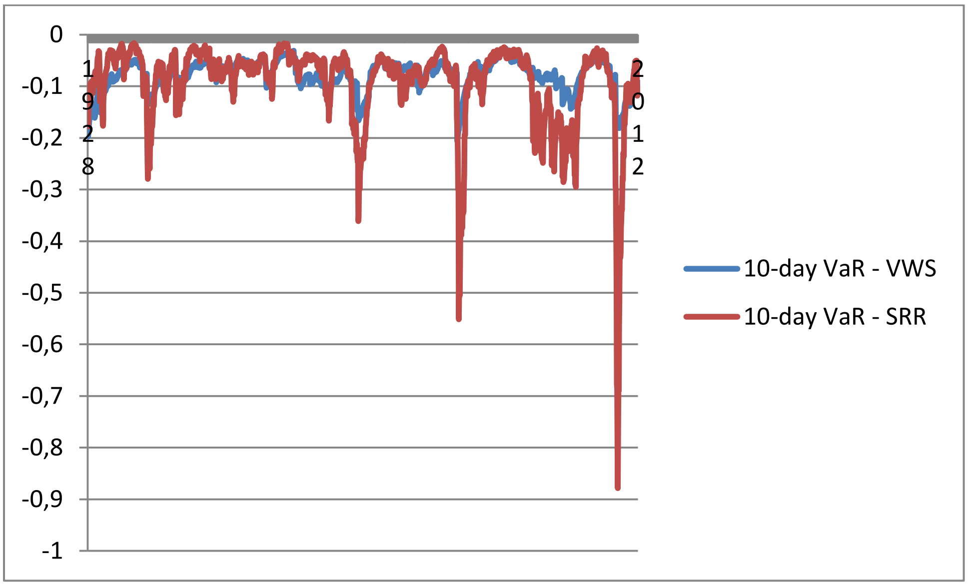


Table 4e



I will now go through our results, based on the questions originally raised in section 1.2.

4.1 Does the SRR live up to the criteria recommended by Basel for evaluating risk measurements through backtesting?

Below, in tables 5a-5e are the results of applying the Basel criteria to each of our models and their SSR. As the Basel criteria involves calculating the number of exceptions per 250 observations, the data has been divided into 7 groups, the first one consisting of the first 250 (10-day) observations, the next one of the next 250, and so on. A further explanation of how to interpret these tables is provided after table 5e.

Table 5a

Excep. per 250 obs.	Parametric Normal		SRR	
Period	#	Color	#	Color
1	1	Green	7	Yellow
2	3	Green	5	Yellow
3	4	Green	9	Yellow
4	6	Yellow	12	Red
5	2	Green	4	Green
6	1	Green	2	Green
7	10	Red	4	Green

Table 5b

Excep. per 250 obs.	Parametric Student t		SRR	
Period	#	Color	#	Color
1	0	Green	4	Green
2	1	Green	3	Green
3	2	Green	6	Yellow
4	2	Green	8	Green
5	1	Green	2	Green
6	1	Green	8	Yellow
7	4	Green	6	Yellow

Table 5c

Excep. per 250 obs.	BHS		SRR	
Period	#	Color	#	Color
1	0	Green	2	Green
2	1	Green	3	Green
3	4	Green	7	Yellow
4	5	Yellow	11	Red
5	2	Green	2	Green
6	2	Green	0	Green
7	5	Yellow	1	Green

Table 5d

Excep. per 250 obs.	AWS		SRR	
Period	#	Color	#	Color
1	1	Green	1	Green
2	2	Green	0	Green
3	7	Yellow	6	Yellow
4	7	Yellow	22	Red
5	4	Green	2	Green
6	4	Green	2	Green
7	7	Yellow	11	Red

Table 5e

Excep. per 250 obs.	VWS		SRR	
Period	#	Color	#	Color
1	3	Green	7	Yellow
2	2	Green	3	Green
3	4	Green	9	Yellow
4	3	Green	5	Yellow
5	4	Green	2	Green
6	4	Green	2	Green
7	4	Green	1	Green

Exceptions were counted per period of 250 days. The non-overlapping periods were numbered from 1 to 7, and the number is given in the left-most column. The column labeled "#" gives the number of exceptions in that period and the columns labeled "Color" gives the Basel color-code corresponding to that the number of exceptions. This is done both for the main model and for the SRR-scaling. Thus, for example, in 5e, in period 3, we had 4 exceptions for VWS, resulting in the color green, and 9 exceptions for the SRR-scaling, resulting in the color red.

