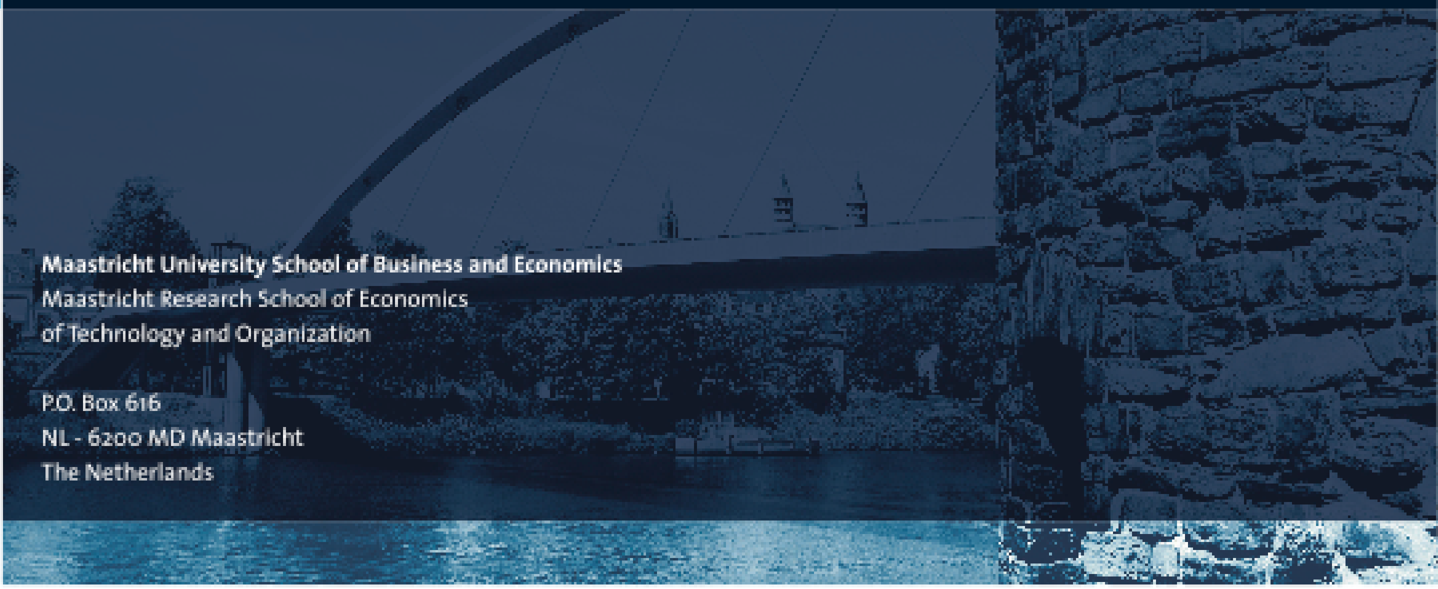
Maastricht University

Bertrand Candelon, Gilbert Colletaz,
Christophe Hurlin, Sessi Tokpavi

Backtesting Value-at-Risk: A GMM Duration-based Test

RM/09/051

METEOR



Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

Backtesting Value-at-Risk: A GMM Duration-based Test*

Bertrand Candelon†, Gilbert Colletaz‡, Christophe Hurlin§, Sessi Tokpavi¶

2nd November 2009

Abstract

This paper proposes a new duration-based backtesting procedure for VaR forecasts. The GMM test framework proposed by Bontemps (2006) to test for the distributional assumption (i.e., the geometric distribution) is applied to the case of VaR forecast validity. Using simple J-statistics based on the moments defined by the orthonormal polynomials associated with the geometric distribution, this new approach tackles most of the drawbacks usually associated with duration based backtesting procedures. In particular, it is among the first to take into account problems induced by the estimation risk in duration-based backtesting tests and to offer a sub-sampling approach for robust inference derived from Escanciano and Olmo (2009). An empirical application of the method to Nasdaq returns confirms that using the GMM test has major consequences for the ex-post evaluation of risk by regulation regulatory authorities.

Keywords: Value-at-Risk, backtesting, GMM, duration–based test, estimation risk.
J.E.L Classification : C22, C52, G28

*The authors thank Jose Olmo and Enrique Sentana for comments on the paper, as well as the participants of the "Methods in International Finance Network" (Barcelona, 2008); the 2^{nd} International Financial Research Forum, entitled "Risk Management and Financial Crisis" (Paris, March 2009); the $5^{ème}$ Journée d'économétrie "Développements récents de l'économétrie appliquée à la finance" (2008, Paris) and the 64^{th} European Meeting of the Econometric Society. The paper was partially written during Christophe Hurlin's visit to Maastricht University through the visiting professorship program of METEOR. Usual disclaimers apply.

†b.candelon@algec.unimaas.nl. Maastricht University, Department of Economics. The Netherlands.

‡gilbert.colletaz@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.

§christophe.hurlin@univ-orleans.fr, University of Orléans, Laboratoire d'Economie d'Orléans (LEO), France.

¶sessi.tokpavi@u-paris10.fr , University of Paris X, Economix, France.

1 Introduction

The recent Basel II agreements have left open the possibility for financial institutions to develop and apply their own internal model of risk management. Value-at-Risk (VaR hereafter), which measures the quantile of the distribution of gains and losses over a target horizon, constitutes the most popular measure of risk. Consequently, regulatory authorities need to set up adequate *ex-post* techniques to validate or invalidate the amount of risk taken by financial institutions. The standard assessment method for VaR consists of backtesting or reality check procedures. As defined by Jorion (2007), backtesting is a formal statistical framework that verifies whether actual trading losses are in line with projected losses. This approach involves a systemic comparison of the history of model-generated VaR forecasts with actual returns and generally relies on testing over VaR violations (also called the Hit).

A violation is said to occur when *ex-post* portfolio returns are lower than VaR forecasts. Christoffersen (1998) argues that a VaR with a chosen coverage rate of $\alpha\%$ is valid as soon as VaR violations satisfy both the hypothesis of unconditional coverage and independence. The hypothesis of unconditional coverage means that the expected frequency of observed violations is precisely equal to $\alpha\%$. If the unconditional probability of violation is higher than $\alpha\%$, then the VaR model understates the portfolio's actual level of risk. The opposite finding, with too few VaR violations, would signal an overly conservative VaR measure. The hypothesis of independence means that if the model of VaR calculation is valid, then violations must be distributed independently. In other words, clusters must not appear in the violation sequence. Both assumptions are essential to characterizing VaR forecast validity: only hit sequences that satisfy each of these conditions (and hence the conditional coverage hypothesis) can be presented as evidence of a useful VaR model.

Although the literature about conditional coverage is quite recent, various tests on independence and unconditional coverage hypotheses have already been developed (see Campbell, 2007 for a survey). Most of them directly employ the violation process.[1] Yet another line of thinking within the literature uses the statistical properties of the duration between two consecutive hits. The baseline idea is that if the VaR one period ahead is correctly specified for a coverage rate α, then the durations between two consecutive hits must have a geometric distribution, with a probability of success equal to $\alpha\%$. On these grounds, Christoffersen and Pelletier (2004) proposed a test of independence. Their duration-based backtesting test specifies a duration distribution that nests the geometric one and allows for duration dependence. The independence hypothesis can thus be tested by means of simple likelihood ratio (LR) tests. This general duration-based approach to backtesting is very appealing (see Haas, 2007) given its ease of use and that it provides a clear-cut interpretation of parameters. Nevertheless, it must be stressed that this approach requires one to specify a particular distribution under the alternative hypothesis, which is not always easy to do. Consequently, duration-based backtesting methods have relatively little power for realistic sample sizes (Haas, 2005). For these reasons, actual duration-based backtesting procedures are not very popular among practitioners. The aim of this paper is to improve these procedures and make them more appealing for practitioners.

