The Risk Map:
A New Tool for Backtesting Value-at-Risk Models

Gilbert Colletaz* Christophe Hurlin* Christophe Pérignon**

March 2011

Abstract

This paper presents a new tool for validating Value-at-Risk (VaR) models. This tool, called the *Risk Map*, jointly accounts for the number and the magnitude of the VaR exceptions and graphically summarizes all information about the performance of a risk model. It relies on the concept of VaR "super exception", which is defined as a situation in which the trading loss exceeds both the standard VaR and a VaR defined at an extremely low coverage probability. We then formally test whether the sequence of exceptions and super exceptions passes standard model validation tests. We show that the Risk Map can be used to backtest VaR for market risk, credit risk, or operational risk, to assess the performance of a margining system on a derivatives exchange, and to validate systemic risk measures (e.g. CoVaR).

JEL classification: G21, G28, G32

Keywords: Financial risk management, Risk model validation, Market risk, Default risk, Systemic risk

* University of Orléans, Laboratoire d'Economie d'Orléans, France. ** HEC Paris, France. We are grateful to Laurent Frésard, Andras Fulop, Bertrand Maillet, and Olivier Scaillet, seminar participants at ESSEC, HEC Paris, and University of Geneva, as well as participants at the 2010 EconomiX Worshop in Financial Econometrics for their comments. Emails: gilbert.colletaz@univ-orleans.fr, christophe.hurlin@univ-orleans.fr, perignon@hec.fr. Contact Author: Christophe Pérignon, Finance Department, HEC Paris, 1 Rue de la Libération, 78351 Jouy-en-Josas, France. Tel: (+33) 139 67 94 11, Fax: (+33) 139 67 70 85.

1 Introduction

The need for sound risk management has never been more essential than in today's financial environment. Of paramount importance for risk managers and regulators is the ability to detect misspecified risk models as they lead to misrepresention of actual risk exposures. In this paper, we focus on the most popular risk measure, the Value-at-Risk (VaR), which is defined as an extreme quantile of a return or profit-and-loss (P&L) distribution. In practice, quantile-based risk measures are used for instance to construct porfolios (Basak and Shapiro, 2001), quantify banks exposures to, and capital requirements for, market, credit, and operational risks (Jorion, 2007), set margin requirements for derivatives positions (Figlewski, 1984), and measure the systemic-risk contribution of a financial institution (Adrian and Brunnermeier, 2010).

In this paper, we present a new tool, called the *Risk Map*, for validating VaR-type models. To grasp the intuition of our approach, consider two banks that both have a one-day Value-at-Risk (VaR) of \$100 million at the 1% probability level. This means that each bank has a one percent chance of losing more than \$100 million over the next day. Assume that, over the past year, each bank has reported three VaR exceptions, or days when the trading loss exceeds its VaR, but the average VaR exceedance is \$1 million for bank A and \$500 million for bank B. In this case, standard backtesting methodologies would indicate that the performance of both models is equal (since both models lead to the same number of exceptions) and acceptable (since the annual number of exceptions is close enough to its target value of 2.5). The reason is that current backtesting methodologies only focus on the number of VaR exceptions and totally disregard the magnitude of these exceptions (Berkowitz, 2001, and Stulz, 2008).

However in practice, market participants do care about the magnitude of VaR exceptions. It is indeed the severity of the trading losses, and not the exceptions per se, that jeopardize banks' solvency. First, risk managers systematically try to learn from past exceptions to improve the forecasting ability of their risk models. Second, investors and other financial statement users assess the risk management skills of a given financial institution based on its backtesting results. Finally, banking regulators may want to penalize more heavily – in terms of capital requirements – a bank that experiences extremely large exceptions than a bank that experiences moderate exceptions. To the best of our knowledge, there is no general hypothesis-testing framework available in the literature that accounts for both the number and the magnitude of VaR exceptions.[1] Our objective is to fill this gap.

The Risk Map approach jointly accounts for the number and the magnitude of the VaR exceptions. The basic intuition is that a large trading loss not only exceeds the regular VaR defined with a probability α (e.g. 1%) but is also likely to exceed a VaR defined with a much lower probability α' (e.g. 0.2%). On this ground, we define a VaR exception as $r_t < -VaR_t(\alpha)$, where r_t denotes the P&L, and a VaR "super exception" as $r_t < -VaR_t(\alpha')$, with α' much smaller than α.[2] We then formally test whether the sequence of exceptions and super exceptions satisfy standard backtesting conditions and we graphically summarize all information about the performance of a risk model in the Risk Map.

There are several advantages to the Risk Map approach. First, it preserves the *simplicity* of the standard validation techniques, such as the unconditional coverage test (Kupiec, 1995) or the z-test (Jorion, 2007, page 144), while also accounting for the magnitude of the losses. Thus, the Risk Map approach is a generalization of the "traffic light" system (Basel Committee on Banking Supervision, 1996) which remains the reference backtest methodology for many national banking regulators. Second, it is a *formal* hypothesis testing framework. Third, the Risk Map approach is very *general* and can be applied to any VaR-type models. For instance, it can be used to backtest the market VaR of a single asset, portfolio, trading desk, business line, bank, mutual fund, or hedge fund (Berkowitz, Christoffersen and Pelletier, 2011). It can also assess the validity of the credit-risk VaR (Lopez and Saidenberg, 2000) or the operational-risk VaR (Dahen and Dionne, 2010) of a financial institution. Furthermore, the Risk Map approach can allow a derivatives exchange to check the validity of its margining system. Finally, we show that the Risk Map can be used to backtest the systemic risk measure recently proposed by Adrian and Brunnermeier (CoVaR, 2010) as it is defined as the conditional quantile of a bank stock return. The Risk Map is, to the best of our knowledge, the first method allowing one to backtest a systemic risk measure.

The outline of the paper is as follows. In the next section, we present our backtesting

[1] Berkowitz (2001) presents a backtesting method that accounts for the magnitude of the VaR exceedance under the normality assumption.

[2] By convention, VaR is expressed as a positive value. When only $VaR_t(\alpha)$ is known, the corresponding time series for $VaR_t(\alpha')$ can be generated by calibration. This is done by extracting the conditional variance of the P&L from $VaR(\alpha)$ and then plugging it into the formula for $VaR(\alpha')$. See Section 2.3 for details.

framework. In Section 3, we present several applications of the Risk Map methodology that sequentially deal with market risk, default risk, and systemic risk. We summarize and conclude our paper in Section 4.

2 Backtesting Framework

2.1 Background

Let r_t denote the return or P&L of a portfolio at time t and $VaR_{t|t-1}(\alpha)$ the ex-ante one-day ahead VaR forecast for an α coverage rate conditionally on an information set $\mathcal{F}_{t-1}$. If the VaR model is adequate, then the following relation must hold:

$$\Pr[r_t < -VaR_{t|t-1}(\alpha)] = \alpha. \tag{1}$$

Let $I_t(\alpha)$ be the hit variable associated with the ex-post observation of a VaR(α) violation at time t:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} . \tag{2}$$

Backtesting procedures are typically only based on the violation process $\{I_t(\alpha)\}_{t=1}^{T}$. As stressed by Christoffersen (1998), VaR forecasts are valid if and only if this violation sequence satisfies the Unconditional Coverage (UC) hypothesis.[3] Under the UC hypothesis, the probability of an ex-post return exceeding the VaR forecast must be equal to the α coverage rate:

$$\Pr[I_t(\alpha) = 1] = \mathbb{E}[I_t(\alpha)] = \alpha. \tag{3}$$

A key limitation of this approach is that it is unable to distinguish between a situation in which losses are below but close to the VaR (e.g. bank A in the introduction) and a situation in which losses are considerably below the VaR (e.g. bank B). A solution proposed

[3] Validation tests are also based on the independence hypothesis (IND), under which VaR violations observed at two different dates for the same coverage rate must be distributed independently. Formally, the variable $I_t(\alpha)$ associated with a VaR violation at time t for a coverage rate α should be independent of the variable $I_{t-k}(\alpha)$, $\forall k \neq 0$. In other words, past VaR violations should not be informative about current and future violations. When the UC and IND hypotheses are simultaneously valid, VaR forecasts are said to have a correct Conditional Coverage, and the VaR violation process is a martingale difference, with $\mathbb{E}[I_t(\alpha) - \alpha | \mathcal{F}_{t-1}] = 0$.

by Lopez (1999a,b) consists in considering the excess losses defined as:

$$L_t(\alpha) = \begin{cases} (r_t + VaR_{t|t-1}(\alpha))^2 & \text{if } r_t < -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} . \tag{4}$$

Lopez proposes various heuristic criteria in order to assess the magnitude of these excess losses. However, such criteria convey only limited information since no normative rule can be deduced for the magnitude of these excess losses.

