An evaluation of quantitative standards for the implementation of internal market risk model by commercial banks.

Dr. Joshua W. Chesoli, PhD
Kisii University,
School of Business and Economics,
Department of Accounting and Finance,
P.o. Box 408-40200, Kisii, Kenya.

Email: chesolij@yahoo.com

Abstract

The purpose of this paper is to evaluate the quantitative standards laid down under the second Basel Accords for the implementation of internal market risk models by banks. The paper surveys available research to evaluate the standards. The standards don't prescribe a VaR method despite evidence that volatility of financial returns is conditional and financial returns are fat tailed. The requirement of a minimum historical period also runs contrary to the finding that volatility is time varying and clustered resulting in banks being able to use weighting schemes conservatively only. The minimum horizon of ten days requires use of a scaling rule that is not accurate. The 99% confidence level requirement increases the inaccuracy when using a normal assumption on fat tailed data. The minimum updation period and minimum historical period requirements effectively smooth the market risk charge over and above the smoothing by the requirement of averaging VaR resulting in unresponsive market risk charges. The regulatory back testing framework is based on unconditional coverage and does not penalize clustered VaR exceptions.

Key Words: Basel accord, GARCH, Historical simulation, Market risk, Value-at-risk, Volatility, Conditional volatility, Back testing.

1. Introduction

This paper evaluates the quantitative standards laid down under the second and third Basel Accord (Basel Committee on Banking Supervision, 2006 and Basel Committee on Banking Supervision, 2009) for the implementation of internal market risk models by banks. The evaluation is based on available research.

The quantitative standards laid down under the second Basel Accord for the implementation of internal market risk models by banks directly impact the choice of the VaR methods that banks use. It is critical that regulators, investors and bankers have the ability to assess the impact that the bank's choices will have on the final risk measure. This, unfortunately, is severely constrained by the paucity of available research. This paper discusses available research and points out directions for research that can help in assessing the impact of the choices made by the bank on its capital charge.

Evaluation of the standards is also important because they are driven by regulatory objectives that may be different from those of the banker or investor. For example, the objective of the regulator may be to lay down the minimum acceptable standards. The minimum standards may not correspond to the best standards. At the same time the objective of the bank will be influenced by the use that the internal models are put to. If the internal models are used to calculate the market risk capital requirements, the bank will want to meet the minimum standards required by regulations with the lowest market risk charge (to operate with lower capital without incurring any penalties).

Basel Committee on Banking Supervision (2011) also reviews the available research on the quantitative standards. However, only three aspects are covered by them, namely (1) time horizon over which VaR is estimated; (2) the recognition of time- varying volatility in VaR risk factors; and (3) VaR backtesting. This paper discusses eight out of the eleven quantitative standards including these three. The revisions to the quantitative standard made under the Basel III provisions (Revisions to the Basel II Market Risk Framework, 2009) have also been incorporated in this paper.

It is useful here to map the process of market risk charge generation because the final market risk capital charge is a combination of data inputs, models, assumptions and calculations. A bank has to make multiple choices resulting in a unique market risk generating model and every choice is constrained by the quantitative standards. Figure 1 maps the process and its inputs. Usually the process of market risk estimation starts with a volatility forecast. An exception to this is the historical simulation method. The methods that use a volatility forecast can be divided into historical time series based methods and implied volatility methods. The historical time series methods can be classified further as conditional and unconditional methods. The unconditional methods generate forecasts under the assumption that returns are independently and identically distributed and include the variance-covariance and Monte Carlo simulation. Conditional methods account for volatility clustering and include ARCH/GARCH models. The generation of VaR from a volatility forecast is the next step in the process outlined in figure 1. The VaR can be derived from the volatility forecast based on a variety of distributional assumptions. The most popular are the normal and the Student's t. The VaR forecast is averaged and scaled up in line with the Basel recommendations to yield a market risk measure. The level of scaling depends on the results of a regulatory back testing procedure. This paper is structured to reflect the process outlined in figure 1. The first section discusses the quantitative standards relating to the volatility forecast. The second and third sections discuss the standards relating to the VaR estimate and the final risk charge respectively.

There are 11 quantitative standards ((a) to (k)) laid down by the second Basel Accord. Of these (g), (h) and (k) are not evaluated. (g) relates to correlations across risk factors and is not discussed since most of the research surveyed relates to single risk factors. (h) is left out since it relates to option positions and (k) because it relates to specific risk.