Relying on the GMM framework of Bontemps and Meddahi (2005, 2006), we derive test statistics similar to J-statistics relying on particular moments defined by the orthonormal polynomials associated with the geometric distribution. This duration-based backtest considers discrete lifetime distributions, which we expect to improve power and size compared to those of competitors using continuous approximations, such as that of Christoffersen and Pelletier (2004). In sum, the present approach appears to have several advantages. First, it provides a unified framework that we can use to separately investigate the unconditional coverage hypothesis, the independence assumption and the conditional coverage hypothesis. Secondly, the optimal weight matrix of our test is known and does not need to be estimated. Thirdly, the GMM statistics can be

[1] Examples include Christoffersen's test (1998) based on the Markov chain, the hit regression test developed by Engle and Manganelli (2004) that relies on a linear auto-regressive model and the tests by Berkowitz et al. (2005) based on tests of martingale difference.

numerically computed for almost all realistic backtesting sample sizes. Fourth, elaborating on a study by Escanciano and Olmo (2008), this paper is the first to control for estimation uncertainty using a subsampling approach for backtesting duration-based tests. Fifth, in contrast with the LR tests, this technique does not impose a particular distribution under the alternative hypothesis. Finally, some Monte-Carlo simulations indicate that, for realistic sample sizes, our GMM test has good power properties as compared to other backtests, especially those based on an LR approach.

The paper is organized as follows. In section 2, we present the main VaR assessment tests and, more specifically, the duration-based backtesting procedures. Section 3 presents our GMM duration-based test. In section 4, we present the results of various Monte Carlo simulations in order to illustrate the finite sample properties of the proposed test. Section 5 is devoted to the issue of parameter uncertainty and robust inference. In section 6, we present an empirical application of the method using daily Nasdaq returns. Finally, the last section concludes.

2 Environment and testable hypotheses

Let r_t denote the return of an asset or a portfolio at time t and $VaR_{t|t-1}(\alpha)$ the *ex-ante* VaR forecast obtained conditionally on an information set $\mathcal{F}_{t-1}$ and for an $\alpha\%$ coverage rate. If the VaR model is valid, then the following relation must hold:

$$\Pr[r_t < VaR_{t|t-1}(\alpha)] = \alpha, \;\; \forall t \in \mathbb{Z}. \tag{1}$$

Let $I_t\left(\alpha\right)$ be the hit variable associated with the *ex-post* observation of an $\alpha\%$ VaR violation at time t, *i.e.*:

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

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

- The unconditional coverage (UC hereafter) hypothesis: the probability of a *ex-post* return exceeding the VaR forecast must be equal to the α coverage rate
$$\Pr\left[I_t(\alpha)=1\right]=\mathbb{E}\left[I_t(\alpha)\right]=\alpha. \tag{3}$$
- The independence (IND hereafter) hypothesis: VaR violations observed at two different dates for the same coverage rate must be distributed independently. Formally, the variable $I_t(\alpha)$ associated with a VaR violation at time t for an $\alpha\%$ coverage rate should be independent of the variable $I_{t-k}(\alpha)$, $\forall k \neq 0$. In other words, past VaR violations should not be informative of current and future violations[2].

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

$$\mathbb{E}\left[I_t(\alpha)-\alpha \mid \Omega_{t-1}\right]=0. \tag{4}$$

It is worth noting that equation (4) implies that the violation sequence $\{I_t(\alpha)\}$ is a random sample from a a Bernoulli distribution with a success probability equal to α

$$\{I_t(\alpha)\} \text{ are } i.i.d. \text{ Bernoulli random variables } (r.v.). \tag{5}$$

This last property is at the core of most of the backtests of VaR models presented in the literature (Christoffersen, 1998; Engle and Manganelli, 2004; Berkowitz, Christoffersen and Pelletier, 2009; etc.). However, as suggested by Christoffersen and Pelletier (2004), another appealing way of testing (5) is reliance on the duration between two consecutive violations. Formally, let us denote d_i the duration between two consecutive violations as

$$d_i=t_i-t_{i-1}, \tag{6}$$

[2] The independence property of violations is an essential property because it is related to the ability of a VaR model to accurately model the higher-order dynamics of returns. In fact, that which does not satisfy the independence property can lead to clusters of violations (for a given period) even if it has the correct average number of violations. Thus, there must be no dependence in the hit variable, regardless of the considered coverage rates.

where t_i denotes the date of the i^{th} violation. Under the CC hypothesis, the duration variable $\{d_i\}$ follows the pattern of a geometric distribution with parameter α and a probability mass function given by

$$f(d;\alpha) = \alpha(1-\alpha)^{d-1} \qquad d \in \mathbb{N}^*. \tag{7}$$

The geometric distribution characterizes the memory-free property of the violation sequence $\{I_t(\alpha)\}$, which means that the probability of observing a violation today does not depend on the number of days that have elapsed since the last violation. Exploiting (7), development of a likelihood ratio (LR) test for the null of the CC hypothesis is straightforward. The general idea is to specify a lifetime distribution that nests the geometric distribution, so that the memoryless property can be tested by means of LR tests. In this line of thinking, Christoffersen and Pelletier (2004) propose the first duration-based test. They used the exponential distribution under the null hypothesis, which is the continuous analogue of the geometric distribution with a probability density function defined as follows:

$$g(d;\alpha) = \alpha \exp(-\alpha d). \tag{8}$$

with $\mathbb{E}(d) = 1/\alpha$ because the CC hypothesis implies that a mean duration between two violations is equal to $1/\alpha$. Under the alternative hypothesis, they postulate a Weibull distribution for the duration variable with distribution function

$$h(d;a,b) = a^b b d^{b-1} \exp\left[-(ad)^b\right]. \tag{9}$$

Because the exponential distribution corresponds to a Weibull distribution with a flat hazard function, *i.e* $b = 1$, the test for IND (Christoffersen and Pelletier, 2004) is then simply as follows:

$$H_{0,IND}: \; b = 1. \tag{10}$$

In a recent work, Berkowitz et al.(2009) extended this approach to consider the CC hypothesis; that is,

$$H_{0,CC}: b = 1, \; a = \alpha, \tag{11}$$

They also propose the corresponding LR test. Nevertheless, as stressed by Haas (2005), relying on the continuous approximation of the geometric distribution is not entirely satisfying and can have major consequences for the finite sample properties of the duration-based backtests. He then motivates the use of discrete lifetime distributions instead of continuous ones, arguing that the parameters of the distribution have a clear-cut interpretation in terms of risk management. He also conducts Monte-Carlo experiments showing that the backtesting tests based on discrete distribution exhibit a higher power than the continuous competitor tests.

