



CAEPR Working Paper  
#004-2010

# Backtesting Value-at-Risk Models: A Multivariate Approach

Cristina Danciulescu  
Indiana University

**April 3 2010**

This paper can be downloaded without charge from the Social Science Research Network electronic library at: <http://ssrn.com/abstract=1591049>.

The Center for Applied Economics and Policy Research resides in the Department of Economics at Indiana University Bloomington. CAEPR can be found on the Internet at: <http://www.indiana.edu/~caepr>. CAEPR can be reached via email at [caepr@indiana.edu](mailto:caepr@indiana.edu) or via phone at 812-855-4050.

©2008 by NAME. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

# Backtesting Value-at-Risk Models: A Multivariate Approach

Cristina Danciulescu  
Indiana University Bloomington

This draft, April 3rd, 2010

## Abstract

The purpose of this paper is to develop a new and simple backtesting procedure that extends the previous work into the multivariate framework. We propose to use the multivariate Portmanteau statistic of Ljung-Box type to *jointly* test for the absence of autocorrelations and cross-correlations in the vector of hits sequences for different positions, business lines or financial institutions. Simulation exercises illustrate that this shift to a multivariate hits dimension delivers a test that increases significantly the power of the traditional backtesting methods in capturing systemic risk: the building up of positive and significant hits cross-correlations which translates into simultaneous realization of large losses at several business lines or banks. Our multivariate procedure is addressing also an operational risk issue. The proposed technique provides a simple solution to the Value-at-Risk(VaR) estimates aggregation problem: the institution's global VaR measure being either smaller or larger than the sum of individual trading lines' VaRs leading to the institution either under- or over- risk exposure by maintaining excessively high or low capital levels. An application using Profit and Loss and VaR data collected from two international major banks illustrates how our proposed testing approach performs in a realistic environment. Results from experiments we conducted using banks' data suggest that the proposed multivariate testing procedure is a more powerful tool in detecting systemic risk if it is combined with multivariate risk modeling i.e. if covariances are modeled in the VaR forecasts.

*Keywords and Phrases:* Risk Management, Value-at-Risk, Backtesting, Multivariate Testing, Systemic Risk, Operational Risk.

*JEL classification code:* C12, C32, C52, G28, G32.

1 2

---

<sup>1</sup>Corresponding author: Cristina Danciulescu, Department of Economics, Indiana University, Wylie Hall 105, 100 S. Woodlawn, Bloomington, IN 47405. E-mail: cdanciul@indiana.edu.

<sup>2</sup>I am indebted to my adviser, Professor Juan Carlos Escanciano, for his valuable advice, guidance and support during the formation of this paper. I am, also, thankful to Christophe Perignon and Daniel R. Smith for sharing their collected data from the paper "The Level and Quality of Value-at-Risk Disclosure by Commercial Banks" to be used in the application section of this paper, Craig W. Holden, Joon Y. Park, Yoosoon Chang, Yoon-Jin Lee, and Econometrics seminar participants at Indiana University Bloomington for helpful comments. All errors are mine.

# 1 Introduction

Trading accounts of large financial institutions have grown very rapidly and became progressively more complex. Nowadays, their global portfolios contain several thousands positions with thousands of market risk factors as for example interest rates, equity, exchange rates, over-the-counter derivatives, commodity prices, etc.<sup>3</sup> In this context, in order to properly manage market risks, major trading banks developed risk measurement models that aggregate these risks in current positions. For assessing risk, these models employed a standard risk metric, Value-at-Risk (VaR), which is the amount lost on a portfolio or investment with a given small probability over a fixed period of time. In statistical terms, VaR is a quantile measure which, at a given confidence level  $\alpha$ , describes the loss that can occur due to the exposure to the market risks over a given time period. The popularity of this measure among practitioners was due to its conceptual simplicity: the summary of many complex bad outcomes in a single monetary account.

From the regulation side, VaR models have been sanctioned by the Basel Committee in 1996 for determining the market risk capital requirements through the internal models. The measure was motivated by the proliferation of the so-called off-balance-sheet products in the banking sector in early '90, and the necessity to find and implement risk measures that could potentially allow proper risk management for these new products. Since then, VaR has become the standard measure for the financial market risk. Regulations stipulate that estimates are to be calculated for a 99 percent lower critical value of the bank's aggregate trading Profit and Losses (P&L) with a one-day horizon. The forecasts provide a lower bound on aggregate trading P&L that should not be breached more than 1 day in 100.

The daily VaR estimates are maintained by the banks for the purpose of forecast evaluation or "backtesting". The backtesting procedure is the standard assessment of VaR models consisting in estimating ex-post the precision of the VaR forecasts. Regulatory authorities require that VaR estimates be calculated with the same risk model used for internal measurement of trading risk. Regulation does not recommend any particular backtesting procedure, though the choice of the validation technique is a key issue for the financial institutions risk management and the financial stability in general.

Traditional backtesting methods consider only univariate VaR sequences either for individual trading lines or for the financial institution global portfolio. The current practice to obtain such a global VaR measure is by estimating VaRs for each portfolio or trading line, then sum all trading lines' VaRs.<sup>4</sup> This practice is employed due to the infeasibility of using structural models to accurately

---

<sup>3</sup>Berkovitz and O'Brien (2007) documented using daily U.S. bank data that banks' trading positions are complex and affected by non-standard risk factors, frequently rebalanced, and very different across banks.

<sup>4</sup>Perignon& Smith (2008 b) reports that banks routinely disclose their aggregate firm-level VaRs and an increasing number of banks started recently disclosing individual VaRs for each broad risk category: equity, interest rate, commodity, credit spread, foreign exchange, etc.

measure the joint distribution of all market risk factors, as well as the relationships among the risk factors and trading positions. Large banks deal on the regular basis with a very large number of positions/risk factors and they need to generate daily forecasts.<sup>5</sup> However the problem with such aggregation, as showed in Artzner et al. (1999) with examples in McNeil et. al (2005), is that the subadditivity property fails to hold for the VaR measure when the assets making up the positions' portfolios have skewed distributions, a situation that can occur when there are defaultable bonds or options in the portfolios.<sup>6</sup> <sup>7</sup>On the other hand, the solution proposed by Artzner et al. (1999) to subadditivity property failure, Expected Shortfall risk measure(ES), is difficult to backtest in practice.  
8

The purpose of this paper is to propose the implementation of a new and simple multivariate VaR backtesting technique able to overcome the VaRs aggregation problem. We propose to implement a multivariate backtesting procedure applied at once to hits collected from several subgroups of positions or trading lines, where a hit or a violation corresponds to a situation in which ex-post portfolio returns are lower than VaR forecasts. More precisely, we implement a Multivariate Portmanteau test statistic of Ljung-Box type applied to hits collected from several business lines. Our proposed backtesting procedure has the advantage of exploiting a larger information set being able to capture potential business lines' contagion or commonality in risks without the need to resort on a large and infeasible structural risk model. The method allows all the relationships among portfolios or trading lines to be tested jointly where joint testing is consistent with the notion that spillovers are the impact of global news on each market. Moreover, the proposed multivariate testing technique is easy applicable from the practitioners' point of view. This paper shows that this shift to a multivariate hits dimension delivers a test that increases the power of the traditional backtesting methods in assessing the accuracy of VaR forecasts in the presence of systemic risk where systemic risk should be understood as the building up of positive and significant hits cross-correlations which translates into simultaneous realization of large losses at several business lines or banks. From the operational risk point of view, the multivariate procedure makes an accurate assessment of the market risks the financial institution is exposed by avoiding under- or over-risk exposure and hence maintaining excessively high or low capital levels due to the trading lines VaRs' subadditivity property failure. Instead of adding ex-ante

---

<sup>5</sup>Andersen, Bollerslev, Christoffersen and Diebold (2007) documented that the size and complexity of banks trading positions make parametric VaR methods hard to implement in practice. As many banks report to be dealing with thousands of risk factors, they choose not to attempt to estimate time-varying volatilities and covariances for the risk factors.

<sup>6</sup>Perignon& Smith (2008 b) found that the aggregate banks' VaR may be either less or more than the sum of their individual VaRs, hence individual VaRs are informative. In support to their findings, authors cite Deutsche Bank 2005 annual report: "Simply adding the Value-at-Risk figures of the individual risk classes to arrive at an aggregate Value-at-Risk measure would imply the assumption that the losses in all risk categories occur simultaneously".

<sup>7</sup>As mentioned by Perignon& Smith (2008 b), Basel Committe on Banking Supervision (1996) allows banks to have discretion in recognizing empirical correlations within and across broad risk categories when computing their aggregate VaR.

<sup>8</sup>ES is defined as the mean exceedance given the VaR is violated. Backtesting ES is difficult due to the fact that a breach of VaR is rare in practice.

business lines' VaRs (percentiles) in order to obtain the bank's global VaR measure the test is adding ex-post hits' autocovariances and cross-covariances which are expectations hence additive. Therefore the multivariate testing procedure coupled with multivariate risk modeling might offer an optimal solution to the operational risk problem.

We firstly introduce our proposed procedure formally, then investigate the size and power performances of the proposed method through several Monte Carlo simulations. We set up several simulation designs in order to investigate extensively the test's power performance when spillovers among time series occur through various channels.<sup>9</sup> Under the specifications and parameterizations considered in this paper, we found that the multivariate testing procedure is more powerful than its univariate counterpart when cross-correlations among trading lines' hits are positive and significant which is the case of a systemic risk development. We also found that the univariate test is more powerful than the multivariate one when cross-correlations among trading lines' hits are negative which, from the operational risk point of view, suggests that the univariate test creates an under-risk exposure hence a loss in profitability for the financial institution by not taking into consideration potential negative co-movements or risk diversification among its trading lines.

An application using data from two major international banks investigated how our proposed backtesting method performs in a realistic environment. From the application part we found that, using a multivariate generalized autoregressive conditional heteroscedastic model, BEKK (1,1,1)<sup>10</sup> to obtain the banks' VaR forecasts instead of the Historical Simulation method that the two banks used, the multivariate test becomes significant at 1% and 5% over certain trading days rolling windows while the univariate tests do not. The result is consistent with our Monte Carlo findings which implied that the multivariate procedure is more powerful than the univariate one in assessing the underlying market risks a bank is exposed when markets co-move. On the other hand, with our proposed more powerful multivariate backtesting technique it is still hard to reject Historical Simulation obtained VaR forecasts. This might be due to the restriction this technique imposes on the estimation. Historical Simulation method assumes that assets are independent and identically distributed (i.i.d.) which is not the case of the financial data. We also found that, though we have an identified event in the data, the 2001 9/11 event, the multivariate test does not become significant over the respective trading window or year but two years latter. Our intuition for getting this result is that this might be due to the presence of forward looking components as for example bonds in trading lines' portfolios. Our work in progress is addressing this issue by incorporating market expectations or market sentiment in risk models.

An important consequence of using our backtesting approach is that capital requirements will be

---

<sup>9</sup>We refer to time series spillovers as defined in Hong et al. (2009), i.e. the risk of a given asset depends on the previous risk of other asset. For more details regarding time series spillovers and their connections with the time series covariances and correlations see Hong et al. (2009).

<sup>10</sup>See Engle and Kroner (1995) for model description. Multivariate GARCH models specify the risk of one asset as depending dynamically on its own past risk as well as the past risk of other assets.

increased only when the dependencies/positive correlations among trading lines, hence hits sequences, become sufficiently important to be taken into consideration given a certain coverage probability  $\alpha$ . Therefore trading floor risk managers will not have to face excessive idle capital problem, on one hand, while, on the other hand, they will have higher chance to avoid huge losses and failures due to the systemic risk. In our set up, trading line managers get an informational advantage that comes from exploiting the multivariate framework.

The remainder of this paper is organized as following. In Section 2 we describe the environment and introduce our multivariate proposed backtesting procedure. The size and the power performance of the test are examined by Monte Carlo simulations in Section 3. Section 4 applies our test to real banks' data, and Section 5 concludes.

## 2 A multivariate approach to backtesting procedure: General theory

This section defines the VaR problem in the context of a financial institution with multiple business lines or trading positions, formulates the institution forecast evaluation problem, then introduces formally the proposed multivariate backtesting method.

### 2.1 Financial institution with multiple trading lines: Environment description

Within the lines of Escanciano and Olmo (2009 a& b), we formalize the financial institution with multiple trading lines problem as follows. Suppose that  $Y_t^h$ , is the  $h$ -th trading line return time series of a certain financial institution where  $h = 1, \dots, H$ , and assume that at time  $t - 1$  the information set of this trading line  $h$  is given by  $W_{t-1}^h$ . Let  $\mathfrak{F}_{t-1}^h$  be the  $\sigma$ -algebra generated by  $W_{t-1}^h$ . Assuming that the conditional distribution of  $Y_t^h$  given  $W_{t-1}^h$ ,  $F_t^h(\cdot, \theta_0^h, W_{t-1}^h)$ , is continuous with a strictly increasing cumulative distribution function (c.d.f.), we define the  $\alpha$ -th conditional VaR of  $Y_t^h$  given  $W_{t-1}^h$  as the  $\mathfrak{F}_{t-1}^h$  measurable function  $q_\alpha^h(W_{t-1}^h)$  satisfying the equation:

$$P(Y_t^h \leq q_\alpha^h(W_{t-1}^h) | W_{t-1}^h) = \alpha, \quad (1)$$

almost surely (a.s.),  $\alpha \in (0, 1)$ ,  $\forall t \in Z$ .

In this paper we will consider only parametric VaR models, meaning that  $q_\alpha^h(W_{t-1}^h) = m_\alpha^h(W_{t-1}^h, \theta_0^h)$  a.s. for some  $\theta_0^h \in \Theta$ , where  $m_\alpha^h(W_{t-1}^h, \theta_0^h)$  is the parametric VaR model, i.e. the inverse of  $F_t^h(\cdot, \theta_0^h, W_{t-1}^h)$  at the level of  $\alpha$  with respect to the first argument.

Equation (1) implies that the parametric VaR for trading line  $h$ ,  $m_\alpha^h(W_{t-1}^h, \theta_0^h)$ , is correctly specified if and only if

$$E[I_{t,\alpha}^h(\theta_0^h) | W_{t-1}^h] = \alpha, \quad (2)$$

a.s. for some  $\theta_0^h \in \Theta$  where  $I_{t,\alpha}^h = 1(Y_t^h \leq m_\alpha^h(W_{t-1}^h, \theta^h))$ , and  $1(A)$  is the indicator function, i.e.  $1(A) = 1$  if the event  $A$  occurs and 0 otherwise. The variable  $I_{t,\alpha}^h$  is called “hit” or “exceedance”.

## 2.2 Forecast evaluation problem

Traditional backtesting procedures are based on testing some implications of equation (2) for the individual trading lines, h, or bank’s aggregate portfolio. If we define the financial institution aggregate returns as  $Y_t = \sum_{h=1}^H Y_t^h$  and the aggregate  $\alpha$ -conditional VaR of  $Y_t$  given  $W_{t-1}$ , the institution information set, as  $q_\alpha(W_{t-1}) = \sum_{h=1}^H q_\alpha^h(W_{t-1}^h)$ , then equation (2) for the aggregates becomes

$$E[I_{t,\alpha}(\theta_0)|W_{t-1}] = \alpha, \quad (3)$$

a.s. for some  $\theta_0 \in \Theta$ , where  $I_{t,\alpha} = 1(\sum_{h=1}^H Y_t^h \leq \sum_{h=1}^H q_\alpha^h(W_{t-1}^h))$ . However, since subadditivity property does not necessarily hold for VaR measure, as already investigated by Artzner et al. (1999), this type of aggregation is problematic.