Another approach is to consider the average loss beyond the VaR using for instance the concept of expected shortfall:

$$ES_t(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} F^{-1}(p)\, dp \tag{5}$$

where $F(.)$ denotes the cumulative distribution function of the P&L (Artzner et al., 1999). The basic idea is that the difference between the loss and the expected shortfall should be zero on days when the VaR is violated. This alternative approach is however difficult to implement in practice since expected shortfalls are typically not disclosed by banks and backtesting only uses the observations where the loss exceeded the VaR, which leads to extremely small samples (Christoffersen, 2009).

Differently, we propose a VaR backtesting methodology that is based on the number and the severity of VaR exceptions, without relying on either ad-hoc loss functions nor expected shortfall. This approach exploits the concept of "super exception", which is defined as a loss greater than $VaR_t(\alpha')$ whereas the coverage probability α' is much smaller than α (e.g. $\alpha = 1\%$ and $\alpha' = 0.2\%$). As an illustration, we show in Figure 1 the joint evolution of the daily P&L, $VaR(\alpha)$, and $VaR(\alpha')$ for a hypothetical bank. We see that, as expected, $-VaR(\alpha')$ is systematically more negative than $-VaR(\alpha)$ as $VaR(\alpha')$ is measured further left in the tail of the P&L distribution. For this bank, we obtain four exceptions (i.e., $r_t < -VaR_{t|t-1}(\alpha)$) and three super exceptions (i.e., $r_t < -VaR_{t|t-1}(\alpha')$).

< Insert Figure 1 here >

One can similarly defined a hit variable associated with $VaR_t(\alpha')$:

$$I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -VaR_{t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \alpha' < \alpha. \tag{6}$$

The defining feature of our approach is to account for both the frequency and the magnitude of trading losses. The intuition of our test is the following. If the frequency of super exceptions is abnormally high, this means that the magnitude of the losses with respect to $VaR_t(\alpha)$ is too large.

For both VaR exceptions and super exceptions, we propose to use a standard backtesting procedure. Consider a time series of T VaR forecasts for an α (respectively α') coverage rate and let N (respectively N') be the number of associated VaR violations:

$$N = \sum_{t=1}^{T} I_t(\alpha) \qquad N' = \sum_{t=1}^{T} I_t(\alpha') \ . \tag{7}$$

If we assume that the $I_t(.)$ variables are *i.i.d.*, then under the UC hypothesis, the total number of VaR exceptions follows a Binomial distribution:

$$N \sim B(T, \alpha) \tag{8}$$

with $E(N) = \alpha T$ and $V(N) = \alpha(1-\alpha)T$. Thus, it is possible to test the UC hypothesis for the VaR expectation as:

$$H_0 : \mathbb{E}\left[I_t(\alpha)\right] = \alpha \tag{9}$$

$$H_1 : \mathbb{E}\left[I_t(\alpha)\right] \neq \alpha. \tag{10}$$

Under H_0, the corresponding log-likelihood ratio statistics is defined as:

$$LR_{UC}(\alpha) = -2\ln\left[(1-\alpha)^{T-N}\alpha^{N}\right] + 2\ln\left[\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right] \xrightarrow[T\to\infty]{d} \chi^2(1). \tag{11}$$

From the critical region, we can deduce conditions on the value of N that allow us not to reject the null of UC. Let us denote l the nominal size of the test, to not reject the null of UC, the number of violations must satisfy the following restriction:

$$\left(1-\frac{N}{T}\right)\ln\left(\frac{1-N}{T-\alpha T}\right) + \frac{N}{T}\ln\left(\frac{N}{\alpha T}\right) > \frac{G^{-1}(1-l)}{2T} \tag{12}$$

where $G(.)$ denotes the c.d.f. of the $\chi^2(1)$ distribution. There is no analytical expression of this condition for N but Jorion (2007) provides numerical values for different values of T and α, for a nominal size of 5%.

A similar validation test can be defined for super exceptions:

$$H_0 : \mathbb{E}\left[I_t\left(\alpha'\right)\right] = \alpha' \tag{13}$$

$$H_1 : \mathbb{E}\left[I_t\left(\alpha'\right)\right] \neq \alpha'. \tag{14}$$

An LR test statistic $LR_{UC}\left(N'\right)$ can be defined as in (11), except that we use both the super exception coverage rate α' and the corresponding number of hits N'.[4]

2.2 The Risk Map

The goal of the Risk Map is to present the backtesting results for a given VaR series in a graphical way. A first way to construct a Risk Map is to jointly display the non-rejection zones for the $LR_{UC}\left(\alpha\right)$ and $LR_{UC}\left(\alpha'\right)$ tests. For instance, we report in Figure 2 the non-rejection zones for $\alpha = 1\%$, $\alpha' = 0.2\%$, and a sample size of $T = 500$. If the number of $VaR\left(1\%\right)$ exceptions is between 2 and 9, we cannot reject the risk model at the 95% confidence level. However, if the number of super exceptions is greater than 3, we reject the validity of the VaR model since it leads to too many super exceptions. It is then possible to put any risk model on the risk map and see which ones lay in the global non-rejection area, i.e., $2 \leqslant N \leqslant 9$ and $N' \leqslant 3$. For instance, we see that the VaR model of the hypothetical bank presented in Figure 1 is rejected at the 95% confidence level for the reason that it leads to too many super exceptions.

< **Insert Figure 2 here** >

Up to now, we have investigated this double validation process (loss frequency + loss magnitude) in a disjointed way. It balls down to consider the $LR_{UC}\left(\alpha\right)$ and $LR_{UC}\left(\alpha'\right)$ tests independently. However, such a multiple testing approach does not allow us to control for the nominal size of the test, i.e., the probability of rejecting a valid model. An alternative approach is to jointly test the number of VaR exceptions and super exceptions:

$$H_0 : \mathbb{E}\left[I_t\left(\alpha\right)\right] = \alpha \text{ and } \mathbb{E}\left[I_t\left(\alpha'\right)\right] = \alpha'. \tag{15}$$

The corresponding test is then directly derived from the test proposed by Pérignon and Smith (2008) and consists in a multivariate unconditional coverage test. Associated with

[4] Alternatively, we could consider a unilateral test in which $H_1 : \mathbb{E}\left[I_t\left(\alpha'\right)\right] > \alpha'$.

both $VaR_{t|t-1}(\alpha)$ and $VaR_{t|t-1}(\alpha')$ is an indicator variable for revenues falling in each disjoint interval:

$$J_{1,t} = I_t(\alpha) - I_t(\alpha') = \begin{cases} 1 & \text{if } -VaR_{t|t-1}(\alpha') < r_t < -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$J_{2,t} = I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -VaR_{t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

and $J_{0,t} = 1 - J_{1,t} - J_{2,t} = 1 - I_t(\alpha)$. The $\{J_{i,t}\}_{i=0}^{2}$ are Bernoulli random variables equal to one with probability $1-\alpha$, $\alpha - \alpha'$, and α', respectively. However they are clearly not independent since only one J variable may be equal to one at any point in time, $\sum_{i=0}^{2} J_{i,t} = 1$. We can test the joint hypothesis (15) of the specification of the VaR model using a simple Likelihood Ratio test. Let us denote $N_{i,t} = \sum_{t=1}^{T} J_{i,t}$, for $i = 0, 1, 2$, the count variable associated with each of the Bernoulli variables. The Pérignon and Smith (2008) multivariate unconditional coverage test is a likelihood ratio test LR_{MUC} that the empirical exception frequencies significantly deviate from the theoretical ones. Formally, it is given by:

$$\begin{aligned} LR_{MUC}(\alpha, \alpha') &= 2\left[N_0 \ln\left(\frac{N_0}{T}\right) + N_1 \ln\left(\frac{N_1}{T}\right) + N_2 \ln\left(\frac{N_2}{T}\right)\right. \\ &\quad \left. - \left[N_0 \ln(1-\alpha) + N_1 \ln(\alpha - \alpha') + N_2 \ln(\alpha')\right]\right] \underset{T\to\infty}{\xrightarrow{d}} \chi^2(2). \end{aligned} \tag{18}$$