2. Quantitative Standards Relating to Volatility Forecasts

The standards (d), (e) and (f) relate to the generation of the volatility forecast. Each standard is reproduced followed by an evaluation. (f) is discussed first since it is relevant to the remaining two standards.

> *(f) No particular type of model is prescribed. So long as each model used captures all the material risks run by the bank, as set out in qualitative standards (specification of*

risk factors), banks will be free to use models based, for example, on variance-covariance matrices, historical simulations, or Monte Carlo simulations.

There are two aspects to the evaluation of this standard. The first is the relative acceptability of a model across banks. The second is the quality of forecasts from alternative models and a reconciliation of this with the banks' choice.

Perignon and Smith (2006) survey the VaR disclosures of a cross section of 60 US, Canadian and large international banks over 1996-2005 and report that 73 percent of banks that disclosed their VaR methodology used historical simulation. The second most popular was Monte Carlo simulation (14%). A survey report on Indian banks (Transition to the Internal Models Approach for Market Risk – a Survey Report, 2010) states that of the 30 banks that participated in the survey 67% of the banks used historical 'methods' (it is not clear if this is simulation), 5% used Monte Carlo simulation and 3% used variance-covariance models. Thus, historical simulation is the preferred method. In case of volatility forecasts the preferred methods are unconditional methods based on historical time series.

While this standard does not prescribe a method, the widespread popularity of historical simulation warrants discussion. Apart from being the simplest model, the historical model has the advantage of being able to deal with portfolios of risk factors without having to explicitly model correlations. Since it is entirely empirical and does not use any distributional assumption it is better able to model the fat tails in the data. As a result it has been found to perform best in back tests of the type prescribed by the Basel committee. Hendricks (1996), Jackson et al (1997), Vlaar (2000), Boudoukh et al (1998) find that historical simulation (HS) provides superior coverage (fractions of exceptions to VaR reported in back test) compared to the EQMA and EWMA approaches. Ouyang (2009) finds the HS superior to GARCH (with normality assumption) in coverage. Sharma (2012) conducts a literature survey of performance of VaR methods and finds that HS has best performance in regulatory type back tests but fails in rigorous tests of conditional coverage. In contrast to its popularity in the industry it is virtually ignored by researchers who have focused on more sophisticated models resulting in a chasm between research and reality. Of the volatility forecasting models, the GARCH, in particular the EGARCH, is found to perform better than others, especially for stock market and short horizons, by Poon and Granger (2003) in a review of volatility forecasting models.

There are numerous gaps in available research that stem from the divergence between research and practice. For example, there is a paucity of studies that compare the performance of the historical simulation method with the other methods. Most research focuses on sophisticated approaches such as GARCH, ignoring the humble, but popular, historical simulation. Secondly, in the evaluation of performance of alternate methods the back testing framework prescribed by the Basel accord is rarely used. Most researchers use traditional measures of forecast accuracy such as mean squared errors (see standard (j) in section III for a list of forecasting accuracy measures used by researchers). Thus, it is difficult to judge the impact that the choice of a model will have on the regulatory back test. The regulatory back testing approach of measuring unconditional coverage is itself flawed in comparison to conditional coverage methods and research using conditional coverage methods is even fewer (see standard (j)).

(d) The choice of historical observation period (sample period) for calculating value-at-

risk will be constrained to a minimum length of one year. For banks that use a weighting scheme or other methods for the historical observation period, the "effective" observation period must be at least one year (that is, the weighted average time lag of the individual observations cannot be less than 6 months).

The objective behind this standard is to set a minimum standard in selection of data. The basic question raised by this standard is - 'how is more data better than less'? There are multiple possible advantages of more data - more reliable risk measures is one, smoother VaR forecasts is another. It is also argued that a longer period VaR is more likely to incorporate a stressed episode and therefore result in higher VaR and better coverage.

As pointed out by Hoppe (1998) more data is **unlikely** to give more reliable forecasts if volatility is time varying and clustered. Evidence of volatility clustering has been reported widely by researchers. The Autoregressive Conditional Heteroscedasticity (ARCH) specification of Engel (1982) and its parsimonious representation – the GARCH proposed by Bollerslev (1986), which model volatility clustering, have found to fit the data well, by Engle (2001), for example.