However, some limitations may explain the lack of popularity of duration-based backtesting tests among practitioners. First, they exhibit low power for realistic backtesting sample sizes. For instance, in some GARCH-based experiments, Haas (2005) found that for a backtesting sample size of 250, the LR independence tests have an effective power that ranges from 4.6% (continuous Weibull test) to 7.3% (discrete Weibull test) for a nominal coverage of 1%VaR. Similarly, for coverage of 5%VaR, the level of power only reaches 14.7% for the continuous Weibull test and 32.3% for the discrete Weibull test. In other words, when VaR forecasts are not valid, LR tests do not reject VaR validity in 7 cases out of 10 at best. Secondly, duration-based tests do not allow formal separate tests for (*i*) unconditional coverage, the (*ii*) conditional coverage assumption or eventually (*iii*) the independence assumption within a unified framework.[3]

3 A GMM duration-based test

In this paper, we propose a new duration-based backtesting test able to tackle these issues. Based on a GMM approach and orthonormal polynomials, our test is in line with the distributional testing procedures recently proposed by Bontemps and Meddahi (2005, 2006) and Bontemps (2006). Our approach presents several advantages. First, it is extremely easy to implement because it consists of a simple GMM moment condition test. Second, it allows for optimal treat-

[3] Unlike in the other approaches based on violations processes (Christoffersen, 1998 or Engle and Manganelli, 2004).

ment of the problem associated with parameter uncertainty. Third, the choice of moment conditions enables us to develop separate tests for the UC, IND and CC assumptions, which was not possible with the existing duration-based tests. Finally, Monte-Carlo simulations will show that this new test has relatively good power properties. Our approach is further discussed in the following section.

3.1 Orthonormal Polynomials and Moment Conditions

In the continuous case, it is well known that the Pearson family of distributions (Normal, Student, Gamma, Beta, Uniform) can be associated with some particular orthonormal polynomials that have an expectation equal to zero. These polynomials can be used as special moments to test for a distributional assumption. For instance, the Hermite polynomials associated with the normal distribution are employed to test for normality (Bontemps and Meddahi, 2005). In the discrete case, orthonormal polynomials can be defined for distributions belonging to the Ord family (Poisson, Binomial, Pascal, hypergeometric). The orthonormal polynomials associated with the geometric distribution (7) are defined[4] as follows:

Definition 1 *The orthonormal polynomials associated with a geometric distribution with a success probability β are defined by the following recursive relationship $\forall d \in \mathbb{N}^*$:*

$$M_{j+1}(d;\beta) = \frac{(1-\beta)(2j+1) + \beta(j-d+1)}{(j+1)\sqrt{1-\beta}} M_j(d;\beta) - \left(\frac{j}{j+1}\right) M_{j-1}(d;\beta), \tag{12}$$

for any order $j \in \mathbb{N}$, with $M_{-1}(d;\beta) = 0$ and $M_0(d;\beta) = 1$. If the true distribution of D is a geometric distribution with a success probability β, then it follows that

$$\mathbb{E}[M_j(d;\beta)] = 0 \qquad \forall j \in \mathbb{N}^*, \forall d \in \mathbb{N}^*. \tag{13}$$

Our duration-based backtest procedure utilizes these moment conditions. More precisely, let us define $\{d_1; ...; d_N\}$ as a sequence of N durations between

[4] These polynomials can be viewed as a particular subset of the Meixner orthonormal polynomials associated with a Pascal (negative Binomial) distribution.

VaR violations, computed from the sequence of hit variables $\{I_t(\alpha)\}_{t=1}^{T}$. Under the conditional coverage assumption, the durations d_i, $i = 1, .., N$, are *i.i.d.* and have a geometric distribution with a success probability equal to the coverage rate α. Hence, the null of CC can be expressed as follows:[5]

$$H_{0,CC}: \ \mathbb{E}\left[M_j\left(d_i;\alpha\right)\right] = 0, \quad j = \{1, .., p\}, \tag{15}$$

where p denotes the number of moment conditions.

This framework allows one to separately test for the UC and IND hypothesis. First, the null UC hypothesis can be expressed as

$$H_{0,UC}: \ \mathbb{E}\left[M_1\left(d_i;\alpha\right)\right] = 0. \tag{16}$$

Indeed, under UC, the expectation of the duration variable is equal to $1/\alpha$. Because $M_1(d;\alpha) = (1-\alpha d)/\sqrt{1-\alpha}$, verification that the condition $\mathbb{E}\left[M_1(d;\alpha)\right] = 0$ is equivalent to the UC condition $\mathbb{E}(d_i) = 1/\alpha$, $\forall d_i$ is straightforward.

Second, a separate test for the IND hypothesis can also be derived. It consists of testing the hypothesis of a geometric distribution (implying the absence of dependence) with a success probability equal to β, where β denotes the true violation rate, which is not necessarily equal to the coverage rate α. This independence assumption can be expressed with the following moment conditions:

$$H_{0,IND}: \ \mathbb{E}\left[M_j\left(d_i;\beta\right)\right] = 0 \quad j = 1, .., p, \tag{17}$$

In this case, the expectation of the duration variable is equal to $1/\beta$ as soon as the first polynomial $M_1(d;\beta)$ is included in the set of moments conditions. Therefore, under $H_{0,IND}$, the duration between two consecutive violations has a geometric distribution, and the correct UC is not valid if $\beta \neq \alpha$.

[5] It is possible to test the conditional coverage assumption by considering at least two moment conditions, even if they are not consecutive, as soon as the first condition $\mathbb{E}\left[M_1(d_i)\right] = 0$ is included in the set of moments. For instance, it is possible to test the CC with the following:

$$H_{0,CC}: \ \mathbb{E}\left[M_j(d_i)\right] = 0 \quad j = \{1, 3, 7\} \tag{14}$$

For the sake of simplicity, we exclusively consider the cases in which moment conditions are consecutive polynomials in the remainder of the paper.

3.2 Empirical Test Procedure

It turns out that VaR forecast tests can be tested within the well-known GMM framework. As observed by Bontemps (2006), the orthonormal polynomials present the great advantage that their asymptotic matrix of variance covariance is known. Indeed, in an *i.i.d.* context, the moments are asymptotically independent with unit variance. As a result, the optimal weight matrix of the GMM criteria is simply an identity matrix, and the implementation of the backtesting test becomes very easy. Let us denote $J_{CC}(p)$ as the CC statistic test associated with the p first orthonormal polynomials.

Proposition 2 *Assume that the duration process* $\{d_i : 1 \leq i\}$ *is stationary and ergodic. Under the null hypothesis (15) of conditional coverage (CC), we have*

$$J_{CC}(p) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\alpha) \right)^{\intercal} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\alpha) \right) \underset{N\to\infty}{\overset{d}{\longrightarrow}} \chi^2(p) . \quad (18)$$

where $M(d_i;\alpha)$ *denotes a* $(p,1)$ *vector whose components are the orthonormal polynomials* $M_j(d_i;\alpha)$ *, for* $j = 1,..,p$ *and* α *denotes the coverage rate* α*.*

The proof follows from Lemma 4.2. in Hansen (1982). Note that among the assumptions used by Hansen (1982) to derive the asymptotic distribution of the over-identified restrictions test statistic, the only relevant ones in this framework are the stationarity and ergodicity of the process that defines the moment conditions (here, the duration variable).

