Review of Futures

Volume 20 Special Edition

Research funding for this Special Edition was made possible by a generous donation from the Clearing Corporation Charitable Foundation to the Institute for Financial Markets.

IFM THE INSTITUTE FOR FINANCIAL MARKETS

In Cooperation with The IFM

MARGIN BACKTESTING

Christophe Hurlin and Christophe Pérignon*

This paper presents a validation framework for collateral requirements or margins on a derivatives exchange. It can be used by investors, risk managers, and regulators to check the accuracy of a margining system. The statistical tests presented in this study are based either on the number, frequency, magnitude, and timing of margin exceedances, which are defined as situations in which the trading loss of a market participant exceeds his or her margin. We show that these validation tests can be implemented at the individual level or at the global exchange level.

What makes derivatives exchanges so special is the extremely low default risk that market participants are exposed to. Collateral requirements or margins are the major tools to protect derivatives users against the default of their counterparties. The challenge faced by derivative exchanges is to set margins high enough to mitigate default risk but not so high as to shy traders away and damage liquidity. The goal of this paper is to design a methodological framework allowing risk managers and regulators to check the validity of the margins charged to derivatives users. It consists of a series of diagnostic tools allowing one to detect misspecified models that lead to margins that are either excessively conservative or lenient. Checking the validity of a margining system is particularly important nowadays as more and more over-the counter (OTC) derivatives products are migrating to clearing platforms (Duffie and Zhu 2010).[1]

There are two types of margining systems used in practice: the Standard Portfolio Analysis of Risk (hereafter SPAN) system and the Value-at-Risk (hereafter VaR) model. Both margining systems consider a series of scenarios representing potential one-day ahead changes in the underlying assets' price and volatility and

---

1. The clearing activity consists in confirming, matching, and settling all trades on an exchange. In order to reduce the risk of non-performance, exchange-traded derivatives are guaranteed against counterparty failure by a central counterparty clearing house. On most derivatives exchanges, only a subset of market participants (i.e., the clearing members) can directly trade with the clearing house whereas all non-clearing member participants have to trade through a designated clearing member.

---

*Christophe Hurlin is a professor at University of Orléans, Laboratoire d'Economie d'Orléans. Email: christophe.hurlin@univ-orleans.fr
Christophe Pérignon (the corresponding author) is an associate professor in the Department of Finance, HEC Paris. E-mail: perignon@hec.fr.

Acknowledgements: The authors acknowledge the grant made possible through the Institute for Financial Markets.

Keywords: collateral requirements, futures markets, tail risk, derivatives clearing
JEL Classification: G13

generate simulated distributions of potential profit-and-loss (hereafter P&L) for derivatives users. Under SPAN, the system selects for each position the largest loss across all considered scenarios, combines financial instruments within the same underlying asset, and total margin is given by the sum of the risk of all underlying assets less some diversification adjustments (CFTC 2001; Chicago Mercantile Exchange 2009). Differently, VaR margins are set such that the probability of the loss on the entire derivatives portfolio exceeding the margin is equal to a pre-specified level, such as 1% (Knott and Mills 2002; Cruz Lopez, Harris, and Pérignon 2011).

On a regular basis, the risk-management department of the clearing-house and the regulatory agencies check the validity of the margining system. In particular, they make sure that the hypothetical shocks used in the scenarios are extreme enough and that the estimation of the derivative prices is reliable. Of particular concern is a situation in which margins are 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 (Jones and Pérignon 2012).

While the performance of the SPAN system has been investigated in a number of papers (Kupiec 1994; Kupiec and White 1996; Eldor, Hauser, and Yaari 2011), VaR margins have not to our knowledge been investigated in the academic literature. This increasingly-popular modeling approach offers several advantages though. First, as it is based on a quantile, it allows derivatives exchanges to pick the level of tail risk that best fits with their risk tolerance. A second advantage is that quantile-based margins are less sensitive to simulation design than maximum-based margins, such as SPAN margins. Most importantly for this study, quantile-based margins can be validated ex-post using formal backtesting methodologies. For instance, as an α% quantile is by definition exceeded α% of the time, one can check whether in reality α% VaR margins are indeed exceeded α% of the time.

Compared to market risk VaR (Jorion 2007; Christoffersen 2009a), which is used by banks to monitor their trading risk and compute capital requirements, the estimation of VaR margin is much simpler. In general, the quantile of the return at time t cannot be estimated without making some strong assumptions about the underlying distribution. Specifically, since there is only one return observation on each date, it is usually assumed that the returns are independently and identically distributed over time. Under these assumptions, VaR can be estimated from the historical path of past returns. In the context of VaR margin; however, the situation is quite different because P&L observations are simulated at time t. This is an ideal situation from an econometric point of view because the quantile of the P&L distribution can be directly estimated without making any assumptions regarding its behavior over time.

