



**HAL**  
open science

# The Risk Map: A New Tool for Validating Risk Models

Gilbert Colletaz, Christophe Hurlin, Christophe Pérignon

► **To cite this version:**

Gilbert Colletaz, Christophe Hurlin, Christophe Pérignon. The Risk Map: A New Tool for Validating Risk Models. 2012. halshs-00746273

**HAL Id: halshs-00746273**

**<https://shs.hal.science/halshs-00746273>**

Preprint submitted on 28 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Risk Map:

## A New Tool for Validating Risk Models

Gilbert Colletaz\*    Christophe Hurlin\*    Christophe Pérignon\*\*

October 2012

### Abstract

This paper presents a new method to validate risk models: the *Risk Map*. This method jointly accounts for the number and the magnitude of extreme losses and graphically summarizes all information about the performance of a risk model. It relies on the concept of a *super exception*, which is defined as a situation in which the loss exceeds both the standard Value-at-Risk (VaR) and a VaR defined at an extremely low coverage probability. We then formally test whether the sequences of exceptions and super exceptions are rejected by standard model validation tests. We show that the Risk Map can be used to validate market, credit, operational, or systemic risk estimates (VaR, stressed VaR, expected shortfall, and CoVaR) or to assess the performance of the margin system of a clearing house.

**JEL classification:** G21, G28, G32

**Keywords:** Financial Risk Management, Tail Risk, Basel III

**Companion Website:** [www.RunMyCode.org/CompanionSite/site2](http://www.RunMyCode.org/CompanionSite/site2)

---

\*University of Orléans, Laboratoire d'Economie d'Orléans, France. \*\* Deloitte - Société Générale Chair in Energy and Finance, HEC Paris, France. We are extremely grateful to an anonymous referee, John Cotter, Jorge Cruz Lopez, Laurent Frésard, Andras Fulop, Thomas Gilbert, Bertrand Maillet, Olivier Scaillet, and Stefan Straetmans, seminar participants at ESSEC, Free University of Berlin, HEC Paris, Maastricht University, Skema Business School, University College Dublin, University of Geneva, University of Glasgow, University of Lille 2, as well as to participants at the 2010 EconomiX Workshop in Financial Econometrics, 2011 Econometric Society European Meeting, 2011 French Finance Association Spring Meeting, 2011 SMU-ESSEC Symposium on Empirical Finance and Financial Econometrics, and 5th Annual Methods in International Finance Network Workshop for their comments. Pérignon gratefully acknowledges the financial support of the HEC Research Foundation. Emails: [gilbert.colletaz@univ-orleans.fr](mailto:gilbert.colletaz@univ-orleans.fr), [christophe.hurlin@univ-orleans.fr](mailto:christophe.hurlin@univ-orleans.fr), [perignon@hec.fr](mailto:perignon@hec.fr). Contact Author: Christophe Pérignon, Finance Department, HEC Paris, 1 Rue de la Libération, 78351 Jouy-en-Josas, France. Phone: +33 139 67 94 11.

# 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 misrepresentation of actual risk exposures (Blöchlinger, 2011). In this paper, we focus on a popular family of risk models, namely the tail risk models. These models can generate a variety of risk measures including the Value-at-Risk (VaR), which is defined as an extreme quantile of a return or profit-and-loss (P&L) distribution, as well as stressed VaR, expected shortfall, and CoVaR. In practice, tail risk models are used to measure downside risk (Bali, Demirtas and Levy, 2009), construct portfolios (Basak and Shapiro, 2001), quantify financial institutions' exposures to, and capital requirements for, market, credit, and operational risks (Jorion, 2007), set margin requirements for derivatives users (Booth et al., 1997), and measure the systemic-risk contribution of a financial institution (Adrian and Brunnermeier, 2011).

In this paper, we present a new tool, called the *Risk Map*, for validating risk 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 \$999 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  $2.5 = 1\%$  of 250 trading days). 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 and regulators do care about the magnitude of their losses.<sup>1</sup> It is indeed the severity of the trading losses, and not the exceptions per se, that jeopardize the solvency of financial institutions. For instance, 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. Furthermore, it makes a big difference whether the margin of a given derivative market

---

<sup>1</sup>In a recent survey on trading risk, members of the Basel Committee state that "account for the severity of losses beyond the confidence threshold [...] is especially important for regulators, who are [...] concerned about exactly these losses" (Basel Committee on Banking Supervision, 2011b).

participant is exceeded by a small or by a large amount as it creates a shortfall of the same amount that the clearing house must cover. 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 extreme losses. The objective of this paper 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 loss not only exceeds the regular VaR defined with a probability  $\alpha$  (e.g. 1%) but is also likely to exceed a VaR defined with a much lower probability  $\alpha'$  (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  $\alpha'$  much smaller than  $\alpha$ . As an illustration, we show in Figure 1 the joint evolution of the daily P&L,  $VaR(\alpha)$ , and  $VaR(\alpha')$  for a hypothetical portfolio. 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 portfolio, there are four exceptions and three super exceptions.

< **Insert Figure 1** >

In practice, the choice of the probability  $\alpha'$  is key. We show in this paper that this parameter can be either chosen freely by the user or defined endogenously. In the former case, the  $\alpha'$  parameter must reflect the loss aversion of the investor whereas in the latter case,  $\alpha'$  is set such that  $VaR(\alpha')$  corresponds either to the stressed VaR or to the expected shortfall.

In order to validate the risk model, we formally test whether the sequences of exceptions and super exceptions satisfy standard backtesting conditions. Formally, we need to test the joint null hypothesis that the probability of having an exception is  $\alpha$  and that the probability of having a super exception is  $\alpha'$ . Several statistical strategies can be used to test this null hypothesis (e.g. log-likelihood ratio, hit regression test) and some statistical tests are readily available in the literature. We then report the p-value of the statistical test in a three-dimensional graph, which we call the Risk Map, that graphically summarizes all information about the performance of a risk model. To sum up, what the Risk Map approach allows us to do is to simplify an intricate problem defined over a {loss frequency, loss severity} domain to an easier problem defined over a {loss frequency, loss frequency} domain.

There are several advantages to the Risk Map approach. First, it preserves the *simplicity* of the standard validation techniques (Kupiec, 1995), while still accounting for the magnitude of the losses. Thus, the Risk Map approach is a three-dimensional generalization of the "Traffic Light" system (Basel Committee on Banking Supervision, 2006, 2011a) which remains the reference backtest methodology for banking regulators.<sup>2</sup> Second, it is a *formal* hypothesis testing framework that provides p-values and rejection ranges. Indeed, it allows us to jointly test the null hypothesis that both the numbers of VaR exceptions and super exceptions are accurate. Third, the Risk Map approach is *general* and can be used with any tail risk model as it only relies on the sequence of exceptions and super exceptions. In particular, no assumptions need to be made regarding the distribution of the P&L. For instance, it can be used to backtest the market VaR of a single asset, portfolio, trading desk, business line, bank, insurance company, mutual fund, or hedge fund (Berkowitz, Christoffersen and Pelletier, 2011). It also permits to jointly validate the standard and stressed VaRs that banks must compute under Basel III, as well as expected shortfalls. Furthermore, the Risk Map can be used to backtest credit-risk VaRs (Lopez and Saidenberg, 2000), operational-risk VaRs (Dahen and Dionne, 2010), or VaR-based margins for derivative users (Cruz Lopez et al., 2011). Finally, we show that the Risk Map can be used to validate the systemic risk measure recently proposed by Adrian and Brunnermeier (CoVaR, 2011) as it is defined as the conditional quantile of a bank asset 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 describe our model validation framework. In Section 3, we present several applications of the Risk Map methodology that sequentially deal with market risk, systemic risk, and margin requirements. We summarize and conclude our paper in Section 4.

---

<sup>2</sup>Accounting for VaR exceedances is particularly important in the context of the Basel III regulation (Basel Committee on Banking Supervision, 2011a). Indeed, under Basel III, capital requirements for market risk depend on both the VaR and stressed VaR (calibrated to historical data from a continuous 12-month period of significant financial stress) of the bank:

$$c = \max \{ VaR_{t|t-1}; m \cdot VaR_{avg} \} + \max \{ sVaR_{t|t-1}; m_s \cdot sVaR_{avg} \}.$$

As the value of the multiplicative factors  $m$  and  $m_s$  only depends on the backtesting results of the VaR, but not of the stressed VaR, it is particularly important to account for the magnitude of the VaR exceedances.

## 2 Validation 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  $\alpha$  coverage rate conditionally on an information set  $\mathcal{F}_{t-1}$ . In practice, VaR is computed using either non-parametric techniques, such as Historical Simulation (Pritsker, 2006), or parametric techniques, such as Monte Carlo simulation (Broadie, Du and Moallemi, 2011). If the VaR model is adequate, then the following relation must hold:

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

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

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

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.<sup>3</sup> Under the UC hypothesis, the probability of an ex-post return exceeding the VaR forecast must be equal to the  $\alpha$  coverage rate:

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

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).<sup>4</sup> A solution proposed by Lopez (1999a,b) consists in considering the excess losses over and above the

---

<sup>3</sup>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  $\alpha$  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 (CC), and the VaR violation process is a martingale difference, with  $\mathbb{E} [I_t(\alpha) - \alpha | \mathcal{F}_{t-1}] = 0$ . For a test of the CC hypothesis, see Christoffersen (1998), Christoffersen and Pelletier (2004), Engle and Manganelli (2004), and Berkowitz, Christoffersen and Pelletier (2011).

<sup>4</sup>Other standard backtesting methods also disregard the magnitude of the losses beyond the VaR (Christoffersen and Pelletier, 2004; Engle and Manganelli, 2004; Gaglianone et al., 2011; Berkowitz, Christoffersen and Pelletier, 2011).