On principle, I will consider a model as unacceptable if even one case of red is detected, in any of the seven periods, even though the Basel-criteria were for the result of one single 250-day period alone. Obviously, repeating the testing over several such periods increases the likelihood of at least one period producing the color red even for a correct model, thus allowing us to exercise some leniency regarding this principle (not much though, as 7 samples is still quite limited) if, say, the result in the other periods was exceptionally good, thus indicating that the one red instance was a statistical anomaly. Due to the lack of formal guidelines for testing of several periods, intuitive judgment must be employed here. It must be stressed though, and this is of paramount importance, that bad performance might be due to the main-model being bad, and not because of the inappropriateness of SRR-scaling. This point will also be discussed, though again rather intuitively.

- The Normal model: Seeing as we have one case of red the conclusion must be, based on the Basel criteria and barring some strong reason to exercise leniency that something is wrong here. Furthermore, seeing as in period 3, with 9 exceptions, we were on the verge of another red, and also in period 1 we had 7 exceptions, the red color in period 4 does not seem like an anomaly. Going to the question of whether this is due to main-model failure or SRR failure, we note that the main-model also yielded an instance of red, in the seventh period, and in that case the SRR scaling resulted in the color green. Taken at face value, this would indicate that the main-model is itself unfit, and surely on a theoretical plane we'd be hard pressed to defend the use of scaling, itself theoretically dubious, of equally dubious 1-day estimates (assuming the main-model performs equally bad on the 1-day horizon as on the 10-day ditto). Now, one of the main points of this paper is to let practical results determine model usage, but here, seeing as the SRR also fails, the operative conclusion must be to reject the combination of the Normal and the SRR.

- The Student-t model: Four green and three yellow periods, a result that should raise some questions, but certainly no dismissal. Seeing as the main-model performed very well across the seven periods, green all through, the combination of the Student-t and the SRR, at least for thumb-rule purposes, seems promising. For usage in circumstanced requiring more exactness, the three yellow periods would make this author feel some un-ease.
- The BHS: The SRR yielded one red grade, in period four, and as the BHS generally performed well although not without question marks, we'd be hard pressed to blame this on main-model failure. However, seeing as the SRR result in the fourth period deviated so significantly from that of the other periods (though period 3 did produce 7 exceptions, which is not so great), I'd still raise a question mark to basing a dismissal on the red instance in period 4. Here, we might actually have strong reason to exercise leniency. Still, obviously, we cannot give thumbs-up either, and so we'd be best off leaving the question of the BHS – SRR combination for thumb-rule usage unanswered from the result of Basel-test, though for more critical purposes than thumb-rule estimates, the conclusion must be a rejection.
- The AWS: One yellow, four green but two red makes for a dismissal of the AWS. Considering that question marks were raised even about the main-model, with three yellow periods, the combination AWS and SRR seems not to be recommended.
- The VWS: Again, as for the Student-t model, no red periods, and at that for a main-model performing excellently. Four green and three yellow for the SRR does mean questions have to be raised, but certainly no dismissal. Again, for thumb-rule purposes, the combination VWS and SRR seems promising whilst for more critical purposes a certain feeling un-ease descends upon you.

We had three well-performing main models here: the BHS, the Student-t and the VWS, and SRR-scaling performed quite well, at least on the two latter. The fact that the SRR for the VWS and the Student-t were without red periods would certainly make us prefer them to the BHS, though considering that not all was green in the these two models either, the conclusion must be that questions are indeed to be raised concerning the SRR even for the best performing combinations, though I'd say nothing definite should be deduced regarding, certainly not for thumb-rule usage.

4.2 Further descriptive statistics

In table 6 below all results pertaining to the next paragraphs this section will be presented:

Table 6

		Model data							Corresponding SRR scaling data					
Model	# of observations	# of exceptions (1)	% of excep. (2)	Size of biggest excep. in % (3)	Ave. excep. size in % (4)	% of higher estimated losses in the main than in the SRR estimate	p-value of the Christoff. Test (5)	% of deviation from 1% target (6)	(1)	(2)	(3)	(4)	(5)	(6)
Parametric normal	1853	31	1.67	24.07	2.80	71.40	0.008	0.67	46	2.48	23.60	2.04	0.000	1.48
Parametric Student-t	1853	14	0.76	22.36	3.85	71.13	0.27	0.24	26	1.40	22.02	1.97	0.10	0.40
BHS	1842	22	1.19	24.55	3.48	58.10	0.416	0.19	26	1.41	17.02	1.67	0.09	0.41
AWS	1842	35	1.90	23.87	2.43	55.19	0.000	0.90	48	2.60	25.14	2.72	0.000	1.60
VWS	1808	26	1.44	23.31	2.17	55.31	0.079	0.44	29	1.60	14.50	1.63	0.018	0.60