If reliability is taken to mean better performance in regulatory back tests, then a study by Ouyang (2009) directly addresses this question. He uses daily returns of the Shanghai Synthesized Index and the Shenzhen Component Index to examine the performance of five different 99% VaR methods (Exponentially Weighted Moving Average (EXMA), Equally Weighted Moving Average (EQMA), GARCH(1,1), Historical Simulation (HS) and an extreme value model) using two lengths of observation periods - one and three years. The standard deviations of the VaR estimates of EQMA and HS are significantly higher using one year compared to 3 year data. Conversely, the GARCH and EWMA approaches have lower scatter in their VaRs when using three year data. All models perform the best in regulatory back tests with one year observation periods. Their performance deteriorates with the 3 year window.

While shorter observation periods give more responsive VaR measures, one problem with more responsive VaR measures is that they result in pro cyclicity of the market risk charge - generating lower capital requirements in bull runs and higher in bear phases.

Though this standard was designed to lay down minimum lengths of historical observation periods it had an unintended consequence of impacting the bank's choice of volatility forecasting model because it effectively ruled out the use of conditional volatility models. For example, the popular exponential weighted average model (EWMA) proposed by Risk Metrics was found by Jorion (2002a) to have a weighted average time lag of only 16.7 days compared to the minimum six month lag required by the standard. He found that the requirement of a minimum "effective" observation period destroyed the advantage from using a weighting scheme that is responsive to recent volatility changes, such as the EWMA. Jorion (2002a) concluded that the Basel II rule of "effective" observation period being at least one year constrained banks to use slow moving models in order to generate smoother capital requirements. The second important point brought out by this study is that a VaR method that doesnot use a weighting scheme (such as the EQMA or HS) may be as responsive to recent volatility if the observation period used by it is short. Thus, there is a trade off between conditional volatility schemes, length of observation period and frequency of updating which needs to be researched to understand its impact on performance of VaR methods. The assumption that longer periods are more likely to contain episodes of stressed volatility and generate higher VaRs as a consequence also needs to be researched.

The Senior Supervisors Group (Senior Supervisors Group, 2008) constituted in the aftermath of the market turbulence of 2007/2008 questioned the use of longer observation periods. It noted that the dependence of VaR models on historical data from benign periods resulted in under estimates of VaR. It suggested shorter horizon historical data, giving greater weights to recent data and more frequent updation of datasets as remedies. As a result this standard was amended by Basel III by adding the following qualification: A bank may calculate the value-at-risk estimate using a weighting scheme that is not fully consistent with (d) as long as that method results in a capital charge at least as conservative as that calculated according to (d). The qualification introduced by the Basel III revision at least allows the use of conditional volatility models provided they result in a higher capital charge than the minimum standard. At the same time this also means a bank can continue to use unconditional models as long as they meet the minimum length of historical observation period.

> *(e) Banks should update their data sets no less frequently than once every month (once every three months earlier, before Basel III revision) and should also reassess them whenever market prices are subject to material changes. This updating process must be flexible enough to allow for more frequent updates. The supervisory authority may also require a bank to calculate its value-at-risk using a shorter observation period if, in the supervisor's judgement, this is justified by a significant upsurge in price volatility.*

This is clearly a minimum standard that forces banks to update their data sets at least monthly. A more frequently updated database will definitely result in better volatility forecasts unlike in the case of historical observation periods where it is unclear what the impact of lengthening/ shortening the historical period will be.

For a bank that uses a variance-covariance approach this standard means that the bank can continue with a fixed volatility forecast for a whole month. Any change in VaR during this interval will be caused entirely by changes in portfolio composition. The volatility forecast will differ on a daily basis in case a conditional volatility model is being used. However, conditional volatility models (ARCH/GARCH) usually incorporate a mean reverting feature that will pull the forecast to a long run value over time, as the forecast moves further into the future.

If the bank does not update the data more frequently than once a month, this standard has a smoothing or slowing effect on the volatility forecast, VaR and market risk charges. According to Jorion (2002a) the aim of standards (d) and (e) is to produce a smooth capital requirement and not necessarily an accurate risk measure. However, this reasoning does not stand in the light of the fact that the calculation of market risk charges itself has a smoothing mechanism inbuilt (see standard (i) in section III). Moreover, the market risk charge may not be a binding constraint for a bank which may have a buffer built into the capital level. Thus, a smooth market risk charge need not be achieved at the cost of accuracy.

An interesting research question that arises is how much does a rolling window add to the accuracy of a model. In other words, is a simple model such as the unconditional historical simulation approach with a rolling window more responsive to recent volatility changes than a sophisticated conditional volatility model with a data set updated once a month? A second research question is to find the trigger point and condition for the supervisory intervention to reduce the length of the observation period in high volatility conditions.

3. Quantitative Standards relating to VaR Estimates

(a) "Value-at-risk" must be computed on a daily basis.