VaR and defining a loss function:

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

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 possibility is to complement the VaR data with some extra risk estimates. A first approach is to consider the average loss beyond the VaR using the concept of expected shortfall:

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

where  $F(\cdot)$  denotes the cumulative distribution function of the P&L (Artzner et al., 1999). Alternatively, Berkowitz (2001) suggests testing the entire P&L density – see Christoffersen (2009) for a discussion of the backtesting of ES or density forecasts.

Differently, we propose a model validation methodology that is based on the number and the severity of VaR exceptions.<sup>5</sup> Our approach exploits the concept of a *super exception*, which we define as a loss greater than  $VaR(\alpha')$ , with  $\alpha'$  much smaller than  $\alpha$  (e.g.  $\alpha = 1\%$  and  $\alpha' = 0.2\%$ ). One can similarly define a hit variable associated with  $VaR_t(\alpha')$ :

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

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  $VaR(\alpha)$  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  $\alpha$  (respectively  $\alpha'$ ) coverage rate and let  $N$  (respectively  $N'$ ) be the number of associated VaR violations:

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

---

<sup>5</sup>We show in Section 3.1.3 that our method can also be used with expected shortfalls.

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

$$N \sim B(T, \alpha) \quad (8)$$

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

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

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

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

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

where the  $LR_{UC}$  statistic is asymptotically chi-square with one degree of freedom (Jorion, 2007). A similar validation test can be defined for super exceptions:

$$H_0 : \mathbb{E}[I_t(\alpha')] = \alpha' \quad (12)$$

$$H_1 : \mathbb{E}[I_t(\alpha')] \neq \alpha'. \quad (13)$$

An LR test statistic  $LR_{UC}(N')$  can be defined as in equation (11), except that we now use the coverage rate  $\alpha'$  and the corresponding number of hits  $N'$ .

## 2.2 The Risk Map

The goal of the Risk Map is to present the backtesting results for a given risk model in a graphical way. A first approach is to jointly display the non-rejection zones for the  $LR_{UC}(\alpha)$  and  $LR_{UC}(\alpha')$  tests. For instance, with 500 observations, if the number of  $VaR(1\%)$  exceptions is between two and nine, we cannot reject the risk model at the 95% confidence level. Similarly, if the number of super exceptions is strictly greater than two, we reject the validity of the VaR model at a similar confidence level for the reason that it leads to too many super exceptions. It is then possible to check whether any risk model lays inside the global non-rejection area, i.e.,  $2 \leq N \leq 9$  and  $N' \leq 3$ . It is important to notice that in this first approach, we investigate this double validation process (loss frequency + loss magnitude) in a disjointed way. It boils down to consider the  $LR_{UC}(\alpha)$  and  $LR_{UC}(\alpha')$  tests independently, which 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:

$$H_0 : \mathbb{E}[I_t(\alpha)] = \alpha \quad \text{and} \quad \mathbb{E}[I_t(\alpha')] = \alpha'. \quad (14)$$

This joint null hypothesis can be tested using either a multivariate version of the unconditional coverage test ( $LR_{UC}$ ) or of the hit regression test of Engle and Manganelli (2004). We follow Pérignon and Smith (2008) and define several indicator variables for revenues falling in each disjoint interval:

$$J_{1,t} = I_t(\alpha) - I_t(\alpha') = \begin{cases} 1 & \text{if } -VaR_{t|t-1}(\alpha') < r_t < -VaR_{t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

$$J_{2,t} = I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -VaR_{t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

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  $\alpha'$ , 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 (14) 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. This 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:

$$LR_{MUC}(\alpha, \alpha') = -2 \ln \left[ (1 - \alpha)^{N_0} (\alpha - \alpha')^{N_1} (\alpha')^{N_2} \right] + 2 \ln \left[ \left( \frac{N_0}{T} \right)^{N_0} \left( \frac{N_1}{T} \right)^{N_1} \left( \frac{N_2}{T} \right)^{N_2} \right] \xrightarrow{T \rightarrow \infty} \chi^2(2). \quad (17)$$

Under the null (14) of joint conditional coverage for the VaR exceptions and super exceptions, the  $LR_{MUC}$  statistic is asymptotically chi-square with two degrees of freedom.<sup>6</sup> The p-values of the test are displayed for different combinations of number of exceptions  $N$  and super exceptions  $N'$  in Figure 2. Alternatively, a Risk Map can be constructed based on the rejection zones for different confidence levels (Figure 3). Note that the cells below the diagonal are not colored as they correspond to situations in which the number of super

---

<sup>6</sup>The p-values of the test can be based on either the asymptotic distribution or the finite sample distribution, which can be generated by simulation. See Section 2.3 for more discussion on finite sample properties.

exceptions exceeds the number of exceptions, which is of course impossible. If the  $(N, N')$  pair corresponds to a green cell, we conclude that we cannot reject the null hypothesis  $\mathbb{E}[I_t(\alpha)] = \alpha$  and  $\mathbb{E}[I_t(\alpha')] = \alpha'$  at the 95% confidence level. If  $(N, N')$  falls 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. Note also that, for technical reasons, the p-values for the first-column cells ( $N' = 0$ ) and diagonal cells ( $N' = N$ ) are computed from a univariate  $LR_{UC}$  test because the multivariate test cannot be computed in these cases. Similarly, when  $N' = N = 0$ , none of the LR test can be computed.

**< Insert Figures 2 and 3 >**

We illustrate the difference between the Risk Map and standard validation tests using a simple example. Consider a sample of 500 daily VaR forecasts generated from a hypothetical VaR model. Out of these 500 days, there are 10 exceptions. According to the  $LR_{UC}$  test, such model would be rejected as the **95%** non-rejection range for this test is  $2 \leq N \leq 9$ . Differently, with the Risk Map, a model with 10 exceptions would not be rejected at the 95% confidence level as long as there are either 1, 2, or 3 three super exceptions. However, the model would be rejected if the number of super exceptions is  $N' = 0$  or  $N' > 3$ .

What the Risk Map does is to test whether both  $N$  and  $N'$  are at an acceptable level. The intuition is as follows: if there are too few super exceptions, it means that even if the model is valid with respect to the  $\alpha$  coverage rate, it turns out not to be valid with respect to lower coverage rates. In that sense, the model is considered to be globally invalid. For instance, if among the 10 exceptions, there is no super exception, the VaR model overestimates the risk for the  $\alpha'$  coverage rate, which casts doubt on the overall performance of the risk model. On the contrary, if there are four or more super exceptions, the model is also rejected as it underestimates  $VaR(\alpha')$ . In the latter case, even if the VaR model is validated on the basis of the  $VaR(\alpha)$  exceptions, the magnitude of these exceptions is too large.

### **2.3 Extended Risk Map and Finite Sample Properties**

The Risk Map can be enriched with information about empirical frequencies of VaR (super) exceptions under the assumption that the VaR model is valid. This additional layer of

information complements the p-values and is independent of the statistical test used to build the Risk Map. To do so, we draw  $T$  observations from a uniform distribution over  $[0, 1]$ . The simulated violation process  $I_t(\alpha)$  is then defined by a binary variable that takes a value of one if the simulated value is less than  $\alpha$  and zero otherwise. The simulated super violation process  $I_t(\alpha')$  is defined in the same way with a cut-off of  $\alpha'$ . Theoretically, these violation processes have the same properties as the processes obtained by comparing valid ex-ante VaR forecasts with ex-post losses. We repeat the experiment 10,000 times and for each simulation we count the number of violations ( $N$ ) and super violations ( $N'$ ).

The upper Risk Map in Figure 4 displays the empirical frequencies of each pair  $(N, N')$  for  $T = 500$ ,  $\alpha = 1\%$ , and  $\alpha' = 0.2\%$ . This extended Risk Map provides useful information about the validity of the risk model. Consider for instance a model that yields 13 exceptions and 3 super exceptions. Figure 4 indicates that this case has only 3 chances out of 10,000 to arise with a valid risk model. Differently, an outcome like  $(N = 5, N' = 1)$  has more than 700 chances out 10,000 to occur with a valid model, which suggests that the analyzed model is more likely to be valid.

**< Insert Figure 4 >**

This extended Risk Map can also be seen as a graphical representation of the finite sample properties of the statistical test used to build the Risk Map (i.e., the  $LR_{MUC}$  test in our case). The statistical test determines the shape of the non-rejection area, which is the green zone if we consider a 5% level of nominal risk. Logically, if the test is well sized, most  $(N, N')$  observations fall inside the non-rejection area. More precisely, the sum of empirical frequencies of pairs of violations/super violations that are outside the non-rejection area corresponds to the empirical size of the test (defined as the risk of rejecting a valid model) for a finite sample size of 500. If the test is well sized, 95% of  $(N, N')$  observations remain inside the non-rejection area.

To further illustrate the size performance of the Risk Map, we report in Panel A of Table 1 the empirical sizes considering successively the cases when  $[\alpha = 5\%, \alpha' = 1\%]$ ,  $[\alpha = 2\%, \alpha' = 0.4\%]$ , and  $[\alpha = 1\%, \alpha' = 0.2\%]$ , as well as various sample sizes. The reported empirical sizes correspond to the rejection rates calculated over 10,000 simulations for a nominal size of 5%. We can observe that the size of the Risk Map is quite close to the nominal size, even for small samples.