Under the null (15) of joint conditional coverage for the VaR exceptions and super exceptions, the LR_{MUC} statistic is asymptotically chi-square with two degrees of freedom. This joint test allows to construct a Risk Map based either on the p-value of the test (Figure 3) or the rejection zone for a given confidence level (Figure 4). The Risk Map in Figure 4 should be read as follows. If the numbers of violations N and super exceptions N' correspond to a green cell, we conclude that we cannot reject the joint hypothesis $\mathbb{E}[I_t(\alpha)] = \alpha$ and $\mathbb{E}[I_t(\alpha')] = \alpha'$ at the 90% confidence level. If the $(N; N')$ pair falls in the yellow zone, we can reject the null at the 90% but not at the 95% confidence level. Similarly, in the orange zone, we can reject the null at the 95% but not at the 99% confidence level. Finally, a red cell implies that we can reject the null hypothesis at the 99% confidence level.

< Insert Figures 3 and 4 here >

2.3 Computation of VaR(α')

In most applications, Risk Map users will have both data on $VaR(\alpha)$ and $VaR(\alpha')$. Nevertheless, a Risk Map can still be generated when only $VaR(\alpha)$, and not $VaR(\alpha')$, is available. This is for instance the case when a bank only discloses its VaR at the 1% level. To overcome this problem, we propose a calibration procedure allowing us to extract $VaR(\alpha')$ from $VaR(\alpha)$, with $\alpha' < \alpha$. The main elements of our procedure are (1) the data generating process (DGP) of the P&L, (2) the internal VaR model used by the bank, and (3) the auxiliary model that we use to generate the $VaR(\alpha')$ estimates.

We assume that the P&L distribution is a member of the location scale family and, for simplicity, that it is centered, i.e., $\mathbb{E}(r_t) = 0$.[5] Under these assumptions, the conditional VaR can be expressed as an affine function of the conditional variance of the P&L, denoted h_t:

$$VaR_{t|t-1}(\alpha;\beta) = \sqrt{h_t} F^{-1}(\alpha;\beta) \tag{19}$$

where $F^{-1}(\alpha;\beta)$ denotes the α-quantile of the conditional standardized P&L distribution. We assume that this distribution is parametric and depends on a set of parameters β. In order to estimate $VaR(\alpha')$, we have to estimate both the distributional parameters β and the conditional variance h_t.

A naive way to compute $VaR(\alpha')$ is to use the QML estimate of the conditional variance and the conditional quantile:

$$VaR_{t|t-1}(\alpha';\widehat{\beta}) = \sqrt{\widehat{h}_t} F^{-1}\left(\alpha';\widehat{\beta}\right) \tag{20}$$

where $\widehat{h}$ and $\widehat{\beta}$ are obtained conditionally to a particular specification of the conditional variance model (auxiliary model). However, in this case, the potential error of specification of the auxiliary model strongly affects $VaR(\alpha')$ through both the conditional variance and the quantile. In practice, the variance effect is likely to be much more pernicious. It is indeed well known in the GARCH literature that conditional volatility forecasts are less sensitive to the choice of the GARCH model than to the choice of the conditional distribution.

Alternatively, we propose an original approach that does not rely on the QML estimate $\widehat{h}_t$.

[5] When $\mathbb{E}(r_t) \neq 0$, the estimated VaR can simply be deduced from (19) by adding the unconditional average return.

Given the VaR forecast reported by the bank $VaR(\alpha)$, we define an implied P&L conditional variance as:

$$\sqrt{\widetilde{h}_t} = \frac{VaR_{t|t-1}(\alpha)}{F^{-1}\left(\alpha; \widehat{\beta}\right)}. \tag{21}$$

We here proceed by analogy with the option pricing literature, in which implied volatility is extracted from option prices (also see Taylor, 2005). $VaR(\alpha')$ is then defined as:

$$VaR_{t|t-1}(\alpha', \widehat{\beta}) = \sqrt{\widetilde{h}_t} F^{-1}\left(\alpha'; \widehat{\beta}\right) = VaR_{t|t-1}(\alpha) \frac{F^{-1}\left(\alpha'; \widehat{\beta}\right)}{F^{-1}\left(\alpha; \widehat{\beta}\right)}. \tag{22}$$

Interestingly, we notice that the auxiliary model is only used to get an estimate of the conditional quantile $F^{-1}\left(\alpha; \widehat{\beta}\right)$. As a result, this process mitigates as much as possible the impact of a misspecification of the auxiliary model on the VaR estimates. Implementing this calibration method requires two ingredients: (1) an auxiliary model for the conditional volatility h_t, such as a GARCH or stochastic volatility model and (2) a conditional distribution for the P&L which depends on a set of parameters β.

In the Appendix, we use a Monte Carlo study to assess the empirical performance of our calibration procedure. We use a GARCH model as the auxiliary model and a t-distribution for F. In that case, the set of parameters β simply corresponds to the degree of freedom of the t-distribution, which can be estimated by QML. Overall, we find that this calibration procedure leads to reliable estimates for $VaR(\alpha')$.

3 The Risk Map at Work

3.1 Market Risk

A natural application of the Risk Map is to backtest the VaR of a bank. It is indeed key for both risk managers and banking regulators to check the validity of banks' VaR engines as market risk charges, and in turn capital requirements, depend on banks' VaR internal estimates (Jorion, 2007). Berkowitz and O'Brien (2002) show that VaR estimates of leading US banks tended to be conservative during the late nineties. Using P&L data from four business lines in a large international commercial bank, Berkowitz, Christoffersen, and Pelletier (2011) find evidence of volatility dynamics and clustering in VaR exceptions. However, none of the prior empirical literature accounts for the magnitude of the VaR exceedances.

We use actual VaR and P&L for a large European bank, namely La Caixa, which is the third largest Spanish bank. We use daily one-day ahead VaR(1%) and daily P&L for that bank over the period 2007-2008. In line with the recommendations of the Basel Committee on Banking Supervision (2009), the P&L data does not include fees and commissions, neither intraday trading revenues (Frésard, Pérignon, and Wilhelmsson, 2011). We plot in the upper part of Figure 5, the bank VaR along with the actual P&L. As our sample period includes the beginning of the recent financial crisis, there is a clear regime shift in the variability of the trading revenues. We see that the volatility spiked after the end of the first semester of 2007. Similarly, the VaR(1%) jumped from around 2 millions euros during 2007-Q1-Q2 to around 4 millions afterwards.

We extract the series for VaR(0.2%) from the VaR(1%) series using the calibration procedure presented in Section 2.3. Over this sample period, there were 13 VaR exceptions ($N = 13$) and three super exceptions ($N' = 3$). The numbers of violations define the coordinates of the point associated with the risk model used by the bank in our Risk Map representation. The lower part of Figure 5 displays the point associated with La Caixa on the Risk Map (blue cross at (3;13)). We conclude that we can reject the validity of the VaR model of the bank since the observation falls outside the non-rejection zone colored in green. The corresponding p-value of the Pérignon and Smith (2008) test is equal to 0.0108.

3.2 Default Risk

VaR models are also used to set margin and collateral requirements on derivatives markets. Margins are key to protect derivatives users against the default of their counterparties. The difficult trade-off faced by the derivative exchange is to set margins high enough to mitigate default risk but not too high since this would shy traders away and damage liquidity.