4.2.1 Does the SRR pass the Christoffersen-test?

- The Normal model: The fact that the main-model didn't itself pass the test, on the standard 1% level, it should be of no surprise that the SRR didn't, and nothing can be deduced from this as to the accuracy of the SRR scaling, though its suitability is obviously as doubtful as the main-model.
- The Student-t model: Here both the main-model and the Student-t passes with excellence, passing even the 10% level.
- The BHS: The main-model did pass with margin here, while SRR-scaling failed on the 10% level but passed at the 5% level.
- The AWS: Similarly to the normal, both the main-model and the SRR were soundly rejected and the conclusion is therefore the same as for the normal above.
- The VWS: Here, the main-model passed at the 5% level but not on the 10% level, while the SRR passed on the 1% level but not on the 5% level.

Summarizing, the results of the Normal and the AWS are soundly negative, while for the Student-t the SRR seems promising. The BHS must be seen as the second-best model here, while for the VWS the results do raise some questions. Taken together with the results of the Basel-test, the Student-t/SRR emerges as the best-performing combination.

4.2.2 How does SRR-scaling compare to directly calculating the h-day VaR the same way 1-day VaR was calculated, the "main-model" in an intuitive sense?

Both regarding the Basel- and the Christoffersen test, each main-model performed better than its SRR alternative. However, we should note a certain discrepancy here; while for the Basel test the results are quite similar for all models in that the main-model clearly outperforms SRR scaling, for the Christoffersen test, there is an exception: the Student-t. Though also for the Student-t, the main-model performs better than the SRR, both pass the highest of the standard levels, the 10% level, while for the BHS and the VWS this is not true. This, again, gives strength to the combination Student-t/SRR as best performing candidate. Furthermore, looking at how close the SRR gets to the target 1% of exceptions and comparing to its main-model alternative, we turn to table 6, and the columns names (6). Now, for all models the SRR produced a percentage of exceptions that lies further away from the target 1% than the main-model alternative, thus being less accurate. However, defining 1% of deviation as "insignificant" (statistically, that is, though money-wise it might be extremely significant) for no model did SRR scaling deviate significantly from the main-model estimate. However, if we look at deviation from the 1% target, then for the Normal and the AWS the SRR scaling deviated more than 1% from the target, while the main-model did not. This last aspect is of course the most crucial, and the whole purpose is to predict those 1% worse cases. The conclusion must therefore be that as far as the intuitive aspects discussed here, the SRR doesn't not perform significantly worse that the main model for the Student-t, the BHS or the VWS, but does perform significantly worse for the Normal and the AWS in that for these the main model is not significantly off target while the SRR is.

4.2.3 Does the main model consistently over- or underestimate risk as compared to SRR-scaling?

We see in table 6 above (the seventh column) that the SRR consistently underestimated risk as compared to its main-method alternatives; this can also be observed in clearly in the 10-day VaR plots of tables 4a-4c where the red SRR line consistently tends to be above the blue main-model line, though it might be slightly less obvious in tables 4d and 4e. This could be a voiced as an argument against the SRR from the point of view of the regulator, whose interest is mainly to limit risk-taking, though from the point of view of, say, bank management, the argument might go the opposite way, so as to maximize lending.

4.2.4 Are the sizes of these exceptions generally significantly greater for either the SRR or its main-model alternative?

Here, we're investigating the tail to the left of the VaR estimate, something which VaR doesn't really address. Still, obviously it's an important question and even if we don't value this point to the extent of looking for alternative risk-measures other than VaR (though we certainly might want to do that, as do the theoreticians of Extreme Value Theory), it would certainly be preferable to use a VaR model that comes as close as possible to the actual losses. In table 6, in the columns named (4), we see that the SRR actually beat all its main model alternatives, the exception being the AWS which anyway, together with the Normal, is not to be recommended for usage with the SRR (or alone, for that matter).