< **Insert Table 1** >

In the same manner, the Risk Map can be used to illustrate the power of the test (defined as the probability to reject an invalid model). Here, the larger the number of exceptions and super exceptions that fall outside the non-rejection zone, the more powerful the test. In this case, two elements are required: a data generating process for the P&L and a risk model to produce the VaR forecasts. Let us consider the following experiment. We assume that the P&L are generated from a Markov Switching GARCH model (Bauwens, Preminger and Rombouts, 2010). This particular model allows both the conditional mean and the volatility to change from one state to the next:

$$r_t = \mu_{s_t} + \sigma_t u_t \quad (18)$$

$$\sigma_t^2 = \omega_{s_t} + \alpha_{s_t} (r_{t-1} - \mu_{s_{t-1}})^2 + \beta_{s_t} \sigma_{t-1}^2 \quad (19)$$

where  $u_t$  is *i.i.d.N.*  $(0, 1)$  and  $s_t$  denotes a two-state hidden Markov Chain with a transition matrix defined by the probabilities  $\{\eta_{ij} = \Pr(s_t = i | s_{t-1} = j)\}$ . The parameters values are those estimated by Bauwens, Preminger and Rombouts (2010) for the S&P 500 index between 2001 and 2007.<sup>7</sup> To illustrate the power of the test, we assume that the VaR model is misspecified and that VaR forecasts are computed according to the Historical Simulation method. This method, widely used by professionals, simply consists in calculating the empirical quantile of the past P&L over a rolling window (set here to 500 observations). The VaR forecasts are clearly invalid in this case, especially in periods of crisis, in the sense that they do not react fast enough to changes in the P&L dynamics (Pritsker, 2006). We simulate 10,000 P&L sequences of 1,000 observations each, and for each of them we compute 500 Historical Simulation out-of-sample VaR(1%) and VaR(0.2%) forecasts. Corresponding violations and super violations are then obtained by comparing these forecasts to the ex-post losses. Similar to what we did in the case of the valid model, we count for each simulation the number of exceptions ( $N$ ) and super exceptions ( $N'$ ).

The lower Risk Map in Figure 4 displays the empirical frequencies of each couple  $(N, N')$  for this invalid model. The Risk Map reveals that most of the  $(N, N')$  couples are outside the non-rejection area (95% confidence level). This configuration clearly indicates that the VaR model employed is misspecified. In this case, the sum of empirical frequencies of pairs

---

<sup>7</sup>The parameter values are  $\omega_1 = 0.419$ ,  $\omega_2 = 1.988$ ,  $\alpha_1 = 0.014$ ,  $\alpha_2 = 0.115$ ,  $\beta_1 = \beta_2 = 0$ ,  $\mu_1 = 0.046$ ,  $\mu_2 = -0.040$ ,  $\eta_{11} = 0.994$ , and  $\eta_{22} = 0.986$ .

of violations/super violations that are outside the non-rejection area corresponds to the empirical power of the LR test.

Similar to what we did for size, we report in Panel B of Table 1 the empirical power of the Risk Map and of the  $LR_{UC}$  test for various sample sizes and coverage rates  $(\alpha, \alpha')$ . The empirical power corresponds to the rejection rates (for a nominal size equal to 5%) based on 10,000 replications of the experiment previously described (MS GARCH DGP and VaR-HS). For comparison purposes, we report in Panel C the power of the  $LR_{UC}$  test for the 1%, 2%, and 5% coverage rates. Two conclusions can be drawn from this experiment. First, the power of the Risk Map generally increases with the sample size and the coverage rate (the latter result being true only for moderately sized samples, as in Berkowitz, Christoffersen and Pelletier, 2011). Second, the power of the Risk Map is at least as good as the one of the  $LR_{UC}$  test.

## 2.4 Computation of $VaR(\alpha')$

In most applications, Risk Map users, namely risk managers and regulators, will have both data on  $VaR(\alpha)$  and  $VaR(\alpha')$ . Nevertheless, we show in this subsection that a Risk Map can still be generated when only  $VaR(\alpha)$  is available. This could for instance be the case when a bank only discloses to the public 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 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.,  $\mathbb{E}(r_t) = 0$ .<sup>8</sup> Under these assumptions, the conditional VaR can be expressed as an affine function of the conditional variance of the P&L, denoted  $h_t$ :

$$VaR_{t|t-1}(\alpha; \beta) = -\sqrt{h_t}F^{-1}(\alpha; \beta) \quad (20)$$

where  $F^{-1}(\alpha; \beta)$  denotes the  $\alpha$ -quantile of the conditional standardized P&L distribution. We assume that this distribution is parametric and depends on a set of parameters  $\beta$ .

A first way to compute  $VaR(\alpha')$  is to use the Quasi-Maximum Likelihood (QML) esti-

---

<sup>8</sup>When  $\mathbb{E}(r_t) \neq 0$ , the estimated VaR can simply be deduced from (20) by adding the unconditional average return.

mates of the conditional variance and of the conditional quantile:

$$VaR_{t|t-1}(\alpha'; \widehat{\beta}) = -\sqrt{\widehat{h}_t} F^{-1}(\alpha'; \widehat{\beta}) \quad (21)$$

where  $\widehat{h}$  and  $\widehat{\beta}$  are the QML estimates. Alternatively, we propose an approach that does not rely on the QML estimate  $\widehat{h}_t$ . Given the VaR forecast reported by the bank,  $VaR(\alpha)$ , we define an implied P&L conditional variance as:

$$\sqrt{\widetilde{h}_t} = -\frac{VaR_{t|t-1}(\alpha)}{F^{-1}(\alpha; \widehat{\beta})}. \quad (22)$$

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

$$VaR_{t|t-1}(\alpha', \widehat{\beta}) = -\sqrt{\widetilde{h}_t} F^{-1}(\alpha'; \widehat{\beta}) = VaR_{t|t-1}(\alpha) \frac{F^{-1}(\alpha'; \widehat{\beta})}{F^{-1}(\alpha; \widehat{\beta})}. \quad (23)$$

Interestingly, we notice that the auxiliary model is only used to get an estimate of the conditional quantile  $F^{-1}(\alpha; \widehat{\beta})$ . 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  $\beta$ .

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  $\beta$  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  $VaR(\alpha')$ . Furthermore, we implement this calibration procedure on real data in Section 3.1.1.

### 3 The Risk Map at Work

In this section, we present a variety of applications for the Risk Map. The different areas of application include market risk modeling, systemic risk measurement, and margins for derivatives users. From one application to the next, it is only the definition of the  $r$  variable

that is changing. It denotes successively the P&L of a bank trading portfolio, the asset return of a financial institution and of the financial system, and the change in a derivatives (portfolio) price, respectively.

## 3.1 Market Risk

### 3.1.1 Bank VaR

A natural application of the Risk Map is to backtest the VaR of a financial firm. It is indeed key for both risk managers and banking regulators to check the validity of banks' VaR engines as capital requirements depend on banks' VaR internal estimates (Jorion, 2007, and footnote 2). 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. Frésard, Pérignon and Wilhelmsson (2011) report that during the global financial crisis, the number of VaR exceptions increased sharply, with some financial institutions experiencing more than 50 VaR(1%) exceptions per year. However, prior empirical literature did not consider 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.<sup>9</sup> We use daily one-day ahead VaR(1%) and daily P&L for that bank over the period 2007-2008. 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 volatility spiked after the end of the first semester of 2007. Similarly, the VaR(1%) jumped from around 2 million euros during 2007-Q1-Q2 to around 4 million afterwards. We provide some summary statistics on the P&L and VaR time series in Table 2.

< **Insert Table 2** >

VaR forecasts defined at a lower coverage rate are required to implement the Risk Map methodology but are not disclosed by the bank. As a result, we use the calibration

---

<sup>9</sup>The VaR and P&L figures correspond to the entire trading portfolio of the bank. VaR and P&L data have been extracted directly from the firm's annual reports, as in Frésard, Pérignon and Wilhelmsson (2011).

procedure presented in Section 2.4 to extract the VaR(0.2%) series from the VaR(1%) series. We use a t-GARCH as the auxiliary model for the P&L and report the estimated parameters in Panel B of Table 2. Out of the four parameters, only the degree of freedom  $\hat{v}$  is needed to estimate the quantile of the conditional distribution. As shown in equation (22), the implied conditional variance  $\tilde{h}_t$  is given by the ratio of VaR(1%) and the quantile of the conditional distribution. Given the implied variance  $\tilde{h}_t$ , we then compute VaR(0.2%) following equation (23). We display in Panel A of Table 2 some summary statistics on VaR(0.2%). We notice that the ratio of means of VaR(0.2%) and VaR(1%) is 1.73, which corresponds to the ratio of quantiles defined in equation (23). Interestingly, we see that VaR(0.2%) has the same skewness and kurtosis as VaR(1%) which suggests that our calibration procedure preserves the main statistical properties of the original VaR( $\alpha$ ).

< **Insert Figure 5** >

Over this sample period, there were 13 VaR exceptions ( $N = 13$ ) and three super exceptions ( $N' = 3$ ) for La Caixa. The number 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 (see the 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  $LR_{MUC}$  test is 0.0108.

One may wonder whether our conclusion is affected by the value of the probability  $\alpha'$ . If we assume that there is a continuum of VaR( $\alpha'$ ) associated with a continuum of  $\alpha'$ , the number of super exceptions would remain the same for a range of value for  $\alpha'$ . For La Caixa, this range is pretty large as it goes from  $\alpha/10$  to  $\alpha/4.3$ . Furthermore, we plot in Figure 6 both the number of super exceptions and the p-value of the  $LR_{MUC}$  test for different  $\alpha'$  coefficients. The main take away from this figure is that the p-value remains remarkably stable, which confirms the robustness of our results. Note also that the question about the value of  $\alpha'$  is not relevant anymore under Basel III. Indeed, as shown in the next two subsections,  $\alpha'$  becomes in this case endogenous.