The test statistic for UC, denoted as J_{UC}, is obtained as a special case of the J_{CC} statistic when one considers only the first orthonormal polynomial–*i.e.* when $M(d_i;\alpha) = M_1(d_i;\alpha)$. J_{UC} is then equivalent to $J_{CC}(1)$ and verifies

$$J_{UC} = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_1(d_i;\alpha) \right)^2 \underset{N\to\infty}{\overset{d}{\longrightarrow}} \chi^2(1). \quad (19)$$

Finally, the statistic for IND, denoted as J_{IND}, can be expressed as follows

$$J_{IND}(p) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\beta) \right)^{\intercal} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\beta) \right) \underset{N\to\infty}{\overset{d}{\longrightarrow}} \chi^2(p) . \quad (20)$$

where $M(d_i;\beta)$ denotes a $(p,1)$ vector with components that are the orthonormal polynomials $M_j(d_i;\beta)$, for $j=1,..,p$, evaluated for a success probability equal to β.

However, in this case, the true VaR violations rate β (which may be different from the coverage rate α fixed by the risk manager in the model) is generally unknown. Consequently, the independence test statistic must be based on orthonormal polynomials that depend on estimated parameters, *i.e.* instead of having $M_j(d_i;\beta)$ where β is known, we have to deal with $M_j\left(d_i;\widehat{\beta}\right)$ where $\widehat{\beta}$ denotes a square-N-root-consistent estimator of β. It is well known that replacing the true value of β with its estimates $\widehat{\beta}$ may change the asymptotic distribution of the GMM statistic. However, Bontemps (2006) shows that the asymptotic distribution remains unchanged if the moments can be expressed as a projection onto the orthogonal of the score. Appendix A shows that the moment conditions defined by the Meixner orthonormal polynomials satisfy this property. Thus, it can be concluded that the asymptotic distribution of the GMM statistic $J_{IND}(p)$ based on $M_j\left(d_i;\widehat{\beta}\right)$ is similar to the one based on $M_j(d_i;\beta)$.

$$J_{IND}(p)=\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}M\left(d_i;\widehat{\beta}\right)\right)^{\intercal}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}M\left(d_i;\widehat{\beta}\right)\right)\underset{N\to\infty}{\overset{d}{\longrightarrow}}\chi^2(p-1), \tag{21}$$

where $M\left(d_i;\widehat{\beta}\right)$ denotes the $(1,p)$ vector defined as $\left(M_1\left(d_i;\widehat{\beta}\right)...M_p\left(d_i;\widehat{\beta}\right)\right)$. Note that in this case, the first polynomial $M_1\left(d_i;\widehat{\beta}\right)$ is strictly proportional to the score used to define the maximum likelihood estimator $\widehat{\beta}$ and thus $M_1\left(d_i;\widehat{\beta}\right)=0$. Therefore, the degree of freedom of the J-statistic must be adjusted accordingly.

4 Small Sample Properties

In this section, we use Monte Carlo simulations to illustrate the finite sample properties (empirical size and power) of the conditional coverage test statistic $J_{CC}(p)$. However, it is worth noting that one of the main issues in the literature on VaR assessment is the relative scarcity of violations. Indeed, even with one year of daily returns, the number of observed durations between two consecutive violations may often be dramatically small (in particular for a 1% coverage rate), and this situation can lead to small sample bias. For this reason, the size of the test should be controlled using, for example, the Monte Carlo (MC) testing approach of Dufour (2006)–as done, for example, in Christoffersen and Pelletier (2004) and Berkowitz et al. (2009).

4.1 Empirical Size Analysis and Numerical Aspects

To illustrate the size performance of our duration-based test using a finite sample, we generated a hits sequence of violations by taking independent draws from a Bernoulli distribution, considering successively $\alpha = 1\%$ and $\alpha = 5\%$ for the VaR nominal coverage. Several sample sizes T ranging from 250 (which roughly corresponds to one year of trading days) to $1,500$ were also used. The durations were computed using the simulated hits sequence and reported empirical sizes correspond to the rejection rates calculated over $10,000$ simulations for a nominal size equal to 5%.

Insert Table 1

Table 1 reports the empirical sizes of the $J_{CC}(p)$ test statistic for different values of p the number of orthonormal polynomials. For the purpose of comparison, we also display results for the duration-based CC test statistics of Berkowitz et al. (2009). Recall that their test statistic, denoted as LR_{CC}, is designed to test the exponential distribution of the duration variable within a likelihood ratio framework. We also present the results of the CC test in Christoffersen (1998), denoted as LR_{CC}^{Markov}. This CC test is currently one of the most used in empirical studies. It is directly based on the violation process (and not on duration) and employs a Markov chain approach.[6]

For a 5% VaR and whatever the choice of p, the empirical size of the J_{CC} test is below the nominal size but relatively close to 5%, even for small sample

[6] We are grateful to an anonymous referee for this suggestion.

sizes. On the contrary, we verify that both LR tests are over-sized. For a 1% VaR, the J_{CC} test is undersized for a finite sample but converges to the nominal size when T increases. However, recall that under the null hypothesis in a sample with $T = 250$ and a coverage rate equal to 1%, the expected number of durations between two consecutive hits ranges from two to three. This scarcity of violations explains why the empirical size of our asymptotic test is different from the nominal size in small samples.

It is important to note that these rejection frequencies are only calculated for the simulations providing a J_{CC} as well as the LR test statistics. Indeed, for a realistic backtesting sample size (for instance, $T = 250$) and a coverage rate of 1%, many simulations do not deliver a statistic. The implementation of LR_{CC} test statistic requires at least one non-censored duration and an additional possible censored duration (*i.e.* two violations). Our GMM test statistic also requires at least two violations because it can be computed only by using uncensored durations. Indeed, if the duration is censored or truncated, then the polynomial $M_j(d_i, \alpha)$ do not have a zero expectation. This is particularly clear for the first polynomial $M_1(d_i, \alpha)$: if duration d_i is truncated or censored, then its unconditional expectation $E(d_i)$ is different from $1/\alpha$ under $H_{0,CC}$, and so $\mathbb{E}[M_1(d_i; \alpha)] = (1 - \alpha\mathbb{E}(d_i)) / \sqrt{1-\alpha}$ is also different from zero.

In order to assess the influence of these truncated durations on the finite sample of our test, we report in Table 1 the size of J_{CC}^{cens} test statistic calculated using both uncensored durations and censored durations (observed before the first VaR violation and after the latest VaR violation). We observe that the sizes are then comparable to those previously obtained and that the difference tends to disappear as T increases and the weight of the two truncated durations decreases.

Insert Table 2

Table 2 reports the feasibility ratios, *i.e.* the fraction of simulated samples where the LR, the J_{CC}^{cens} and the J_{CC} tests are feasible. Theoretically, the feasibility ratios should be exactly the same for our J-statistic (based on uncensored durations) and for the Christoffersen's LR test. However, we can observe that the feasibility ratios of the J-statistic are slightly superior to those of the LR test for a small samples size. Indeed, for some simulations in which there are only two violations (i.e., three durations), the numerical optimization of the likelihood function for the Weibull distribution[7] under H_1 cannot be achieved, and then the LR cannot be computed. In contrast, the J statistic does not require any optimization and so can always be computed. These cases are relatively rare but explain the difference between feasibility ratios.