5 Conclusions and final remarks

5.1 The main questions

5.1.1 Can the SRR serve as a sufficient replacement for more sophisticated models?

The VWS was included as a proxy for more sophisticated models and as has been concluded, certain question marks have to be raised regarding the VWS/SRR combination. There were in fact question marks also about the best-performing combination: the Student-t/SRR. However, I believe it'd be a mistake to therefore draw the conclusion that SRR scaling is not to be employed as a replacement for the more sophisticated models or to be used in critical circumstances. Why? Well, first of all we'd obviously have to test these more sophisticated models themselves. Do they themselves perform without giving rise to question marks? Are they at all "better" than the models tested here? If not, and if SRR scaling could not be dismissed for the Student-t and the VWS, perhaps it cannot be dismissed for these sophisticated models either. Furthermore, there is a deeper reason why not to infer from the result of one combination onto another. I will now explain why. The whole case for the SRR is empirical to the point of being anti-theoretical, meaning that there is no theoretical grounds for the SRR (having dismissed the i.i.d. the assumption regarding high-frequency asset returns) and any defense of the SRR would have to be based on humbly acknowledging that we know little of the mechanisms driving asset returns, and that our approach should be based on statistical fit rather than imposing any theory on observed reality. Therefore, inferring that what was found regarding SRR performance on one model onto another is not to be done, for perhaps the enigmatic DGP generating the returns is such that one main-model is fit for SRR-scaling while another is not. The answer to the above question will therefore be left answered.

5.1.2 Can the use of the SRR be recommended, or at least defended, as a thumb-rule for VaR estimation?

Here I think the answer must be a yes. In no way could the SRR be described as "completely off" as compared to its main-model alternatives, as discussed in section 4.2.2 above, and thus for thumb-rule usage the SRR can certainly be employed. This SRR sanctioning, however, is obviously only applicable when the main-model is "good enough". In light of the results presented in this paper, the combinations Student-t/SRR and VWS/SRR should be accepted for thumb-rule usage, and possibly even the BHS/SRR.

5.2 Ideas for future research

To perform similar research to this on:

- Other assets
- Other models
- Time frames between one day and 10 days.

Furthermore, it would obviously be nice if we could escape the clash between theory and empirical results, and reconcile the relatively satisfactory performance of the SRR with a suitable theoretical framework, however daunting such a task might seem.

6 Bibliography

Allen, L., Boudoukh, J. and Saunders, A (2004).	Understanding Market, Credit and Operational Risk, Blackwell Publishing Ltd.
Basel Committee on Banking Supervision (1996)	Amendment to Capital Accord to accommodate Market Risks.
Basel Committee on Banking Supervision (1996)	Supervisory framework for the use of backtesting in with the internal models approach to market risk capital requirements
Berkowitz, J. and O'Brian, J. (2002)	How accurate are Value-at-risk models at commercial banks, The Journal of Finance, Volume LVII, number 3.
Christoffersen, P.F. (1998)	Evaluating interval forecasts, International Economic Review 39, p. 841-862
Christoffersen P. F., Diebold, F., Schuermann, T. (1998)	Horizon Problems and Extreme Events in Financial Risk Management, Economic Policy Review, Federal Reserve Bank of New York, October, p. 109-118.
Diebold, F., Hickman, A., Inoue, A. and Schuermann, T. (1996)	Converting 1-Day Volatility to h-Day volatility: Scaling by $\sqrt{h}$ is worse than you think, Wharton Financial Institutions Center, Working Paper 97-34.
Dowd, K., Blake, D. and Cairns, A. (2003)	Long term Value at risk, CRIS discussion paper series.
Dowd, K., Blake, D. and Cairns, A. (2000)	Extrapolating VaR by the Square Root Rule, Financial Engineering News, August, pp. 3,7
Dowd, K. (2005)	Measuring Market Risk, Second edition, John Wiley & Sons Ltd.
Dowd, K. (2000)	Assessing VaR accuracy, Derivatives quarterly 6 (3), 61-63
Dowd, K. and Cotter, J. (2007)	Evaluating the Precision of Quantile Based Risk Estimators, Munich Personal RePEc Archive 2007, online at (January 2012) http://mpra.ub.uni-muenchen.de/3504/
Drost, F. C. and Nijman, T. J. (1993)	Temporal aggregation of GARCH processes, Econometrica 61, 1993, p. 909-927
Engle, R. F. (1982)	Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation, Econometrica 50, 1982, p. 987-1007
Nilsson, Birger (2012)	Lecture Notes, Lund University, School of Economics and management, available by e-mail from the author: birger.nilsson@nek.lu.se
Provizionatou, V., Markose S. and Menkens, O. (2005)	Center for Computational Finance and Economic Agents, Economic Department, University of Essex
Yahoo Finance time-series of daily DJI-index prices, available during March 2012	http://finance.yahoo.com/q/hp?s=%5EDJI+Historical+Prices