< **Insert Figure 6** >



In the context of banking regulation, the Risk Map can be used to generalize the "Traffic Light" system of the Basel Committee. Under current regulation, part of the capital requirement for banks comes from the market risk exposure of the banks. More specifically, the Market Risk Charge depends on the  $VaR(1\%)$  and a multiplicative factor which is set to 3 plus a penalty term  $\pi$ . The value of the applicable penalty only depends on the number of exceptions:  $\pi = 0$  for  $N < 5$ ,  $\pi = 0.4$  for  $N = 5$ ,  $\pi = 0.5$  for  $N = 6, \dots$ , and  $\pi = 1$  for  $N > 10$ . We suggest generalizing this approach by accounting jointly for the number and the magnitude of the exceptions within the framework of the Risk Map. We present in Table 3 the value of the penalty term for different number of exceptions and super exceptions and different sample sizes ( $T = 250$  and  $T = 500$ ). The no-penalty zone corresponds to the green zone of the Risk Map and the maximum penalty zone corresponds to the red zone (see Figure 3). Between these two zones, the bank incurs an increasing penalty, which is obtained through linear interpolation. As an illustration, given the backtesting results in Figure 5, the penalty coefficient that would apply to La Caixa is 0.7.

< **Insert Table 3** >

### 3.1.2 Bank Stressed VaR

Under Basel III, all financial institutions with material trading activities must compute both their VaR using recent data and their stressed VaR (sVaR) using data from a particularly volatile period (Basel Committee on Banking Supervision, 2011a). As shown in equation (20), the VaR is an affine function of the conditional variance of the P&L, denoted  $h_t$ . Differently, the sVaR depends on a conditional variance of the P&L measured over a particularly volatile period, denoted  $H_t$ :

$$sVaR_{t|t-1}(\alpha; B) = -\sqrt{H_t}F^{-1}(\alpha; B) \tag{24}$$

where  $B$  denotes the parameters of the P&L distribution  $F$  in the high-volatility period.

The Risk Map can be used to validate globally the VaR engine of the bank using both VaR and sVaR figures. Note that this is, to the best of our knowledge, the first attempt to validate a sVaR, which is the preferred market-risk measure under Basel III. To do so, we define the coverage rate  $\alpha'$  for which  $VaR(\alpha'; \beta)$  is as close as possible to  $sVaR(\alpha; B)$ :

$$\hat{\alpha}' = \arg \min_{\{\alpha'\}} \left\{ \frac{1}{T} \sum_{t=1}^T [VaR_{t|t-1}(\alpha'; \beta) - sVaR_{t|t-1}(\alpha; B)]^2 \right\}_{t=1}^T. \quad (25)$$

In this case, we define a super exception as a loss exceeding the stressed VaR and consider the associated hit variable:

$$I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -sVaR_{t|t-1}(\alpha; B) \\ 0 & \text{otherwise} \end{cases}. \quad (26)$$

The relevant joint null hypothesis remains equation (14). A particularly appealing feature of applying the Risk Map to stressed VaRs is that the probability  $\alpha'$  becomes endogenous.

### 3.1.3 Expected Shortfall

Similarly to what we did with stressed VaR, the risk model can be validated using both the VaR and the expected shortfall ES. The latter risk measure is defined as the average loss beyond the  $VaR$  (see equation 5) and is produced by the same risk model as the VaR. First, we identify the  $VaR(\alpha')$  that corresponds to the  $ES(\alpha)$ . As  $VaR(\alpha') = F^{-1}(\alpha')$ , then  $\alpha' = F(ES(\alpha))$ .

When the P&L distribution is known, backing out the  $\alpha'$  probability is straightforward. For instance, if we assume that the unconditional distribution of the P&L is a  $t$ -distribution, then the expected shortfall is:

$$ES(\alpha; v) = - \left( \frac{v}{v-1} \right) \left( 1 + \frac{1}{v} F^{-1}(\alpha; v) \right) \frac{f(\alpha; v)}{F(\alpha; v)} \quad (27)$$

where  $f(x; v)$  and  $F(x; v)$  respectively denote the pdf and the cdf of the  $t$ -distribution with  $v$  degrees of freedom (Zhu and Galbraith, 2010). For  $\alpha = 1\%$  and  $v = 2$ , we obtain  $\alpha' = F(ES(\alpha; v); v) = 0.25\%$ , which is close to our benchmark value of 0.2%.

Alternatively, when the P&L distribution is unknown,  $\alpha'$  can be estimated by:

$$\hat{\alpha}' = \arg \min_{\{\alpha'\}} \left\{ \frac{1}{T} \sum_{t=1}^T [VaR_{t|t-1}(\alpha'; \beta) - ES_{t|t-1}(\alpha; \beta)]^2 \right\}_{t=1}^T. \quad (28)$$

Then we define the  $ES$ -based hit variable as:

$$I_t(\alpha') = \begin{cases} 1 & \text{if } r_t < -ES_{t|t-1}(\alpha; \beta) \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

and, again, consider the null hypothesis in equation (14).

## 3.2 Systemic Risk

Since the recent financial crisis, the quest for measuring and forecasting systemic risk has never been more popular. Of particular importance is the quantification of the marginal contribution of systemically important financial institutions to the overall risk of the system. While many methodologies have been recently proposed to measure systemic risk (Acharya et al. 2010; Adrian and Brunnermeier, 2011; Engle and Brownless, 2011), there is to the best of our knowledge no ex-post validation method 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 and Brunnermeier (2011) and define the CoVaR measure as the VaR of the financial system conditional on institutions being under distress. Formally, CoVaR is defined as the  $\alpha$ -quantile of the conditional probability distribution of the financial system asset returns  $r_j$ :

$$\Pr [r_j \leq -CoVaR^{j|\mathbb{C}(r_i)}(\alpha) \mid \mathbb{C}(r_i)] = \alpha \quad (30)$$

where  $\mathbb{C}(r_i)$  denotes a conditioning event concerning firm  $i$ . One possible conditioning event is a situation in which the loss of firm  $i$  exceeds its VaR:

$$\Pr [r_j \leq -CoVaR^{j|i}(\alpha) \mid r_i \leq -VaR^i(\alpha)] = \alpha. \quad (31)$$

Given this definition, it is obvious that CoVaR can be backtested within the Risk Map framework. Just like with VaR, we need to analyze the frequency of the conditional probability and the magnitude of the losses in excess of the CoVaR. The latter will provide us with some crucial information about the resiliency of the financial system when a particular firm is in financial distress.

The Risk Map framework allows one to validate the CoVaR in two dimensions: 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:

$$I^j(q) = \begin{cases} 1 & \text{if } r_j < -CoVaR^{j|i}(q) \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } q = \alpha \text{ or } q = \alpha' \quad (32)$$

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

$$\mathbb{E} [I^j(q) \mid r_i \leq -VaR^i(q)] = q. \quad (33)$$

It is possible to transform the conditional expectation (33) into an unconditional one since the condition (33) implies  $\mathbb{E} [I^j(q) \times I^i(q)] = q^2$  for  $q = \alpha$  or  $q = \alpha'$ , where  $I_t^i(\alpha)$  is the standard VaR( $\alpha$ ) hit variable for institution  $i$ :

$$I^i(\alpha) = \begin{cases} 1 & \text{if } r_i < -VaR^i(\alpha) \\ 0 & \text{otherwise} \end{cases}. \quad (34)$$

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

$$H_0 : \mathbb{E} [I^j(\alpha) \times I^i(\alpha)] = \alpha^2 \text{ and } \mathbb{E} [I^j(\alpha') \times I^i(\alpha)] = \alpha'^2. \quad (35)$$

Under the null, conditional on the distress of institution  $i$ , the probability to observe a loss in the financial system larger than the CoVaR( $\alpha$ ) is precisely equal to  $\alpha$ , and the probability to observe a *CoVaR super exception* should not exceed  $\alpha'$ .

The CoVaR Risk Map can be constructed by analogy with the VaR Risk Map. Given a sequence of estimated conditional CoVaR 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  $\{I_t^j(q)\}$ . The corresponding test statistic can then directly be derived from the  $LR_{MUC}(\alpha, \alpha')$  test shown in (18). In this case,  $N_1 = \sum_{t=1}^T [I_t^j(\alpha) - I_t^j(\alpha')] \times I_t^i(\alpha)$ ,  $N_2 = \sum_{t=1}^T I_t^j(\alpha') \times I_t^i(\alpha)$ , and  $N_0 = T - N_1 - N_2$ . The non-rejection area of the test can be represented as in Figure 3. 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  $i$ . Adrian and Brunnermeier (2011) call this difference  $\Delta CoVaR$ :

$$\Delta CoVaR^{j|C(r_i)}(\alpha) = CoVaR^{j|r_i \leq -VaR^i(\alpha)}(\alpha) - CoVaR^{j|r_i \leq Median^i}(\alpha). \quad (36)$$

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  $CoVaR^{j|r_i \leq Median^i}$ .

### 3.3 Margin Requirements

Tail risk 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 so high as to 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:

$$\Pr[r_t < -C_{t|t-1}(\alpha)] = \Pr[r_t > C_{t|t-1}(\alpha)] = \alpha. \quad (37)$$

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 modeling 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 (Cruz Lopez et al., 2011, and Hurlin and Pérignon, 2012). 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 over-the-counter (OTC) (Duffie and Zhu, 2011).