This quantitative standard specifies that VaR should be computed on a daily basis. The VaR will differ from day to day under two circumstances. The first when daily portfolio changes are incorporated in VaR calculations. The decision of whether to use a static portfolio rests with the bank. The second is when VaR is forecast using a rolling window of data or is output from conditional volatility ARCH/GARCH type of models. There is nothing in the standards that force a bank to use a rolling window or conditional volatility models. Hence, the bank can continue to use a static portfolio composition and static volatility forecast till the time it is required to mandatorily update its data set. Effectively this is a redundant standard when combined with the remaining quantitative standards. All it says is that the VaR should be calculated daily. It does not say that the VaR should change daily.

(b) In calculating the value-at-risk, a 99th percentile, one-tailed confidence interval is to be used.

This standard does not give the bank any leeway in choice of confidence level. A higher confidence level would give a higher VaR and market risk charge. However, banks are not allowed to choose a higher (or lower) level than 99%.

The specification of the 99th percentile has ramifications on the performance of the VaR model used. Equity markets have been found to have fat tails with negative skewness compared to the normal (see for example Engle (2001) who took a ten year sample of Nasdaq and Dow Jones daily returns and found that the actual distribution has a 1 percent quantile standard deviation multiplier of 2.844 (compared to the normal distribution's multiplier of 2.326). Hull and White (1998) examine daily exchange rates for 12 major currencies between January 4, 1988 and August 15, 1997 and find that percentage changes of all exchange rates exhibit fatter tails than the normal. What this means is that VaR calculations based on a normal distribution may not perform as well as those based on fat tailed distributions like student's t especially at 99% confidence levels. Also models that avoid distributional assumptions, like the historical simulation method will produce more accurate VaR measures than those based on normal distributions. The relative advantage of the methods that incorporate fat tails may not hold at a lower level of percentile, say 95%. For example, Sheedy (2008) analysed daily return data ranging from 16 to 30 years for five equity indices (S&P500, FTSE100, HSI, Nikkei and ASX200) and reported that the normal VaR models reported higher number of exceptions (performed poorly) compared to the Student's t VaR models for long equity portfolios and 99% confidence levels. This was true of both conditional and unconditional volatility models. However, the unconditional normal performed better than the unconditional Student's t for long equity and 95% confidence levels. Hendricks (1996) compared equally weighted moving average (EQMA), exponentially weighted moving average (EWMA), and historical simulation value-at-risk models using foreign exchange portfolios and found that the EQMA method provides excess coverage for 95% confidence levels while the historical simulation method provides the best conditional coverage (fewest exceptions) at the 99% level.

Thus, a bank with a model based on normal distributional assumption will generally report a lower VaR at the 99% level compared to a bank with a model based on a fat tailed distribution or using a historical simulation method. Since the quantitative standards donot place any restriction on the type of model used to forecast volatility or on the distributional assumption used to generate VaR (see quantitative standard (f)) it appears that banks have an incentive to use variance-covariance or Monte Carlo simulations combined with normal distributional assumptions provided they do not have to pay the penalty for a model that fails the back test.

A number of research questions are raised by this standard. For example, does the normal distributional assumption result in a saving of market risk charges, after accounting for the penalty? What is the trade off between cost of capital in case a model generates higher capital requirement versus regulatory penalties in case a VaR method fails the regulatory back test?

It has also been suggested that a 99% interval is too wide and does not allow the early detection of high volatility conditions. In effect this is another standard that ensures a smoother (though more conservative) VaR measure. Research on this issue is not available.

> *(c) 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 ten trading days. Banks may use value-at-risk numbers calculated according to shorter holding periods scaled up to ten days by the square root of time rule.*

The horizon can be interpreted as the time the portfolio remains frozen. Alternately, it can be interpreted as the time taken to liquidate or hedge the portfolio .

The first observation is that since volatility over longer horizons is higher, the bank may seek to operate with a ten day horizon to economize on capital requirements (unless its portfolio size or composition is changing in a way that lowers capital charges over longer horizons). Secondly, since the VaR using a ten day holding period will usually be calculated assuming that the portfolio remains frozen for ten days, the VaR of a bank with a portfolio that changes on a more frequent (say daily) basis will not be representative of the true VaR. In particular, if the portfolio is growing quickly, the VaR will be underestimated by the assumption of fixed portfolio composition. From the point of view of the regulator this could result in an understatement of VaR and market risk capital and attention will need to be given to the actual turnover level of the portfolio. Thus, a supervisor will have to reconcile the horizon chosen by the bank with the level of portfolio turnover and the time required to actually liquidate the positions.