The most popular implication of the equation (2) for the univariate case is explored by Christoffersen (1998) which is

$$E[I_{t,\alpha}(\theta_0)|\tilde{I}_{t,\alpha}(\theta_0)] = \alpha, \quad (4)$$

a.s. for some  $\theta_0$ , where  $\tilde{I}_{t,\alpha}(\theta_0) = (I_{t-1,\alpha}(\theta_0), I_{t-2,\alpha}(\theta_0)\dots)'$ .

This condition is equivalent to  $\{I_{t,\alpha}(\theta_0)\}$  being independent and identical distributed (*i.i.d.*) Bernoulli random variable with parameter  $\alpha$ , ( $Ber(\alpha)$ ). Therefore, the problem of evaluating the accuracy of VaR forecasts can be reduced to the problem of examining the unconditional coverage and independence properties of the univariate hits sequence,  $\{I_{t,\alpha}(\theta_0)\}$ . Testing for  $E[I_{t,\alpha}(\theta_0)] = \alpha$  is called the unconditional backtesting and testing for  $\{I_{t,\alpha}(\theta_0)\}$  being *i.i.d.* is called the independence test.

Berkovitz et al. (2006) outlined an unified approach of VaR assessment based on the fact that the unconditional coverage and independence hypotheses are both consequences of the martingale difference hypothesis for the hits process. They noted that the univariate de-meaned hits sequence,  $\{I_{t,\alpha}(\theta_0) - \alpha\}$ , forms a martingale difference sequence (m.d.s), and this implies that the hits sequence is uncorrelated at all leads and lags. On this basis, authors proposed a univariate test of the Ljung-Box type that considers the nullity of the first  $K$  autocorrelations for the hits sequence.

If we denote by  $\gamma_k$  the univariate hits sequence autocorrelation of order  $k$ , then to test if  $\gamma_k = 0$  holds for the first  $K$  autocorrelations, we have

$$LB(K) = T(T+2) \sum_{k=1}^K \frac{\hat{\gamma}_k^2}{T-k} \quad (5)$$

which is, under some regularity condition<sup>11</sup>, asymptotically a  $\chi^2$  with  $K$  degrees of freedom as  $T \rightarrow \infty$ .

<sup>11</sup>See Box and Pierce (1970) and Ljung and Box (1978).

This procedure, which considers the empirical autocorrelation of order  $K$  for the hits sequence, is an improvement compared with Christoffersen (1998) test which only considered the autocorrelation of order one.

Our paper's main assumption is that, if past hits from one trading line  $h$  is in the information set of the others, i.e.  $I_{t-k}^h \in W_{t-1}^i, \forall k \geq 1, \forall i, \forall h, i \neq h$ , with  $i, h \in H$  and  $I_{t,\alpha}^i = 1(Y_t^i \leq m_\alpha^i(W_{t-1}^i, \theta^i))$ , then the joint VaRs validation for the  $H$  trading lines using a multivariate version of the Ljung-Box test statistic will significantly improve the validity checking of the models. More specifically, if instead of using for testing univariate hits sequences from each trading line  $h \in H$  or the bank's aggregate hits sequence, we stack all hits sequences in an  $H$  dimensional vector, i.e.  $I_{t,\alpha}(\theta_0) = [I_{t,\alpha}^1(\theta_0^1), \dots, I_{t,\alpha}^H(\theta_0^H)]'$ , then the problem of evaluating the accuracy of VaR forecasts imply testing *jointly* for the unconditional coverage and independence properties of the  $H$  dimensional hits vector,  $I_{t,\alpha}(\theta_0)$ , for some  $\theta_0 = [\theta_0^1, \dots, \theta_0^H]' \in \Theta$ . The unconditional coverage test implies testing for  $E[I_{t,\alpha}(\theta_0)] = \alpha$ , where  $\alpha$  here denotes the vector of coverage probabilities. The independence property implies checking, in addition to the previous used tests, for

$$E[(I_{t,\alpha}^i(\theta_0^i) - \alpha)(I_{t-k,\alpha}^h(\theta_0^h) - \alpha)] = 0, \quad (6)$$

$\forall i = 1, \dots, H, \forall h = 1, \dots, H$  where  $i \neq h$  and  $k = 1, \dots, K$  lags.

In other words, this means that, if each trading line  $VaR^h = m_\alpha^h(W_{t-1}^h, \theta^h)$  model is correctly specified, and there is no commonality in risks, then past observations from a business line hits sequence should not help predict future violations of itself or violations for other business lines.

### 2.3 Backtesting procedure using a multivariate Portmanteau test statistic

Our proposed backtesting method is based on the multivariate Ljung-Box statistic. The test takes into consideration *both the autocorrelations and cross-correlations among hits sequences* for trading lines under consideration or supervision. The procedure is a joint test for the unconditional coverage and independence properties using violations from several business lines at once, hence *exploiting a larger information set than the previous methods*.

Let  $I_{t,\alpha}(\theta_0)$  be the  $H$ -dimensional vector of the trading lines violations series as defined in the previous section. If we denote by  $\Gamma_k$  the population covariance matrix,  $\Gamma_k = E[(I_{t,\alpha}(\theta_0) - \alpha)(I_{t-k,\alpha}(\theta_0) - \alpha)']$ , and by  $D$  an  $H \times H$  diagonal matrix with standard deviation of  $I_{t,\alpha}^h(\theta_0^h)$  on the main diagonal, then by analogy with the univariate case, we can define the lag- $k$  cross-correlation matrix of  $I_{t,\alpha}(\theta_0)$  as

$$\rho_k = D^{-1}\Gamma_k D^{-1}, \quad (7)$$

with its  $(i, h)$ -th element given by

$$\rho_k^{ih} = \frac{\gamma_k^{ih}}{\sqrt{\gamma_0^{ii}\gamma_0^{hh}}}, \quad (8)$$



$\forall i = 1, \dots, H, \forall h = 1, \dots, H$  trading lines, and  $\forall k = 1, \dots, K$  lags. When  $k = 0$  we get the contemporaneous cross-correlation matrix of  $I_{t,\alpha}(\theta_0)$ .

The multivariate testing procedure is carried out in out-of-sample exercises. The forecast environment can be described as following. Let  $Y_t = \{Y_t^1, \dots, Y_t^H\}$ , and suppose that  $\{Y_t, Z_t\}_{t=1}^T$  of size  $T \geq 1$  are used to evaluate  $Var = \{Var^1, \dots, Var^H\}$  forecasts, where here  $Z_t$  denotes other economic and financial variables from the information set. Assuming that the first  $R$  observations in each trading line sample are used to estimate the parameters for the respective VaR model, then it remains  $P = T - R$  predictions to be evaluated for each  $h$  trading line. West and McCracken (1998) considered, for example, three forecasting schemes: recursive, rolling, and fixed. They differ depending on how  $\theta_0^h$  are estimated. In the recursive scheme, the estimators  $\hat{\theta}_t^h$  are computed with all the sample available up to time  $t$ . In the rolling scheme only the last  $R$  values of the series are used to estimate  $\hat{\theta}_t^h$ , which means that they are constructed from the sample  $s = t - R + 1, \dots, T$ . In the fixed scheme the parameters are not updated when new observations become available, meaning that  $\hat{\theta}_t^h = \hat{\theta}_R^h$ , for all  $t, R \leq t \leq T$ . In the current set up we will only consider the fixed forecasting scheme for the sake of computational simplicity.

In the backtesting context, the  $ih$ -th element of the hits covariance matrix at different lags  $k$  is defined as

$$\xi_{P,k}^{ih} = Cov(I_{t,\alpha}^i(\theta_0^i), I_{t-k,\alpha}^h(\theta_0^h)), k \geq 1, \quad (9)$$

$\forall i = 1, \dots, H, \forall h = 1, \dots, H$ , and can be consistently estimated under  $E[I_{t,\alpha}^i(\theta_0^i)] = \alpha, E[I_{t,\alpha}^h(\theta_0^h)] = \alpha$  by

$$\gamma_{P,k}^{ih} = \frac{1}{P-k} \sum_{t=R+k+1}^T [(I_{t,\alpha}^i(\theta_0^i) - \alpha)(I_{t-k,\alpha}^h(\theta_0^h) - \alpha)], k \geq 1. \quad (10)$$

Analogously, the sample covariance of the multivariate hits vector is given by

$$\Gamma_{P,k} = \frac{1}{P-k} \sum_{t=R+k+1}^T [(I_{t,\alpha}(\theta_0) - \alpha)(I_{t-k,\alpha}(\theta_0) - \alpha)], k \geq 1. \quad (11)$$

Alternatively, if we use the hits correlation matrix for testing, its  $ih$ -th element is defined in our backtesting framework as

$$\rho_{P,k}^{ih} = \frac{\gamma_{P,k}^{ih}}{\sqrt{\gamma_{P,0}^{ii} \gamma_{P,0}^{hh}}}. \quad (12)$$

The univariate Ljung-Box statistic applied to univariate hits sequences can be generalized to the multivariate case. The implementation of the multivariate test consists, for a given lag length  $K \geq 1$ , in testing the null hypothesis corresponding to the joint nullity for correlation of order  $k$  in the hits vector  $I_t(\theta_0)$ , where  $k = 1, \dots, K$ .

The null hypothesis of the test statistic is

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0, \quad (13)$$

and the alternative hypothesis is

$$H_1 : \rho_k \neq 0, \quad (14)$$

for some  $k = 1, 2, \dots, K$ .

However, in practice tests for (13) are based on estimates of the relevant parameters such as

$$\hat{\gamma}_{P,k}^{ih} = \frac{1}{P-k} \sum_{t=R+k+1}^T [(I_{t,\alpha}^i(\hat{\theta}_{t-1}^i) - \alpha)(I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - \alpha)], \quad (15)$$

$$\hat{\Gamma}_{P,k} = \frac{1}{P-k} \sum_{t=R+k+1}^T [(I_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha)(I_{t-k,\alpha}(\hat{\theta}_{t-k}) - \alpha)], \quad (16)$$

respectively

$$\hat{\rho}_{P,k}^{ih} = \frac{\hat{\gamma}_{P,k}^{ih}}{\sqrt{\hat{\gamma}_{P,0}^{ii} \hat{\gamma}_{P,0}^{hh}}}. \quad (17)$$

The proposed multivariate Portmanteau statistic tests for the absence of autocorrelations and cross-correlations between pairwise hits sequences, jointly, and in terms of sample covariance matrices takes the following form:

$$Q_H(K) = P(P+2) \sum_{k=1}^K \frac{1}{P-k} \text{tr}(\hat{\Gamma}_{P,k} \hat{\Gamma}_{P,0}^{-1} \hat{\Gamma}_{P,k} \hat{\Gamma}_{P,0}^{-1}), \quad (18)$$

where  $P$  is the size of the predicted interval,  $K \geq 1$  is the considered lag length,  $H$  is the dimension of the vector of hits considered,  $I_t(\theta)$ , and  $\text{tr}(A)$  is the trace of the matrix  $A$ .

In terms of sample correlation matrices, the test statistic  $Q_H(K)$  can, also, be written as

$$Q_H(K) = P(P+2) \sum_{k=1}^K \frac{1}{P-k} \hat{r}'_{P,k} (\hat{\rho}_{P,0}^{-1} \otimes \hat{\rho}_{P,0}^{-1}) \hat{r}_{P,k}, \quad (19)$$

where  $\hat{r}_{P,k} = \text{vec}(\hat{\rho}'_{P,k})$ ,  $\text{vec}(A)$  denotes the vectorization of the matrix  $A$  and  $\otimes$  denotes Kronecker product operation.

The modification of this test statistic recommended for samples of moderate size by Li and McLeod is given by

$$Q_H^*(K) = P \sum_{k=1}^K \frac{1}{P-k} \hat{r}'_{P,k} (\hat{\rho}_{P,0}^{-1} \otimes \hat{\rho}_{P,0}^{-1}) \hat{r}_{P,k} + \frac{H^2 K(K+1)}{2P}, \quad (20)$$

which is asymptotically equivalent to  $Q_H(K)$ .

## 2.4 Asymptotic theory

In order to derive the limit distribution for our proposed test under the null hypothesis, we need to impose a set of regularity conditions on the data generating process for  $Y_t^h$ , the VaR models  $m_\alpha^h(W_{t-1}^h, \theta^h)$ , the parameter estimators  $\hat{\theta}_t^h$ , and the ratio between the size of the estimation sample,  $R$ , versus the prediction sample,  $P$ ,  $\pi = \frac{P}{R}$ . A detailed description of the conditions and assumptions we make can be found in the Appendix.

The derivation of the limit distribution of the test is complicated by the fact that we do not observe the true parameters value,  $\theta_0^h$ , hence we have to estimate them. For the consequences of ignoring parameter uncertainty and the ways to correct the limit distributions of the current backtesting methods in use see Escanciano&Olmo (2008 a&b). Alternatively, one can proceed assuming that the estimation of  $\theta_0^h$  by  $\hat{\theta}_t^h$  has no effect on inference as the existing literature assumed with the exception of Escanciano&Olmo (2009 a&b). Note that this assumption is valid only if the sample size used for estimating the parameters,  $R$ , is much larger than the prediction sample,  $P$ . Under this circumstance, replacing  $\theta_0^h$  with  $\hat{\theta}_t^h$  has no impact on the limit distribution of the test,  $Q_H(K)$ . In this paper, we derive the limit distribution of  $Q_H(K)$  under this assumption. This assumption greatly simplifies the construction and implementation of the proposed multivariate test because we do not need to know the asymptotic expansion of  $\hat{\theta}_t^h$  and can choose any  $\sqrt{T}$ -consistent estimator.

**Theorem 1:** *Under the Assumptions A1-A5 in the Appendix, under  $H_0$*

$$Q_H(K) \rightarrow_d \chi^2(KH^2), \quad (21)$$

as  $T \rightarrow \infty$ , where  $K$  is the lag length and  $H$  is the number of trading lines considered.

## 3 Monte Carlo simulations

In this section we examine the finite sample performance of our proposed test through several Monte Carlo simulations. The aim of the exercises is to assess the empirical size (probability of incorrectly rejecting the null hypothesis) and power (probability of rejecting a false null) for the multivariate test. We used several data generating processes (DGPs) so that we can investigate extensively what are the potential gains and drawbacks from applying the proposed multivariate testing procedure under various realistic environments.

In our Monte Carlo experiments we investigate both the influence of the lag order  $K$  and out-of-sample size choices  $P$ . For the sake of computational simplicity we report results only for the fixed forecasting scheme with  $\pi = \frac{P}{R} = 0.05$ <sup>12</sup> <sup>13</sup>, where  $R$  is the in-sample size, and  $P$  is the out-of-sample

<sup>12</sup>Our choice value for  $\pi$  is motivated by our assumptions, see the Appendix.

<sup>13</sup>See, also, Escanciano and Olmo (2009 a) for how different values for  $\pi$  can affect the traditional backtesting

size to be forecasted. For all simulations we considered the out-of-sample sizes  $P = 250, 500$  for which the in-sample sizes implied by imposing  $\pi = \frac{P}{R} = 0.05$  are  $R = 5000, 10000$ . The choices for the lag lengths are  $K = 1, 5, 10, 15$ .<sup>14</sup>

The first Monte Carlo design follows the one proposed by Christoffersen (1998). He modeled the violations process by a Markov chain with transition probabilities:

$$\Pi = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}. \quad (22)$$

Under the null hypothesis ( $H_0$ ), the violations have a constant conditional mean which implies the linear restriction  $p_{00} = p_{10} = \alpha$ . Hence, the probability of having a violation at time  $t$  is equal with  $\alpha$ , the coverage rate, no matter the state at  $t - 1$ . Under the alternative hypothesis ( $H_1$ ),  $p_{ij} = P[I_{t,\alpha}(\theta_0) = j | I_{t-1,\alpha}(\theta_0) = i] \neq \alpha$ . The Markov chain reflects only the existence of a correlation of order one in the process of hits sequence,  $I_t(\theta_0)$ . This means that the probability of having a violation (respective not having one) for the current period depends only on the occurrence of a violation or not for the same level of coverage  $\alpha$  in the previous period.