To show how the Risk Map can be used in the context of a clearing house, let  $\omega_{i,t-1}$  be the vector of positions of clearing member  $i$  at the end of day  $t - 1$ :

$$\omega_{i,t-1} = \begin{bmatrix} \omega_{i,1,t-1} \\ \vdots \\ \omega_{i,D,t-1} \end{bmatrix} \quad (38)$$

where  $D$  is the number of derivatives contracts (e.g. futures, options, CDS, and swaps) cleared by this clearing house and  $i = 1, \dots, N$ . 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 value and volatility of the underlying assets. For each scenario, the value of the portfolio is recomputed, or marked-to-model, using derivatives pricing formulas, and the associated hypothetical P&L is computed:

$$\tilde{r}_{i,t} = \begin{bmatrix} \tilde{r}_{i,t}^1 \\ \vdots \\ \tilde{r}_{i,t}^S \end{bmatrix}. \quad (39)$$

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:

$$\Pr[\tilde{r}_{i,t}^s < -C_{i,t|t-1}^{VaR}(\alpha)] = \alpha \quad (40)$$

with  $s = 1, \dots, S$ .<sup>10</sup> 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, they need 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

---

<sup>10</sup>For instance, the European Central Bank's Recommendations for Central Counterparties (2009) state that "Margin requirements [...] should be sufficient to cover losses that result from at least 99 % of the price movements over an appropriate time horizon."

or for all the clearing members pooled together.

In addition to the VaR-based margins presented in equation (40), clearing houses have recently started using expected shortfall-based margins  $C^{ES}$ , which are defined as:<sup>11</sup>

$$C_{i,t|t-1}^{ES}(\alpha) = E[\tilde{r}_{i,t}^s | \tilde{r}_{i,t}^s < -C_{i,t|t-1}^{VaR}(\alpha)]. \quad (41)$$

The margin model of the clearing house can then be tested using both  $C^{VaR}$  and  $C^{ES}$  by following the procedure described in Section 3.1.3.

## 4 Conclusion

In finance, risk models must capture both (1) the frequency of the losses and (2) the magnitude of these losses. Surprisingly, current model validation techniques completely neglect the magnitude dimension of financial risk. In this paper, we have proposed a validation framework allowing risk managers and regulators to assess the validity of a risk model by accounting for both the number and the magnitude of extreme losses. We have introduced the concept of a super exception and designed a testing procedure that combines information about both exceptions and super exceptions. The main advantages of the Risk Map is that it is as simple to use as standard validation techniques; it is a formal hypothesis testing framework; and it can be applied to any tail risk model. The Risk Map framework can be handy in validating market risk, credit risk, or operational risk estimates, systemic risk measures such as CoVaR, or the margin system of a clearing house.

The Risk Map approach may prove particularly effective in banking. Indeed, as a generalization of the system currently used by banking regulators to validate banks' risk models, the Risk Map could help detecting misspecified risk models and penalizing banks that experience VaR exceptions that are too frequent and/or too large. In this case, bank capital requirements would directly be affected by the conclusions of the Risk Map analysis.

In order to ease the implementation of the present methodology, we have created a website that automatically generates Risk Maps:

[www.RunMyCode.org/CompanionSite/site2](http://www.RunMyCode.org/CompanionSite/site2)

---

<sup>11</sup>See the Options Clearing Corporation (OCC) website for a recent example of expected shortfall-based margins: <http://www.optionsclearing.com/risk-management/margins/>

Users only have to upload their risk estimates and P&L data and to indicate the respective coverage rates. All calculations are done in a cloud computer and results are directly displayed to the users.



## Appendix: Monte Carlo Experiment

We assess the accuracy of the calibration procedure for  $VaR(\alpha')$  presented in Section 2.4 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. 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:

$$r_t = \sigma_t z_t \sqrt{\frac{v-2}{v}} \quad (\text{A1})$$

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

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

Parameterization of the coefficients and initial conditions are deduced from maximum-likelihood estimated parameters for the S&P 500 index daily returns over the period 08/07/2010 to 01/07/2011. The parameter values are  $\omega = 4.9453e^{-7}$ ,  $\gamma = 0.1012$ ,  $\pi = 0.8550$ , and  $v = 4.1561$ . The initial condition  $\sigma_1^2$  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. As a result, unconditional coverage is satisfied 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 it generates too many extreme losses (see Panel A of Table A1). Then, the third experiment considers the case when both the internal VaR models and the auxiliary model are misspecified.

< **Insert Table A1** >

Recall that the aim of the experiments is to assess 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  $VaR_{t|t-1}^0(\alpha')$ , to the estimated one, denoted  $VaR_{t|t-1}(\alpha', \hat{\beta})$ . 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  $VaR_{t|t-1}^0(\alpha')$  and  $I_t(\alpha', \widehat{\beta})$  the estimated hit process associated to  $VaR_{t|t-1}(\alpha', \widehat{\beta})$ . There are two types of errors that can occur: false positive (type 1 error) and false negative (type 2 error). We evaluate the quality of our approach by using three indicators. The True Positive Rate (TPR) denotes the fraction of super exceptions that occur 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:

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

The False Positive Rate (FPR) 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).

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

Finally, we consider the True Negative Rate (TNR), which measures the fraction of non super VaR exceptions that are correctly identified (complementary of the type 2 error).

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

### 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:

$$VaR_{t|t-1}^0(\alpha) = -\sqrt{\frac{\beta - 2}{\beta}} \sigma_t F^{-1}(\alpha; \beta) \quad (\text{A6})$$

where  $F(\cdot; \beta)$  denotes the c.d.f. of the  $t(v)$  distribution. Let us denote  $VaR_{t|t-1}^0(\alpha')$  the true conditional VaR (unobservable) for the coverage rate  $\alpha'$ . 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,  $\left\{ VaR_{t|t-1}^0(\alpha) \right\}_{t=1}^T$ , we apply our calibration procedure to estimate the VaR for a coverage rate  $\alpha'$ . In this first experiment, we consider an auxiliary model also defined as a

$t(\beta)$ -GARCH(1, 1) model:

$$r_t = \sigma_t z_t \sqrt{\frac{\beta - 2}{\beta}} \quad (\text{A7})$$

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

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:

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

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 whereas the auxiliary model is a  $t$ -GARCH(1,1) model. We see in Panel C of Table **A1** that even in a situation in which the auxiliary model is wrong, our calibration procedure allows us to extract a reliable estimate of  $VaR(\alpha')$ .

### Experiment 3: Invalid Internal and Auxiliary Models

In the third experiment, neither the internal model nor the auxiliary model conform with the DGP, which is assumed to be a MS-GARCH model described by equations (18) and (19). We assume that the bank computes its VaR through Historical Simulation and we use a GARCH(1,1) model as the auxiliary model. In that case,  $VaR(\alpha')$  is simply defined as:

$$VaR_{t|t-1}(\alpha', \widehat{\beta}) = VaR_{t|t-1}(\alpha) \frac{\Phi^{-1}(\alpha')}{\Phi^{-1}(\alpha)} \quad (\text{A10})$$

where  $\Phi(\cdot)$  is the cdf of the standard normal distribution. Results in Panel D indicate that

the calibration procedure works fine in this case too. Note that this does not imply that our calibrated  $\text{VaR}(\alpha')$  is always a good proxy for the true  $\text{VaR}(\alpha')$  in term of bias. It only means that the errors in the  $\text{VaR}(\alpha')$  estimates are not sufficiently important to have any material impact on the number of super exceptions. Even with a MS-GARCH model, the TPR is close to 95% and the TNR close to 99%.