The second problem with this standard relates to the volatility forecasts over the horizon. Firstly, most research literature on volatility forecasting uses a horizon of a single day. Thus, there is very little evidence on the accuracy of volatility forecasts for horizons longer than a day. Secondly, the accuracy of volatility forecasts is dependant not only on the horizon but also on the model used. Christoffersen and Diebold (2000) analyze daily stock market returns for the U.S. S&P 500, the German DAX, the U.K. FTSE, and the Japanese TPX. They aggregate non-overlapping returns at daily, weekly, two weekly and four weekly levels and use a model free method to assess the forecastability of volatility. They report that at aggregation levels of less than ten trading days (two weeks) volatility is significantly

forecastable. The converse is true for aggregation levels of more than ten days. Thus, a horizon of more than ten days will provide a poor VaR estimate.

While banks can aggregate returns over ten days to calculate VaR, the length of historical time period required to generate acceptable number of non-overlapping ten day samples will be very large. A solution to this problem is to upscale the VaR for a daily holding period by the square root of time rule. Unfortunately this method is not theoretically correct if log price changes follow a GARCH process. Diebold et al (1996) outline that the square root of time rule (for volatility and VaR) holds only when log price changes are independently and identically distributed. However, if they follow a GARCH process, the square root of time rule can be highly misleading. Provizionatou, Markose and Menkens (2005) give alternatives to the rule and find the alternatives outperform the square root of time rule in back testing.

There is a dearth of research on the topic of volatility forecastability over variety of horizons. The methods of measuring forecastability have also not been researched. Most research studies evaluate the accuracy of volatility forecasts from models, making the performance of the forecasts model dependant. At the same time there is little research examining the question of performance of different models over different horizons since most research studies on comparative performance look at horizons of a single day. Also the measures of accuracy of volatility forecasts used by researchers are not the ones used by the back testing framework of the Basel accord for VaR estimates (see standard (j) in section III).

4. Quantitative standards relating to calculation of the market risk capital charge

(i)Each bank must meet, on a daily basis, a capital requirement expressed as the higher of (i) its previous day's value-at-risk number measured according to the parameters specified in this section and (ii) an

average of the daily value-at-risk measures on each of the preceding sixty business days, multiplied by a multiplication factor. [i]In addition, a bank must calculate a 'stressed value-at-risk' measure. This measure is intended to replicate a value-at-risk calculation that would be generated on the bank's current portfolio if the relevant market factors were experiencing a period of stress; and should therefore be based on the 10-day, 99th percentile, one-tailed confidence interval value-at-risk measure of the current portfolio, with model inputs calibrated to historical data from a continuous 12-month period of significant financial stress relevant to the bank's portfolio. The period used must be approved by the supervisor and regularly reviewed. As an example, for many portfolios, a 12-month period relating to significant losses in 2007/2008 would adequately reflect a period of such stress; although other periods relevant to the current portfolio must be considered by the bank.

The aim of the averaging condition in this standard is to generate a smooth capital charge. But combined with the conditions on minimum historical period and frequency of updation this standard can substantially slow the response of capital charges to actual risk requirements.

Jorion (2002a) examined the role of VaR models in increasing the volatility in markets during 1998 when the Russian default triggered heightened volatilty. He examined the VaR generated by exponential weighted moving average (EWMA, $\lambda = 0.992$) and that from equally weighted moving average model (EQMA) with a moving window of 250 trading days. Since the market risk charge itself is calculated as the average of past 60 day VaRs, the capital requirements are smoother compared to the VaR. The final effect of having a minimum "effective" observation period and of averaging for calculation of market risk charge was that in 1998 the increase in the market risk capital charge was 'barely noticeable' for both the EWMA ($\lambda = 0.992$) model and the EQMA with a moving window of 250 trading days. This lead him to conclude that the market risk capital charge (based on VaR calculations) could not have caused an increase in volatility in 1998. It should be noted that the observation of the market risk capital charge being 'barely noticeable' was based on a daily rolling window. A static window or less frequent updation would have generated an even less noticeable response of the charge to volatility spikes.