4.2 Empirical Power Analysis

We now investigate the power of the test for different alternative hypotheses. Following Christoffersen and Pelletier (2004), Berkowitz et al.(2009) or Haas (2005), the DGP under the alternative hypothesis assumes that returns, r_t, are issued from a GARCH$(1,1)$-$t\,(d)$ model with an asymmetric leverage effect. More precisely, it corresponds to the following model:

$$r_t = \sigma_t \; z_t \; \sqrt{\frac{v-2}{v}}, \tag{22}$$

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

$$\sigma_t^2 = \omega + \gamma \sigma_{t-1}^2 \left(\sqrt{\frac{v-2}{v}} z_{t-1} - \theta \right)^2 + \beta \sigma_{t-1}^2. \tag{23}$$

The parameterization of the coefficients is similar to that proposed by Christoffersen and Pelletier (2004) and used by Haas (2005)–*i.e.* $\gamma = 0.1$, $\theta = 0.5$, $\beta = 0.85$, $\omega = 3.9683e^{-6}$ and $d = 8$. The value of ω is set to target an annual standard deviation of 0.20, and the global parameterization implies a daily volatility persistence of 0.975.

Using the simulated Profit and Loss (P&L thereafter) distribution issued from this DGP, it is then necessary to select a method of forecasting the VaR.

[7] The corresponding codes are based on the function wblfit of Matlab 7.10. The feasibility ratio varies with the choice of initial conditions. The reported results correspond to the initial conditions defined by default in Matlab. All programs are available at $http:$ $http://www.univ-orleans.fr/deg/masters/ESA/CH/churlin_R.htm$

This choice is of major importance for the power of the test. Indeed, it is necessary to choose a VaR calculation method that is not adapted to the P&L distribution, as that adaptation would violate efficiency– *i.e* the nominal coverage and\or independence hypothesis. Of course, we expect that the larger the deviation from the nominal coverage and\or independence hypothesis, the higher the power of the tests. For comparison purposes, we consider the same VaR calculation method as used by Christoffersen and Pelletier (2004), Berkowitz et al.(2009) or Haas (2005)–*i.e.* the Historical Simulation (HS). As in Christoffersen and Pelletier (2004), the rolling window Te is taken to be either 250 or 500. Formally, HS-VaR is defined by the following relation:

$$VaR_{t|t-1}(\alpha) = percentile\left(\{r_j\}_{j=t-T_e}^{t-1}, 100\alpha\right). \tag{24}$$

HS easily generates VaR violations. In Figure 1, observed simulated returns r_t for a given simulation and VaR-HS are plotted. Violation clusters are evident, whether for 1% VaR or for 5% VaR .

Insert Figure 1

For each simulation, the zero-one hit sequence I_t is calculated by comparing the *ex post* returns r_t to the *ex ante* forecast $VaR_{t|t-1}(\alpha)$, and the sequence of durations Di (or Y_i) between violations is calculated from the hit sequence. From this duration sequence, the test-statistics $J_{CC}(p)$ for $p = 1, ..., 5$ and the Berkowitz et al. (2009) (LR_{CC}) and Christoffersen (1998) (LR_{CC}^{Markov}) tests are implemented. The empirical power of the tests is then deduced from rejection frequencies based on 10,000 replications. However, as previously mentioned, the use of asymptotic critical values (based on a χ^2 distribution) induces important size distortions, even for a relatively large sample. Thus, given the scarcity of violations (particularly for a 1% coverage rate), it is particularly important to control for the size of the backtesting tests. As usual in this literature, the Monte Carlo technique proposed by Dufour (2006) is implemented (see Appendix B).

Insert Table 3

Table 3 reports the rejection frequencies (the nominal size is fixed at 5%) of the tests for 1% and 5% VaR. We report the power of our test for various values of the number of moment conditions, p. We observe that except for $T = 250$, power is increasing with p. This result illustrates that the Bontemps's framework is not robust to any specification under the alternative hypothesis if one uses only a small number of polynomials. Each test based on a specific polynomial is robust against the alternatives for which the corresponding moment has some expectation different from zero. Therefore, the tests will be robust only if we consider a sufficient number of polynomials. In our simulations, the power is optimized by considering three moment conditions in the case of the 5% VaR, whereas five Meixner polynomials are required for a 1% VaR. To illustrate this point, the empirical power is plotted for different number of moment conditions in Figure 2.

Insert Figure 2

In all cases, the power of the GMM-based backtesting test J_{CC} is greater than that of the Berkowitz et al. (2009) test regardless of the considered sample size. The gain provided by our test is especially noticeable for the more interesting cases from a practical point of view; that is, those with small sample size and $\alpha = 1\%$. For $T = 250$, the power of our test is twice the power of the standard LR test. Besides, the power of the GMM duration-based test is always higher than that of the Markov chain LR test, which is one of the most often used backtests. This property constitutes a key point in promoting the empirical popularity of duration-based backtesting tests. Comparison with the test for UC is not possible because traditional duration-based tests do not provide such information.[8] Nevertheless, its power is always relatively high and in all cases is larger than 17%.

These simulation experiments confirm that the GMM-based duration test has improved power compared to traditional duration-based tests. IT also

[8] As already noted, traditional duration-based tests do not provide a separate test for UC.

provides a separate test for CC, UC, and IND hypotheses. Our initial objectives are thus fulfilled.

4.3 Discrete versus continuous distribution

When one compares our GMM duration-based test with other duration-based backtesting procedures, the differences observed in finite sample properties may have different origins: (i) the use of a discrete distribution instead of a continuous approximation (Haas, 2005), (ii) and the use of an M-test approach instead of the traditional LR one. In order to assess the relative importance of these two channels, we now propose to consider an extension of our GMM testing procedure based on the exponential distribution. Through a comparison of this GMM test to the GMM test proposed in section 3, it will be possible to evaluate the influence on the duration process of the choice of a discrete distribution versus a continuous approximation.

As in Christoffersen and Pelletier (2004), we assume that under the null hypothesis of CC, the duration d_i between two violations has an exponential distribution with a rate parameter equal to α and a pdf defined by (8). As previously mentioned, the Pearson family of distributions, including the exponential distribution, can be associated with some particular orthonormal polynomials with an expectation equal to zero (Bontemps and Meddahi, 2006). For the exponential distribution, these are known as Laguerre polynomials.

Definition 3 *The orthonormal polynomials associated with an exponential distribution with a rate parameter β are defined by the following recursive relationship $\forall d \in \mathbb{N}^*$:*

$$L_{k+1}(d;\beta) = \frac{1}{k+1}\left[(2k+1-\beta d)\ \ L_k(d;\beta) - k\ L_{k-1}(d;\beta)\right], \qquad (25)$$

for any order $j \in \mathbb{N}$, with $L_0(d;\beta) = 0$ and $M_1(d;\beta) = 1 - \beta d$. If the true distribution of D is an exponential distribution with a rate parameter β, then it follows that

$$\mathbb{E}\left[L_j(d;\beta)\right] = 0 \qquad \forall j \in \mathbb{N}^*, \forall d \in \mathbb{N}^*. \qquad (26)$$

Hence, in this context, the null of CC can be expressed as follows:

$$H_{0,CC}: \ \mathbb{E}\left[L_j\left(d_i;\alpha\right)\right]=0, \quad j=\{1,..,p\}\,, \tag{27}$$

where p denotes the number of moment conditions, whereas the UC hypothesis corresponds to the nullity of the expectation of the first Laguerre polynomial.

$$H_{0,UC}: \ \mathbb{E}\left[L_1\left(d_i;\alpha\right)\right]=0. \tag{28}$$

It is then possible to define appropriate J-test statistics, as in section 3. Let us denote $J_{CC}^{\exp}(p)$ as the CC statistic test associated with the p first orthonormal Laguerre polynomials and $J_{UC}^{\exp}$ as the UC statistic equal to $J_{CC}^{\exp}(1)$.