For our multivariate case we generated, under the null, two uncorrelated Markov chains for which the violations for both have a constant conditional mean which implies the linear restriction  $p_{00}^1 = p_{10}^1 = p_{00}^2 = p_{10}^2 = \alpha$ , where the superscript indicates the chain. Under the alternative hypothesis we maintained the linear restriction but we generated correlated chains with cross-correlation set at 0.9. We considered  $\alpha = 1\%$ ,  $5\%$  and  $10\%$  corresponding to hits sequences with shortfall probabilities or risk levels of  $1\%$ ,  $5\%$  and  $10\%$ .

The Monte Carlo algorithm's main steps in this design are as following:

1. Generate  $R + P$  observations for each  $h$  binary (hits) sequence (we will get a matrix of  $H$  hits),
2. Implement the proposed multivariate test for the obtained matrix of hits sequences and compare it with a  $\chi^2(H^2K)$ ,
3. Implement the univariate test for each univariate hits sequence and compare with a  $\chi^2(K)$ ,
4. Repeat the previous steps for  $l$  times and calculate the rejection rates. Rejection rates are calculated over  $l = 1000$  Monte Carlo trials.

The advantage of considering this design is that estimation effects do not affect the limit distribution of the tests. The draw back is that it captures only the correlation of order 1.

Table 1 displays simulation results for the size of the multivariate test and its univariate counterpart. For the size of the test, if the asymptotic distribution is accurate in the sample sizes considered, the rejection frequencies should be close to the nominal size of the test which we set to be either

---

procedures' asymptotic distribution.

<sup>14</sup>The choices for lag lengths are in line with Chitturi (1974) assumption. The author derived the asymptotic distribution of the multivariate test under the assumption that  $P \gg K$ .

$\tau = 5\%$  or  $\tau = 10\%$ . We found that the size of the multivariate test is close to the nominal values considered of 5% and 10% for the risk levels  $\alpha = 5\%$  and  $\alpha = 10\%$ . On the other hand the test is oversized for  $\alpha = 1\%$  risk level. This result is not new in the literature. Escanciano and Olmo (2008 a) obtained similar results for their Monte Carlo experiments and they suggested that this problem may be intrinsic to VaR inferences at low quantile levels and not to the existence of the estimation risk. This problem is also investigated extensively in Danciulescu (2010 b).

Table 2 shows results for the power simulations under this Monte Carlo design. Our results illustrate a significant increase in power in moving from the univariate to a multivariate testing environment. The power of the test is increasing in the out-of-sample size  $P$  and decreasing with the lag length  $K$  over all significance and risk levels considered. The latter is expected since all the correlation is at the first lag.

For our second Monte Carlo exercise we employed Hong et al. (2009) nested GARCH DGP to simulate returns for  $h$  trading lines. We decided to use this DGP since GARCH is the most common specification for modeling financial returns. In our simulations we considered only  $H = 2$  as the authors did. Hong et al. (2009) DGP allowed us to disentangle and investigate separately the channels through which spillovers among the business lines' returns may occur. In this paper we investigated only the spillovers that may occur through the time series' means and variances.

The nested GARCH DGP is specified as following:

$$Y_{ht} = \beta_{h1}Y_{1t-1} + \beta_{h2}Y_{2t-1} + u_{ht}, \quad (23)$$

$$u_{ht} = \sigma_{ht}\varepsilon_{ht}, \quad (24)$$

$$\sigma_{ht}^2 = \gamma_{h0} + \gamma_{h1}\sigma_{ht-1}^2 + \gamma_{h2}u_{1t-1}^2 + \gamma_{h3}u_{2t-1}^2, \quad (25)$$

$$\varepsilon_{ht} \sim m.d.s., \quad (26)$$

We assumed that innovations,  $\varepsilon_{ht}$ , are *i.i.d*  $t_\nu$  standardized disturbances i.e.  $\varepsilon_{ht} = \sqrt{\frac{\nu-2}{\nu}} * v_{ht}$  with  $v_{ht}$  distributed as a Student-t with  $\nu$  degrees of freedom for  $h = 1, 2$ .

Using Hong et al. (2009) DGP we investigated both the size and the power of the proposed multivariate test. The values of the parameters are obtained by fitting GARCH models to the banks' daily returns data we used in the application part of the paper.

We assessed the size of the test under the null (H0) using the following parameter values:

$$(\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0, 0, 0.05, 0.88, 0.01, 0), \quad (27)$$

$$(\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0, 0.15, 0.73, 0, 0.1). \quad (28)$$

The innovation processes,  $\varepsilon_{ht}$ , are assumed to follow a Student-t distribution with  $\nu_h = 5$  degrees of freedom for  $h = 1, 2$ . Our choice for the innovations' distribution and degrees of freedom parametrization is motivated by Perignon & Smith (2009 a) estimation results. Using data from around the World

fifty major banks they found that a Student-t distribution with between 5 and 8 degrees of freedom is the best choice to account for the observed data leptokurtosis.

For this Monte Carlo design, the algorithm's main steps are as follows:

1. Using Hong et al. (2009) DGP and the true parameter values, generate  $R + P$  observations for the  $H$  trading lines,
2. Using the first  $R$  (in-sample) observations generated, estimate the parameters of the model by quasi maximum likelihood method (QMLE),
3. Using the estimated parameters we generate the out-of-sample  $P$  observations from univariate GARCH DGPs,
4. Get the hits sequence for each individual series of financial return generated. We will get a matrix of  $H$  hits ( $H = 2$  in our bivariate case),
5. Implement the proposed multivariate test for the obtained matrix of hits sequences and compare it with a  $\chi^2(H^2K)$ ,
6. Implement the univariate test for the univariate hits sequence of interest and compare it with a  $\chi^2(K)$ ,
7. Repeat the previous steps for  $l$  times and calculate the rejection rates.

As in our previous Monte Carlo design we considered  $l = 1000$ .

Table 3 shows the results for the size of the test using the above DGP and parametrization. We found that the multivariate test is slightly over-sized for this specification compared to the nominal values considered of 5% and 10% for the risk levels  $\alpha = 5\%$  and  $\alpha = 10\%$  while the test is undersized for the risk level  $\alpha = 1\%$ . As in the previous Monte Carlo design for size we refer to Danciulescu (2010 b) work as a potential explanation for these size distortions.

Using Hong et al. (2009) DGP with different parameter values we conducted Monte Carlo experiments to compare the empirical powers of the multivariate and univariate methods for rejecting some alternatives to the null. We employed parameterizations that allowed us to investigate separately the power of the test when spillovers between financial returns come through their mean (main body of the distributions) or through their variance (tails of the distributions).

We used the following parameter values to investigate the power of the test under the alternative (H1) when there are spillovers between returns through their mean:

$$(\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0, 0.7, 0.05, 0.88, 0.01, 0), \quad (29)$$

$$(\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0, 0.15, 0.73, 0, 0.1), \quad (30)$$

and when there are spillovers between returns through their variance:

$$(\beta_{11}, \beta_{12}, \gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = (0, 0, 0.05, 0.88, 0.01, 0.7), \quad (31)$$

$$(\beta_{21}, \beta_{22}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \gamma_{23}) = (0, 0, 0.15, 0.73, 0, 0.1). \quad (32)$$

For all cases, the innovations processes  $\varepsilon_{ht}$  are assumed to follow a Student-t distribution with  $\nu_h = 5$  degrees of freedom for  $h = 1, 2$ .

In tables 4 and 5 we report the rejection probabilities at 5% and 10% significance level for the last two parameterizations. The simulation results for the power show that in both cases, spillovers between returns through their mean and variance, we get a significant gain in power for the multivariate test versus the univariate one for the lag lengths 1, 10 and 15, while at the lag 5 the univariate test is more powerful than the multivariate one. The power of the test is also increasing in the out-of-sample size  $P$  over all significance levels considered. These findings suggest that the multivariate test is more powerful in capturing the co-movements (positive and significant cross-correlations) among the trading lines hits sequences. Also, by capturing the negative movements among trading lines (negative cross-correlations), our results suggest the multivariate test is a better choice from the operational risk point of view. Proposed procedure makes an accurate assessment of the market risks the financial institution avoiding under-risk as well as over-risk exposure and consequently maintaining excessively high or low capital levels with negative implications as decrease in bank's profitability or its failure.

The economic intuition of this result is also of interest. A potential explanation of the negative cross-correlation at lag 5 might be due to the fact that banks might use assets (bonds) with different maturities for hedging the risk in their portfolios/business lines or different trading strategies. For example one trading line/bank is shorting the risky assets at day 5 while the other is holding risky assets longer in its portfolio.

The third Monte Carlo design investigates the power of the multivariate testing procedure using data generated from a bivariate BEKK process as introduced by Engle and Kroner (1995). This specification is recommendable for modeling the dynamic correlation structure among different trading lines (risk categories) as recommended in the Amendment of the Basel Accord (1996).<sup>15</sup> Moreover, BEKK model is a more realistic representation of the financial markets environment with spillovers among time series that occur through various channels without the possibility to clearly identify them.

In this set up financial returns are modeled as a multivariate stochastic vector process  $\{Y_t\}$  with dimension  $H \times 1$  such that  $E(Y_t) = 0$ . The vector  $Y_t$  is assumed to be conditionally heteroscedastic:

$$Y_t = M_t^{1/2} \varepsilon_t, \quad (33)$$

given the information set  $W_{t-1}$ , where  $W_{t-1}$  denotes the information set generated by the observed series  $\{Y_{t-1}\}$  up to and including time  $t-1$ . The  $H \times H$  matrix  $M_t = [m_{ijt}]$  is the conditional covariance matrix of  $Y_t$ , and  $\varepsilon_t$  is an i.i.d. vector error process.

The matrix process  $M_t$  for the BEKK model has the form:

$$M_t = A_0 A_0' + \sum_{k=1}^K \sum_{i=1}^q A_{ki}' Y_{t-i} Y_{t-i}' A_{ki} + \sum_{k=1}^K \sum_{j=1}^p B_{kj}' M_{t-j} B_{kj}, \quad (34)$$

---

<sup>15</sup>See Perignon & Smith (2008 b).

where  $A_{ki}$ ,  $B_{kj}$ , and  $A_0$  are  $H \times H$  parameter matrices, with  $A_0$  lower triangular. The decomposition of the constant term into a product of two triangular matrices is to ensure positive definiteness of  $M_t$ . In our case  $Y_t$  is a vector of  $H$  log-returns corresponding to  $H$  trading lines.

Since the number of parameters in a full BEKK model is  $(p+q)KH^2 + H(H+1)$ , in order to reduce the computational burden, we employed in our Monte Carlo simulations a bivariate BEKK(1,1,1), hence  $H = 2$ .

As in the case of Hong et al. (2009) design, the BEKK DGP is parametrized using values that we got by fitting the model to the banks' data:

$$(a_{0,11}, a_{0,21}, a_{0,22}, a_{1,11}, a_{1,12}, a_{1,21}, a_{1,22}, b_{11}, b_{12}, b_{21}, b_{22}) = (14.8511, 0.6318, 1.0809, 0.2525, -0.2308, -0.3709, 0.0807, 0.3503, 0.6730, -0.4592, 0.3663) \quad (35)$$

Innovations are assumed to follow a bivariate Student-t distribution with  $\nu = 5$  degrees of freedom, and the variance-covariance matrix elements  $\sigma_{11} = \sigma_{22} = 1$  and  $\sigma_{12} = \sigma_{21} = 0.4$ .

The simulation environment follows the same steps and considerations as in the case of Hong et al. (2009) DGP. The fitted risk model for each generated time series from the BEKK DGP is a GARCH(1,1).

Table 6 displays the results for this Monte Carlo experiment. Using BEKK design we get similar results for power as in the Hong et al. (2009) Monte Carlo design.

In summarizing our Monte Carlo results we conclude that our paper contributes to the exiting literature by showing that in moving to a multivariate backtesting procedure from a univariate method we get a power improvement in capturing the co-movements (positive cross-correlations) among the trading lines or financial institutions while avoiding financial institution under-risk exposure by capturing the negative movements (negative cross-correlations) among her trading lines. Hence the multivariate technique represents an improvement in testing the accuracy of VaR forecasts.

## 4 Application

To illustrate how our new proposed backtesting procedure works in a real data environment, we apply the test to two international major banks' P&L and VaR data. The sample was made available to us by Christophe Perignon and Daniel R. Smith, who developed in Perignon& Smith (2009 a) a method to extract one day-ahead VaRs and daily trading revenues data from the graphs disclosed by the banks. The authors selected a sample of five large banks from five different countries and collected annual 10-K forms from the SEC-EDGAR website, annual reports from the banks' websites or hardcopies from the banks, for the period 1995-2005. They focused on the largest banks since presumably large banks devote the most resources to computing VaRs. <sup>16</sup>

---

<sup>16</sup>See Perignon and Smith (2009 a) paper's Appendix for a detailed description of the data extraction procedure.



From their sample of five banks we chose for our analysis Bank of America and Credit Suisse First Boston Bank. In their annual statements the two financial institutions report actual revenues that are affected by their intraday trades. Banks' trading revenues are based on position values recorded at the close of the day and represent the banks' consolidated trading activities. The usual activities include trading in interest rate, foreign exchange, equity assets, liabilities, and derivative contracts. Perignon&Smith (2009 a) reports that, for these two banks, it is not stated explicitly if their trading revenues are inflated or not by trading fees or commissions, which may create some distortions in backtesting. The banks' VaRs are calculated for a one-day-ahead-horizon and a 99% confidence level for profit and losses (P&L), that is the 1% lower tail of their  $P&L$  distributions.

Figure 1 shows the graphs of the daily trading revenues and one day-ahead 99% VaRs for our selected banks. From the graphs we observe that there are fewer exceptions or days when the actual loss is greater than the VaR consistent with the one percent coverage probability. Bank of America had four exceptions and Credit Suisse First Boston bank had six exceptions over the period considered. Because there are around 1000 observations in the sample, the expected number of exceptions is 10 for both banks. The two banks differ, also, in the magnitude of violations. As one can notice from figure 1 the magnitude of violations for Credit Suisse First Boston is much higher than the ones for Bank of America.

Figure 3 plots the violation sequences for the two banks obtained using their reported  $P&L$  and VaRs. The graph suggests that there is at best a weak relationship between the two banks' violation sequences. Results for cross correlation between the two banks trading revenue and VaRs displayed in table 11 supports the inference made using figure 3. The correlation of their daily P&L and VaR is low. This low correlation may reflect the difference in portfolios' composition between the two banks.

Tables 9 and 10 present the summary statistics for daily P&L and VaR data for the two banks under our investigation as reported in Perignon&Smith (2009). From the descriptive statistics one can notice significant differences in average P&L, standard deviation and kurtosis between the two banks. The magnitude of trading activity is almost three times larger for Bank of America versus Credit Suisse First Boston Bank. The average daily P&L for Bank of America was 13.8698 million dollars while the average daily P&L for Credit Suisse First Boston bank was 5.0318 million dollars. Trading revenues for both banks are highly volatile with extreme profits and losses, right skewed, and exhibit ARCH effects. The P&L for Credit Suisse First Boston bank displays excess kurtosis relative to the normal distribution. The Dickey-Fuller test indicates that both banks' trading revenues are stationary. There is also evidence of modest autocorrelation around 5 to 10 % for revenue series of both banks. The summary statistics for their VaR figures shows that they are strongly autocorrelated. The methodology used by banks to construct their VaRs is Historical Simulation. Histograms of P&L and VaRs for the two banks are presented in Figure 4.