While this standard provides for the higher of the previous day's VaR and a multiple of the 60 day average to be used for the market charge, the previous day VaR has been found by researchers to be rarely binding. For example, Jorion (2002a) calculates the required change in return measured in volatility multiples needed for the previous day's VaR to be binding, assuming a constant portfolio composition. He finds that a return 32.9 times volatility is required for a EWMA ($\lambda = 0.992$) model and 46.5 times volatility for an EQMA is required for the previous day's VaR to be binding. Effectively this means that the previous day's VaR can be binding only if the portfolio composition changes dramatically (for e.g. a portfolio size growth of three times). Thus, this condition does not appear to be designed to increase the responsiveness of the VaR to sudden spikes in volatility.

However, there are caveats to the research results discussed above. Firstly, emerging market volatility may generate more instances of large single day movements compared to developed market data used for the study, increasing the plausibility of the previous day's VaR being binding. Also the spike in volatility required for the previous day's VaR to be binding drops as the length of the historical observation period shortens. Thus, for some conditional volatility models it is possible that the previous day's VaR may be binding. Both these are researchable questions.

The Basel III revision to this framework adds the requirement of a stressed value-at-risk measure. Since this is an add-on and is based on a stressed scenario, it will more than double the required market charge. This will have a number of consequences for the bank's choice of parameters for its VaR model. For example, a longer historical observation period that includes a stressful event would effectively result in double counting of the stressed value-at-risk charge. Thus, banks will prefer to use the shortest observation period allowed. Conversely, in the immediate aftermath of a stressful scenario a longer period may generate a lower charge. Supervisors may need to keep this in mind while reviewing decisions to change the length of historical observation period.

This also increases the penalty for a model that performs poorly on a back test since the multiplicative factor is applied to both the VaR and the stressed VaR. Thus, the incentive to choose a conservative model that generates fewer exceptions will be higher.

Research is required to examine the impact of the averaging required in this condition combined with the requirements of standards (d) and (e) especially on the responsive of the market risk charge to volatility changes.

> *(j) The multiplication factor will be set by individual supervisory authorities on the basis of their assessment of the quality of the bank's risk management system, subject to an absolute minimum of 3. Banks will be required to add to this factor a "plus" directly related to the ex-post performance of the model, thereby introducing a built in positive incentive to maintain the predictive quality of the model. The plus will range from 0 to 1 based on the outcome of so-called "back testing." If the back testing results are satisfactory and the bank meets all of the qualitative standards, the plus factor could be zero. The back testing results applicable for calculating the plus are based on value-at-risk only and not stressed value-at-risk.*

While rigorous measures of model accuracy are available in case of volatility forecasts (Poon and Granger (2003) survey the volatility forecast evaluation measures and report that the popular ones are *Mean Error* (ME), *Mean Square Error* (MSE), *Root Mean Square Error*

(RMSE), *Mean Absolute Error* (MAE), and *Mean Absolute Percent Error* (MAPE)), the focus of the Basel back testing framework is not the volatility forecast but the VaR. Unfortunately, researchers have focused largely on measuring the volatility forecast accuracy and not much work is available on the performance of VaR measures. VaR measures that have been used by researchers are unconditional and conditional coverage, mean relative bias, root mean squared relative bias, annualized percentage volatility, average multiple of tail event to risk measure, maximum multiple of tail event to risk measure, correlation between risk measure and absolute value of outcome, and capital shortfall (see Hendricks (1996), Jackson et al (1997) and Kupiec (1995)). Tests of the null hypothesis of adequate conditional and unconditional coverage have been suggested by Christoffersen (1998).

The back tests specified by the Basel accord are based on the number of exceptions reported by the VaR models of the banks. Exceptions are the times the trading outcome exceeds the VaR. At a 99% confidence level a model should generate 2 exceptions in a 200 trading day sample. The back test is calibrated to a one day holding period in order to avoid the complications from the changes in portfolio composition over a ten day holding period and has to be carried out on a quarterly basis using the most recent twelve months (250 days) of data. The number of exceptions has been categorized into three zones, namely green, yellow and red. This categorization is based on a balancing of type I and type II errors. Up to four exceptions fall in the green zone. There is no plus factor for models in the green zone. The range from five to nine exceptions constitutes the yellow zone. Outcomes in this range are plausible for both accurate and inaccurate models, and result in a supervisory review of the banks model and a plus factor of between 0.4 and 0.85. Outcomes in the red zone (ten or more exceptions) lead to an automatic rejection of the bank's model and a plus factor of 1.

A basic problem with the back testing approach is the short period of 250 days specified. Jackson et al (1997) report that a parametric VaR models for the actual trading portfolio of a bank moved frequently from the green to the yellow zone from period to period when a back testing window of 250 days was used. Moreover, these tests have low power in small samples such as one year (Campbell, 2007) impairing their ability to reject a flawed VaR model.