Insert Table 4

In Table 4, we report a comparison of the 5% power of both statistics $J_{CC}^{\exp}(p)$ and $J_{CC}(p)$. The experiment design is exactly the same as that described in section 4.2. We can observe that whatever the sample size and whatever the VaR coverage rate (1% or 5%), the finite sample power of the J_{CC} tests is very close to that of its continuous analogue $J_{CC}^{\exp}$. The only exception is the case in which T is equal to 250 and the coverage rate is equal to 1%. In this case, the power of the $J_{CC}^{\exp}$ test is even greater than the power of the J_{CC} tests. This result seems to prove that, at least in our experiment, the gain in power (compared to standard LR backtesting tests) is mainly due to the use of a GMM approach. Unlike for the results achieved by Haas (2005), the use of a discrete or continuous distribution does not appear to change the finite sample properties of our test.

5 Parameter uncertainty and robust inference

This last section is devoted to a discussion[9] of the effect of parameter uncertainty on statistical inference through our three tests statistics $J_{CC}(p)$, $J_{UC}(p)$ and $J_{IND}\left(p;\widehat{\beta}\right)$. Indeed, as shown by Escanciano and Olmo (2009), the use

[9] We are grateful to an anonymous referee and to the editor for this suggestion.

of standard backtesting procedures to assess VaR models on an out-of-sample basis can be misleading because these procedures do not consider the impact of parameter uncertainty or estimation risk. They denote $q_{t|t-1}(\alpha)$ the true conditional $\alpha\%$-VaR of r_t, *i.e.*.$\Pr\left(r_t \leq q_{t|t-1}(\alpha)\right) = \alpha$, $\forall t \in \mathbb{Z}$, and consider a given VaR model $\mathcal{M} = \left\{VaR_{t|t-1}(\alpha;\theta) : \theta \in \Theta \subset \mathbb{R}^p, \forall t \in \mathbb{Z}\right\}$, where θ is a vector of parameters that can be either finite-dimensional for parametric VaR models or infinite-dimensional for semi-parametric or non-parametric VaR models. Escanciano and Olmo (2009) note that inference within the VaR model $\mathcal{M}$ is heavily based on the hypothesis that $q_{t|t-1}(\alpha) \in \mathcal{M}$, *i.e.* if there exists some $\theta^0 \in \Theta$ such that $VaR_{t|t-1}(\alpha;\theta^0) = q_{t|t-1}(\alpha)$ *almost surely (a.s.).* Therefore, the candidate model $\mathcal{M}$ is correctly specified if and only if

$$\Pr\left(r_t \leq VaR_{t|t-1}(\alpha;\theta^0)\right) = \alpha,\ a.s. \text{ for some } \theta^0 \in \Theta,\ \forall t \in \mathbb{Z}, \text{ or} \tag{29}$$

$$E\left[I_t\left(\alpha;\theta^0\right)\right] = \alpha \quad a.s. \text{ for some } \theta^0 \in \Theta,\ \forall t \in \mathbb{Z}. \tag{30}$$

In this context, the correct CC hypothesis defined through equation (5) must be expressed as

$$\left\{I_t\left(\alpha;\theta^0\right)\right\} \text{ are } i.i.d. \text{ Bernoulli } r.v. \text{ for some } \theta^0 \in \Theta,\ \forall t \in \mathbb{Z}. \tag{31}$$

and the duration between two consecutive violations $d_i\left(\theta^0\right) = t_i\left(\theta^0\right) - t_{i-1}\left(\theta^0\right)$ follows a geometric distribution with parameter α, *i.e.*,

$$f\left(d_i\left(\theta^0\right)\right) = \alpha\left(1-\alpha\right)^{d_i\left(\theta^0\right)-1} \text{ for some } \theta^0 \in \Theta,\ d_i\left(\theta^0\right) \in \mathbb{N}^*. \tag{32}$$

The main point of this discussion is that the asymptotic distributions (see, equations 18,19 and 21) of our three test statistics $J_{CC}(p)$, $J_{UC}(p)$ and $J_{IND}\left(p;\widehat{\beta}\right)$ should be written strictly as follows:

$$J_{CC}\left(p;\theta^0\right) = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M\left(d_i\left(\theta^0\right);\alpha\right)\right)^{\intercal}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M\left(d_i\left(\theta^0\right);\alpha\right)\right) \underset{N\to\infty}{\overset{d}{\longrightarrow}} \chi^2(p), \tag{33}$$

$$J_{UC}\left(p;\theta^0\right) = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M_1\left(d_i\left(\theta^0\right);\alpha\right)\right)^2 \underset{N\to\infty}{\overset{d}{\longrightarrow}} \chi^2(1), \text{ and} \tag{34}$$

$$J_{IND}\left(p;\widehat{\beta},\theta^0\right) = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M\left(d_i\left(\theta^0\right);\widehat{\beta}\right)\right)^{\intercal}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M\left(d_i\left(\theta^0\right);\widehat{\beta}\right)\right) \xrightarrow[N\to\infty]{d} \chi^2\left(p-1\right). \quad (35)$$

Nevertheless, the above test statistics are not operational because θ^0 is not known. In practice, one must replace θ^0 with a consistent estimator using available data. Formally, the sample with size T is divided into an in-sample portion of size R and an out-of-sample portion of size P, with $T = R+P$. The P VaR forecasts are produced using a fixed, rolling or recursive forecasting scheme. For example, the fixed forecasting scheme involves estimating the parameters θ only once on the first R observations and using these estimates to produce all of the VaR forecasts for the out-of-sample period. Denote $VaR_{t|t-1}(\alpha;\widehat{\theta}_R)$, $t = R+1,...,T$, the P conditional VaR forecasts, $I_t\left(\alpha;\widehat{\theta}_R\right)$, $t = R+1,...,T$, the sequence of the hit variable, and $d_i(\widehat{\theta}_R)$, $i = 1,...,N$, the durations between violations. Then, the three test statistics can be computed by replacing θ^0 with $\widehat{\theta}_R$ in equations (33), (34) and (35).

5.1 A subsampling approach

Uncertainty about the value of $\widehat{\theta}_R$ could affect the asymptotic distributions of the tests statistics. In the framework of hypothesis testing, this problem is referred to as parameter uncertainty or estimation risk. As previously mentioned, in the GMM framework, the problem of parameter uncertainty can be handled by finding moments that are robust against estimation risk[10]. However, the above results are valid only under the assumption that the moments are smooth in the parameters. Unfortunately, under the present setup, this requirement is violated because the duration variable depends on the hit variable, which is not differentiable with respect to the vector of parameters θ.

A possible solution for dealing with the issue of parameter uncertainty is the use of robust backtesting procedures that entail (block) bootstrap or sub-

[10] These moments can be either residuals of the projection of the moments on the score function of the stochastic process that defines the moments, or they can be obtained through a suitable transformation of the original moments that guarantees the orthogonality of the score function (see Bontemps and Meddah, 2006).

sampling approximations of the true test statistic distributions. Escanciano and Olmo (2009) advocate the use of subsampling approximation in the context of VaR backtesting, arguing that it is a general resampling method that is consistent under a minimal set of assumptions, including cases where the (block) bootstrap is inconsistent. Therefore, following Escanciano and Olmo (2008), we deal with the issue of parameter uncertainty by using subsampling to approximate the true distribution of the test statistic $J_{CC}(p, \widehat{\theta}_R)$. The basic idea, described in detail in Politis, Romano and Wolf (2001), is to approximate the sampling distribution of a statistic based on the values of the statistic computed over smaller subsets of the data.