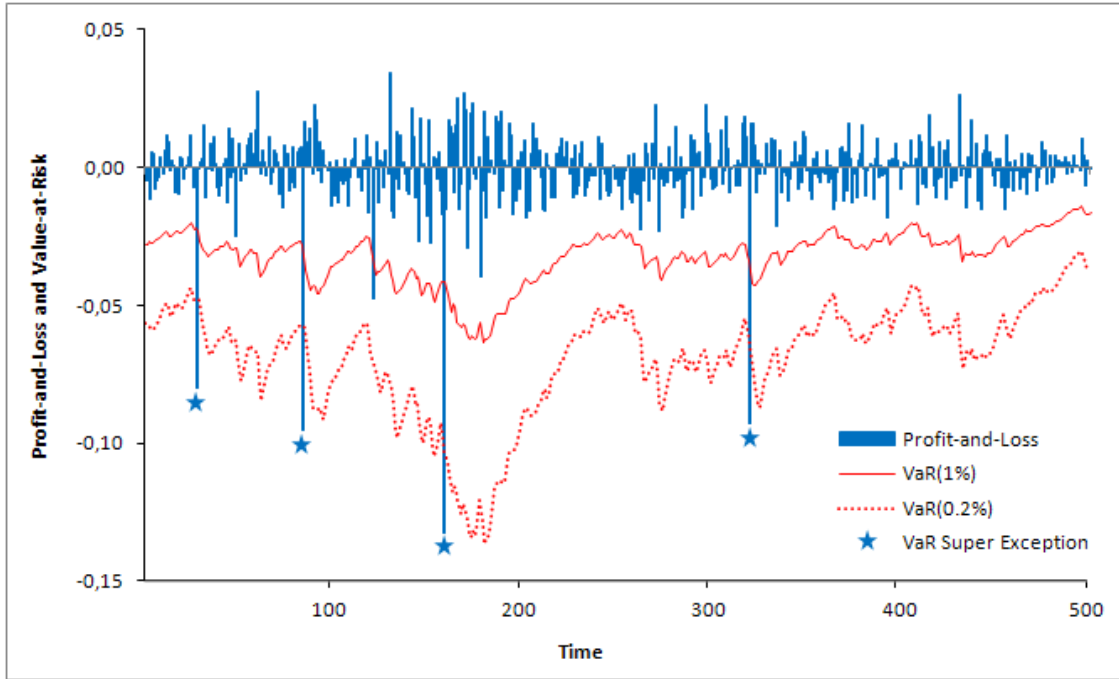
## References

- [1] Acharya, V. V., L. Pedersen, T. Philippon, M. Richardson (2010), "Measuring Systemic Risk," Working Paper, New York University.
- [2] Adrian, T., M. K. Brunnermeier (2011), "CoVaR," Working Paper, Princeton University and Federal Reserve Bank of New York.
- [3] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 203-228.
- [4] Bali, T. G., K. O. Demirtas, H. Levy (2009), "Is There an Intertemporal Relation between Downside Risk and Expected Returns?," *Journal of Financial and Quantitative Analysis* 44, 883-909.
- [5] Basel Committee on Banking Supervision (2006), "International Convergence of Capital Measurement and Capital Standards: A Revised Framework," Bank for International Settlements.
- [6] Basel Committee on Banking Supervision (2011a), "Revisions to the Basel II Market Risk Framework," Bank for International Settlements.
- [7] Basel Committee on Banking Supervision (2011b), "Messages from the Academic Literature on Risk Measurement for the Trading Book," Bank for International Settlements.
- [8] Bauwens, L., A. Preminger, J. V. K. Rombouts (2010), "Theory and Inference for a Markov Switching GARCH Model," *Econometrics Journal*, 13, 218-244 .
- [9] Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business and Economic Statistics*, 19, 465-474.
- [10] Berkowitz, J., P. F. Christoffersen, D. Pelletier (2011), "Evaluating Value-at-Risk Models with Desk-Level Data," *Management Science*, 57, 2213-2227.
- [11] Berkowitz, J., J. O'Brien (2002), "How Accurate Are Value-At-Risk Models at Commercial Banks?," *Journal of Finance*, 57, 1093-1111.
- [12] Blöchlinger, A. (2011), "Validation of Default Probabilities," *Journal of Financial and Quantitative Analysis*, forthcoming.
- [13] Broadie, M., Y. Du, C. C. Moallemi (2011), "Efficient Risk Estimation via Nested Sequential Simulation," *Management Science*, 57, 1172-1194.
- [14] Brownlees, C. T., R. F. Engle (2011), "Volatility, Correlation and Tails for Systemic Risk Measurement," Working Paper, New York University.
- [15] Brunnermeier, M. K., L. H. Pedersen (2009), "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201-2238.

- [16] 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.
- [17] Christoffersen, P. F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841-862.
- [18] Christoffersen, P. F. (2009), "Backtesting," in Encyclopedia of Quantitative Finance, R. Cont (Editor), Wiley.
- [19] Christoffersen, P. F., D. Pelletier (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, 2, 84-108.
- [20] Cotter, J. (2001), "Margin Exceedences for European Stock Index Futures using Extreme Value Theory," *Journal of Banking and Finance*, 25, 475-502.
- [21] Cruz Lopez, J. A., J. H. Harris, C. Pérignon (2011), "Clearing House, Margin Requirements, and Systemic Risk," *Review of Futures Markets*, 19, 39-54.
- [22] Dahan, H., G. Dionne (2010), "Scaling Models for the Severity and Frequency of External Operational Loss Data," *Journal of Banking and Finance*, 34, 1484-1496.
- [23] Duffie, D., H. Zhu (2011), "Does a Central Clearing Counterparty Reduce Counterparty Risk?," *Review of Asset Pricing Studies*, 1, 74-95.
- [24] Eisenberg, L., T. H. Noe (2001), "Systemic Risk in Financial Systems," *Management Science*, 47, 236-249.
- [25] Engle, R. F., S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367-381.
- [26] European Central Bank (2009), "Recommendations for Securities Clearing and Settlement Systems in the European Union, ESCB/CESR," Eurosystem Document.
- [27] Frésard, L., C. Pérignon, A. Wilhelmsson (2011), "The Pernicious Effects of Contaminated Data in Risk Management," *Journal of Banking and Finance*, 35, 2569-2583.
- [28] Gaglianone, W. P., L. R. Lima, O. Linton, D. R. Smith (2011), "Evaluating Value-at-Risk Models via Quantile Regression," *Journal of Business and Economic Statistics*, 29, 150-160.
- [29] Hurlin, C., C. Pérignon (2012), "Margin Backtesting," *Review of Futures Markets*, 20, 179-194.
- [30] Jorion, P. (2007), *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, 3rd Edition.
- [31] Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73-84.

- [32] Lopez, J. A. (1999a), "Regulatory Evaluation of Value-at-Risk Models," *Journal of Risk*, 1, 37-64.
- [33] Lopez, J. A. (1999b), "Methods for Evaluating Value-at-Risk Models," *Federal Reserve Bank of San Francisco Economic Review*, 2, 3-17.
- [34] Lopez, J. A., M. R. Saldenber (2000), "Evaluating Credit Risk Models," *Journal of Banking and Finance*, 24, 151-167.
- [35] Pérignon, C., D. R. Smith (2008), "A New Approach to Comparing VaR Estimation Methods," *Journal of Derivatives*, 16, 54-66.
- [36] Pritsker, M. (2006), "The Hidden Dangers of Historical Simulation," *Journal of Banking and Finance*, 30, 561-582.
- [37] Stulz, R. (2008), "Risk Management Failures: What Are They and When Do they Happen?," *Journal of Applied Corporate Finance*, 20, 39-48.
- [38] Taylor, J. W. (2005), "Generating Volatility Forecasts from Value at Risk Estimates," *Management Science*, 51, 712-725.
- [39] Zhu, D., J. W. Galbraith (2010), "A Generalized Asymmetric Student-t Distribution with Application to Financial Econometrics," *Journal of Econometrics*, 157, 297-305.

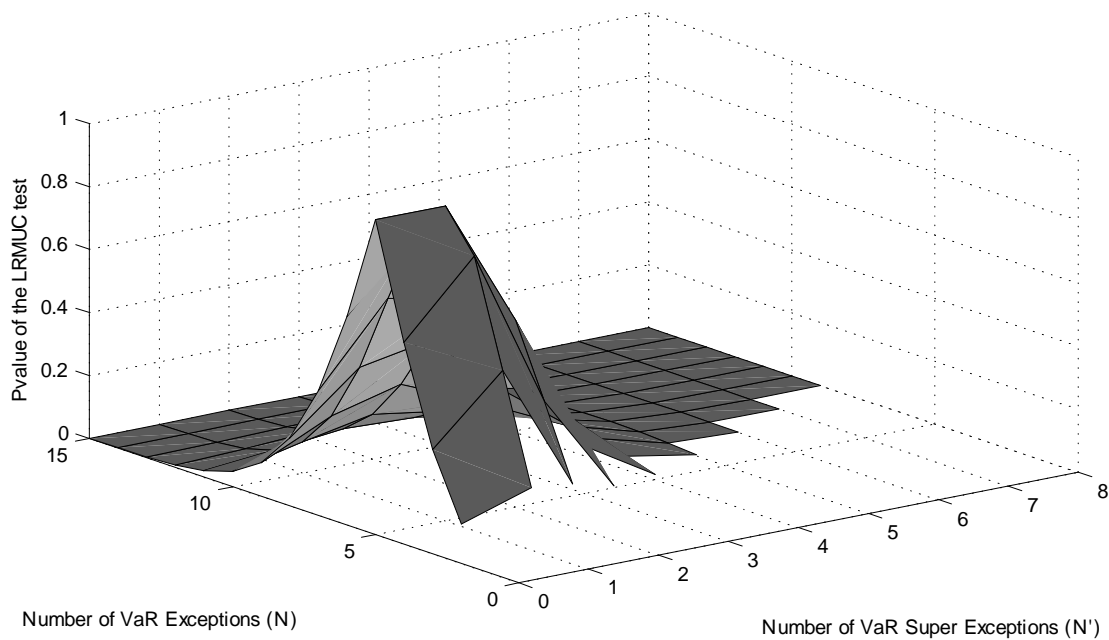
Figure 1: VaR Exception vs. VaR Super Exception



Notes: This figure displays the daily P&L,  $VaR(\alpha)$ , and  $VaR(\alpha')$  for a hypothetical portfolio, with  $\alpha = 1\%$  and  $\alpha' = 0.2\%$ . 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 five exceptions and four super exceptions, which are highlighted with stars.