The initial margin C for one futures contract (long or short position) must be set so that the probability of a futures price F_t change, $r_t = F_t - F_{t-1}$, exceeding the margin is equal to a prespecified level:

$$\Pr[r_t < -C_{t|t-1}(\alpha)] = \Pr[r_t > C_{t|t-1}(\alpha)] = \alpha. \tag{23}$$

Depending on the expected volatility, the derivatives exchange frequently adjusts the level of the margin, as shown by Brunnermeier and Pedersen (2009, Figure 1) for the S&P 500 futures. The empirical literature has considered a variety of distributions for futures price changes and volatility dynamics (Booth et al., 1997, Cotter, 2001). The Risk Map approach can be used to test whether actual margins or optimal margins according to a given modelling technique generate too many margin exceedances and too many margin "super exceedances". The analysis would have to be conducted separately for the left and right tails.

A related problem is the determination of the margin requirements for the clearing members of a given clearing house. In this case again, VaR models are one of the two main techniques used to set margins – the other one being the SPAN system. The issue of validating clearing members' margins takes nowadays central stage as clearing houses have moved to clear new products that used to be traded over-the-counter (OTC) (Duffie and Zhu, 2010).

Let $\omega_{i,t-1}$ being the vector of positions of clearing member i at the end of day $t-1$:

$$\omega_{i,t-1} = \begin{bmatrix} \omega_{i,1,t-1} \\ \vdots \\ \omega_{i,D,t-1} \end{bmatrix} \tag{24}$$

where D is the number of derivatives contracts (futures and options) traded on this exchange and $i = 1, ..., N$. We assume that these contracts are written on U different underlying assets. To arrive at a margin for this portfolio, the clearing house considers a series of S scenarios representing potential changes in the level and volatility of the underlying assets. For each scenario, the value of the portfolio is recomputed, or marked-to-model, using futures and option pricing formulas, and the associated hypothetical P&L is computed:

$$\widetilde{r}_{i,t} = \begin{bmatrix} \widetilde{r}_{i,t}^{1} \\ \vdots \\ \widetilde{r}_{i,t}^{S} \end{bmatrix}. \tag{25}$$

Notice that this simulation-based technique allows the clearing house to account for diversification among underlying assets and maturities, which reduces collateral requirements. From this simulated distribution of P&L, the clearing house can set the margins for clearing member i such that:

$$\Pr[\widetilde{r}_{i,t} < -C_{i,t|t-1}(\alpha)] = \alpha. \tag{26}$$

The clearing house will proceed in the same way for the $N-1$ other clearing members and only those who will be able to pile up this amount of collateral on their margin accounts will be allowed to trade on the next day. On a regularly basis, the risk-management department of the clearing-house and the regulatory agencies check the validity of the margining system. In particular, one needs to check whether the hypothetical shocks used in the scenarios are extreme enough or whether the estimation of the derivative prices is reliable. Of particular concern is a situation in which the collateral is set at too low a level. In this case, a default by a clearing member following a big trading loss would lead to a massive shortfall, which may propagate default within the clearing system (Eisenberg and Noe, 2001). The evaluation of the margining system can be conducted using the Risk Map approach. In this particular case, the analysis can be conducted by clearing member or for all the clearing members pooled together.

It is interesting to notice that the setting above also applies to OTC derivatives trading. Indeed, OTC market participants also require collateral from their counterparties in order to mitigate default risk exposure. Consider the case of a prime broker that needs to decide on the amount of cash and securities that a given hedge fund needs to maintain as collateral. The collateral requirement depends on the outstanding positions of the hedge fund with respect to the prime broker (e.g. short positions in exotic options and some interest rate swaps). Just like in the clearing house example, the prime broker can apply a set of scenarios to the hedge fund's positions to generate a distribution of hypothetical P&L, and set the collateral as a quantile of this distribution, as in (26). In this case too, a Risk Map could be used to backtest the collateral requested by this prime broker.

3.3 Systemic Risk

Since the recent financial crisis, the quest for measuring, and forecasting, systemic risk has never been as popular. Of particular importance is the quantification of the marginal contribution of systemically important financial institutions to the overall risk of the system. Acharya et al. (2009) suggest that each firm should pay for its own systemic risk contribution through capital requirements, taxes, and required purchase of insurance against aggregate risk (also see Brunnermeier et al. 2009). While many methodologies have been recently proposed to measure systemic risk (Engle and Brownless, 2010; Acharya et al. 2010; Adrian et Brunnermeier, 2010), there is to the best of our knowledge no ex-post validation methods for systemic risk measures. In this section, we show that the Risk Map approach can be used to backtest systemic risk measures.

We follow Adrian et Brunnermeier (2010) and define the CoVaR measure as the VaR of the financial system conditional on institutions being under distress. Formally, CoVaR is defined as the α-quantile of the conditional probability distribution of the financial system returns r_j:

$$\Pr\left[r_j \leq -CoVaR^{j|\mathbb{C}(r_i)}(\alpha) \;\middle|\; \mathbb{C}(r_i) \right] = \alpha \tag{27}$$

where $\mathbb{C}(r_i)$ denotes a conditioning event concerning institution i. For technical reasons (CoVaR is estimated by quantile regression), the authors consider the conditioning event $\{r_i = -VaR^i(\alpha)\}$. However, CoVaR can also be defined using $\{r_i \leq -VaR^i(\alpha)\}$ as the conditioning event. In that case, financial distress for institution i is defined as a situation in which the losses exceed the VaR and the definition of CoVaR becomes:

$$\Pr\left[r_j \leq -CoVaR^{j|i}(\alpha) \;\middle|\; r_i \leq -VaR^i(\alpha) \right] = \alpha. \tag{28}$$

Given this definition, it is obvious that CoVaR can be backtested within the Risk Map framework. Just like with VaR, we need to analyse the frequency of the conditional probability and the magnitude of the losses in excess of the CoVaR. The later will provide us with some crucial information about the resiliency of the financial system when a particular firm is in financial distress. Note also that the loss in excess of the CoVaR can be studied using the concept of co-expected shortfall (Adrian and Brunnermeier, 2010), which is the expected shortfall of the system conditional on $\mathbb{C}(r_i)$. However, backtesting (co-)expected shortfalls remains very difficult in practice (see Section 2.1).

We suggest using the Risk Map framework to backtest the CoVaR in both dimensions (i.e., number and severity of CoVaR exceptions). We define a CoVaR exception and super exception as $r_j < -CoVaR^{j|i}(\alpha)$ and $r_j < -CoVaR^{j|i}(\alpha')$, respectively. By analogy with VaR, we define the following hit variables:

$$I^j(q) = \begin{cases} 1 & \text{if } r_j < -CoVaR^{j|i}(q) \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } q = \alpha \text{ or } q = \alpha'. \tag{29}$$

with $\alpha' < \alpha$. By definition of the CoVaR exception (and super exception), we have:

$$\mathbb{E}\left[I^j(q) \mid r_i \leq -VaR^i(q)\right] = q. \tag{30}$$

It is possible to transform the conditional expectation (30) into an unconditional one, since the condition (30) implies $\mathbb{E}\left[I^j(q) \times I^i(q)\right] = q$ for $q = \alpha$ or $q = \alpha'$, where $I_t^i(\alpha)$ the standard VaR(α) hit variable for the institution i:

$$I^i(\alpha) = \begin{cases} 1 & \text{if } r_i < -VaR^i(\alpha) \\ 0 & \text{otherwise} \end{cases}. \tag{31}$$

So, the CoVaR-Risk Map can be defined as the non-rejection area of the joint test:

$$H_0 : \mathbb{E}\left[I^j(\alpha) \times I^i(\alpha)\right] = \alpha \text{ and } \mathbb{E}\left[I^j(\alpha') \times I^i(\alpha)\right] = \alpha'. \tag{32}$$

Under the null, conditionally on the distress of financial institution i, the probability to observe a loss in the financial system larger than the CoVaR(α) is precisely equal to α, and the probability to observe an "super loss" should not exceed α'.