While the back testing approach penalizes large number of exceptions, it does not penalize the observation of exceptions in clusters i.e., it is a test for unconditional coverage. The observation of exceptions in clusters should warrant a higher risk charge since it is an additional risk to solvency. None of the three most popular approaches, i.e. the historical simulation, variance-covariance and the Monte Carlo simulation approaches account for volatility clustering and are likely to generate clustered exceptions. Models that account for volatility clustering should not result in exception clustering.

Sheedy (2008) took daily return data ranging from 16 to 30 years for five equity indices (S&P500, FTSE100, HSI, Nikkei and ASX200) to evaluate the conditional coverage of VaRs estimated from conditional and unconditional models. She finds clear evidence that unconditional models fail the test of conditional coverage. The best models are those that combine a heavy tailed distribution (such as historical simulation or Student's t) with a conditional volatility forecast, such as GARCH.

5. Conclusions

This paper evaluates the quantitative standards laid down under the second and third Basel

Accord for the implementation of internal market risk models by banks using available published research. The first observation is the paucity of research that studies the impact of the quantitative standards. For example, research on the performance of volatility forecasts looks at single day horizons while the minimum horizon specified by the quantitative standards is ten days. There is also disconnect between research and industry practices. For example research has focused on sophisticated volatility forecasting methods whereas the model used most by the industry is historical simulation. In summary there is almost no research whatsoever on the implications of the quantitative standards on performance of alternate models.

The evaluation reveals that there is a lack of coherence among the standards owing to the objectives behind the standards being unclear. The specification of a minimum observation period of one year results in a smoother and less responsive VaR estimate and rules out conditional volatility models. After the introduction of a stressed VaR charge that incorporates historical stressful periods and given the averaging built into the calculation of the market risk charge the rationale of continuing with this standard is unclear. Similarly, the reasoning behind the standard of a minimum horizon of ten days is unclear. The standards do not say anything explicitly about the frequency with which banks must update their portfolio compositions. It appears that the minimum horizon allows banks to continue with frozen portfolio compositions for anything between ten days to a month (minimum frequency of data set updation). Similarly, the requirement of minimum monthly updation of data sets results in smoother and slower VaR estimates. The problem is compounded by virtual absence of research on the impact of choice of observation period, horizon and frequency of updation, either independently or in combination, on the performance of alternate market risk models.

Some of the standards laid down contradict the findings of available research. Conditional volatility models are ruled out by the minimum historical period requirement even though they are supported by research. The specification of a minimum horizon of ten days runs contrary to research results that donot find volatility forecastability beyond ten days. The use of the 99th percentile has implications for the distributional assumptions used by banks in light of fat tails observed in financial time series data. The square root of time scaling rule runs contrary to the findings of volatility clustering in research. Lastly, the back testing framework for the VaR models is based on unconditional coverage and may be inadequate given the evidence favouring volatility clustering.

References

Basel committee on Banking Supervision (2006) International Convergence of Capital Measurement and Capital Standards. [Online] Available: http://www.bis.org/publ/bcbs128latest.pdf (November 30, 2011).

Basel Committee on Banking Supervision (2009) Revisions to the Basel II market risk framework. [Online] Available: http://www.bis.org/publ/bcbs158.pdf (November 19, 2011).

Basel Committee on Banking Supervision (2011) Messages from the Academic Literature on Risk Measurement for the Trading Book. Working Paper No. 19, [Online] Available: http://www.bis.org/publ/bcbs_wp19.pdf (November 01, 2011).

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of*

Econometrics. 31(3), 307-328.

Boudoukh, J, M. Richardson, and R. Whitelaw (1998). The Best of Both Worlds. *Risk.* May, 64-67.

Campbell, S.D. (2007). A Review of Backtesting and Backtesting Procedures, *The Journal of Risk.* 9, 2, 1-17. Christoffersen, P and F Diebold. (2000). How relevant is volatility forecasting for financial risk management? *The Review of Economics and Statistics*. 82(1), 12–22.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39: 841-62.

Diebold, Francis X., Andrew Hickman, Atsushi Inoue and Til Schuermann. (1996). Converting 1-Day Volatility to h-Day Volatility: Scaling by √h is Worse than You Think. Wharton Financial Institutions Center, Working Paper No. 97-34.

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987-1008.

Engle, R.F. (2001). GARCH 101: The use of ARCH/GARCH models in Applied Econometrics. *Journal of Economic Perspectives.* 15(4), 157-168.