Our main contribution to the literature on derivatives margins is to present a backtesting framework for derivatives margins. It consists of a series of hypotheses that must be validated by a well-functioning margin model. Then, we propose a series of statistical tests that aim to test these hypotheses in order to detect misspecified margining models. We show that these validation tests can be implemented either at the indivual investor level or at the global exchange level. In this framework, not only can we find out whether a model is misspecified but we can also unmask the reasons of rejection of a misspecified model. Finally, in order to ease the implementation of the backtesting methodologies presented in this paper, we created a website on which users can freely upload their margins and P&L data and run the associated computer codes (www.RunMyCode.org).

The outline of the paper is the following. In Section I, we discuss how to estimate VaR margins and present the main testable hypotheses. In Section II, we show how to test these hypotheses in order to validate or invalidate a given margining model. We present in Section III some statistical test that aim to validate the margining model at the exchange level. Section IV summarizes and concludes our paper.

I. MARGIN ESTIMATION AND TESTABLE HYPOTHESES

A. Margin Estimation

For retail investors, margins are typically set at the contract level (e.g., \$1,000 for any long or short position in a given futures contract). Depending on the expected volatility, the derivatives exchange can adjust the level of the margin, as shown by Brunnermeier and Pedersen (2009, Figure 1) for the S&P 500 futures. Differently, for large market participants such as clearing members, margins are computed at the portfolio level in order to account for diversification effects and are adjusted daily. The VaR margin B_i is set such that there is a probability α that the loss on the derivative position exceeds the margin:

$$\Pr\left[V_{i,t} < -B_{i,t|t-1}(\alpha)\right] = \alpha \tag{1}$$

where V_i denotes the P&L of investor i, and α is called the coverage rate. Let $\omega_{i,t-1}$ be the vector of positions of clearing member i at the end of day t–1:

$$\omega_{i,t-1} = \begin{bmatrix} \omega_{i,1,t-1} \\ \vdots \\ \omega_{i,D,t-1} \end{bmatrix} \tag{2}$$

where D is the number of derivatives contracts (futures and options) traded on this exchange and $i = 1, ..., N$. To arrive at a margin for this portfolio, the clearing house considers a series of S scenarios representing potential one-day ahead 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:

$$v_{i,t} = \begin{bmatrix} v_{i,t}^{1} \\ \vdots \\ v_{i,t}^{S} \end{bmatrix}. \tag{3}$$

Given the simulated path $\{v_{i,t}^{s}\}_{s=1}^{S}$, the VaR margin for clearing member i is given by:

$$\widehat{B}_{i,t} = \text{percentile}\left(\{v_{i,t}^{s}\}_{s=1}^{S}, 100\alpha\right). \tag{4}$$

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.

B. Backtesting VaR Margin

Traditionally the quality of the forecast of an economic variable is assessed by comparing its *ex-post* realization with the *ex-ante* forecast value. The comparison of the various forecast models is thus generally made by using a criterion such as the Mean Squared Error criterion or standard information criteria (AIC and BIC). However, this approach is not suitable for VaR margin forecasts because the true quantile of the P&L distribution is not observable. That is why VaR assessment is generally based on the concept of margin exceedance (also called hit, violation, or exception).

For a given clearing member *i*, a margin exceedance is said to occur if the *ex-post* realization of the P&L at time t, $V_{i,t}$, is more negative than the *ex-ante* VaR margin forecast. Let $I_t(\alpha)$ be a binary variable associated with an α% VaR margin at time *t* (we omit the index *i* for simplicity):

$$I_t(\alpha) = \begin{cases} 1 & \text{if } V_{i,t} < -B_{i,t|t-1}(\alpha) \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

As stressed by by Christoffersen (1998, 2009b), VaR forecasts are valid if and only if the violation process $I_t(\alpha)$ satisfies the following two hypotheses:

- The Unconditional Coverage (hereafter UC) hypothesis: The probability of an *ex-post* return exceeding the VaR forecast must be equal to the α coverage rate:

$$\Pr[I_t(\alpha) = 1] = \mathbb{E}[I_t(\alpha)] = \alpha. \tag{6}$$

- The Independence (hereafter IND) hypothesis: VaR margin violations observed at two different dates for the same coverage rate must be distributed independently. Formally, the variable $I_t(\alpha)$ associated with a margin exceedance at time t for an α% coverage rate should be independent of the variables $I_{t-k}(\alpha)$, $\forall k \neq 0$. In other words, past VaR violations should not be informative about current and future violations.