We applied our backtesting procedure to the banks' observed sequences of  $P&L$  and VaRs considering a 250-day moving window. Therefore, with our available data, we repeated the procedure for

a total of four different periods. That is, for the second period we considered the forecasting period from  $P = 251$  to  $P = 500$ , for the third period from  $P = 501$  to  $P = 750$ , and for the last period from  $P = 751$  to  $P = 1000$ .

Table 7 reports statistics for both the multivariate and univariate tests and the number of exceedences (Vio) for the four windows. We found no rejections at 1% and 5% either for the multivariate test or for the univariate one over the windows and lag lengths considered.

We repeated the testing procedures using the observed  $P\&L$  but with VaR forecasts obtained by fitting a BEKK (1,1,1) as risk model. To the best of our knowledge this is the first paper to use a multivariate risk model for obtaining the VaR forecasts. The advantage of using a multivariate risk model is capturing in VaRs not only business lines or banks' conditional variances but also their covariances. Table 8 reports the testing results. We found that for the rolling window  $P = 501 : 750$  the multivariate test is significant at 1% for lags  $K = 1, 5, 10$  while the univariate tests are not. The total number of violations obtained using BEKK(1,1,1) risk model at 1 % risk level is comparable to the one obtained using banks' reported method which is Historical Simulation.

The results obtained from our two exercises have several implications. First, they are consistent with our Monte Carlo simulations, which found that the multivariate test is more powerful than the univariate test in capturing co-movements (positive and significant cross-correlations). Second, consistent with Berkovitz et al. (2006) results, we found that it is difficult to reject historical simulation obtained VaRs even with the multivariate technique and this might be due to their design. Third, the multivariate testing procedure is a more powerful tool, able to capture the systemic risk if cross-correlations are modeled in the VaR forecasts. Fourth, multivariate risk modeling combined with multivariate testing procedures might be the best approach from the operational risk point of view since financial institutions and regulators would like to avoid over- as well as under-risk exposures. Multivariate methods are able to capture markets' or institutions' co-movements hence avoiding institutions' failures at the macro level if systemic risk builds up or a decrease in profitability at the micro level if the negative correlations among trading lines are not considered. Fifth, the presence of forward looking components in trading lines' or banks' portfolios, as for example bonds, suggests the incorporation of market expectations or market sentiment in risk models for accurately timing risk spillovers and contagion periods. As one could notice from table 8, though the 2001 9/11 event belongs to the first rolling window, the multivariate test becomes significant over the third rolling window, in 2003.

The economic significance of backtesting methods and VaR forecasts derives from the fact that results are used to determine the minimum regulatory capital requirements which must be met by banks to guard against credit and market risks. The Basel Accord stipulates that a bank daily capital charge must be set at the higher of the previous day's VaR or the average VaR over the last 60 business days multiplied by a factor,  $mf_t$ , i.e.  $CR_t = mf_t * \max\{VaR_t^{0.01}, \frac{1}{60} \sum_{i=1}^{60} VaR_{t-i}^{0.01}\}$ . The multiplicative factor,  $mf_t$ , is determined by the results from backtesting banks' models on the

previous  $T$  days, where  $T = 250$  in Basel Accord, but  $mf_t$  must not be lower than 3. Basel Accord imposes penalties in the form of a higher multiplicative factor on banks which use models that lead to a greater number of violations than would be expected given the specified risk level of 1 %. A high capital charge is undesirable, other things equal, as it reduces banks' profitability. Table 12 displays the penalties imposed for a given number of violations over a 250 trading days period.

Table 13 shows our calculations for the two banks' mean daily capital charges using the two risk models to obtain the VaR forecasts. The results for capital charges are comparable for the two risk modeled. This outcome suggests that employing a dynamic model which accounts for the time varying assets' correlation, as for example BEKK, does not lead to higher capital requirements for banks on one hand, while giving a higher chance to capture potential trading lines commonality in risks (the case of positive cross-correlations) or avoiding under-risk exposure (the case of negative correlations) on the other.

Capital requirement calculations revealed also a drawback of the risk models in use which is their backward looking feature. The lowest capital charges occurred during the period with the highest number of violations for both banks. This might have happened due to the fact that capital charges are based only on backtesting results and VaR forecasts from the previous periods in which volatilities and the number of violations were lower due maybe to more favorable market conditions. This outcome in particular is very important to be investigated in future research since insufficient capital buffers to cover the realized losses may lead to banks or financial institutions' failures.

## 5 Conclusion

In this paper we proposed a new backtesting technique that exploits the informational advantage of the multivariate framework. The test is easy to implement and simulation results conducted over a relevant number of sample sizes, number of lags and specifications showed that the proposed multivariate backtesting procedure represents an improvement versus its univariate counterpart in assessing the accuracy of VaR forecasts. Multivariate testing technique allows all the relationships among trading lines to be tested jointly revealing a considerable increase in power for cross-market spillovers. Joint testing is consistent with the notion that spillovers are due to the impact of global news in each market and use of multivariate data allows analyzing markets' interactions simultaneously.

An application of our proposed procedure to two major international banks real data confirmed the Monte Carlo results. Our findings imply that a partial disaggregation and analysis of risk on classes of risks or trading lines is recommendable to a full financial institution risk aggregation as a way to capture the complexity of financial linkages.

The results from the application part of the paper revealed, also, a drawback of the risk models in use, which is their backward looking feature, with important implications for correct calculation of banks' capital requirements and identification of the risk spillover periods. Our current work in progress explores in this direction.

## 6 Appendix

### ASSUMPTIONS AND MATHEMATICAL PROOFS

ASSUMPTIONS:

Assumptions 1-5 under which Theorem 1 is derived are similar with the ones in Escanciano and Olmo (2008 a and b).

Let the family of conditional distributions be defined as:

$$F_x(y) = P(Y_t \leq y | W_{t-1} = x), \quad (36)$$

and let  $f_x(y)$  be the associated conditional densities. We define the  $\alpha$ -mixing coefficients as

$$\alpha(m) = \sup_{n \in \mathbb{Z}, B \in \mathfrak{F}_n, A \in \mathfrak{P}_{n+m}} \sup |P(A \cap B) - P(A)P(B)|, m \geq 1 \quad (37)$$

where the  $\sigma$ -fields  $\mathfrak{F}_n$  and  $\mathfrak{P}_n$  are  $\mathfrak{F}_n = \sigma(X_t, t \leq n)$ , respectively  $\mathfrak{P}_n = \sigma(X_t, t \geq n)$ , and  $X_t = (Y_t, Z_t)'$ . Mixingness is the property that ensures dependence dies out with horizon.

For each trading line  $h \in H$  we assume the following:

**Assumption 1:**  $\{Y_t^h, Z_t^h\}_{t \in \mathbb{Z}}$  are strictly stationary and strong mixing process with mixing coefficients satisfying  $\sum_{j=1}^{\infty} (\alpha(j))^{1-\frac{2}{d}} < \infty$  with  $d > 2$ .

**Assumption 2:** The family of distribution functions  $\{F_x^h, x \in R^{d_w}\}$  has Lebesgue densities  $\{f_x^h, x \in R^{d_w}\}$  that are uniformly bounded:

$$\sup_{x \in R^{d_w}, y \in R} |f_x^h(y)| \leq C \quad (38)$$

and equicontinuous: for every  $\varepsilon > 0$ , there exists a  $\delta > 0$  such that

$$\sup_{x \in R^{d_w}, |y-z| \leq \delta} |f_x^h(y) - f_x^h(z)| \leq \varepsilon. \quad (39)$$

**Assumption 3:** The model  $m_\alpha^h(W_{t-1}^h, \theta^h)$  is continuously differentiable in  $\theta^h$  (a.s.) with derivatives  $g^h(W_{t-1}^h, \theta^h)$  such that  $E[\sup_{\theta^h \in \Theta_0} |g^h(W_{t-1}^h, \theta^h)|^2] < C$ , for a neighborhood  $\Theta_0$  of  $\theta_0^h$ .

**Assumption 4:** The parameter space  $\Theta$  is compact in  $\mathbb{R}^p$ . The true parameter  $\theta_0^h$  belongs to the interior of  $\Theta$ . The estimator  $\hat{\theta}_t^h$  satisfies the asymptotic expansions  $\hat{\theta}_t^h - \theta_0^h = H^h(t) + o_P(1)$ ,

where  $H^h(t)$  are a  $px1$  vector such that  $H^h(t) = R^{-1} \sum_{s=1}^R l^h(Y_s^h, W_{s-1}^h, \theta_0^h)$  for the fixed forecasting scheme.

We assume that  $E[l^h(Y_s^h, W_{s-1}^h, \theta_0^h) | W_{t-1}^h] = 0$  a.s., and  $V^h = E[l^h(Y_s^h, W_{s-1}^h, \theta_0^h) l^{h'}(Y_s^h, W_{s-1}^h, \theta_0^h) | W_{t-1}^h]$  exist and are positive definite.

Moreover,  $l^h(Y_t^h, W_{t-1}^h, \theta^h)$  are continuous (a.s.) in  $\theta^h$  in  $\Theta_0$  and  $E[\sup_{\theta^h \in \Theta_0} |l^h(Y_t^h, W_{t-1}^h, \theta^h)|^2] \leq C$  are small neighborhoods around  $\theta_0^h$ .

**Assumption 5:**  $R, P \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $\lim_{n \rightarrow \infty} \frac{P}{R} = \pi$ ,  $0 \leq \pi < \infty$ .

PROOF OF THEOREM 1:

We provide a sketch of proof for Theorem 1 using empirical process theory as in Escanciano and Olmo (2008 a) and a variation of weak convergence theorem as developed in Delgado and Escanciano (2007).

For notation simplicity we denote, for  $\theta^h \in \Theta$ ,

$$F_{t-1}^h(\theta^h) = F_{W_{t-1}^h}^h(m_\alpha^h(W_{t-1}^h, \theta^h)), \quad (40)$$

and

$$f_{t-1}^h(\theta^h) = f_{W_{t-1}^h}^h(m_\alpha^h(W_{t-1}^h, \theta^h)). \quad (41)$$

As in Escanciano and Olmo (2008 a) we define the process

$$K_n^h(c) = \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [I_{t,\alpha}^h(\theta_0^h + c(t-1)^{-1/2}) - F_{t-1}^h(\theta_0^h + c(t-1)^{-1/2})], \quad (42)$$

indexed by  $c \in C_D$ , where  $C_D = \{c \in \mathbb{R}^p : c \leq D\}$ , and  $D > 0$  is an arbitrary but fixed constant.

Lemma A1 in Escanciano and Olmo (2008 a) states that, under the Assumptions A1-A5, the process  $K_n^h(c)$  is asymptotically tight with respect to  $c \in C_D$ . Moreover it can be shown that for each  $c \in C_D$ ,

$$E[|K_n^h(c) - K_n^h(0)|^2] = o_P(1). \quad (43)$$

The last equation and the asymptotically tightness of  $K_n^h(c)$  imply that if  $\hat{c}$  is bounded in probability, i.e.  $\hat{c} = O_P(1)$ , then

$$|K_n^h(\hat{c}) - K_n^h(0)| = o_P(1). \quad (44)$$

Next we apply these results with  $\hat{c} = \max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t^h - \theta_0^h)$  where  $R$  denotes the in-sample size. We need to prove that, under the fixed forecasting scheme we considered in this paper,  $\max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t^h - \theta_0^h) = O_P(1)$  holds. Since for the fixed forecasting scheme we have

$$|\max_{R \leq t \leq n} (\frac{\sqrt{t}}{R}) \sum_{s=1}^R l^h(Y_t^h, W_{t-1}^h, \theta_0^h)| \leq |(\frac{1}{\sqrt{R}}) \sum_{s=1}^R l^h(Y_t^h, W_{t-1}^h, \theta_0^h)| = O_P(1), \quad (45)$$

then equation (44) holds for  $\hat{c} = \max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t^h - \theta_0^h)$ , and hence we have

$$\left| \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [I_{t,\alpha}^h(\hat{\theta}_{t-1}^h) - F_{t-1}^h(\hat{\theta}_{t-1}^h)] - \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [I_{t,\alpha}^h(\theta_0^h) - F_{t-1}^h(\theta_0^h)] \right| = o_P(1), \quad (46)$$

where  $P$  denotes the out-of-sample size.

This imply the decomposition

$$\begin{aligned} \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [I_{t,\alpha}^h(\hat{\theta}_{t-1}^h) - \alpha] &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [I_{t,\alpha}^h(\theta_0^h) - F_{t-1}^h(\hat{\theta}_{t-1}^h)] + \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{t-1}^h(\hat{\theta}_{t-1}^h) - F_{t-1}^h(\theta_0^h)] + \\ &\quad \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{t-1}^h(\theta_0^h) - \alpha] + o_P(1). \end{aligned} \quad (47)$$

Using the Mean Value Theorem, and since expectation and differentiation can be interchanged, we have

$$\begin{aligned} A_{1n} &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \{F_{t-1}^h(\hat{\theta}_{t-1}^h) - E[F_{t-1}^h(\hat{\theta}_{t-1}^h)] - F_{t-1}^h(\theta_0^h) + E[F_{t-1}^h(\theta_0^h)]\} = \\ &\quad \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \{g_\alpha^{h'}(W_{t-1}^h, \tilde{\theta}_{t-1}^h) f_{t-1}^h(\tilde{\theta}_{t-1}^h) - E[g_\alpha^{h'}(W_{t-1}^h, \tilde{\theta}_{t-1}^h) f_{t-1}^h(\tilde{\theta}_{t-1}^h)]\} (\hat{\theta}_{t-1}^h - \theta_0^h), \end{aligned} \quad (48)$$

where  $\tilde{\theta}_{t-1}^h$  is between  $\hat{\theta}_{t-1}^h$  and  $\theta_0^h$ .

Assumptions A2 and A3 imply that  $E[\sup_{\theta^h \in \Theta_0} |g_\alpha^{h'}(W_{t-1}^h, \theta^h) f_{t-1}^h(m_\alpha^h(W_{t-1}^h, \theta^h))|] < C$ . Therefore by uniform law of large numbers of Jennrich (1969, Theorem 2) and  $\max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t^h - \theta_0^h) = O_P(1)$ , we have that  $A_{1n} = o_P(1)$ .

Similarly,

$$\begin{aligned} \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \{E[F_{t-1}^h(\hat{\theta}_{t-1}^h)] - E[F_{t-1}^h(\theta_0^h)]\} &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^n E[g_\alpha^{h'}(W_{t-1}^h, \theta_0^h) f_{t-1}^h(\theta_0^h)] (\hat{\theta}_{t-1}^h - \theta_0^h) + \\ \frac{1}{\sqrt{P}} \sum_{t=R+1}^n \{E[g_\alpha^{h'}(W_{t-1}^h, \tilde{\theta}_{t-1}^h) f_{t-1}^h(\tilde{\theta}_{t-1}^h)] - E[g_\alpha^{h'}(W_{t-1}^h, \theta_0^h) f_{t-1}^h(\theta_0^h)]\} &= B_{1n} + B_{2n}. \end{aligned} \quad (49)$$

By uniform law of large numbers and  $\max_{R \leq t \leq n} \sqrt{t}(\hat{\theta}_t^h - \theta_0^h) = O_P(1)$ , we have that  $B_{2n} = o_P(1)$  holds. Hence,

$$\left| \frac{1}{\sqrt{P}} \sum_{t=R+1}^n [F_{t-1}^h(\hat{\theta}_{t-1}^h) - F_{t-1}^h(\theta_0^h)] - E[g_\alpha^{h'}(W_{t-1}^h, \theta_0^h) f_{t-1}^h(\theta_0^h)] \frac{1}{\sqrt{P}} \sum_{t=R+1}^n H^h(t-1) \right| = o_P(1). \quad (50)$$

s

In order to proceed with the proof of theorem 1, for each  $i, h \in H$ ,  $i \neq h$ , we define de process:

$$K_{n,k}^{ih}(c) = \frac{1}{\sqrt{P}} \sum_{t=R+k+1}^n \{I_{t,\alpha}^i[\theta_0^i + c(t-1)^{-1/2}] - F_{t-1}^i[\theta_0^i + c(t-1)^{-1/2}]\} I_{t-k,\alpha}^h[\theta_0^h + c(t-k-1)^{-1/2}], \quad (51)$$

indexed by  $c \in \mathcal{C}_D$ , where  $\mathcal{C}_D = \{c \in \mathbb{R}^p : |c| \leq D\}$ ,  $k \geq 1$ , and  $D > 0$  is an arbitrary but fixed constant.