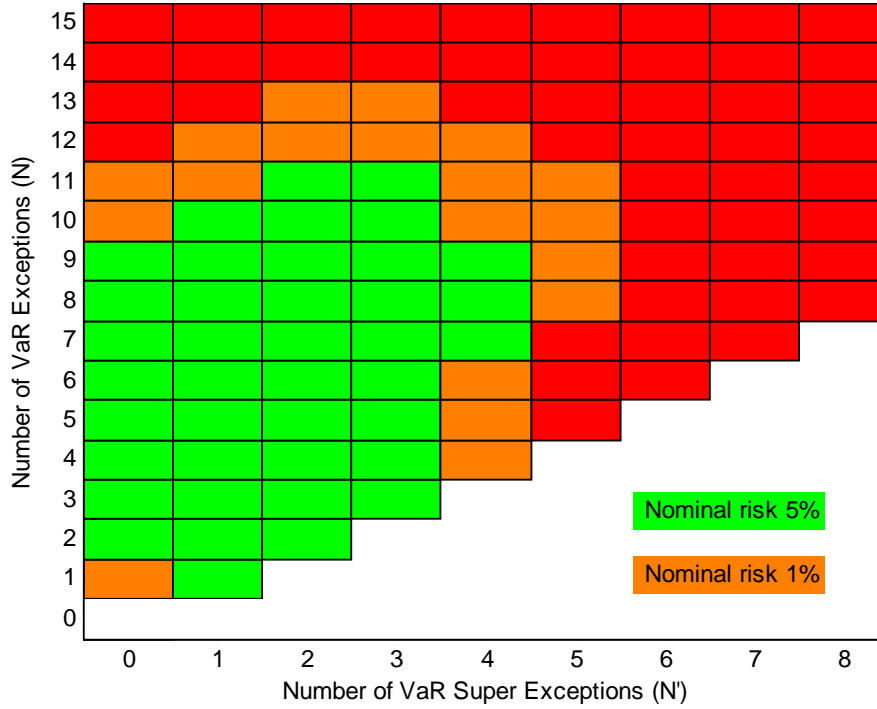


**Figure 2: P-values of the Multivariate Unconditional Coverage Test**



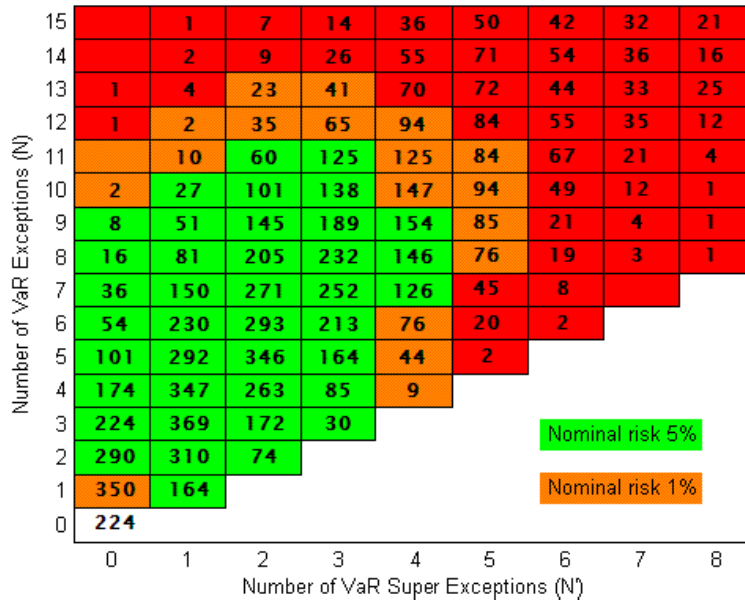
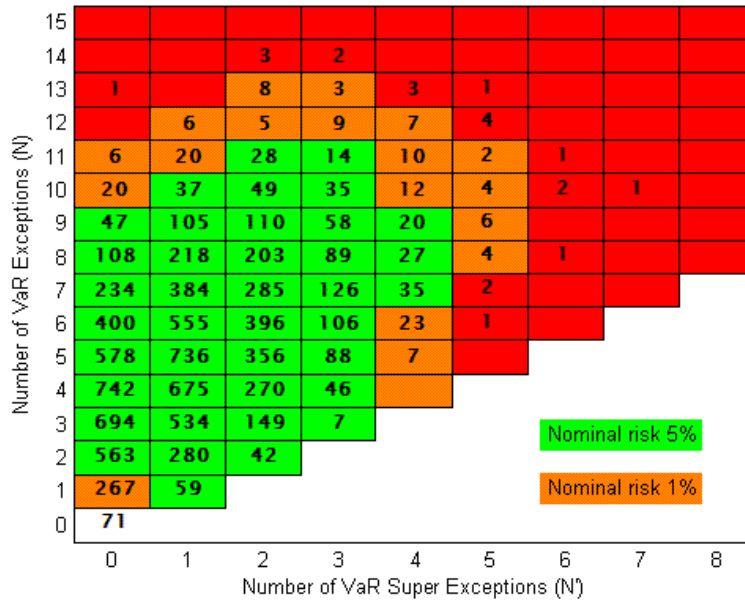
Notes: This figure displays the p-value of a multivariate unconditional coverage test,  $LR_{MUC}(\alpha, \alpha')$  for different numbers of exceptions ( $N$ ) and super exceptions ( $N'$ ). The parameter values are  $\alpha = 1\%$ ,  $\alpha' = 0.2\%$ , and  $T = 500$ .

Figure 3: The Risk Map



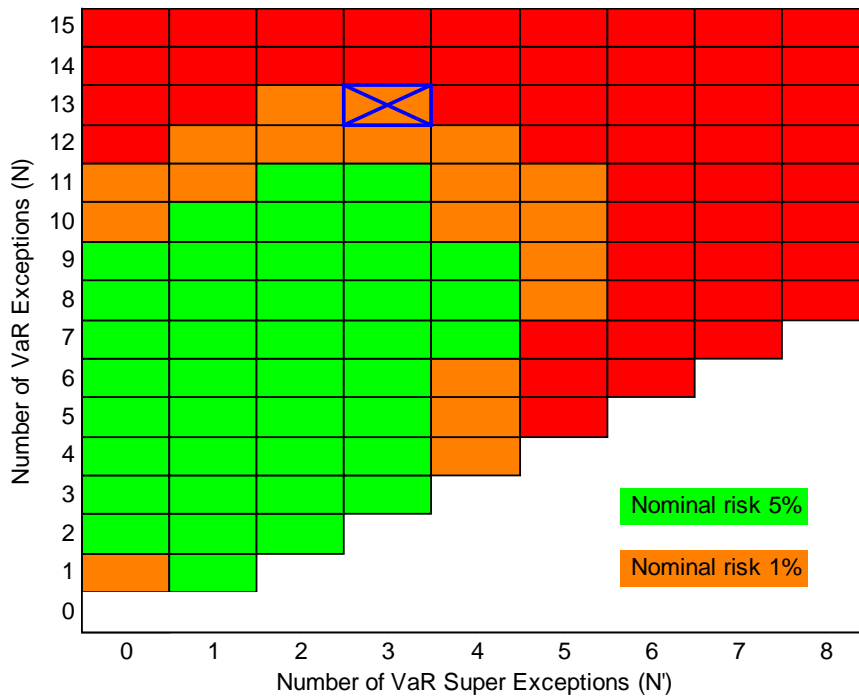
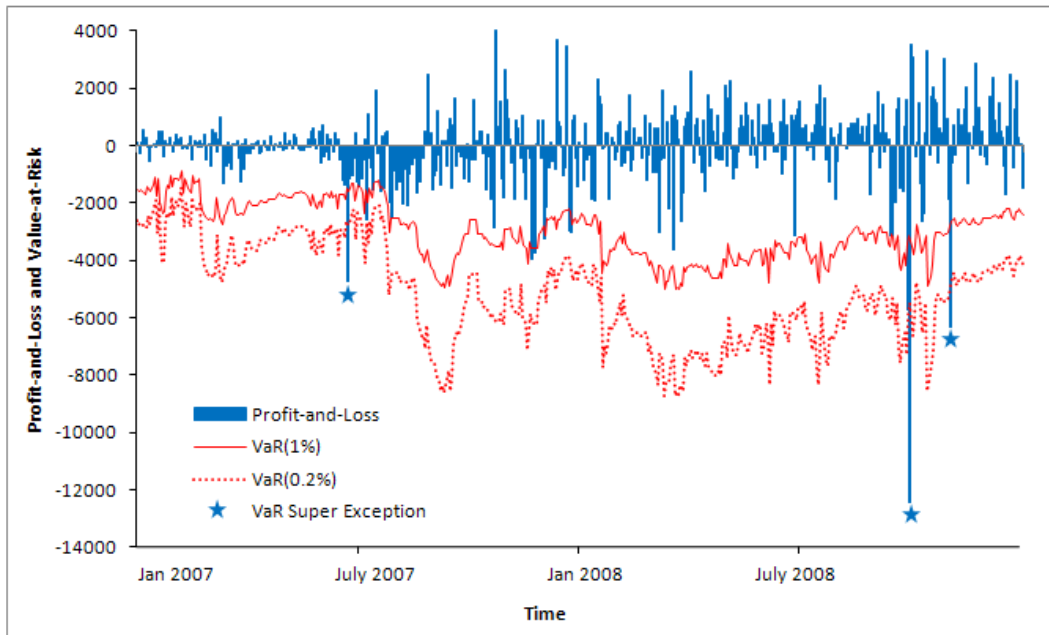
Notes: This figure displays a Risk Map based on the p-value of a multivariate unconditional coverage tests,  $LR_{MUC}(\alpha, \alpha')$ , for different numbers of VaR exceptions ( $N$ ) and VaR super exceptions ( $N'$ ). The parameter values are  $\alpha = 1\%$ ,  $\alpha' = 0.2\%$ , and  $T = 500$ . For technical reasons, the p-values for the first-column cells ( $N' = 0$ ) and diagonal cells ( $N' = N$ ) are computed from a univariate  $LR_{UC}$  test as the multivariate test cannot be computed in these cases. When  $N' = N = 0$ , none of the LR test can be computed.

Figure 4: Valid vs. Invalid Model



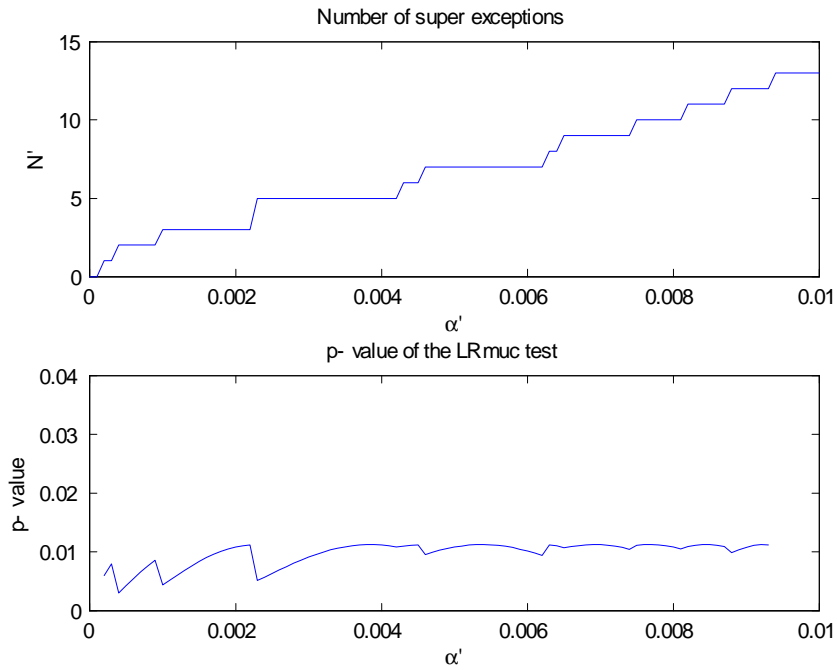
Notes: The upper Risk Map displays the empirical frequencies of each pair  $(N, N')$ , where  $N$  and  $N'$  denote the total number of VaR exceptions and super exceptions, respectively, obtained from 10,000 simulations of a valid VaR model. The lower Risk Map displays the empirical frequencies of the VaR exceptions and super exceptions generated from an Historical Simulation VaR methodology applied to simulated returns drawn according to the MS-GARCH of Bauwens, Preminger and Rombouts (2010). In both cases, parameter values are  $\alpha = 1\%$ ,  $\alpha' = 0.2\%$ , and  $T = 500$ .