Then, the CoVaR Risk Map is similar to the Risk Map for VaR. Given a sequence of estimated conditional CoVaRs for the system, $\left\{CoVaR_{t|t-1}^{j|i}(q)\right\}_{t=1}^{T}$ for $q = \alpha$ and $q = \alpha'$, we compute the sequences of hits $\left\{I_t^j(q)\right\}$. The estimated CoVaR can simply be derived from a a multivariate GARCH model (e.g. DCC) or directly estimated from a quantile regression. The corresponding test statistic can then directly be derived from the test proposed by Pérignon and Smith (2008), $LR_{MUC}(\alpha, \alpha')$ shown in (18). In this case, $N_1 = \sum_{t=1}^{T}\left[I_t^j(\alpha) - I_t^j(\alpha')\right] \times I_t^i(\alpha)$, $N_2 = \sum_{t=1}^{T} I_t^j(\alpha') \times I_t^i(\alpha)$, and $N_0 = T - N_1 - N_2$. The non rejection area of the test can be represented as in Figure 2. For a given financial institution i, the numbers of CoVaR exceptions and CoVaR super exceptions correspond to one particular cell on the Risk Map. It produces a direct diagnostic about the validity of the systemic risk measure, which jointly accounts for the number and the magnitude of the exceptions.

Finally, given the CoVaR definition, it is possible to compute the difference between (1) the VaR of the financial system conditional on the distress of a particular financial institution i and (2) the VaR of the financial system conditional on the median state of institution i. This difference corresponds to the $\Delta CoVaR$, defined as follows:

$$\Delta CoVaR^{j|\mathbb{C}(r_i)}(\alpha) = CoVaR^{j|r_i \leq -VaR^i(\alpha)}(\alpha) - CoVaR^{j|Median^i}(\alpha) \tag{33}$$

The measure $\Delta CoVaR^{j|i}$ quantifies how much an institution contributes to the overall systemic risk. Backtesting $\Delta CoVaR^{j|i}$ can be achieved by applying the Risk Map methodology successively to $CoVaR^{j|r_i \leq -VaR^i(\alpha)}$ and $CoVaR^{j|Median^i}$.

4 Conclusion

In this paper, we have proposed a formal backtesting procedure allowing bank risk managers and regulators to assess the validity of a risk model by accounting for both the number and the magnitude of VaR exceptions. We have introduced the concept of VaR super exceptions and derived a test that combine information about both VaR exceptions and super exceptions. We have shown that the Risk Map framework can be handy in validating market, credit, or operational VaRs, margin requirements for derivatives, or systemic risk measures such as the CoVaR.

The Risk Map approach may prove effective in banking regulation. Indeed, as it is a generalization of the the system currently used by banking regulators ("traffic light" system) to validate banks' risk models, the Risk Map could help detecting misspecified risk models and penalize banks that experience VaR exceptions that are too frequent and too large. In this case, bank capital requirements would be affected by the conclusions of the Risk Map analysis.

Our analysis could be extended in different directions. First, we could implement our methodology with other statistical tests, such as the hit regression test of Engle and Manganelli (2004). Second, we could take into account the model risk specification induced by the estimation of VaR super exceptions on the finite sample and asymptotic distribution of the test statistic (Escanciano and Olmo, 2010a,b).

Appendix: Monte Carlo Experiments

We assess the accuracy of the calibration procedure for $VaR(\alpha')$ presented in Section 2.3 in a series of controlled experiments. The aim of these experiments is to systematically compare the estimated $VaR(\alpha')$ to the true $VaR(\alpha')$. To conduct the experiments, we need to specify (1) the data-generating process (DGP) of the P&L, (2) the internal VaR model used by the bank, and (3) the auxiliary model that we use to generate the VaR estimates. In particular, we check whether our approach is able to accurately estimate $VaR(\alpha')$ when (1), (2) and/or (3) are misspecified.

For the DGP of the P&L, we follow Berkowitz, Christoffersen and Pelletier (2011) and assume that returns r_t are issued from a $t\,(v)$-GARCH$(1,1)$ model:

$$r_t = \sigma_t \; z_t \; \sqrt{\frac{v-2}{v}} \tag{A1}$$

where $\{z_t\}$ is an *i.i.d.* sequence form a Student's t-distribution with v degrees of freedom and where conditional variance is:

$$\sigma_t^2 = \omega + \gamma \; \left(\frac{v-2}{v} \right) z_{t-1}^2 \sigma_{t-1}^2 + \pi \; \sigma_{t-1}^2. \tag{A2}$$

Parameterization of the coefficients and initial condition are deduced from maximum-likelihood estimated parameters for the S&P 500 index daily returns over the period 02/01/1970 to 05/05/2006. The parameter values are $\omega = 7.977e^{-7}$, $\gamma = 0.0896$, $\pi = 0.9098$, and $v = 6$. The initial condition σ_1^2 is set to the unconditional variance.

Using the simulated P&L distribution issued from this DGP, it is then necessary to select a method to forecast the VaR. This method represents the internal VaR model used by the financial institution. We first consider a VaR calculation method that perfectly matches the P&L distribution and therefore it satisfies unconditional coverage for both standard and super exceptions. In a second experiment, we use a method that induces a violation of unconditional coverage for super exceptions: a method that is valid with respect to the current backtesting procedures, but that generates too many extreme losses (see Panel A of Table A1). In this case, we use the Historical Simulation (HS) method.

< **Insert Table A1** >

Recall that the aim of the experiments is to asses the capacity of our method to estimate the $VaR(\alpha')$ produced by the internal model of the bank. For that, it is possible to compare directly the "true" VaR, denoted $VaR^0_{t|t-1}(\alpha')$, to the estimated one, denoted $VaR_{t|t-1}(\alpha', \widehat{\beta})$. However, what is most important is not comparing the VaRs, but the exceptions induced by these VaRs, since backtesting is based on exceptions. Ideally, the timing of the super exceptions should be the same with the true and the estimated $VaR(\alpha')$.

We denote $I_t^0(\alpha')$ the "true" hit process associated to $VaR^0_{t|t-1}(\alpha')$ and $I_t\left(\alpha', \widehat{\beta}\right)$ the estimated hit process associated to $VaR_{t|t-1}(\alpha', \widehat{\beta})$. As usual, two types of errors can occur: the false positive one (type 1 error) and the false negative one (type 2 error). We evaluate the quality of our approach by using three indicators. The True Positive Rate (TPR), also called hit rate or sensitivity, denotes the frequency of having a super exception which occurs concurrently for the true and for the estimated $VaR(\alpha')$. More formally, if we denote T the out-of-sample size of the sample, we define:

$$TPR = \frac{\sum_{t=1}^{T} I_t^0(\alpha') \times I_t\left(\alpha', \widehat{\beta}\right)}{\sum_{t=1}^{T} I_t^0(\alpha')}. \tag{A3}$$

The False Alarm Rate (FPR), also called fall out, represents the frequency of type 1 error. It gives the fraction of estimated super exceptions events that were observed to be non events (false alarms).

$$FPR = \frac{\sum_{t=1}^{T} \left[1 - I_t^0(\alpha')\right] \times I_t\left(\alpha', \widehat{\beta}\right)}{\sum_{t=1}^{T} \left[1 - I_t^0(\alpha')\right]}. \tag{A4}$$

Finaly, we consider the True Negative Rate (TNR), also called specificity. It measures the fraction of non super VaR exceptions which are correctly identified (complementary of the type 2 error).

$$TNR = \frac{\sum_{t=1}^{T} \left[1 - I_t^0(\alpha')\right] \times \left[1 - I_t\left(\alpha', \widehat{\beta}\right)\right]}{\sum_{t=1}^{T} \left[1 - I_t^0(\alpha')\right]}. \tag{A5}$$

Experiment 1: Valid Internal VaR Model

In the first experiment, the internal risk model of the bank corresponds to the true DGP. In this context, the true conditional VaR is defined as:

$$VaR^0_{t|t-1}(\alpha) = \sqrt{\frac{\beta - 2}{\beta}}\, \sigma_t\ F^{-1}(\alpha; \beta) \tag{A6}$$

where $F(.;\beta)$ denotes the c.d.f. of the $t(v)$ distribution. Let us denote $VaR^0_{t|t-1}(\alpha')$ the true conditional VaR (unobservable) for the coverage rate α'. We consider the case where $\alpha = 1\%$ and $\alpha' = 0.2\%$. Theoretically, these VaR forecasts are deduced from a valid internal model and consequently the unconditional coverage assumption is satisfied for all coverage rates.