The UC hypothesis is quite intuitive. Indeed, if the frequency of violations observed over T days is significantly lower (respectively higher) than the coverage rate α, then risk is overestimated (respectively underestimated). However, the UC hypothesis sheds no light on the possible dependence of margin exceedances. Therefore, the independence property of violations is an essential one, because it is related to the ability of a VaR margin model to accurately model the higher-order dynamics of the P&L. In fact, a model that does not satisfy the independence property can lead to clusterings of margin exceedances even if it has the correct average number of violations. Consequently, there must be no dependence in the violations variable, whatever the coverage rate considered.

When the UC and IND hypotheses are simultaneously valid, VaR forecasts are said to have a correct Conditional Coverage (hereafter CC), and the VaR violation process is a martingale difference with:

$$\mathbb{E}\left[I_t(\alpha) - \alpha \mid \Omega_{t-1}\right] = 0. \tag{7}$$

This last property is at the core of most of the validation tests for VaR models (Christoffersen 1998; Engle and Manganelli 2004; Berkowitz, Christoffersen, and Pelletier 2011). It is worth noting that equation (CC) implies that the violation $I_t(\alpha)$ has Bernoulli distribution with a success probability equal to α:

$$\{I_t(\alpha)\} \text{ are } i.i.d. \text{ Bernoulli}(\alpha). \tag{8}$$

II. TESTS OF MARGIN ACCURACY

A. Frequency of Margin Exceedances

A first way of testing margin accuracy is to test the number or the frequency of margin exceedances. Thus the null hypothesis corresponds to equation (6):

$$H_{0,UC} : \mathbb{E}[I_t(\alpha)] = \alpha. \tag{9}$$

A first statistical test, called the *Z*-test, is based on a normal approximation and the assumption of independence. Consider a sequence $\{I_t(\alpha)\}_{t=1}^{T}$ of T margin exceedances associated to VaR (α%) margins and denote by H the total number of exeedances or hits, $H = \sum_{t=1}^{T} I_t(\alpha)$. If we assume that the variables $I_t(\alpha)$ are *i.i.d.*, then under the null of UC, the total number of hits has a Binomial distribution:

$$H \sim B(T, \alpha) \tag{10}$$

with $\mathbb{E}(H) = \alpha T$ and $\mathbf{V}(H) = \alpha(1\text{-}\alpha)T$. For a large T sample the Binomial distribution can be approximated by a normal distributon and a simple Z-test statistic can be defined as:

$$Z = \frac{H - \alpha T}{\sqrt{\alpha(1-\alpha)T}} \approx N(0,1). \tag{11}$$

Alternatively, Kupiec (1995) and Christoffersen (1998) propose a Likelihood Ratio (hereafter LR) test based on the process of *VaR* margin exceedances $I_t(\alpha)$. Under H_0, the LR statistic is defined as:

$$LR_{UC} = -2\ln\left[(1-\alpha)^{T-H}\alpha^{H}\right] + 2\ln\left[\left(1-\frac{H}{T}\right)^{T-H}\left(\frac{H}{T}\right)^{H}\right] \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(1). \tag{12}$$

Under the null (9), the LR_{UC} statistic converges to a chi-square distribution with two degrees of freedom. The intuition for the LR test is the same as for the Z statistics. The null of UC is not rejected if the empirical frequency of VaR margin exceedances H/T is close enough to the coverage rate α. Jorion (2007) reports some non-rejection regions for the LR_{UC} test. For a 5% nominal size and sample size T = 250, the UC assumption is not rejected if the total number of VaR(1%) violations is strictly smaller than 7. If the sample size is equal to 500, the total number of exceedances must strictly range between 1 and 11.

B. Frequency and Severity of Margin Exceedances

A key limitation of the previous approach is that it is unable to distinguish between a situation in which losses are below but close to the margin and a situation in which losses are considerably below the margin. Colletaz, Hurlin, and Pérignon (2012) propose a backtesting methodology that is based on the number and the severity of VaR exceptions. Their approach exploits the concept of *super exception*, which is defined as a loss greater than a super VaR margin $B_{i,\,t|t-1}(\alpha')$ whereas the coverage probability α' is much smaller than α (e.g., α = 1% and α' = 0.2%). As in Section I.B, we define a hit variable associated with $B_{i,\,t|t-1}(\alpha')$:

$$I_t(\alpha') = \begin{cases} 1 & \text{if } V_{i,t} < -\beta_{i,t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \alpha' < \alpha \tag{13}$$

The defining feature of their approach is to account for both the frequency and the magnitude of VaR margin exceedances. The intuition is the following. If the frequency of super exceptions is abnormally high, this means that the magnitude of the losses with respect to $B_{i,\,t|t-1}(\alpha)$ is too large. For both VaR margin exceptions and super exceptions, they propose to use a standard backtesting procedure. Consider a time series of T VaR margin forecasts for an α (respectively α') coverage rate and let H (respectively H') be the number of associated VaR margin violations:

$$H = \sum_{t=1}^{T} I_t(\alpha) \qquad H' = \sum_{t=1}^{T} I_t(\alpha'). \tag{14}$$