Applying Lemma A1 from Escanciano and Olmo (2008 a) to the process  $K_{n,k}^{ih}(c)$  and following the previous arguments we get

$$\begin{aligned} \sqrt{P-k}(\hat{\xi}_{P,k}^{ih} - \xi_{P,k}^{ih}) &= \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n [F_{t-1}^i(\hat{\theta}_{t-1}^i) I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - F_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^h)] + o_P(1) \\ &= \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n [F_{t-1}^i(\hat{\theta}_{t-1}^i) I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - F_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^h) + \\ &\quad F_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - F_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^h)] + o_P(1), \end{aligned} \quad (52)$$

where  $\xi_k^{ih} = cov[I_{t,\alpha}^i(\theta_0^i), I_{t-k,\alpha}^h(\theta_0^h)]$  at different lags  $k$  with  $k \geq 1$ , which can be consistently estimated by  $\hat{\xi}_{P,k}^{ih} = \frac{1}{P-k} \sum_{t=R+k+1}^n [I_{t,\alpha}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^h) - \alpha^2]$ .

The previous expression can be rearranged as following

$$\begin{aligned} \sqrt{P-k}(\hat{\xi}_{P,k}^{ih} - \xi_{P,k}^{ih}) &= \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n F_{t-1}^i(\theta_0^i) [I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - I_{t-k,\alpha}^h(\theta_0^h)] + \\ &\quad \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) [F_{t-1}^i(\hat{\theta}_{t-1}^i) - F_{t-1}^i(\theta_0^i)] + o_P(1) = \\ &\quad \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n F_{t-1}^i(\theta_0^i) [I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - I_{t-k,\alpha}^h(\theta_0^h)] + \\ &\quad \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n [g_\alpha^{i'}(W_{t-1}^i, \tilde{\theta}_{t-1}^i) f_{t-1}^i(\tilde{\theta}_{t-1}^i) I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h)] (\hat{\theta}_{t-1}^i - \theta_0^i) = \\ &\quad C_{1n} + C_{2n} + o_P(1), \end{aligned} \quad (53)$$

where  $\tilde{\theta}_{t-1}^i$  is between  $\hat{\theta}_{t-1}^i$  and  $\theta_0^i$ .

Since  $F_{t-1}^i(\theta_0^i) = \alpha$  a.s., the previous results imply that

$$C_{1n} = \alpha E[g_\alpha^{h'}(W_{t-k-1}^h, \theta_0^h) f_{t-k-1}^h(\theta_0^h)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^h(t-k-1) + o_P(1), \quad (54)$$

and

$$|C_{2n} - E[g_\alpha^{i'}(W_{t-1}^i, \theta_0^i) f_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^h)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^i(t-k-1)| = o_P(1), \quad (55)$$



Hence, we proved that

$$\begin{aligned} \sqrt{P-k}(\hat{\xi}_{P,k}^{ih} - \xi_{P,k}^{ih}) &= \alpha E[g_\alpha^{h'}(W_{t-k-1}^h, \theta_0^h) f_{t-k-1}^h(\theta_0^h)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^h(t-k-1) + \\ &E[g_\alpha^{i'}(W_{t-1}^i, \theta_0^i) f_{t-1}^i(\theta_0^i) I_{t-k,\alpha}^h(\theta_0^i)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^i(t-k-1) + o_P(1). \end{aligned} \quad (56)$$

Furthermore we define the following quantities

$$\hat{\xi}_{1P,k}^i = \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n [I_{t,\alpha}^i(\hat{\theta}_{t-1}^i) - \alpha], \quad (57)$$

$$\hat{\xi}_{1P,k}^h = \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n [I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - \alpha], \quad (58)$$

and, similarly, define  $\xi_{1P,k}^i$  and  $\xi_{1P,k}^h$  with  $\theta_0^i$  and  $\theta_0^h$  replacing  $\hat{\theta}_{t-1}^i$  and  $\hat{\theta}_{t-k-1}^h$ .

Since

$$\hat{\gamma}_{P,k}^{ih} = \frac{1}{P-k} \sum_{t=k+1}^n [I_{t,\alpha}^i(\hat{\theta}_{t-1}^i) - \alpha][I_{t-k,\alpha}^h(\hat{\theta}_{t-k-1}^h) - \alpha], \quad (59)$$

this implies that

$$\sqrt{P-k} \hat{\gamma}_{P,k}^{ih} = \sqrt{P-k} \hat{\xi}_{P,k}^{ih} - \alpha \hat{\xi}_{1P,k}^i - \alpha \hat{\xi}_{1P,k}^h + 2\alpha^2. \quad (60)$$

The same equality holds for  $\gamma_{P,k}^{ih}$ ,  $\xi_{P,k}^{ih}$ ,  $\xi_{1P,k}^i$ ,  $\xi_{1P,k}^h$

$$\sqrt{P-k} \gamma_{P,k}^{ih} = \sqrt{P-k} \xi_{P,k}^{ih} - \alpha \xi_{1P,k}^i - \alpha \xi_{1P,k}^h + 2\alpha^2. \quad (61)$$

Hence we have that

$$\sqrt{P-k}(\hat{\gamma}_{P,k}^{ih} - \gamma_{P,k}^{ih}) = \sqrt{P-k}(\hat{\xi}_{P,k}^{ih} - \xi_{P,k}^{ih}) - \alpha(\hat{\xi}_{1P,k}^i - \xi_{1P,k}^i) - \alpha(\hat{\xi}_{1P,k}^h - \xi_{1P,k}^h). \quad (62)$$

Previous arguments imply that

$$\hat{\xi}_{1P,k}^i - \xi_{1P,k}^i = E[g_\alpha^{i'}(W_{t-1}^i, \theta_0^i) f_{t-1}^i(\theta_0^i)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^i(t-1) + o_P(1), \quad (63)$$

and

$$\hat{\xi}_{1P,k}^h - \xi_{1P,k}^h = E[g_\alpha^{h'}(W_{t-k-1}^h, \theta_0^h) f_{t-k-1}^h(\theta_0^h)] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^h(t-k-1) + o_P(1). \quad (64)$$

Therefore we get that

$$\sqrt{P-k}(\hat{\gamma}_{P,k}^{ih} - \gamma_{P,k}^{ih}) = E\{g_\alpha^{i'}(W_{t-1}^i, \theta_0^i) f_{t-1}^i(\theta_0^i) [I_{t-k,\alpha}^h(\theta_0^h) - \alpha] \frac{1}{\sqrt{P-k}} \sum_{t=R+k+1}^n H^i(t-k-1)\} + o_P(1). \quad (65)$$

Corollary 3 in Escanciano and Olmo (2008 a) and our Assumption 5 imply that  $\hat{\gamma}_{P,k}^{ih} \rightarrow_d N(0, \alpha^2(1-\alpha)^2)$ .

Moving to the multivariate framework, let  $\{I_{t,\alpha}(\theta_0)\}$  be the  $H$  dimensional vector that collects the hits sequences from all  $H$  trading lines. The multivariate hits process sample autocovariance and autocorrelation are defined as

$$\Gamma_{P,k} = \frac{1}{P-k} \sum_{t=R+k+1}^n (I_{t,\alpha}(\theta_0) - \alpha)(I_{t-k,\alpha}(\theta_0) - \alpha)', \quad (66)$$

and respectively as

$$\rho_{P,k} = \Gamma_{P,k} \Gamma_{P,0}^{-1}, \quad (67)$$

for  $k \geq 1$ , where both  $\Gamma_{P,k}$  and  $\rho_{P,k}$  are  $H \times H$  matrices.

The two matrices can be stacked as  $1 \times H^2$  row vectors with rows stacked one next to the other as following  $\tilde{\Gamma}'_{P,k} = [\gamma_{P,k}^{11}, \dots, \gamma_{P,k}^{1H}, \dots, \gamma_{P,k}^{H1}, \dots, \gamma_{P,k}^{HH}]$  respectively  $\tilde{\rho}'_{P,k} = [\rho_{P,k}^{11}, \dots, \rho_{P,k}^{1H}, \dots, \rho_{P,k}^{H1}, \dots, \rho_{P,k}^{HH}]$ .

Chitturi (1974) showed that for large  $P (\gg k)$  the multivariate autocorrelation vector process has, approximately, a multivariate normal distribution with

$$E[\tilde{\rho}'_{P,k}] = 0, \quad (68)$$

and

$$cov[\tilde{\rho}_{P,k}, \tilde{\rho}'_{P,l}] \doteq \frac{1}{P} (V \otimes V^{-1}) \delta_{k-l}, \quad (69)$$

where  $\doteq$  denotes an approximate relationship,  $\otimes$  denotes the direct product and  $\delta_{k-l}$  denotes Kronecker delta with unity at  $k-l=0$  and zero elsewhere.

Rao (1973, p. 524) showed that if a random vector  $x$  has a multivariate normal distribution  $N(0, Q\Sigma)$  where  $Q$  is idempotent of rank  $p$  and  $\Sigma$  is positive-definite symmetric, then

$$x' \Sigma^{-1} x \sim \chi_p^2. \quad (70)$$

Chitturi (1974) and Hosking (1980, Theorem 2) completed the proofs for the multivariate autocovariance and autocorrelation functions considering AR and ARMA processes<sup>17</sup>, hence the result from our Theorem 1 follows. Q.E.D

---

<sup>17</sup>See Chitturi (1974) and Hosking (1980) for complete proofs.

Table 1: Sizes of the Multivariate Ljung-Box and Univariate Ljung-Box tests for Christoffersen (1998) design. We simulate *i.i.d.* Bernoulli variables with the probability of having a violation at time  $t$  equal with  $\alpha$ , the coverage rate, (i.e.  $p_{00}^1 = p_{10}^1 = p_{00}^2 = p_{10}^2 = \alpha$ ). VaR is computed at  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Size of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|---|--------------------|----------------------|---------|---------|----------|----------|
|   |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                              | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                          | $\tau = 5\%$       | P=250                | 0.0580  | 0.1510  | 0.1840   | 0.1650   |
|   |                    | P=500                | 0.1140  | 0.1900  | 0.1970   | 0.1760   |
| $\alpha = 1\%$                          | $\tau = 10\%$      | P=250                | 0.0660  | 0.1640  | 0.1920   | 0.1710   |
|   |                    | P=500                | 0.1350  | 0.2280  | 0.2270   | 0.1980   |
| $\alpha = 5\%$                          | $\tau = 5\%$       | P=250                | 0.0620  | 0.0760  | 0.0720   | 0.0640   |
|   |                    | P=500                | 0.0510  | 0.0590  | 0.0630   | 0.0610   |
| $\alpha = 5\%$                          | $\tau = 10\%$      | P=250                | 0.0950  | 0.1050  | 0.1050   | 0.1030   |
|   |                    | P=500                | 0.0880  | 0.1000  | 0.1140   | 0.1110   |
| $\alpha = 10\%$                         | $\tau = 5\%$       | P=250                | 0.0490  | 0.0460  | 0.0500   | 0.0390   |
|   |                    | P=500                | 0.0480  | 0.0570  | 0.0590   | 0.0500   |
| $\alpha = 10\%$                         | $\tau = 10\%$      | P=250                | 0.0850  | 0.0930  | 0.0830   | 0.0780   |
|   |                    | P=500                | 0.0940  | 0.1050  | 0.0990   | 0.1000   |

| Size of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|---------------------------------------|--------------------|----------------------|---------|---------|----------|----------|
|                                       |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                            | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                        | $\tau = 5\%$       | P=250                | 0.0190  | 0.0870  | 0.0920   | 0.1190   |
|                                       |                    | P=500                | 0.0400  | 0.1190  | 0.1500   | 0.1380   |
| $\alpha = 1\%$                        | $\tau = 10\%$      | P=250                | 0.0210  | 0.0890  | 0.0940   | 0.1210   |
|                                       |                    | P=500                | 0.0400  | 0.1290  | 0.1590   | 0.1440   |
| $\alpha = 5\%$                        | $\tau = 5\%$       | P=250                | 0.0580  | 0.0750  | 0.0670   | 0.0630   |
|                                       |                    | P=500                | 0.0360  | 0.0660  | 0.0470   | 0.0570   |
| $\alpha = 5\%$                        | $\tau = 10\%$      | P=250                | 0.0770  | 0.1160  | 0.1040   | 0.1030   |
|                                       |                    | P=500                | 0.0750  | 0.0910  | 0.0940   | 0.0920   |
| $\alpha = 10\%$                       | $\tau = 5\%$       | P=250                | 0.0380  | 0.0420  | 0.0490   | 0.0490   |
|                                       |                    | P=500                | 0.0360  | 0.0320  | 0.0510   | 0.0500   |
| $\alpha = 10\%$                       | $\tau = 10\%$      | P=250                | 0.0890  | 0.0800  | 0.0790   | 0.0770   |
|                                       |                    | P=500                | 0.0890  | 0.0730  | 0.0900   | 0.0900   |

Table 2: Powers of the Multivariate Ljung-Box and Univariate Ljung-Box tests for Christoffersen (1998) design. We simulate correlated Bernoulli variables with the probability of having a violation at time  $t$  equal with  $\alpha$ , the coverage rate, (i.e.  $p_{00}^1 = p_{10}^1 = p_{00}^2 = p_{10}^2 = \alpha$ ) and cross-correlation set to 0.9. VaR is computed at  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Power of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                               | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                           | $\tau = 5\%$       | P=250                | 0.3300  | 0.1390  | 0.1230   | 0.1160   |
|  |                    | P=500                | 0.7670  | 0.3560  | 0.2620   | 0.2240   |
| $\alpha = 1\%$                           | $\tau = 10\%$      | P=250                | 0.4670  | 0.1750  | 0.1540   | 0.1440   |
|  |                    | P=500                | 0.8680  | 0.4410  | 0.3350   | 0.2920   |
| $\alpha = 5\%$                           | $\tau = 5\%$       | P=250                | 0.9970  | 0.8990  | 0.6580   | 0.5170   |
|  |                    | P=500                | 1.0000  | 1.0000  | 0.9910   | 0.9670   |
| $\alpha = 5\%$                           | $\tau = 10\%$      | P=250                | 0.9980  | 0.9380  | 0.7580   | 0.6240   |
|  |                    | P=500                | 1.0000  | 1.0000  | 0.9940   | 0.9790   |
| $\alpha = 10\%$                          | $\tau = 5\%$       | P=250                | 1.0000  | 0.9990  | 0.9650   | 0.8640   |
|  |                    | P=500                | 1.0000  | 1.0000  | 1.0000   | 1.0000   |
| $\alpha = 10\%$                          | $\tau = 10\%$      | P=250                | 1.0000  | 1.0000  | 0.9880   | 0.9280   |
|  |                    | P=500                | 1.0000  | 1.0000  | 1.0000   | 1.0000   |

| Power of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                             | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                         | $\tau = 5\%$       | P=250                | 0.0190  | 0.0870  | 0.0920   | 0.1190   |
|  |                    | P=500                | 0.0400  | 0.1190  | 0.1500   | 0.1380   |
| $\alpha = 1\%$                         | $\tau = 10\%$      | P=250                | 0.0210  | 0.0890  | 0.0940   | 0.1210   |
|  |                    | P=500                | 0.0400  | 0.1290  | 0.1590   | 0.1440   |
| $\alpha = 5\%$                         | $\tau = 5\%$       | P=250                | 0.0580  | 0.0750  | 0.0670   | 0.0630   |
|  |                    | P=500                | 0.0360  | 0.0660  | 0.0470   | 0.0570   |
| $\alpha = 5\%$                         | $\tau = 10\%$      | P=250                | 0.0770  | 0.1160  | 0.1040   | 0.1030   |
|  |                    | P=500                | 0.0750  | 0.0910  | 0.0940   | 0.0920   |
| $\alpha = 10\%$                        | $\tau = 5\%$       | P=250                | 0.0380  | 0.0420  | 0.0490   | 0.0490   |
|  |                    | P=500                | 0.0360  | 0.0320  | 0.0510   | 0.0500   |
| $\alpha = 10\%$                        | $\tau = 10\%$      | P=250                | 0.0890  | 0.0800  | 0.0790   | 0.0770   |
|  |                    | P=500                | 0.0890  | 0.0730  | 0.0900   | 0.0900   |