Given the simulated returns path and the VaR displayed by the bank, $\left\{VaR^0_{t|t-1}(\alpha)\right\}_{t=1}^{T}$, we apply our calibration procedure to estimate the VaR for a coverage rate α'. In this first experiment, we consider an auxiliary model also defined as a $t\,(\beta)$-GARCH$(1,1)$ model:

$$r_t = \sigma_t \; z_t \; \sqrt{\frac{\beta - 2}{\beta}} \tag{A7}$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 \left(\frac{\beta - 2}{\beta} \right) z_{t-1}^2 \sigma_{t-1}^2 + \gamma_2 \sigma_{t-1}^2. \tag{A8}$$

The parameters of this auxiliary model are estimated by QML. Let us denote $\widehat{\beta}$ the estimator of the distributional parameter (i.e., the degree of freedom of the t-distribution). Conditionally on this estimated parameter and to the $VaR_{t|t-1}^0(\alpha)$ produced by the bank, we can define the estimated $VaR(\alpha')$ as:

$$VaR_{t|t-1}(\alpha', \widehat{\beta}) = VaR_{t|t-1}^0(\alpha) \frac{F^{-1}\left(\alpha'; \widehat{\beta}\right)}{F^{-1}\left(\alpha; \widehat{\beta}\right)}. \tag{A9}$$

In Panel B of Table A1, the three indicators of the quality of the calibration process are reported. In the second column, the frequency of super VaR exceptions obtained from our estimated 0.2% VaR is reported for various sample sizes. Recall that in this experiment, the unconditional coverage is valid (since the VaR model corresponds to the true P&L DGP), and the super-exception frequency should be equal to 0.2% if the calibration method correctly estimates the VaR. The third, fourth, and fifth columns, respectively, report the True Positive Rate, the False Positive Rate, and the True Negative Rate obtained from 10,000 simulations. Overall, Panel B shows that the performance of our estimation method is very good.

Experiment 2: Invalid Internal VaR Model

In the second experiment, we introduce a discrepancy between the internal risk model used by the bank to generate the VaR figures and the auxiliary model used in the calibration procedure. More specifically, we assume that the bank uses historical simulation (HS) whereas the auxiliary model is a t-GARCH(1,1) model. We see in Panel C of Table 3 that even in a situation in which the auxiliary model is wrong, our calibration procedure allows us to extract a reliable estimate of $VaR\,(\alpha')$.

References

[1] Acharya, V. V., L. Pedersen, T. Philippon, M. Richardson (2009), "Regulating Systemic Risk, in Restoring Financial Stability: How to Repair a Failed System," in *Restoring Financial Stability: How to Repair a Failed System*, V. Acharya and M. Richardson (Editors), John Wiley & Sons.

[2] Acharya, V. V., L. Pedersen, T. Philippon, M. Richardson (2010), "Measuring Systemic Risk," Working Paper, New York University.

[3] Adrian, T., M. K. Brunnermeier (2010), "CoVaR," Working Paper, Princeton University and FRB of New York.

[4] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 203-228.

[5] Basel Committee on Banking Supervision (1996), "Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements," Bank for International Settlements.

[6] Basel Committee on Banking Supervision (2009), "Revisions to the Basel II Market Risk Framework," Bank for International Settlements.

[7] Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465-474.

[8] Berkowitz, J., P. F. Christoffersen, D. Pelletier (2011), "Evaluating Value-at-Risk Models with Desk-Level Data," *Management Science*, forthcoming.

[9] Berkowitz, J., J. O'Brien (2002), "How Accurate Are Value-At-Risk Models at Commercial Banks?," *Journal of Finance*, 57, 1093-1111.

[10] Brunnermeier, M. K., A. Crocket, C. Goodhart, A. Perssaud, H. Shin (2009), The Fundamental Principals of Financial Regulation, 11th Geneva Report on the World Economy.

[11] Brunnermeier, M. K., L. H. Pedersen (2009), "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201-2238.

[12] Booth, G. G., J. P. Broussard, T. Martikainen, V. Puttonen (1997), "Prudent Margin Levels in the Finnish Stock Index Futures Market," *Management Science*, 43, 1177-1188.

[13] Christoffersen P. F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.

[14] Christoffersen, P. F. (2009), "Backtesting," in Encyclopedia of Quantitative Finance, R. Cont (ed.), Wiley.

[15] Cotter, J. (2001), "Margin Exceedences for European Stock Index Futures using Extreme Value Theory," *Journal of Banking and Finance*, 25, 475-502.

[16] Cotter, J., K. Dowd (2006), "Extreme Spectral Risk Measures: An Application to Futures Clearinghouse Margin Requirements," *Journal of Banking and Finance*, 30, 3469-3485.

[17] Dahen, H., G. Dionne (2010), "Scaling Models for the Severity and Frequency of External Operational Loss Data," *Journal of Banking and Finance*, 34, 1484-1496.

[18] Duffie, D., H. Zhu (2010), "Does a Central Clearing Counterparty Reduce Counterparty Risk?," Working Paper, Stanford University.

[19] Engle, R. F., S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367-381.

[20] Escanciano, J. C., J. Olmo (2010a), "Backtesting Parametric Value-at-Risk with Estimation Risk," *Journal of Business and Economics Statistics*, 28, 36-51.

[21] Escanciano, J. C., J. Olmo (2010b), "Robust Backtesting Tests for Value-at-Risk Models," *Journal of Financial Econometrics*, forthcoming.

[22] Eisenberg, L., T. H. Noe (2001), "Systemic Risk in Financial Systems," *Management Science*, 47, 236-249.

[23] Figlewski, S. (1984), "Margins and Market Integrity: Margin Setting for Stock Index Futures and Options," *Journal of Futures Markets*, 4, 385-414.

[24] Frésard, L., C. Pérignon, A. Wilhelmsson (2011), "The Pernicious Effects of Contaminated Data in Risk Management," *Journal of Banking and Finance*, forthcoming.

[25] Gaglianone, W. P., L. R. Lima, O. Linton, D. R. Smith (2010), "Evaluating Value-at-Risk Models via Quantile Regression," *Journal of Business and Economic Statistics*, forthcoming.

[26] Gay, G. D., W. C. Hunter, R. W. Kolb (1986), "A Comparative Analysis of Futures Contract Margins," *Journal of Futures Markets*, 6, 307-324.

[27] Jorion, P. (2007), *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, 3rd Edition.

[28] Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73-84.

[29] Lopez, J. A. (1999a), "Regulatory Evaluation of Value-at-Risk Models," *Journal of Risk*, 1, 37-64.

[30] Lopez, J. A. (1999b), "Methods for Evaluating Value-at-Risk Models," *Federal Reserve Bank of San Francisco Economic Review*, 2, 3-17.