Colletaz, Hurlin and Pérignon (2012) propose a new tool, called the *Risk Map*, which graphically summarizes all information about the performance of a VaR model. It is based on a joint test of the number of VaR exceptions and VaR super exceptions:

$$H_{0,MUC} : \mathbb{E}[I_t(\alpha)] = \alpha \text{ and } \mathbb{E}[I_t(\alpha')] = \alpha'. \tag{15}$$

The corresponding test statistic consists in a multivariate unconditional coverage test. This test is based on three indicator variables:

$$J_{0,t} = 1 - J_{1,t} - J_{2,t} = 1 - I_t(\alpha) \tag{16}$$

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

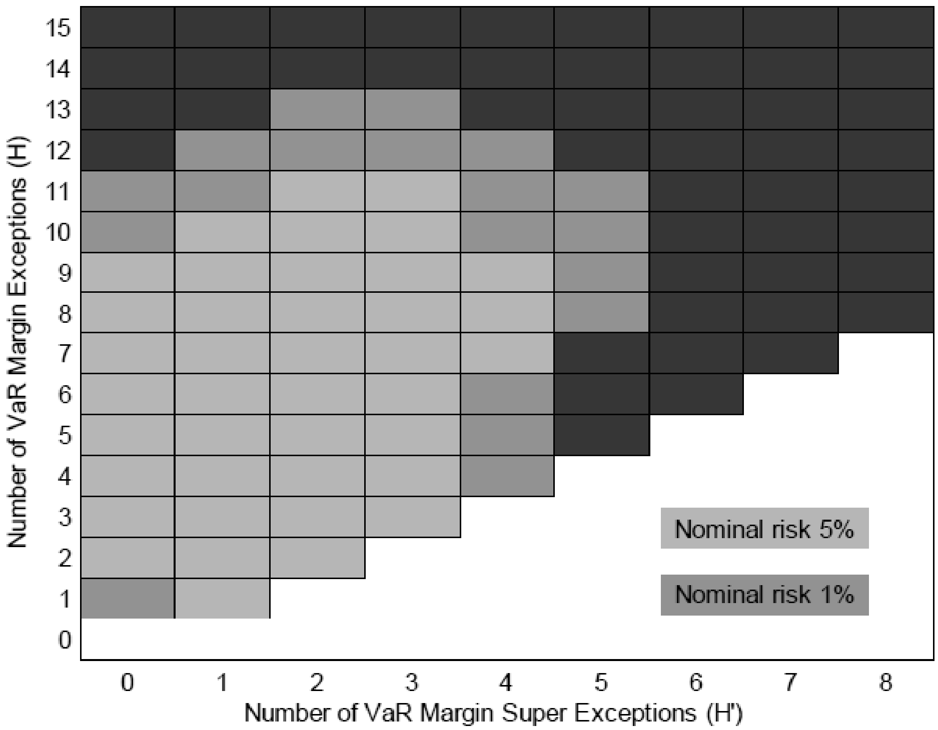
$$J_{2,t} = I_t(\alpha') = \begin{cases} 1 & \text{if } V_{i,t} < -B_{i,t|t-1}(\alpha') \\ 0 & \text{otherwise} \end{cases}. \tag{18}$$

The $\{J_{i,t}\}_{i=0}^{2}$ are Bernoulli random variables equal to one with probability $1-\alpha$, $\alpha-\alpha'$, and α', respectively. Given these definitions, we can test the joint hypothesis (15) using a LR test. Let us denote $H_i = \sum_{t=1}^{T} J_{it}$, for $i = 0, 1, 2$, the count variable associated with each of the Bernoulli variables. The multivariate unconditional coverage test is an LR test that the empirical exception frequencies significantly deviate from the theoretical ones. Formally, it is given by:

$$LR_{MUC}(\alpha,\alpha') = -2\ln\left[(1-\alpha)^{H_0}(\alpha-\alpha')^{H_1}(\alpha')^{H_2}\right] + 2\ln\left[\left(1-\frac{H_0}{T}\right)^{H_0}\left(\frac{H_0}{T}-\frac{H_1}{T}\right)^{H_1}\left(\frac{H_2}{T}\right)^{H_2}\right] \xrightarrow[T\to\infty]{d} \chi^2(2). \tag{19}$$

A Risk Map can be constructed based on the rejection zones for different confidence levels (Figure 1). Note that the cells below the diagonal are not colored as they correspond to situations in which the number of super exceptions exceeds the number of exceptions, which is of course impossible. If the (H, H′) pair corresponds to a light gray 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 (H, H') falls in the gray zone, we can reject the null at the 95% but not at the 99% confidence level.