Table 3: Sizes of the Multivariate Ljung-Box and Univariate Ljung-Box tests using data generated from Hong et al.(2009) DGP. The parametrization considered is:  $\beta_{11} = 0$ ,  $\beta_{12} = 0$ ,  $\gamma_{10} = 0.05$ ,  $\gamma_{11} = 0.88$ ,  $\gamma_{12} = 0.01$ ,  $\gamma_{13} = 0$ ,  $\beta_{21} = 0$ ,  $\beta_{22} = 0$ ,  $\gamma_{20} = 0.15$ ,  $\gamma_{21} = 0.73$ ,  $\gamma_{22} = 0$ ,  $\gamma_{23} = 0.1$ . The innovations are assumed to follow Student T distributions with  $\nu_1 = 5$  and  $\nu_2 = 5$  degrees of freedom. VaR is computed at  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Size of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|---|--------------------|----------------------|---------|---------|----------|----------|
|   |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                              | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                          | $\tau = 5\%$       | P=250                | 0.0270  | 0.0390  | 0.0520   | 0.0580   |
|   |                    | P=500                | 0.0290  | 0.0570  | 0.0790   | 0.0930   |
| $\alpha = 1\%$                          | $\tau = 10\%$      | P=250                | 0.0270  | 0.0390  | 0.0520   | 0.0580   |
|   |                    | P=500                | 0.0290  | 0.0650  | 0.0810   | 0.0950   |
| $\alpha = 5\%$                          | $\tau = 5\%$       | P=250                | 0.0960  | 0.1300  | 0.1220   | 0.1200   |
|   |                    | P=500                | 0.1230  | 0.1680  | 0.1800   | 0.1790   |
| $\alpha = 5\%$                          | $\tau = 10\%$      | P=250                | 0.1120  | 0.1530  | 0.1490   | 0.1410   |
|   |                    | P=500                | 0.1430  | 0.2060  | 0.2090   | 0.2070   |
| $\alpha = 10\%$                         | $\tau = 5\%$       | P=250                | 0.0930  | 0.1330  | 0.1400   | 0.1440   |
|   |                    | P=500                | 0.1090  | 0.1700  | 0.2040   | 0.2110   |
| $\alpha = 10\%$                         | $\tau = 10\%$      | P=250                | 0.1170  | 0.1670  | 0.1820   | 0.1860   |
|   |                    | P=500                | 0.1480  | 0.2190  | 0.2590   | 0.2670   |

| Size of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|---------------------------------------|--------------------|----------------------|---------|---------|----------|----------|
|                                       |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                            | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                        | $\tau = 5\%$       | P=250                | 0.0340  | 0.0780  | 0.1030   | 0.1330   |
|                                       |                    | P=500                | 0.0530  | 0.1450  | 0.1860   | 0.1760   |
| $\alpha = 1\%$                        | $\tau = 10\%$      | P=250                | 0.0340  | 0.0790  | 0.1050   | 0.1350   |
|                                       |                    | P=500                | 0.0560  | 0.1500  | 0.1950   | 0.1800   |
| $\alpha = 5\%$                        | $\tau = 5\%$       | P=250                | 0.0570  | 0.0730  | 0.0740   | 0.0770   |
|                                       |                    | P=500                | 0.0590  | 0.1060  | 0.0990   | 0.1050   |
| $\alpha = 5\%$                        | $\tau = 10\%$      | P=250                | 0.0750  | 0.1140  | 0.1050   | 0.1060   |
|                                       |                    | P=500                | 0.0860  | 0.1390  | 0.1500   | 0.1510   |
| $\alpha = 10\%$                       | $\tau = 5\%$       | P=250                | 0.0490  | 0.0590  | 0.0720   | 0.0620   |
|                                       |                    | P=500                | 0.0510  | 0.0540  | 0.0670   | 0.0710   |
| $\alpha = 10\%$                       | $\tau = 10\%$      | P=250                | 0.0860  | 0.1170  | 0.1110   | 0.1020   |
|                                       |                    | P=500                | 0.1040  | 0.1070  | 0.1150   | 0.1170   |

Table 4: Powers of the Multivariate Ljung-Box and Univariate Ljung-Box tests using data generated from Hong et al.(2009) DGP with spillovers through the mean. The parametrization considered is:  $\beta_{11} = 0$ ,  $\beta_{12} = 0.7$ ,  $\gamma_{10} = 0.05$ ,  $\gamma_{11} = 0.88$ ,  $\gamma_{12} = 0.01$ ,  $\gamma_{13} = 0$ ,  $\beta_{21} = 0$ ,  $\beta_{22} = 0$ ,  $\gamma_{20} = 0.15$ ,  $\gamma_{21} = 0.73$ ,  $\gamma_{22} = 0$ ,  $\gamma_{23} = 0.1$ . The innovations are assumed to follow Student T distributions with  $\nu_1 = 5$  and  $\nu_2 = 5$  degrees of freedom. VaR is computed at  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Power of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                               | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                           | $\tau = 5\%$       | P=250                | 0.1330  | 0.1990  | 0.2350   | 0.2170   |
|  |                    | P=500                | 0.1730  | 0.3160  | 0.3600   | 0.3480   |
| $\alpha = 1\%$                           | $\tau = 10\%$      | P=250                | 0.1360  | 0.2010  | 0.2360   | 0.2220   |
|  |                    | P=500                | 0.1800  | 0.3350  | 0.3640   | 0.3510   |
| $\alpha = 5\%$                           | $\tau = 5\%$       | P=250                | 0.2000  | 0.2980  | 0.2780   | 0.2570   |
|  |                    | P=500                | 0.2920  | 0.4660  | 0.4730   | 0.4550   |
| $\alpha = 5\%$                           | $\tau = 10\%$      | P=250                | 0.2470  | 0.3480  | 0.3270   | 0.3110   |
|  |                    | P=500                | 0.3390  | 0.5280  | 0.5230   | 0.5140   |
| $\alpha = 10\%$                          | $\tau = 5\%$       | P=250                | 0.1840  | 0.2600  | 0.2610   | 0.2380   |
|  |                    | P=500                | 0.2880  | 0.4670  | 0.4860   | 0.4470   |
| $\alpha = 10\%$                          | $\tau = 10\%$      | P=250                | 0.2300  | 0.3190  | 0.3230   | 0.2970   |
|  |                    | P=500                | 0.3650  | 0.5470  | 0.5720   | 0.5270   |

| Power of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                             | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                         | $\tau = 5\%$       | P=250                | 0.0670  | 0.2420  | 0.1920   | 0.2100   |
|  |                    | P=500                | 0.1270  | 0.4000  | 0.3420   | 0.3440   |
| $\alpha = 1\%$                         | $\tau = 10\%$      | P=250                | 0.0710  | 0.2510  | 0.1930   | 0.2120   |
|  |                    | P=500                | 0.1330  | 0.4220  | 0.3500   | 0.3480   |
| $\alpha = 5\%$                         | $\tau = 5\%$       | P=250                | 0.1400  | 0.4260  | 0.2000   | 0.1830   |
|  |                    | P=500                | 0.1870  | 0.6190  | 0.3260   | 0.3140   |
| $\alpha = 5\%$                         | $\tau = 10\%$      | P=250                | 0.1810  | 0.5330  | 0.2410   | 0.2350   |
|  |                    | P=500                | 0.2430  | 0.7180  | 0.4130   | 0.3910   |
| $\alpha = 10\%$                        | $\tau = 5\%$       | P=250                | 0.0960  | 0.4270  | 0.1410   | 0.1230   |
|  |                    | P=500                | 0.1460  | 0.5840  | 0.2360   | 0.2250   |
| $\alpha = 10\%$                        | $\tau = 10\%$      | P=250                | 0.1400  | 0.5750  | 0.2020   | 0.1820   |
|  |                    | P=500                | 0.2110  | 0.6940  | 0.3300   | 0.3090   |

Table 5: Powers of the Multivariate Ljung-Box and Univariate Ljung-Box tests using data generated from Hong et al.(2009) DGP with spillovers through the variance. The parametrization considered is:  $\beta_{11} = 0$ ,  $\beta_{12} = 0$ ,  $\gamma_{10} = 0.05$ ,  $\gamma_{11} = 0.88$ ,  $\gamma_{12} = 0.01$ ,  $\gamma_{13} = 0.7$ ,  $\beta_{21} = 0$ ,  $\beta_{22} = 0$ ,  $\gamma_{20} = 0.15$ ,  $\gamma_{21} = 0.73$ ,  $\gamma_{22} = 0$ ,  $\gamma_{23} = 0.1$ . The innovations are assumed to follow Student T distributions with  $\nu_1 = 5$  and  $\nu_2 = 5$  degrees of freedom. VaR is computed at  $\alpha = 0.01$ ,  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Power of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                               | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                           | $\tau = 5\%$       | P=250                | 0.0860  | 0.1520  | 0.1790   | 0.1680   |
|  |                    | P=500                | 0.1810  | 0.3170  | 0.3400   | 0.3450   |
| $\alpha = 1\%$                           | $\tau = 10\%$      | P=250                | 0.0970  | 0.1610  | 0.1870   | 0.1720   |
|  |                    | P=500                | 0.2000  | 0.3370  | 0.3530   | 0.3600   |
| $\alpha = 5\%$                           | $\tau = 5\%$       | P=250                | 0.1850  | 0.3020  | 0.3220   | 0.2960   |
|  |                    | P=500                | 0.3180  | 0.5290  | 0.5780   | 0.5790   |
| $\alpha = 5\%$                           | $\tau = 10\%$      | P=250                | 0.2230  | 0.3360  | 0.3570   | 0.3230   |
|  |                    | P=500                | 0.3750  | 0.5830  | 0.6260   | 0.6250   |
| $\alpha = 10\%$                          | $\tau = 5\%$       | P=250                | 0.2030  | 0.3530  | 0.3790   | 0.3570   |
|  |                    | P=500                | 0.3700  | 0.6220  | 0.6770   | 0.6760   |
| $\alpha = 10\%$                          | $\tau = 10\%$      | P=250                | 0.2630  | 0.4160  | 0.4290   | 0.4120   |
|  |                    | P=500                | 0.4500  | 0.6750  | 0.7330   | 0.7310   |

| Power of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                             | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                         | $\tau = 5\%$       | P=250                | 0.0230  | 0.1070  | 0.0930   | 0.1010   |
|  |                    | P=500                | 0.0500  | 0.1830  | 0.1780   | 0.2000   |
| $\alpha = 1\%$                         | $\tau = 10\%$      | P=250                | 0.0230  | 0.1080  | 0.0940   | 0.1010   |
|  |                    | P=500                | 0.0500  | 0.1840  | 0.1780   | 0.2000   |
| $\alpha = 5\%$                         | $\tau = 5\%$       | P=250                | 0.1000  | 0.3540  | 0.2560   | 0.2580   |
|  |                    | P=500                | 0.2080  | 0.5800  | 0.4540   | 0.4730   |
| $\alpha = 5\%$                         | $\tau = 10\%$      | P=250                | 0.1180  | 0.3920  | 0.2690   | 0.2770   |
|  |                    | P=500                | 0.2320  | 0.6250  | 0.4850   | 0.4940   |
| $\alpha = 10\%$                        | $\tau = 5\%$       | P=250                | 0.1690  | 0.4960  | 0.3220   | 0.3200   |
|  |                    | P=500                | 0.2650  | 0.7430  | 0.5360   | 0.5660   |
| $\alpha = 10\%$                        | $\tau = 10\%$      | P=250                | 0.2000  | 0.5730  | 0.3590   | 0.3520   |
|  |                    | P=500                | 0.3200  | 0.7970  | 0.5890   | 0.6140   |

Table 6: Powers of the Multivariate Ljung-Box and Univariate Ljung-Box tests using data generated from a bivariate BEKK DGP. The parametrization considered is:  $a_{0,11} = 14.8511$ ,  $a_{0,21} = 0.6318$ ,  $a_{0,22} = 1.0809$ ,  $a_{1,11} = 0.2525$ ,  $a_{1,12} = -0.2308$ ,  $a_{1,21} = -0.3709$ ,  $a_{1,22} = 0.0807$ ,  $b_{11} = 0.3503$ ,  $b_{12} = 0.6730$ ,  $b_{21} = -0.4592$ ,  $b_{22} = 0.3663$ . The innovations are assumed to follow a bivariate Student T distribution with variance covariance matrix:  $\sigma_{11} = \sigma_{22} = 1$ ,  $\sigma_{12} = \sigma_{21} = 0.4$  and  $\nu = 5$  degrees of freedom. VaR is computed at  $\alpha = 0.05$  and  $\alpha = 0.10$ . 1000 Monte Carlo simulations were performed and  $\pi$  is fixed at  $\pi = P/R = 0.05$ .

| Power of the Multivariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                               | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                           | $\tau = 5\%$       | P=250                | 0.1600  | 0.2260  | 0.2350   | 0.1970   |
|  |                    | P=500                | 0.1860  | 0.2480  | 0.2360   | 0.2260   |
| $\alpha = 1\%$                           | $\tau = 10\%$      | P=250                | 0.1880  | 0.2560  | 0.2580   | 0.2120   |
|  |                    | P=500                | 0.2450  | 0.2970  | 0.2850   | 0.2630   |
| $\alpha = 5\%$                           | $\tau = 5\%$       | P=250                | 0.1150  | 0.1400  | 0.1190   | 0.1090   |
|  |                    | P=500                | 0.1080  | 0.1290  | 0.1040   | 0.0980   |
| $\alpha = 5\%$                           | $\tau = 10\%$      | P=250                | 0.1620  | 0.1850  | 0.1650   | 0.1480   |
|  |                    | P=500                | 0.1520  | 0.1820  | 0.1590   | 0.1430   |
| $\alpha = 10\%$                          | $\tau = 5\%$       | P=250                | 0.0840  | 0.0860  | 0.0680   | 0.0720   |
|  |                    | P=500                | 0.0840  | 0.0730  | 0.0650   | 0.0740   |
| $\alpha = 10\%$                          | $\tau = 10\%$      | P=250                | 0.1340  | 0.1300  | 0.1090   | 0.1100   |
|  |                    | P=500                | 0.1180  | 0.1190  | 0.1240   | 0.1110   |

| Power of the Univariate Ljung-Box test |                    |                      |         |         |          |          |
|--|--------------------|----------------------|---------|---------|----------|----------|
|  |                    | Lags $K$             | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Risk level                             | Significance level | Forecast horizon $P$ |         |         |          |          |
| $\alpha = 1\%$                         | $\tau = 5\%$       | P=250                | 0.0690  | 0.2970  | 0.1110   | 0.1030   |
|  |                    | P=500                | 0.1000  | 0.3670  | 0.1330   | 0.1240   |
| $\alpha = 1\%$                         | $\tau = 10\%$      | P=250                | 0.0910  | 0.3760  | 0.1390   | 0.1280   |
|  |                    | P=500                | 0.1360  | 0.5050  | 0.1770   | 0.1600   |
| $\alpha = 5\%$                         | $\tau = 5\%$       | P=250                | 0.0540  | 0.3060  | 0.0570   | 0.0530   |
|  |                    | P=500                | 0.0670  | 0.3840  | 0.0700   | 0.0640   |
| $\alpha = 5\%$                         | $\tau = 10\%$      | P=250                | 0.0850  | 0.4480  | 0.0990   | 0.0910   |
|  |                    | P=500                | 0.1180  | 0.5090  | 0.1320   | 0.1100   |
| $\alpha = 10\%$                        | $\tau = 5\%$       | P=250                | 0.0590  | 0.3340  | 0.0500   | 0.0590   |
|  |                    | P=500                | 0.0620  | 0.3620  | 0.0650   | 0.0600   |
| $\alpha = 10\%$                        | $\tau = 10\%$      | P=250                | 0.1220  | 0.4830  | 0.0940   | 0.0980   |
|  |                    | P=500                | 0.1140  | 0.5360  | 0.1030   | 0.0990   |