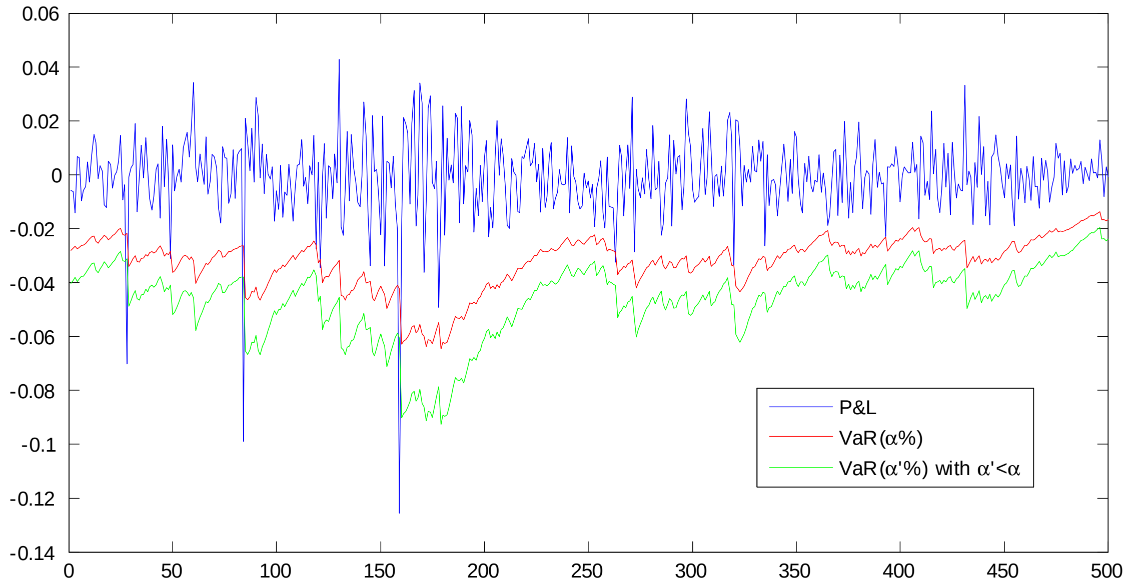
[31] Lopez, J. A., M. R. Saidenberg (2000), "Evaluating Credit Risk Models," *Journal of Banking and Finance*, 24, 151-167.

[32] Pérignon, C., D. R. Smith (2008), "A New Approach to Comparing VaR Estimation Methods," *Journal of Derivatives*, 16, 54-66.

[33] Stulz, R. (2008), "Risk Management Failures: What Are They and When Do they Happen?," *Journal of Applied Corporate Finance*, 20, 39-48.

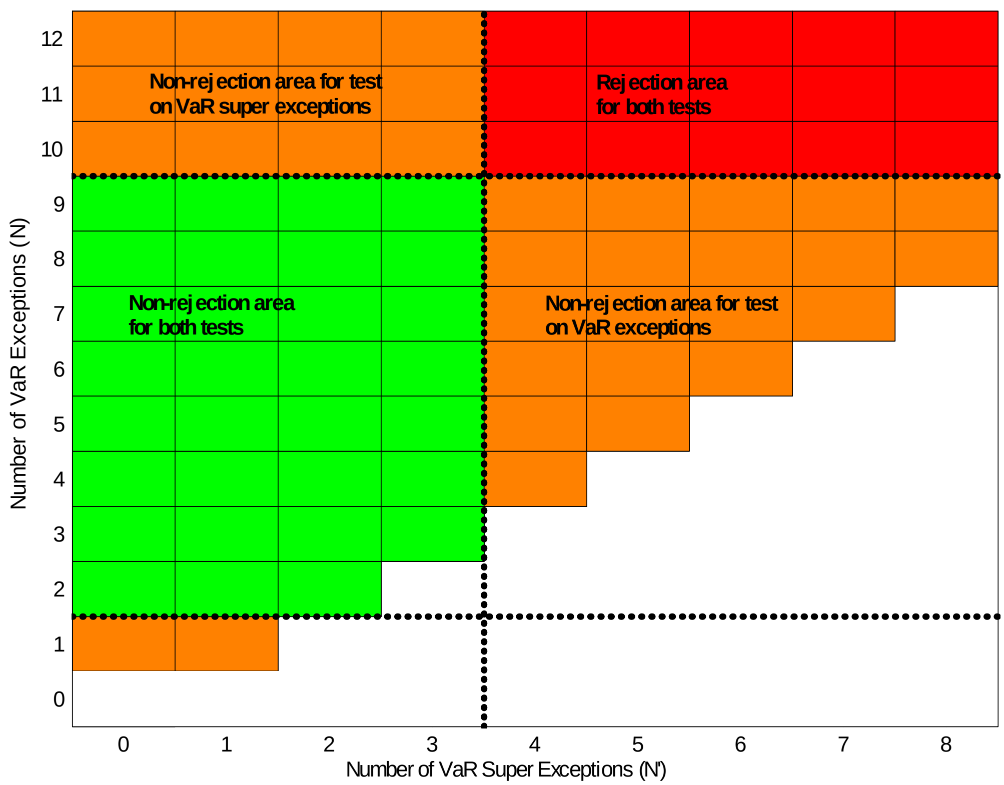
[34] Taylor, J. W. (2005), "Generating Volatility Forecasts from Value at Risk Estimates," *Management Science*, 51, 712-725.

Figure 1: **VaR Exceptions vs. VaR Super Exceptions**



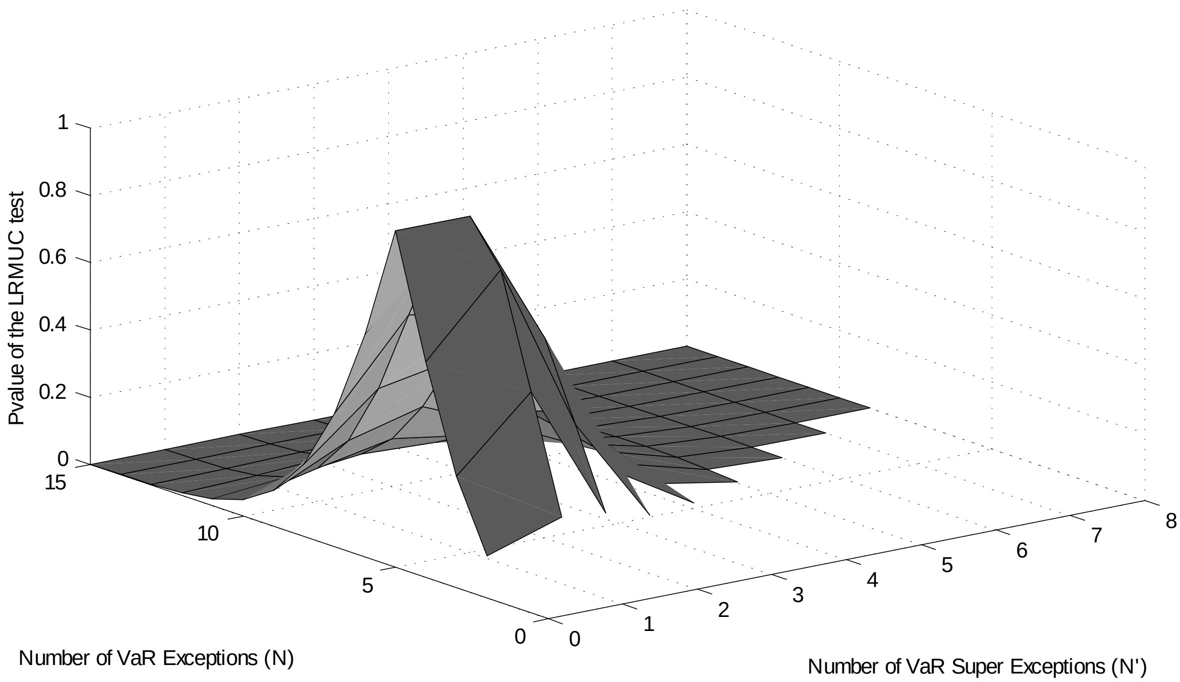
Notes: This figure displays the daily P&L, $VaR\,(\alpha)$, and $VaR\,(\alpha')$ for a hypothetical bank. By definition, the coverage rate α' is smaller than α. Both the P&L and VaR series are simulated using a t-Garch model. A VaR exception is defined as $r_t < -VaR_{t|t-1}(\alpha)$ whereas a super exception is defined as $r_t < -VaR_{t|t-1}(\alpha')$. Over this 500-day sample period, there are four exceptions and three super exceptions.

Figure 2: **Risk Map based on Two Univariate Coverage Tests**



Notes: This figure displays a Risk Map based on two univariate unconditional coverage tests, $LR_{UC}(\alpha)$ and $LR_{UC}(\alpha')$. If the number of $VaR(1\%)$ exceptions N is between 2 and 9, we cannot reject the risk model at the 95% confidence level, and if the number of $VaR(0.2\%)$ exceptions N' is strictly less than 4, we cannot reject the risk model at the 95% confidence level. Thus, the global non-rejection area corresponds to $2 \leqslant N \leqslant 9$ and $N' \leqslant 3$.