Figure 1. Backtesting VaR Margins with 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 margin exceptions (H) and VaR margin super exceptions (H'). Parameter values are α=%, α'=0.2%, and T=500.

Finally, a dark gray cell implies that we can reject the null hypothesis at the 99% confidence level.

C. Independence of Margin Exceedances

The UC property does not give any information about the temporal independence of VaR margin exceedances. However, generating margin exceedances that are temporally independent is an important property for a margining system to have since it suggests that the margin immediately reflects new information. A margining system that violates this property leads to clusters of margin exceedances.[2]

It is important to note that these two VaR margin properties are independent one from the other. At this point, if a VaR margin does not satisfy either one of these two hypotheses, it must be considered as not valid. For example, satisfying the hypothesis of unconditional coverage does not compensate for the possible existence of violations clusters nor the noncompliance with the independence

2. Berkowitz and O'Brien (2002) show that the VaR models used by six large U.S. commercial banks (1) tend to be very conservative, at least when financial markets are not under stress and (2) lead to clusters of VaR exceedances. This second result indicates that risk models fail to forecast volatility changes.

hypothesis. On the contrary, there is CC when the VaR margin satisfies both the UC and IND hypotheses.

1. LR Approach

Christoffersen (1998) proposes an LR test based on the assumption that the process of *VaR* margin exceedances $I_t(\alpha)$ is modeled with the following matrix of transition probabilities:

$$\Pi = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix} \tag{20}$$

where $\pi_{ij} = \Pr[I_t(\alpha) = j \mid I_{t-1}(\alpha) = i]$, that is, probability of being in state j at time t conditioning on being in state i at time $t-1$. Under the null of independence, we have $\pi_{01} = \pi_{11} = \beta$ and:

$$H_{0,IND} : \Pi_\beta = \begin{pmatrix} 1-\beta & \beta \\ 1-\beta & \beta \end{pmatrix} \tag{21}$$

where β denotes a denotes a margin exceedance probability, which can be different from the coverage rate α. What these transition probabilities imply is that the probability of experiencing a margin exceedance in the current period depends on the occurrence or not of a margin exceedance in the previous period. The estimated VaR margin exceedance probability is the empirical frequency of violations, *H/T*. Under the alternative, no restriction is imposed on the Π matrix. The corresponding LR statistic, denoted LR_{IND} is defined by:

$$\begin{aligned} LR_{IND} &= -2\ln\left[\left(1-\frac{H}{T}\right)^{T-H}\left(\frac{H}{T}\right)^{H}\right] \\ &\quad + 2\ln[(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}] \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(1) \end{aligned} \tag{22}$$

where n_{ij} denotes the number of times we have $I_t(\alpha) = j$ and $I_{t-1}(\alpha) = i$, and:

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}} \qquad \hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}}. \tag{23}$$

Finally, it is also possible to test the CC assumption for VaR margins. Under CC:

$$H_{0,CC} : \Pi_\alpha = \begin{pmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{pmatrix} \tag{24}$$

and then:

$$LR_{CC} = -2\ln\left[(1-\alpha)^{T-H}(\alpha)^{H}\right] + 2\ln[(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}] \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(2) \tag{25}$$

The corresponding *LR* statistic, denoted LR_{CC}, is defined by the sum of the LR_{UC} and LR_{IND} statistics. Under the null of CC, it satisfies:

$$LR_{CC} = LR_{UC} + LR_{IND} \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(2). \tag{26}$$

2. Regression-based Tests

Engle and Manganelli (2004) suggest another approach based on a linear regression model. This model links current margin exceedances to past exceedances and/or past information. Let $Hit(\alpha) = I_t(\alpha) - \alpha$ be the demeaned process associated with $I_t(\alpha)$:

$$Hit_t(\alpha) = \begin{cases} 1-\alpha & \text{if } V_{i,t} < -B_{i,t|t-1}(\alpha) \\ -\alpha & \text{otherwise} \end{cases}. \tag{27}$$

Consider the following linear regression model:

$$Hit_t(\alpha) = \delta + \sum_{k=1}^{K} \beta_k Hit_{t-k}(\alpha) + \sum_{k=1}^{K} \gamma_k z_{t-k} + \varepsilon_t \tag{28}$$

where the z_{t-k} variables belong to the information set Ω_{t-1}. For example, one can use lagged P&L, squared past P&L, past margins, and so on. Whatever the chosen specification, the null hypothesis test of conditional efficiency corresponds to testing the joint nullity of all the regression coefficients:

$$H_{0,CC} : \delta = \beta_k = \gamma_k = 0, \quad \forall k = 1,..,K. \tag{29}$$

The independence hypothesis implies that β_k and γ_k coefficients are equal to zero whereas the unconditional coverage hypothesis is verified when δ is null. Indeed, under the null hypothesis, $\mathbb{E}[Hit_t(\alpha)] = \mathbb{E}(\varepsilon_t) = 0$, which implies by definition that $\Pr[I_t(\alpha) = 1] = E[I_t(\alpha)] = \alpha$.