Table 7: Application uses data provided by Christophe Perignon and Daniel R. Smith for P&L and VaR at 1% risk level for Bank of America (BOA) and Credit Suisse First Boston (CSFB). Data contains 1000 observations starting from January 1, 2001 up to December 31, 2004. Panel A displays p-values for the multivariate test, Panel B p-values for the univariate test for BOA and Panel C p-values for the univariate test for CSFB.  $\{*\}$  denotes statistical significance at 5% level and  $\{**\}$  at 1% level. Tests consider rolling windows of 250 trading days. Vio denotes banks' actual number of violations. The expected number of violations at 1% risk level for each 250 trading days rolling window is 2.5 .

| Panel A    |             | Lags $K$                | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
|------------|-------------|-------------------------|---------|---------|----------|----------|
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |         |         |          |          |
| 1          | 2           | P=1:250                 | 0.9998  | 1.0000  | 1.0000   | 1.0000   |
| 0          | 1           | P=251:500               | 1.0000  | 1.0000  | 1.0000   | 1.0000   |
| 3          | 2           | P=501:750               | 0.9986  | 1.0000  | 1.0000   | 1.0000   |
| 0          | 1           | P=751:1000              | 1.0000  | 1.0000  | 1.0000   | 1.0000   |
| Panel B    |             | Lags $K$                | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |         |         |          |          |
| 1          | 2           | P=1:250                 | 0.9487  | 1.0000  | 1.0000   | 1.0000   |
| 0          | 1           | P=251:500               | 1.0000  | 1.0000  | 1.0000   | 1.0000   |
| 3          | 2           | P=501:750               | 0.8456  | 0.9991  | 1.0000   | 1.0000   |
| 0          | 1           | P=751:1000              | 1.0000  | 1.0000  | 1.0000   | 1.0000   |
| Panel C    |             | Lags $K$                | $K = 1$ | $K = 5$ | $K = 10$ | $K = 15$ |
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |         |         |          |          |
| 1          | 2           | P=1:250                 | 0.8971  | 0.9999  | 1.0000   | 1.0000   |
| 0          | 1           | P=251:500               | 0.9487  | 1.0000  | 1.0000   | 1.0000   |
| 3          | 2           | P=501:750               | 0.8971  | 0.9999  | 1.0000   | 1.0000   |
| 0          | 1           | P=751:1000              | 0.9487  | 1.0000  | 1.0000   | 1.0000   |

Table 8: Application uses data provided by Christophe Perignon and Daniel R. Smith for P&L for Bank of America (BOA) and Credit Suisse First Boston (CSFB). Data contains 1000 observations starting from January 1, 2001 up to December 31, 2004. VaRs are obtained by fitting a BEKK(1,1,1) model. The risk level considered is 1 %. Panel A displays p-values for the multivariate test, Panel B p-values for the univariate test for BOA and Panel C p-values for the univariate test for CSFB. {\*} denotes statistical significance at 5% level and {\*\*} at 1% level. Tests consider rolling windows of 250 trading days. Vio denotes banks' actual number of violations. The expected number of violations at 1% risk level for each 250 trading days rolling window is 2.5 .

| Panel A    |             | Lags $K$                | $K = 1$         | $K = 5$         | $K = 10$        | $K = 15$ |
|------------|-------------|-------------------------|-----------------|-----------------|-----------------|----------|
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |                 |                 |                 |          |
| 0          | 2           | P=1:250                 | 1.0000          | 1.0000          | 1.0000          | 1.0000   |
| 0          | 2           | P=251:500               | 1.0000          | 1.0000          | 1.0000          | 1.0000   |
| 1          | 4           | P=501:750               | <b>0.0000**</b> | <b>0.0000**</b> | <b>0.0063**</b> | 0.2434   |
| 0          | 2           | P=751:1000              | 0.9994          | 1.0000          | 1.0000          | 1.0000   |
| Panel B    |             | Lags $K$                | $K = 1$         | $K = 5$         | $K = 10$        | $K = 15$ |
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |                 |                 |                 |          |
| 0          | 2           | P=1:250                 | 1.0000          | 1.0000          | 1.0000          | 1.0000   |
| 0          | 2           | P=251:500               | 0.9487          | 1.0000          | 1.0000          | 1.0000   |
| 1          | 4           | P=501:750               | <b>0.9487</b>   | <b>1.0000</b>   | <b>1.0000</b>   | 1.0000   |
| 0          | 2           | P=751:1000              | 0.9487          | 1.0000          | 1.0000          | 1.0000   |
| Panel C    |             | Lags $K$                | $K = 1$         | $K = 5$         | $K = 10$        | $K = 15$ |
| Vio<br>BOA | Vio<br>CSFB | Forecast<br>horizon $P$ |                 |                 |                 |          |
| 0          | 2           | P=1:250                 | 0.8971          | 0.9999          | 1.0000          | 1.0000   |
| 0          | 2           | P=251:500               | 0.9487          | 1.0000          | 1.0000          | 1.0000   |
| 1          | 4           | P=501:750               | <b>0.7943</b>   | <b>0.9976</b>   | <b>1.0000</b>   | 1.0000   |
| 0          | 2           | P=751:1000              | 0.8456          | 0.9995          | 1.0000          | 1.0000   |

Table 9: This table, from Perignon & Smith (2009), reports VaR backtesting information for Bank of America (BOA) and Credit Suisse First Boston (CSFB). Data is collected by Christophe Perignon and Daniel R. Smith from January 1, 2001 up to December 31, 2004. Banks' total assets (in millions) and ranking (based on total assets) are as of December 31, 2005 and were collected by the authors from Bankersalmanac.com. Figures on trading revenues are expressed in local currencies.

|                           | BOA                   | CSFB                  |
|---------------------------|-----------------------|-----------------------|
| Country                   | U.S.                  | Switzerland           |
| Currency                  | USD                   | CHF                   |
| Total Assets              | 1,082,243             | 1,016,050             |
| Worldwide Rank            | 7                     | 9                     |
| Domestic Rank             | 1                     | 2                     |
| VaR method                | Historical Simulation | Historical Simulation |
| Number of Trading Days    | 1,008                 | 1,031                 |
| % Days with Neg. Revenues | 13.00                 | 37.05                 |
| Expected Exceptions       | 10                    | 10                    |
| Actual Exceptions         | 4                     | 6                     |

Table 10: Banks Profit and Loss (P&L) and VaR Summary Statistics Table from Perignon & Smith (2009). This table displays for each bank the summary statistics for the daily trading revenue (P&L) and VaR data. Data are expressed in local currency and the sample period is 2001-2004. The summary statistics include the first four moments, minimum, and maximum for each variable, the Bera-Jarque normality test, the first-order autocorrelation coefficient, the Augmented Dickey-Fuller test (ADF), the Ljung-Box (LB) autocorrelation test using 12 lags, and the ARCH-12 test, which is a LB test applied to the squared demeaned returns. The ADF test includes an intercept, time trend and twelve lags. {\*} denotes statistical significance at 5% confidence level.

|                  | BOA       |           | CSFB      |           |
|------------------|-----------|-----------|-----------|-----------|
|                  | Daily P&L | Daily VaR | Daily P&L | Daily VaR |
| Mean             | 13.85     | 43.43     | 5.03      | 63.55     |
| Variance         | 222.24    | 144.14    | 369.16    | 215.89    |
| Skewness         | 0.123     | 0.10      | 0.240*    | 0.02      |
| Kurtosis         | 4.93*     | 2.81      | 9.90*     | 2.29      |
| Bera-Jarque Test | 159.11*   | 3.05      | 1,510.30* | 21.94     |
| Autocorrelation  | 0.064     | 0.892     | 0.124     | 0.947     |
| ADF              | -31.09*   | -9.45*    | -40.63*   | -9.53*    |
| LB-12            | 9.38      | 7,246.70  | 66.20*    | 8,563.80  |
| ARCH-12          | 23.73*    | 2,916.90  | 44.99*    | 4,350.40  |
| Minimum          | -57.39    | 11.38     | -105.30   | 27.16     |
| Maximum          | 84.33     | 90.49     | 138.45    | 100.19    |

Table 11: Correlations of Profit and Loss (P&L) and VaR time series across Bank of America (BOA) and Credit Suisse First Boston Bank (CSFB). Data is collected by Christophe Perignon and Daniel R. Smith from January 1, 2001 up to December 31, 2004.

|      | Daily P&L |      | Daily VaR |      |
|------|-----------|------|-----------|------|
|      | BOA       | CSFB | BOA       | CSFB |
| BOA  | 1         |      | 1         |      |
| CSFB | 0.0541    | 1    | 0.1681    | 1    |

Table 12: Table displays Basel Accord penalty zones. The number of violations are given for 250 trading days.  $k$  is the penalty.

| Zone   | Number of violations | Increase in $k$ |
|--------|----------------------|-----------------|
| Green  | 0 to 4               | 0.00            |
| Yellow | 5                    | 0.40            |
|        | 6                    | 0.50            |
|        | 7                    | 0.65            |
|        | 8                    | 0.75            |
|        | 9                    | 0.85            |
| Red    | 10+                  | 1.00            |

Table 13: Table displays the mean daily capital charge (in millions). The daily capital charge is given as the negative of the higher of the previous day VaR or the average VaR over the last 60 business days times  $(3 + k)$ , where  $k$  is the penalty.

| Panel A: Bank of America            |                      |                      |                           |
|-------------------------------------|----------------------|----------------------|---------------------------|
| Model                               | Forecast horizon $P$ | Number of violations | Mean daily capital charge |
| Historical Simulation               | P=1:250              | 1                    | 155.5478                  |
|                                     | P=251:500            | 0                    | 111.6825                  |
|                                     | P=501:750            | <b>3</b>             | <b>86.8456</b>            |
|                                     | P=751:1000           | 0                    | 123.1842                  |
| Model                               | Forecast horizon $P$ | Number of violations | Mean daily capital charge |
| BEKK(1,1,1)                         | P=1:250              | 0                    | 210.3212                  |
|                                     | P=251:500            | 0                    | 150.4526                  |
|                                     | P=501:750            | <b>1</b>             | <b>135.6461</b>           |
|                                     | P=751:1000           | 0                    | 161.6967                  |
| Panel B: Credit Suisse First Boston |                      |                      |                           |
| Model                               | Forecast horizon $P$ | Number of violations | Mean daily capital charge |
| Historical Simulation               | P=1:250              | 2                    | 201.5421                  |
|                                     | P=251:500            | 1                    | 199.4692                  |
|                                     | P=501:750            | 2                    | 164.2718                  |
|                                     | P=751:1000           | 1                    | 153.2637                  |
| Model                               | Forecast horizon $P$ | Number of violations | Mean daily capital charge |
| BEKK(1,1,1)                         | P=1:250              | 2                    | 213.6133                  |
|                                     | P=251:500            | 2                    | 163.2248                  |
|                                     | P=501:750            | <b>4</b>             | <b>108.0322</b>           |
|                                     | P=751:1000           | 2                    | 128.5367                  |

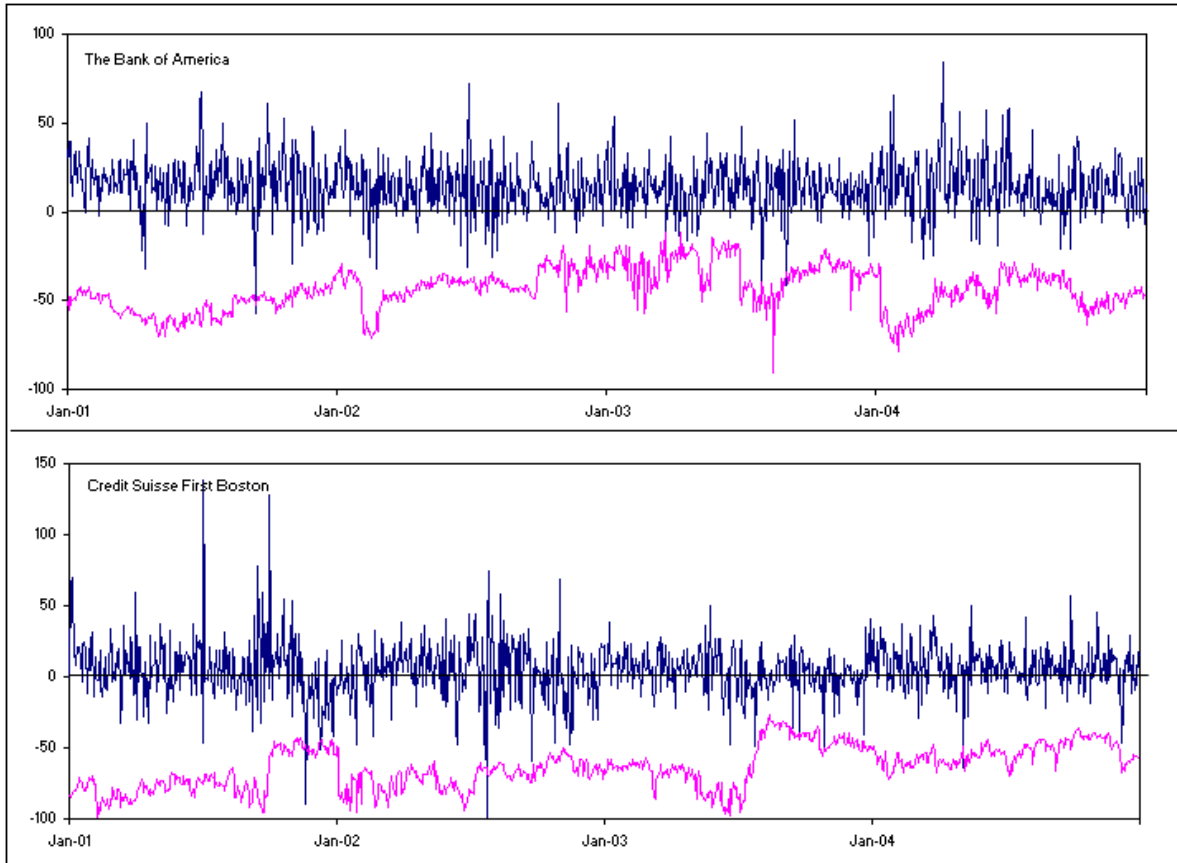


Figure 1: Figure displays the daily VaRs (lower line) and trading revenues (upper line) of Bank of America and Credit Suisse First Boston between January 1, 2001 and December 31, 2004. All values are in millions and are expressed in local currencies. Data from Perignon&Smith (2009).