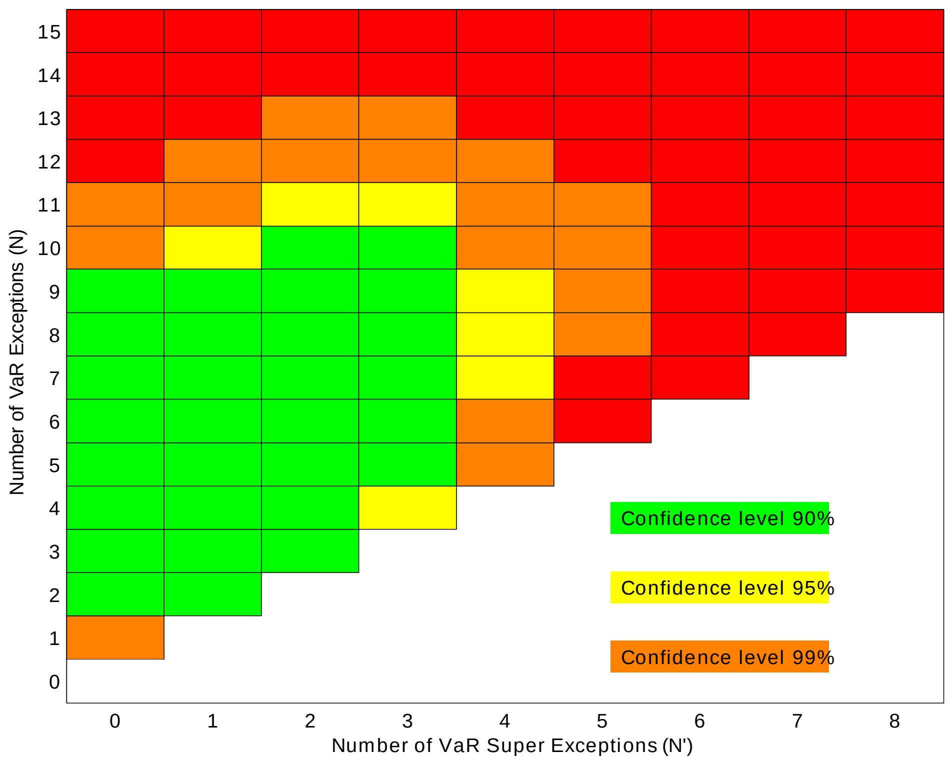
Figure 3: **P-values of the Multivariate Unconditional Coverage Test**



2

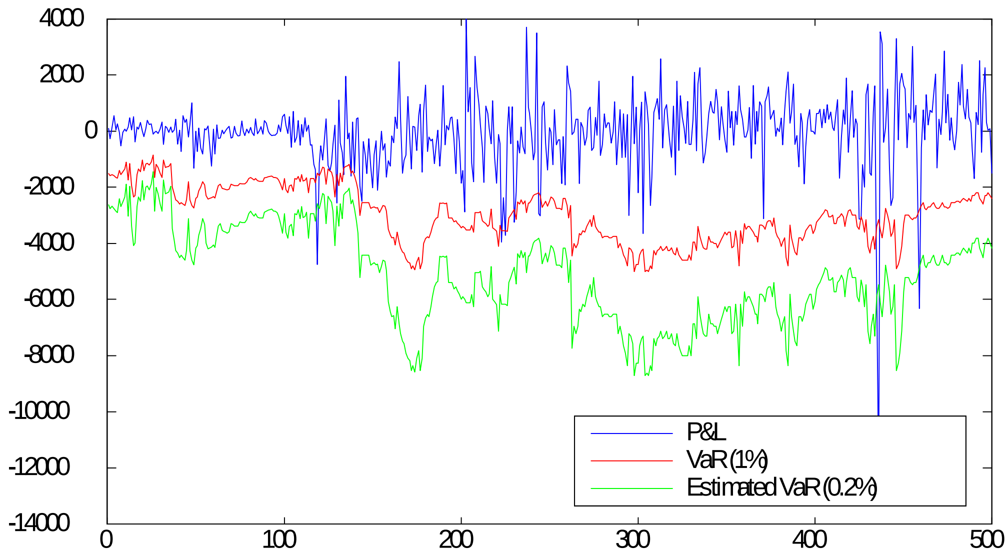
Notes: This figure displays the p-value of a multivariate unconditional coverage tests, $LR_{MUC}(\alpha, \alpha')$ for different numbers of exceptions (N) and super exceptions (N').

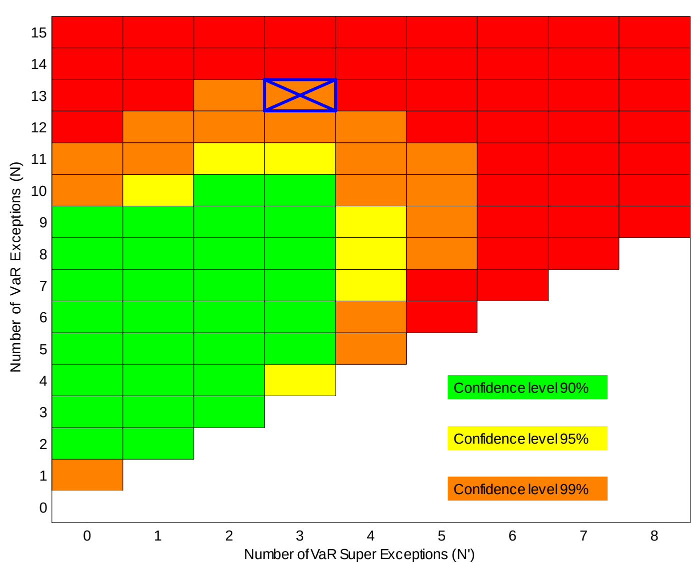
Figure 4: **The Risk Map**



Notes: This figure displays a Risk Map based on the p-value of a multivariate unconditional coverage tests, $LR_{MUC}(\alpha, \alpha')$.

Figure 5: **Backtesting Results for La Caixa (2007-2008)**





Notes: The upper graph displays the daily trading profit-and-loss (P&L), VaR($\alpha = 1\%$), and VaR($\alpha' = 0.2\%$) for La Caixa between January 1, 2007 and December 31, 2008. All figures are in thousands of euros. Over this sample period, there were 13 exceptions and 3 super exceptions. The lower graphs displays the Risk Map for La Caixa. The blue cross corresponds to La Caixa for the period 2007-2008. As the blue cross falls in the orange zone, we can reject the null hypothesis at the 95% but not at the 99% confidence level.

Table A1: Monte Carlo Experiments

Panel A: Summary of the Monte Carlo Experiment Design		
	Experiment 1	**Experiment 2**
P&L Data Generating Process	t-Garch(1,1)	t-Garch(1,1)
Internal VaR Model	t-Garch(1,1)	HS
Auxiliary Model	t-Garch(1,1)	t-Garch(1,1)

Panel B: Experiment 1 - Valid Internal VaR Model

Sample Size	Violation Rate	TPR	FPR	TNR
250	0.0019	0.8801	0.0003	0.9997
500	0.0020	0.9218	0.0002	0.9998
750	0.0021	0.9425	0.0002	0.9998
1,000	0.0020	0.9478	0.0001	0.9999
1,250	0.0020	0.9602	0.0001	0.9999
1,500	0.0020	0.9534	0.0001	0.9999
1,750	0.0020	0.9632	0.0001	0.9999
2,000	0.0020	0.9598	0.0001	0.9999

Panel C: Experiment 2 - Invalid Internal VaR Model

Sample Size	Violation Rate	TPR	FPR	TNR
250	0.0051	0.8441	0.0017	0.9983
500	0.0050	0.8801	0.0015	0.9985
750	0.0051	0.8734	0.0016	0.9984
1,000	0.0048	0.8900	0.0014	0.9986
1,250	0.0053	0.8814	0.0016	0.9984
1,500	0.0051	0.8871	0.0016	0.9984
1,750	0.0050	0.8774	0.0015	0.9985
2,000	0.0052	0.8831	0.0016	0.9984