Denote the vector $\Psi = (\delta \;\; \beta_1 \ldots \beta_K \; \gamma_1 \; \ldots \gamma_K)'$ of the $2K + 1$ parameters in this model and Z the matrix of explanatory variables of model (28), the Wald statistic, denoted DQ_{CC}, in association with the test of CC hypothesis then verifies:

$$DQ_{CC} = \frac{\hat{\Psi}' Z' Z \hat{\Psi}}{\alpha(1-\alpha)} \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(2K+1) \tag{30}$$

where $\hat{\Psi}$ is the OLS estimate of Ψ. Notice that one can also test the UC hypothesis by testing $H_{0,UC}: \delta = 0$ or test the IND hypothesis with $H_{0,\,\mathrm{IND}}: \beta_k = \gamma_k = 0$. A natural extension of the test of Engle and Manganelli (2004) consists in considering a (probit or logit) binary model linking current violations to past ones (Patton 2002; Dumitrescu, Hurlin, and Pham 2012).

3. Autocorrelation Test

Rather than using a regression model, Berkowitz, Christoffersen, and Pelletier (2011) test directly the martingale difference assumption. As under CC, the VaR margin exceedance process $Hit_t(\alpha)$ is a martingale difference; it should be uncorrelated. A natural test is the univariate Ljung-Box test of $H_{0,\,CC}: r_1 = \ldots = r_K = 0$ where r_k denotes the k^{th} autocorrelation:

$$LB(K) = T(T+2)\sum_{k=1}^{K} \frac{\hat{r}_k^2}{T-k} \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(K) \tag{31}$$

where $\hat{r}_k$ is the empirical autocorrelation of order k of the $Hit(\alpha)$ process.

D. Duration between Margin Exceedances

The UC, IND, and CC hypotheses also have some implications on the time between two consecutive VaR margin exceedances. Following Christoffersen and Pelletier (2004), we denote by d_v the duration between two consecutive VaR margin violations:

$$d_v = t_v - t_{v-1} \tag{32}$$

where t_v denotes the date of the v^{th} exceedance. Under CC hypothesis, the duration process d_i has a probability density function given by:

$$f(d_v;\alpha) = \alpha(1-\alpha)^{d_v - 1} \quad d_v \in \mathbb{N}^*. \tag{33}$$

This distribution characterizes the memory-free property of the VaR margin violation process $I_t(\alpha)$, which means that the probability of observing a violation today does not depend on the number of days that have elapsed since the last violation. Note that $\mathbb{E}(d_v) = 1/\alpha$ since the CC hypothesis implies an average duration between two margin exceedances equals to $1/\alpha$. The general idea of the test consists in specifying a distribution that nests equation (33), so that the memoryless property can be tested through parameter restriction. In this line, Christoffersen and Pelletier (2004) use under the null hypothesis the exponential distribution, which is the continuous analogue of the probability density function in equation (33):

$$g(d_v;\alpha) = \alpha \exp(-\alpha d_v). \tag{34}$$

Under the alternative hypothesis, Christoffersen and Pelletier (2004) postulate a Weibull distribution for the duration variable:

$$h(d_v; a, b) = a^b b d_v^{b-1} \exp\left[-(a d_v)^b\right]. \tag{35}$$

As the exponential distribution corresponds to a Weibull distribution with $b = 1$, the test for IND is:

$$H_{0,IND} : b = 1 \tag{36}$$

and for CC is:

$$H_{0,CC} : b = 1,\ a = \alpha \tag{37}$$

Christoffersen and Pelletier (2004) propose the corresponding LR test (see also Haas 2005), and Candelon et al. (2011) derive a GMM duration-based test.

III. TESTS OF GLOBAL VALIDITY

To the best of our knowledge, all empirical studies on VaR backtesting considers individual banks in isolation (Berkowitz and O'Brien 2002; Pérignon and Smith 2010; Berkowitz et al. 2011). The reason for doing so is that financial institutions use different proprietary risk models, which needs to be tested separately. Differently on a derivatives exchange, the margins of all market participants are computed using the same model developed by the clearing house. Hence, this model can be tested globally using information from all market participants, which helps in detecting misspecified models.