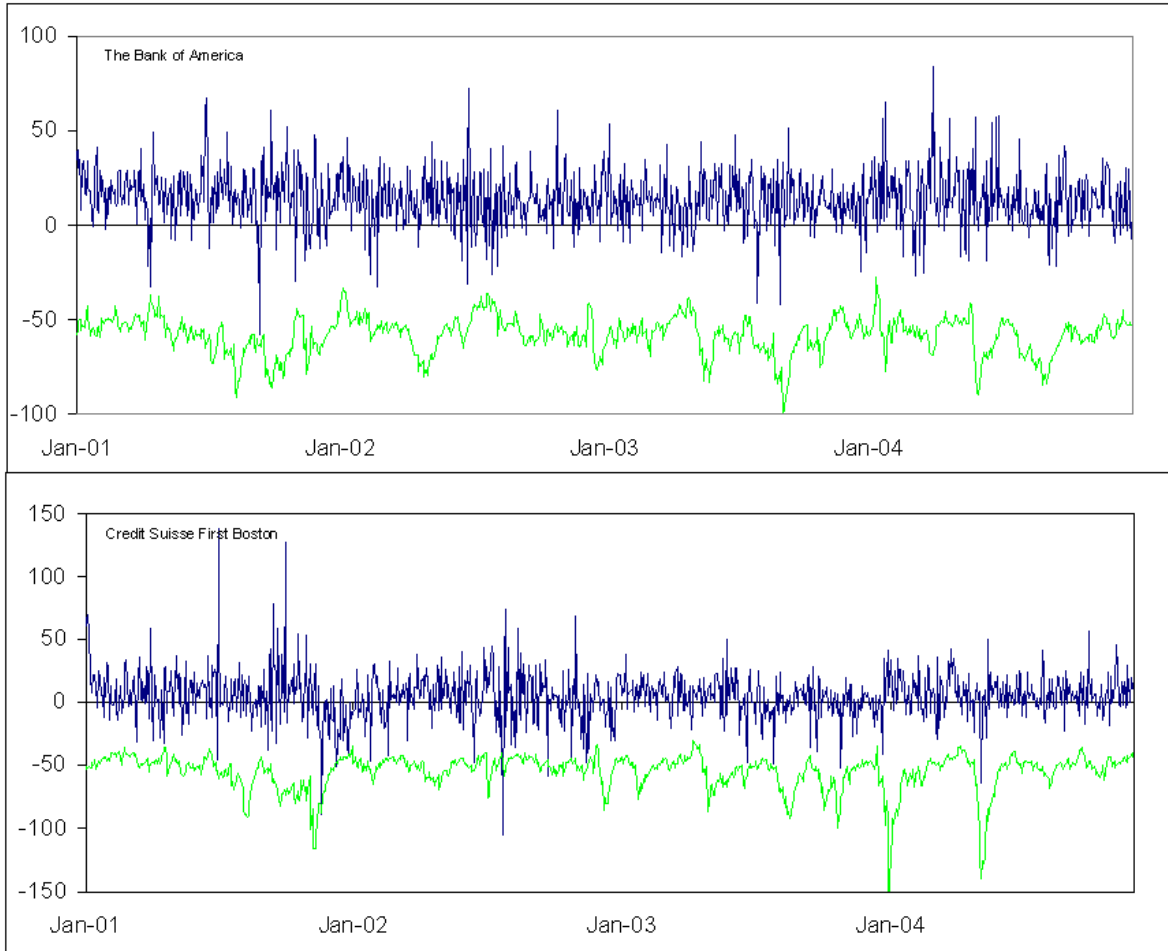


Figure 2: Figure displays the daily BEKK obtained VaRs (lower line) and trading revenues (upper line) of Bank of America and Credit Suisse First Boston between January 1, 2001 and December 31, 2004. All values are in millions and are expressed in local currencies. Trading revenues data are from Perignon&Smith (2009). VaR data is obtained by employing a BEKK(1,1,1) as risk model.

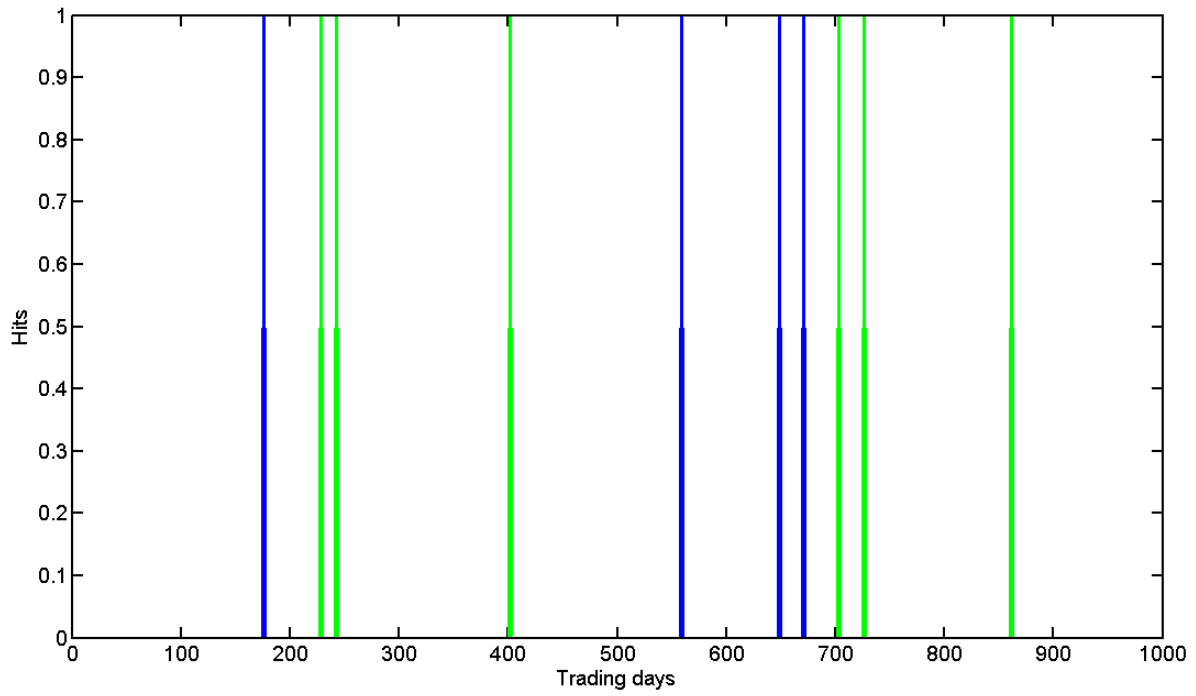


Figure 3: Figure displays the hits sequences for Bank of America (blue line) and Credit Suisse First Boston (green line) for 1000 trading days between January 1, 2001 and December 31, 2004.



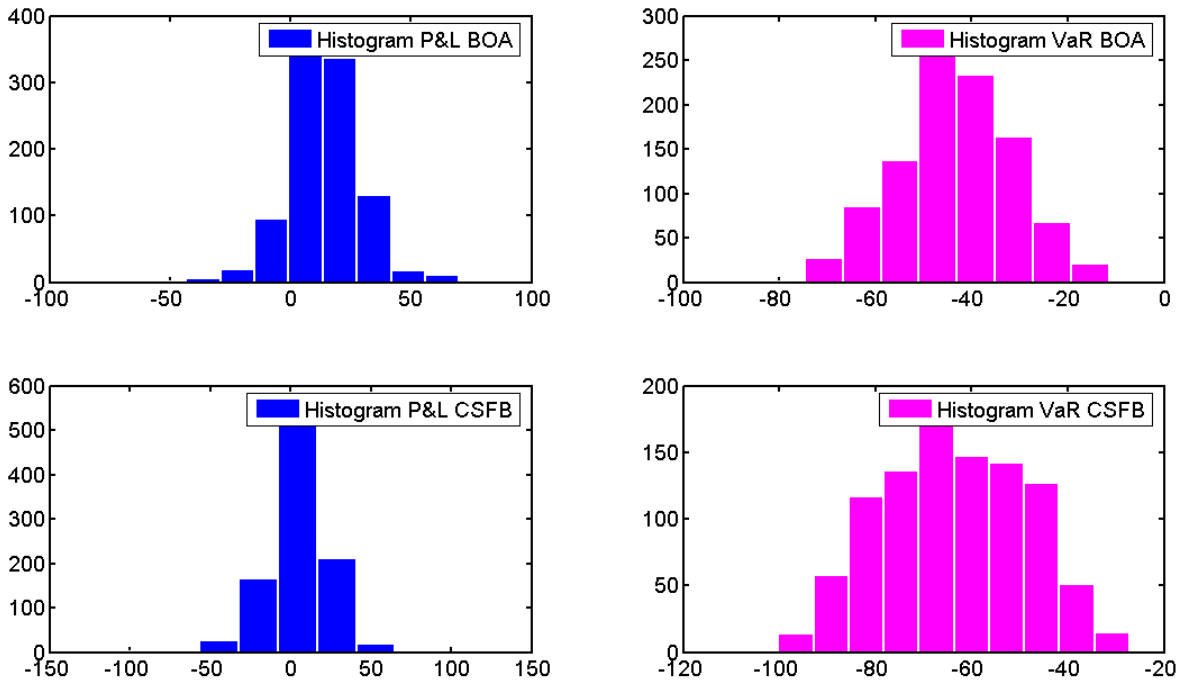


Figure 4: Figure displays the Profit/Loss (P&L) and VaR histograms of Bank of America (BOA) and Credit Suisse First Boston Bank (CSFB) for trading days between January 1, 2001 and December 31, 2004. The scale of the subplots differs across banks and is indicated on the x-axis.

## References

- Aad W. van der Vaart (1998): *Asymptotic Statistics.*, Cambridge University Press
- Aad W. van der Vaart and Jon A. Wellner (1996): *Weak Convergence and Empirical Processes with Applications to Statistics.*, Springer Series in Statistics
- Alexander J. McNeil, Rudiger Frey and Paul Embrechts (2005): *Quantitative Risk Management: Concepts, Techniques, and Tools.*, Princeton University Press, 2005
- Christophe Perignon and Daniel R. Smith (2008): A New Approach to Comparing VaR Estimation Methods., *Journal of Derivatives*, 2008, 54-66
- Christophe Perignon and Daniel R. Smith (2009): The Level and Quality of Value-at-Risk Disclosure by Commercial Banks., *Journal of Banking and Finance*, Forthcoming.
- Christophe Perignon and Daniel R. Smith (2009): Diversification and Value-at-Risk., *Journal of Banking and Finance*, Forthcoming.
- Cristina Danculescu: *Pitfalls and Solutions in Current Risk Management Methodology.*, Indiana University Bloomington, Working Paper
- G.E.P. Box and David A. Pierce (December 1970): Distribution of Residuals Autocorrelation in Autoregressive-Integrated Moving Average Time Series Models., *Journal of the American Statistical Association*, Volume 65, Number 332, Theory and Methods Section, Princeton University Press
- G.M. Ljung and G.E.P. Box (1978): On the lack of fit in the time series models., *Biometrika*, 65, 2, pp. 297-303, Great Britain
- Jeremy Berkowitz, Peter Christoffersen and Denis Pelletier: *Evaluating Value-at-Risk Models with Desk-Level Data.*, *Management Science*, Forthcoming
- Jeremy Berkowitz and James O'Brien (2002): How Accurate are Value-at-Risk Models at Commercial Banks?, *Journal of Finance*, VOL. LVII, NO. 3, June 2002
- Jeremy Berkowitz, James O'Brien (2007): *Estimating Bank Trading Risk: A Factor Model Approach.*, *The Risk of Financial Institutions*, University of Chicago Press
- J.R.M. Hosking (1980): The Multivariate Portmanteau Statistic., *Journal of the American Statistical Association*, September 1980, Volume 75, Number 371, Theory and Methods Section
- Juan Carlos Escanciano and Ignacio N. Lobato: An Automatic Portmanteau Test for Serial Correlation., *Journal of Econometrics* forthcoming, 2009

- Juan Carlos Escanciano and Miguel A. Delgado (2007): Nonparametric tests for conditional symmetry in dynamic models., *Journal of Econometrics*, 141(2007), 652-682
- Juan Calos Escanciano and Jose Olmo (2008): Backtesting Value-at-Risk with Estimation Risk., Forthcoming in *Journal of Business and Economics Statistics*
- Juan Calos Escanciano and Jose Olmo (November 2008): Robust Backtesting Tests For Value-at-Risk Models.,
- Kenneth D. West (September 1996): Asymptotic Inference About Predictive Ability., *Econometrica*, Vol. 64, No.5, 1067-1084
- Kenneth D. West, Michael W. McCracken (March 1998): Regression-Based Tests Of Predictive Ability., NBER Technical Working Paper No. 226
- Larry D. Haugh (June 1976): Checking the Independence of Two Covariance-Stationary Time Series: A Univariate Residual Cross-Correlation Approach., *Journal of the American Statistical Association*, Vol. 71, Number 354, Theory and Methods Section
- Michael W. McCracken (2000): Robust out-of-sample inference., *Journal of Econometrics* 99, 195-223
- Oliver Linton: Topics in Financial Markets: Teaching Material. Discrete time volatility., London School of Economics, Department of Economics
- Peter F. Christoffersen (1998): Evaluating Interval Forecasts., *International Economic Review*, Vol. 39, No. 4, November 1998
- Phillipe Artzner, Freddy Delbaen, Jean-Mark Eber, David Heath (1999): Coherent measures of risk., *Mathematical Finance* 9, 203-228
- Ralf Bender, Ulrich Grouven (1993): On the Choice of the Number of Residual Autocovariances for the Portmanteau Test of Multivariate Autoregressive Models., *Communications in Statistics-Simulation and Computation*, 22, pp. 19-32
- Raffaella Giacomini, Ivana Komunjer (2005): Evaluation and Combination of Conditional Quantile Forecasts., *Journal of Business and Economics Statistics*, 24, 4, 416-431(16)
- Radhakrishna C. Rao (1973): *Linear Statistical Inference and its Applications*. (2nd ed.), New York: John Wiley & Sons
- Ratnam V. Chitturi (1974): Distribution of Residual Autocorrelations in Multiple Autoregressive Schemes., *Journal of the American Statistical Association*, December 1974, Volume 69, Number 348, Theory and Methods Section

- Robert F. Engle, Kenneth F. Kroner (1995): Multivariate Simultaneous Generalized ARCH., *Econometric Theory* (1995), 11:122-150, Cambridge University Press
- Robert F. Engle, Simone Manganelli (2004): CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles., *Journal of Business and Economics Statistics*
- Robert F. Engle, Simone Manganelli (1999): Conditional Value at Risk: Modeling a Time Varying Order Statistic.,
- Robert F. Engle, Simone Manganelli (2001): Value at Risk in Finance., European Central Bank
- Roger Koenker (2005): *Quantile Regression.*, Cambridge University Press
- Tobias Adrian, Markus K. Brunnermeier (2008): CoVaR., Federal Reserve Bank of New York
- Torben G. Andersen, Tim Bollerslev, Peter F. Christoffersen, Francis X. Diebold (2007): Practical Volatility and Correlation Modeling for Financial Market Risk Management., *The Risk of Financial Institutions*, Mark Carey and Rene M. Stulz (Editors), University of Chicago Press
- Yongmiao Hong (1996): Testing for independence between two covariance stationary time series., *Biometrika*, 83, 3, pp. 615-625
- Yongmiao Hong, Yanhui Liu, Shouyang Wang: Granger Causality in Risk and Detection of Risk Transmission Between Financial Markets., *Journal of Econometrics* (2009), Forthcoming