Figure 5: Backtesting Results for a Large Bank



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, which are highlighted with stars. The lower graphs displays the Risk Map for La Caixa. The cross corresponds to La Caixa for the period 2007-2008. As the cross falls inside the orange zone, we can reject the null hypothesis at the 95% but not at the 99% confidence level.

**Figure 6: Changing the Value of the  $\alpha'$  Coverage Rate**



Notes: The upper graph displays the number of super-exceptions  $N'$  obtained for La Caixa for various coverage rates  $\alpha'$ . The lower graph displays the  $p$ -value of the  $LR_{MUC}$  test given  $N'$  and  $\alpha'$ .

**Table 1: Finite Sample Properties of the Risk Map**

<b>Panel A: Empirical Size</b>			
Sample	$\alpha = 5\%, \alpha' = 1\%$	$\alpha = 2\%, \alpha' = 0.4\%$	$\alpha = 1\%, \alpha' = 0.2\%$
500	0.047	0.043	0.046
1,000	0.050	0.041	0.043
2,000	0.054	0.053	0.039
<b>Panel B: Empirical Power of the Risk Map</b>			
Sample	$\alpha = 5\%, \alpha' = 1\%$	$\alpha = 2\%, \alpha' = 0.4\%$	$\alpha = 1\%, \alpha' = 0.2\%$
500	0.393	0.345	0.305
1,000	0.335	0.365	0.391
2,000	0.378	0.518	0.645
<b>Panel C: Empirical Power of the <math>LR_{UC}</math> Test</b>			
Sample	$\alpha = 5\%$	$\alpha = 2\%$	$\alpha = 1\%$
500	0.399	0.361	0.326
1,000	0.275	0.297	0.339
2,000	0.220	0.378	0.434

Notes: Panel A presents the empirical size of the Risk Map for various values of  $(\alpha, \alpha')$  and sample sizes. The empirical size corresponds to the rejection rate calculated over 10,000 simulations for a nominal size equal to 5%. Panel B reports the empirical power of the Risk Map when the P&L are generated from a MS-GARCH model and the VaR forecasts are computed through Historical Simulation. Panel C reports the power of the  $LR_{UC}$  tests for the 1%, 2%, and 5% coverage rates.

**Table 2: P&L and VaR of a Large Bank**

<b>Panel A: Summary Statistics</b>			
	P&L	VaR(1%)	VaR(0.2%)
Mean	-39.45	2,958.10	5,138.05
Median	50.00	3,000.00	5,211.24
Standard-Deviation	1,328.21	979.40	1,710.92
Skewness	-2.08	-0.02	-0.02
Excess Kurtosis	16.66	-0.88	-0.88
Minimum	-12,410	5,000	8,705
Maximum	4,000	850	1,455

<b>Panel B: <math>VaR(\alpha')</math> Estimation</b>		
	Estimates	(Standard-Errors)
Constant	$1.76e^{-5}$	$(8.79e^{-5})$
Garch parameter	0.6689	(0.074)
Arch parameter	0.3311	(0.132)
Degree of freedom	3.1020	(0.627)
Log-likelihood	$-4.180e^3$	

Notes: Panel A presents some descriptive statistics on 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. Panel B presents some detailed information about the estimation of  $VaR(\alpha' = 0.2\%)$ . We report the estimated parameters and standard errors for the auxiliary t-GARCH model, along with the value of the log-likelihood function. For each time series, the number of observations is 500.

**Table 3: Generalized Traffic Light System**

<b>T=250</b>		<b>Number of Super Exceptions (N<sup>2</sup>)</b>						
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Number of Exceptions (N)</b>	<b>0</b>							
	<b>1</b>	0.00	0.00					
	<b>2</b>	0.00	0.00	0.00				
	<b>3</b>	0.00	0.00	0.00	0.00			
	<b>4</b>	0.00	0.00	0.00	0.00	0.40		
	<b>5</b>	0.00	0.00	0.00	0.00	0.49	0.40	
	<b>6</b>	0.40	0.00	0.00	0.40	0.57	0.50	1.00
	<b>7</b>	0.52	0.40	0.40	0.50	0.66	0.60	1.00
	<b>8</b>	0.64	0.55	0.52	0.60	0.74	0.70	1.00
	<b>9</b>	0.76	0.70	0.64	0.70	0.83	0.80	1.00
	<b>10</b>	0.88	0.85	0.76	0.80	0.91	0.90	1.00
	<b>11</b>	1.00	1.00	0.88	0.90	1.00	1.00	1.00
	<b>12</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00

<b>T=500</b>		<b>Number of Super Exceptions (N<sup>2</sup>)</b>						
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Number of Exceptions (N)</b>	<b>0</b>							
	<b>1</b>	0.40	0.00					
	<b>2</b>	0.00	0.00	0.00				
	<b>3</b>	0.00	0.00	0.00	0.00			
	<b>4</b>	0.00	0.00	0.00	0.00	0.00		
	<b>5</b>	0.00	0.00	0.00	0.00	0.00	0.40	
	<b>6</b>	0.00	0.00	0.00	0.00	0.00	0.49	1.00
	<b>7</b>	0.00	0.00	0.00	0.00	0.00	0.57	1.00
	<b>8</b>	0.00	0.00	0.00	0.00	0.00	0.66	1.00
	<b>9</b>	0.00	0.00	0.00	0.00	0.00	0.74	1.00
	<b>10</b>	0.40	0.00	0.00	0.00	0.40	0.83	1.00
	<b>11</b>	0.70	0.40	0.00	0.00	0.60	0.91	1.00
	<b>12</b>	1.00	0.70	0.40	0.40	0.80	1.00	1.00
	<b>13</b>	1.00	1.00	0.70	0.70	1.00	1.00	1.00
	<b>14</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Notes: This table displays the increase in the multiplicative factor  $k$  that is applied to the average VaR to generate the Market Risk Charge ( $MRC = 12.5 \times k \times VaR$ ) within the Basel framework. Unlike the standard Traffic Light system, the Generalized Traffic Light system accounts for both the number of exceptions and super exceptions. By analogy to the standard Traffic Light system, the base multiplicative factor is set to a value of three and can increase up to four. The no penalty zone corresponds to the green zone of the Risk Map and the maximum penalty zone corresponds to the red zone of the Risk Rap. Between these two zones, the bank incurs an increasing penalty which is obtained by linear interpolation. The upper panel is for a sample size of  $T = 250$  whereas the lower panel is for a sample size of  $T = 500$ .



**Table A1: Monte Carlo Experiment**

<b>Panel A: Summary of the Monte Carlo Experiment</b>			
	<b>Experiment 1</b>	<b>Experiment 2</b>	<b>Experiment 3</b>
P&L Data Generating Process	t-Garch(1,1)	t-Garch(1,1)	MS-Garch(1,1)
Internal VaR Model	t-Garch(1,1)	HS	HS
Auxiliary Model	t-Garch(1,1)	t-Garch(1,1)	Garch(1,1)
<b>Panel B: Experiment 1 - Valid Internal VaR Model</b>			
<b>Sample Size</b>	<b>TPR</b>	<b>FPR</b>	<b>TNR</b>
250	0.8763	0.0002	0.9998
500	0.9220	0.0002	0.9998
1,000	0.9535	0.0001	0.9999
1,500	0.9562	0.0001	0.9999
2,000	0.9642	0.0001	0.9999
<b>Panel C: Experiment 2 - Invalid Internal VaR Model</b>			
<b>Sample Size</b>	<b>TPR</b>	<b>FPR</b>	<b>TNR</b>
250	0.7785	0.0006	0.9994
500	0.8134	0.0006	0.9994
1,000	0.8424	0.0007	0.9993
1,500	0.8366	0.0006	0.9994
2,000	0.8160	0.0007	0.9993
<b>Panel D: Experiment 3 - Invalid Internal and Auxiliary Models</b>			
<b>Sample Size</b>	<b>TPR</b>	<b>FPR</b>	<b>TNR</b>
250	0.9515	0.0020	0.9980
500	0.9606	0.0021	0.9979
1,000	0.9543	0.0021	0.9979
1,500	0.9514	0.0021	0.9979
2,000	0.9584	0.0022	0.9978

Notes: This table presents the design and the results of the Monte Carlo simulation. In the three experiments, we vary (1) the data generating process (DGP) of the profit-and-loss data, (2) the internal VaR model used by the bank, and (3) the auxiliary model used to generate the  $VaR(\alpha')$  estimates. In experiment 1, (1), (2), and (3) match perfectly. In experiment 2, only the internal model differ from the DGP, whereas in experiment 3, both internal and external models differ from the DGP. For each experiment, we report the True Positive Rate (TPR; fraction of super exceptions that occur concurrently for the true and for the estimated  $VaR(\alpha')$ ); the False Positive Rate (FPR; fraction of estimated super exceptions events that were observed to be non events); and True Negative Rate (TNR; fraction of non super exceptions that are correctly identified). In each experiment, we vary the sample size from 250 to 2,000 observations and we use 10,000 simulations.