To introduce the notation, let $(r_k, ..., r_{k+b-1})$ be any of the subsamples of size b from the returns $\{r_t\}_{t=1}^{T}$, with $k = 1, ..., T - b + 1$. Divide each subsample into an *in-sample* portion of size R_b and an *out-of-sample* portion of size P_b according to the ratio $\pi = P_b/R_b = P/R$. Let us denote $G_{T,R}(w)$ the *c.d.f* of the test statistic $J_{CC}\left(p, \widehat{\theta}_R\right)$. Then, the sampling distribution of $J_{CC}\left(p; \widehat{\theta}_R\right)$ is approximated by

$$\widehat{G}_{T,R_b}(w) = \frac{1}{T-b+1} \sum_{k=1}^{T-b+1} \mathbf{1}\left(J_{CC}^{(k)}\left(p, \widehat{\theta}_{R_b}\right) \leq w\right) \quad \forall w \in \mathbb{R}^+, \tag{36}$$

For each subsample, the statistic $J_{CC}^{(k)}\left(p, \widehat{\theta}_{R_b}\right)$ is computed by first estimating the vector of parameters θ using $(X_k, ..., X_{k+R_b-1})$and using the estimates $\widehat{\theta}_{R_b}$ to produce the P_b VaR forecasts over the period $t = k + R_b, ..., k + b - 1$. Given the estimated sampling distribution, the critical value for the correct CC is obtained as the $1 - \eta$ quantile of $\widehat{G}_{T,R_b}(w)$ defined as

$$g_{T,R_b}(1-\eta) = \inf\left\{w : \widehat{G}_{T,R_b}(w) \geq 1-\eta\right\}. \tag{37}$$

As a result, one rejects the null hypothesis at the nominal level η, if and only if $J_{CC}\left(p, \widehat{\theta}_R\right) > g_{T,R_b}(1-\eta)$.

Proposition 4 *Assume that* $\{d_{i,\alpha}(\theta) : 1 \leq i,\ \theta \in \Theta\}$ *is stationary and ergodic. Assume also that* $b/T \to 0$ *and* $b \to \infty$ *as* $T \to \infty$*. Under the assumption that the mixing sequence corresponding to* $\{r_t\}$ *converges to 0, then* $g_{T,b}(1-\eta) \to$

$g(1-\eta)$ in probability and $\Pr\left\{J_{CC}\left(p,\widehat{\theta}_R\right) > g_{T,R_b}(1-\eta)\right\} \to \delta$ *as* $T \to \infty$, *where* $g(1-\eta)$ *is the* $(1-\eta)^{th}$ *quantile of* $G_{T,R_b}(w)$.

The proof follows from theorem 5.1. in Politis, Romano and Wolf (2001). The stationarity and ergodicity condition of $d_{i,\alpha}(\theta)$ is required because it ensures (see proposition 2) continuity of the distribution function of $J_{CC}\left(p,\theta^0\right)$, which is in occurrence a chi-square.

5.2 Finite sample properties

For some popular VaR backtests (Kupiec, 1995; Christoffersen, 1998), Escanciano and Olmo (2008) have used Monte Carlo experiments to show the importance of correcting for parameter uncertainty using the above subsampling approximation. We provide similar evidence for our test statistic $J_{CC}\left(p;\widehat{\theta}_R\right)$ using a VaR model in which the true dynamics of r_t are known.

More precisely, let us consider a t-GARCH(1,1) data-generating process for the returns r_t. The parameters of the GARCH(1,1) process are chosen[11] to reflect standard values found in real-time series of financial returns. Then, the VaR model is defined for a given coverage rate $\alpha \in \{1\%, 5\%\}$ by

$$\mathcal{M} = \left\{VaR_{t|t-1}(\alpha;\theta) : \theta \in \Theta \subset \mathbb{R}^4, \forall t \in \mathbb{Z}\right\}, \text{ with} \tag{38}$$

$$VaR_{t|t-1}(\alpha;\theta) = F^{-1}(\alpha)\,\sigma_t\sqrt{\frac{v-2}{v}}, \tag{39}$$

$$\sigma_t^2 = \omega + \gamma\sqrt{\frac{v-2}{v}}z_{t-1} + \beta\sigma_{t-1}^2, \tag{40}$$

where $F(.)$ is the *c.d.f.* of a $t(v)$, and $\theta = (\omega,\gamma,\beta,v)$ the vector of parameters.

The simulation exercise consists of generating returns data from the GARCH process described above and a sample size T equal to $P+R$; in a second stage, the parameters of the model are estimated by QMLE using the first R observations, and the corresponding VaR model is computed for the remaining P out-of-sample observations. For ease of computation, we have implemented a fixed

[11] We consider $\omega = 7.9778e^{-7}$, $\gamma = 0.0896$, $\beta = 0.9098$ and $v = 6.12$. These values correspond to the values estimated over a sample of SP500 daily returns from 02/01/1970 and 05/05/2006.

forecasting scheme for estimating the GARCH parameters, and where the out-of-sample size P equal to 1000 is considerably greater than the in-sample size $R = 500$. The choice of this sample size embodies a mix of absence of estimation risk effects ($P/R < 1$) and meaningful results derived from the subsampling and asymptotic tests (P sufficiently large). Let us denote $VaR_{t|t-1}\left(\alpha; \widehat{\theta}_R\right)$, for $t = R+1, ..., T$, the out-of-sample VaR forecasts and $I_{t,\alpha}(\widehat{\theta}_R)$ the corresponding VaR violations observed *ex-post*. Given the violations sequence, we compute the durations variable $d_{i,\alpha}(\widehat{\theta}_R)$, for $i = 1, ..., N$ and our three test statistics.

Insert Table 5

Table 5 also reports the (uncorrected) empirical sizes, for a nominal size η equal to 5%, of the test statistic $J_{CC}(p) \equiv J_{CC}(p, \widehat{\theta}_R)$, with $p = 2, 3, 5$. This size corresponds to the rejection frequencies over $1,000$ simulations for each test using the asymptotic critical values. For the purpose of comparison, we also display results for the duration-based CC test statistic of Christoffersen and Pelletier (2004) and the Markov CC test in Christoffersen (1998). We verify that the estimation risk, induced by $\widehat{\theta}_R$, creates a size distortion for all three tests, even if J_{CC} tests seem to be less oversized than other LR tests. This distortion is relatively important in the case of $\alpha = 5\%$, but less important for $\alpha = 1\%$, especially for our J_{CC} tests.

The second part of Table 5 presents the rejection frequencies over $1,000$ simulations for each test using the subsampling critical values. Following Escanciano and Olmo (2008), we have used a subsample size $b = \left[KP^{2/5}\right]$, that has been implemented for $K = \{65, 70, 75, 80\}$ and $P = 1,000$. For each value of K, we report the number N_b of subsamples and the size P_b of the out-of-sample portion of the subsamples. We observe that for all tests, the Monte Carlo correction reduces the size distortion, especially for $\alpha = 5\%$. It thus appears the subsampling methods offer a reliable approximation of the asymptotic critical values.