A. Definitions

Let us denote $I_{i,t}(\alpha)$ the VaR margin exceedance for clearing member i at time t. We define the Global Unconditional Coverage (hereafter GUC) hypothesis as a situation where the probability of an *ex-post* loss exceeds the VaR margin forecast is equal to the α coverage rate for all clearing members:

$$H_{0,GUC} : \mathbb{E}[I_{i,t}(\alpha)] = \alpha \quad \forall i = 1,..,N. \tag{38}$$

The GUC means that the frequency of VaR margin exceedances is accurate for all clearing members. Note that it is important not to pool the N margin exceedance processes. Indeed, an under-estimation of the margin for member i could be offset by an over-estimation of the margin for another member j. Thus, the GUC hypothesis requires the UC hypothesis to be valid for all clearing members.

We proceed in a similar way for the Global Independence (hereafter GIND) hypothesis. Under GIND, the VaR margin exceedances observed for all the members at two different dates are independent; that is, $I_{i,t}(\alpha)$ is independent from $I_{i,t-k}(\alpha)$,

$\forall k \neq 0$. Furthermore, $I_{i,t}(\alpha)$ is also independent from past (and future) VaR margin exceedances of other members $I_{j,t-k}(\alpha)$, $\forall k \neq 0$ and $j \neq i$. Notice that we allow for contemporaneous dependencies between VaR margin exceedances of different members.

Finally, the global conditional coverage (GCC) hypothesis corresponds to a case where the N margin exceedance processes are a martingale difference:

$$H_{0,GCC} : \mathbb{E}\left[I_{i,t}(\alpha) | \Omega_{t-1} \right] = \alpha \quad \forall i = 1,..,N \tag{39}$$

where Ω_{t-1} denotes the information set available at time t – 1 for all the members, including past values of VaR margin and VaR margin exceedances of other members j.

A natural test for the GUC hypothesis consists in testing the null (38) against the following alternative:

$$H_{1,GUC} : \mathbb{E}[I_{i,t}(\alpha)] \neq \alpha \quad \text{for } i \in S \tag{40}$$

$$\mathbb{E}[I_{i,t}(\alpha)] = \alpha \quad \text{for } i \in \overline{S} \tag{41}$$

where $dim(S) = N_1$ satisfies $1 < N_1 \leq N$ and $dim(\overline{S}) = N_2$ with $N_1 + N_2 = \mathrm{N}$. Under this alternative, the margin of at least one member does not satisfy the UC hypothesis. Similarly, a natural test of GCC is based on the null (GCC) against the alternative:

$$H_{1,GCC} : \mathbb{E}\left[I_{i,t}(\alpha) | \Omega_{t-1} \right] \neq \alpha \quad \text{for } i \in S \tag{42}$$

$$\mathbb{E}\left[I_{i,t}(\alpha) | \Omega_{t-1} \right] = \alpha \quad \text{for } i \in \overline{S}. \tag{43}$$

B. Testing Strategies

Let us consider an individual test statistic of the UC (or CC) hypothesis, denoted X_i specific to clearing member i. For instance, for the *UC* test, this statistic corresponds to the LR_{UC} statistic or the duration-based LR_{UC} statistic. For the CC test, this statistic corresponds to the LR_{CC} statistic, DQ statistic, or duration-based statistic LR_{CC}. Whatever the chosen test, the individual statistic for member i can be expressed as a non-linear function of the sequence of the margin exceedances of this member, that is, $X_i = g\,(I_{i,1}(\alpha), ..., I_{i,T}(\alpha))$. To test the GUC or GCC null hypothesis, we follow Im, Pesaran, and Shin (2003) and use the average of the individual statistics:

$$\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\sum_{i=1}^{N} g(I_{i,1}(\alpha),..,I_{i,T}(\alpha)). \tag{44}$$

If we assume margin exceedances are cross-sectionally independent, that is, $I_{i,t}$ are independent of $I_{j,s}$ for $i \neq j$ and all (t,s), the $\overline{X}_N$ statistic converges to a normal distribution when T and N grow large. The intuition is as follows. When T tends to infinity, each individual statistic X_i converges to the same distribution. For instance, the LR_{UC} statistic converges to a chi-square distribution. Under the cross-sectional independence assumption, the individual statistics $X_i = g(I_{i,1}(\alpha), ..., I_{i,T}(\alpha))$ are also independent. Thus, the individual statistics X_i are independently and identically distributed. The central limit theorem is then sufficient to show that the cross-sectional average mean $\overline{X}_N$ converges to a normal distribution when N tends to infinity:[3]

$$\overline{X}_N \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} N(0,1). \tag{45}$$

An alternative testing strategy consists in combining the p-values associated with the N individual tests. A Fisher type test is then defined by:

$$P_{X_N} = -2\sum_{i=1}^{N}\log(p_i) \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(2N). \tag{46}$$