Hendricks, D. (1996). Evaluation of Value-at-Risk Models Using Historical Data. *FRBNY Economic Policy Review*. 39-69.

Hull, J. and A. White (1998). Value at Risk when Daily Changes in Market Variables are not Normally Distributed. *Journal of Derivatives*. Spring 1998, 9-19.

Jackson, P., Maude, D. and Perraudin, W. (1997). Bank Capital and Value at Risk. *Journal of Derivatives.* 4 (Spring), 73–89.

Jorion, P. (2002a). Fallacies about the effects of market risk management systems. *Financial Stability Review*,.December, 115-127.

Kalvyas, L. Dritsakis, N., Siriopoulos C. and Grose, C. (2004). Selecting Value at Risk Methods According to their Hidden Characteristics. *Operational Research*. 4(2), 167-189.

Kupiec, Paul H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. Board of Governors of the Federal Reserve System, Mimeographed.

Ouyang, Z. (2009). Model Choice and Value at Risk Estimation. *Qual Quant.* 43,983-991.

Perignon, C., & Smith, D. (2006). The Level and Quality of Value-at-Risk Disclosure by Commercial Banks. [Online] Available: http://www.unifr.ch/controlling/seminar/2007-2008/Perignon_ParisII.pdf, (October 28, 2011).

Poon, S. H. and C. Granger (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature.* 41, 478-539.

Provizionatou, V., S. Markose and O. Menkens (2005). Empirical scaling rules for value-at-risk. [Online] Available: http://www.econ.ku.dk/fru/conference/Programme/friday/a4/provizionatou_empirical%20scaling%20rule.pdf, (October 11, 2011).

Senior Supervisors Group (2008). Observations on Risk Management Practices during the Recent Market Turbulence, New York Federal Reserve. [Online] Available: http://newyorkfed.org/newsevents/news/banking/2008/SSG_Risk_Mgt_doc_final.pdf, (October 15, 2011).

Sharma, M. (2012). The Historical Simulation Method for Value-at-Risk: A Research Based Evaluation of the Industry Favorite. [Online] Available: http://dx.doi.org/10.2139/ssrn.2042594 (April 19, 2012).

Sheedy, E. (2008). Why VaR Models Fail and what can be done. Macquarie University Applied Finance Center Research Paper no. 34.

Transition to the Internal Models Approach for Market Risk – a Survey Report (2000). Aptivaa. [Online] Available: www.aptivaa.com/pdfs/SurveyReport.pdf, August 23, 2011.

Vlaar P. (2000). Value at risk models for Dutch bond portfolios. *Journal of Banking and Finance.* 24, 131–154

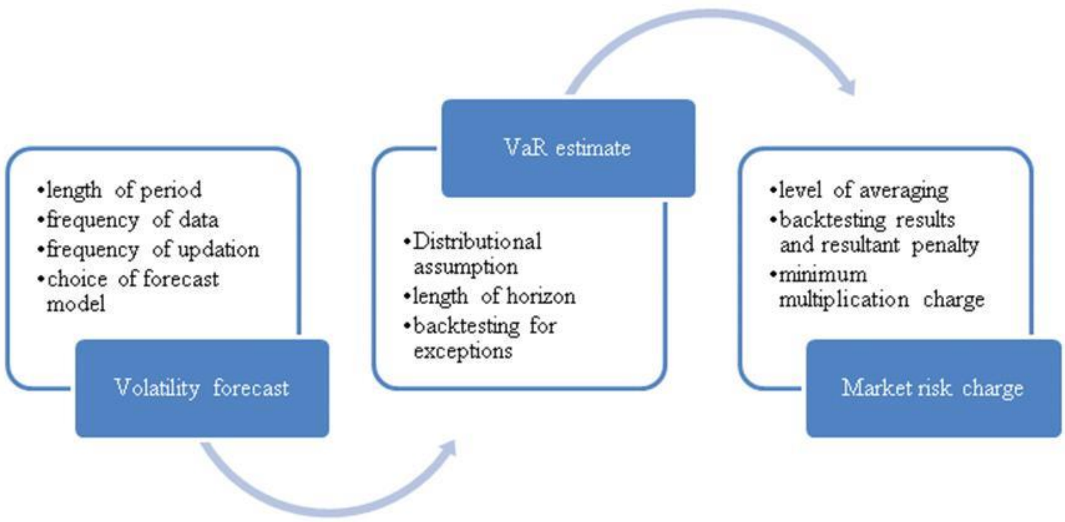

Figure 1 Calculation process of market risk charge