For any statistic X_i, such as LR_{UC}, LR_{CC}, or DQ_{CC}, its p-value is uniformly distributed over [0, 1]. Under the assumption of cross-sectional independence, P_{X_N} has a chi-square distribution with $2N$ degrees of freedom. For large N samples, we can use a standardized statistic:

$$Z_{X_N} = -\frac{\sum_{i=1}^{N}\log(p_i)+N}{\sqrt{N}} \underset{N,T\to\infty}{\overset{d}{\longrightarrow}} N(0,1). \tag{47}$$

IV. CONCLUSION

Having a well-functioning margining system is a prerequisite for any derivatives exchange. It allows the exchange to closely monitor tail risk and make the system resilient. In this paper, we have provided a backtesting framework allowing investors, risk managers and regulators to validate margin models. The statistical tests we have presented capture different facets of the margin model performance including frequency, timing, and magnitude of margin exceedances. Rather than being substitutes, the different statistical tests appear to complement each other and can be used to identify the source(s) of model misspecification.

The quest for the ideal margining system is still ongoing. Market participants and regulators want collateral requirements to be less procyclical in order to prevent liquidity spiral (Brunnermeier and Pedersen 2009). What we show in this paper is a

3. When the contemporaneous exceedances $I_{i,t}$ and $I_{j,t}$ are correlated, the distribution of the average statistic $\overline{X}_N$ can be estimated by bootstrap.

second property that ideal margins should have: Their accuracy should be testable ex-post. Indeed, even the most advanced risk measures are of little help if they cannot be systematically validated.

References

Berkowitz, J., Christoffersen, P.F., and Pelletier, D., 2011, Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science*, forthcoming.

Berkowitz, J. and O'Brien, J., 2002, How Accurate Are Value-At-Risk Models at Commercial Banks? *Journal of Finance*, 57, 1093-1111.

Brunnermeier, M.K. and Pedersen, L.H., 2009, Market Liquidity and Funding Liquidity. *Review of Financial Studies*, 22, 2201-2238.

Candelon B., Colletaz, G., Hurlin, C., and Tokpavi, S., 2011, Backtesting Value-at-Risk: A GMM Duration-based Test. *Journal of Financial Econometrics*, 9, 314-343.

Chicago Mercantile Exchange, 2009, CME Clearing Financial Safeguards.

Christoffersen, P.F., 1998, Evaluating Interval Forecasts. *International Economic Review*, 39, 841-862.

Christoffersen, P.F., 2009a, Value-at-Risk Models. In *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J.-P. Kreiss, and T. Mikosch (Springer Verlag).

Christoffersen, P.F., 2009b, Backtesting. In *Encyclopedia of Quantitative Finance*, edited by R. Cont (John Wiley and Sons).

Christoffersen, P.F. and Pelletier D., 2004, Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, 2, 84-108.

Colletaz, G., Hurlin, C., and Pérignon, C., 2012, The Risk Map: A New Tool for Backtesting Risk Models. *Working Paper, HEC Paris*.

Commodity Futures Trading Commission, 2001, Review of Standard Portfolio Analysis of Risk ("SPAN") Margin System. *Report of the Division of Trading and Markets*.

Cruz Lopez, J.A., Harris, J.H., and Pérignon, C., 2011, Clearing House, Margin Requirements, and Systemic Risk. *Review of Futures Markets*, 19, 39-54.

Duffie, D. and Zhu, H., 2010, Does a Central Clearing Counterparty Reduce Counterparty Risk? *Review of Asset Pricing Studies*, forthcoming.

Dumitrescu E. I., Hurlin, C. and Pham, V., 2012, Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance*, forthcoming.

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

Haas, M., 2005, Improved Duration-based Backtesting of Value-at-Risk. *Journal of Risk*, 8, 17-36.

Im, K.S., Pesaran, M.H., and Shin, Y., 2003, Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*, 115, 53-74.

Jones, R.A. and Pérignon, C., 2012, Derivatives Clearing, Default Risk, and Insurance. *Journal of Risk and Insurance*, forthcoming.

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

Knott, R. and Mills, A., 2002, Modelling Risk in Central Counterparty Clearing Houses: A Review. *Bank of England Financial Stability Review*, December, 162-174.

Kupiec, P., 1994, The Performance of S&P 500 Futures Product Margins under the SPAN Margining System. *Journal of Futures Markets*, 14, 789-811.

Kupiec, P., 1995, Techniques for Verifying the Accuracy of Risk Measurement Models. J*ournal of Derivatives*, 3, 73-84.

Kupiec, P. and White, P., 1996, Regulatory Competition and the Efficiency of Alternative Derivative Product Margining Systems. *Journal of Futures Markets*, 16, 943-968.

Patton, A.J., 2002, Application of Copula Theory in Financial Econometrics. *Ph.D. Dissertation*, University of California, San Diego.

Pérignon, C. and Smith, D. R., 2010, The Level and Quality of Value-at-Risk Disclosure by Commercial Banks, *Journal of Banking and Finance*, 34, 362-377.