6 Empirical Application

To illustrate these new tests, an empirical application is performed, considering three sequences of 5%VaR forecasts on the daily returns of the Nasdaq index. These sequences correspond to three different VaR forecasting methods traditionally used in the literature: a pure parametric method (GARCH model under Student distribution), a non-parametric method (Historical Simulation) and a semi-parametric method based on a quantile regression (CAViaR, Engle and Manganelli, 2004). Each sequence contains 250 successive one-period-ahead forecasts for the period June 22, 2005 to June 20, 2006. The parameters of the GARCH and CAViaR models are estimated according to a rolling windows method with a length [12] fixed to 250 observations.

Insert Table 6

The results obtained using the GMM duration-based tests are reported in Table 6. For each VaR method, we report the UC, CC and IND statistics. For the two last tests, the number of moments p is fixed at 2, 4 and 6. For the sake of comparison, LR_{CC} statistics (Christoffersen and Pelletier, 2004; Berkowitz et al. 2005) are also reported. For all tests, two p-values are reported: the first one corresponds to the size-corrected p-value based on the Dufour's MC procedure (see Appendix B), while the second one corresponds to the p-value based on the subsampling approximation of the true test statistic distributions obtained with $K = 5, p = 1, P = R = 250$ and $P_b = R_b = 98$ (see Section 5). Several comments can be made about these results. First, at a 10% significance level, our unconditional coverage test statistic J_{UC} leads to an unambiguous rejection of the validity of the CAViaR-based VaR (even if the rejection is stronger when we ignore the potential estimation risk). As expected, this result is due to the too-low violation rate associated with this method. Of course, the value of J_{UC} is identical for HS and GARCH because these two methods lead to the same number of hits even if these violations do not occur during the same periods.

[12] The total sample runs from June 20, 2004 to June 20, 2006 (500 observations). The length of the rolling estimation window of the HS is also fixed to 250 observations.

Second, we observe that our GMM independence test (J_{IND}) is able to reject (except in the case $p = 2$) the null for HS-VaR. In contrast, the LR_{IND} test does not reject the null of independence for any of the three VaRs, regardless of the distribution (MC or sub-sampling). Third, the GMM conditional coverage test (J_{CC}) rejects the validity of CAViaR and HS VaR forecasts, unlike standard LR tests. Finally, the GARCH-t(d) emerges as the best method by which to forecast risk: the UC, IND and CC are not rejected. This application shows the ability of our GMM test to discriminate between various VaR models, especially when one takes into account the potential estimation risks.

7 Conclusion

This paper develops a new duration-based backtesting procedure for VaR forecasts. The underlying idea is that if the one-period-ahead VaR is correctly specified, then every period, the duration until the next violation should be distributed according to a geometric distribution with a success probability equal to the VaR coverage rate. On this basis, we adapt the GMM framework proposed by Bontemps (2006) in order to test for this distributional assumption that corresponds to the null of VaR forecast validity. The test statistic is essentially a simple J-statistic based on particular moments defined by the orthonormal polynomials associated with the geometric distribution. This new approach tackles most of the drawbacks usually associated with the duration-based model. First, its implementation is extremely easy. Second, it allows for a separate test for the unconditional coverage, independence and conditional coverage hypotheses (Christoffersen, 1998). Third, Monte-Carlo simulations show that, for realistic sample sizes, our GMM test outperforms traditional duration-based tests. Finally, we pay particular attention to the consequences of the estimation risk for the duration-based backtesting tests and propose a sub-sampling approach for robust inference derived from Escanciano and Olmo (2009).

Our empirical application of the method to Nasdaq returns confirms that using the GMM test leads to main in the *ex-post* evaluation of risk by regulation authorities. Our hope is that this paper will encourage regulatory authorities to use duration-based tests to assess the risk taken by financial institutions. There is no doubt that a more adequate evaluation of risk would decrease the probability of banking crises and systemic banking fragility.

Appendix A: Proof of parameter uncertainty robustness with respect to β

Under the IND hypothesis, the sequence of durations $d = \left\{ d_i \left(\widehat{\theta}_R \right) \right\}_{i=1}^{N}$ is *i.i.d.* geometric with parameter β. The *p.d.f.* of d is

$$f\left(d;\beta\right) = \left(1-\beta\right)^{d-1}\beta \qquad d \in \mathbb{N}^*. \tag{41}$$

The score function is defined as

$$\frac{\partial \ln f\left(d;\beta\right)}{\partial \beta} = \frac{1-\beta d}{\beta\left(1-\beta\right)}. \tag{42}$$

It is straightforward to prove that this score is proportional to the first Meixner polynomial because

$$M_1\left(d;\beta\right) = \frac{1-\beta d}{\sqrt{1-\beta}}, \text{ and} \tag{43}$$

$$\frac{\partial \ln f\left(d;\beta\right)}{\partial \beta} = \frac{M_1\left(d;\beta\right)}{\beta\sqrt{1-\beta}}. \tag{44}$$

Consequently, the orthonormal polynomials with degrees greater than or equal to 2 are also proportional to the score function, and the moments $M_j\left(d;\beta\right)$, $j = 1,...,p$ are robust against estimation risk with respect to β. Indeed, robust moments defined by the projection of the moments on the score function correspond exactly to the initial moments.

Appendix B: Dufour (2006) Monte-Carlo Method

To implement MC tests, first generate M independent realizations of the test statistic–say S_i, $i = 1, \ldots, M$–under the null hypothesis. Denote by S_0 the value of the test statistic obtained for the original sample. As shown by Dufour (2006) in a general case, the MC critical region is obtained as $\hat{p}_M(S_0) \leq \eta$ with $1-\eta$ the confidence level and $\hat{p}_M(S_0)$ defined as

$$\hat{p}_M(S_0) = \frac{M\ \hat{G}_M(S_0)+1}{M+1}, \tag{45}$$

where

$$\widehat{G}_M(S_0) = \frac{1}{M}\sum_{i=1}^{M}\mathbb{I}(S_i \geq S_0), \tag{46}$$

when the ties have zero probability, *i.e.* $\Pr\left(S_i = S_j\right) \neq 0$, and otherwise,

$$\widehat{G}_M(S_0) = 1 - \frac{1}{M}\sum_{i=1}^{M}\mathbb{I}\left(S_i \leq S_0\right) + \frac{1}{M}\sum_{i=1}^{M}\mathbb{I}\left(S_i = S_0\right)\times\mathbb{I}\left(U_i \geq U_0\right). \tag{47}$$

Variables U_0 and U_i are uniform draws from the interval $[0,1]$ and $\mathbb{I}(.)$ is the indicator function. As an example, for the MC test procedure applied to the test statistic $S_0 = J_{CC}\left(p;\widehat{\theta}_R\right)$, we just need to simulate under H_0 M independent realizations of the test statistic (*i.e.*, using durations constructed from independent Bernoulli hit sequences with parameter α), and then apply formulas (45-47) to make inferences at the confidence level $1-\eta$. Throughout the paper, we set M at $9,999$.

References

Berkowitz, J. , Christoffersen, P. F. and Pelletier, D. (2009), "Evaluating Value-at-Risk models with desk-level data", forthcoming in *Management Science.*

Bontemps, C. (2006), "Moment-based tests for discrete distributions", *Working Paper.*

Bontemps, C. and N. Meddahi (2005), "Testing normality: A GMM approach", *Journal of Econometrics*, 124, pp. 149-186.

Bontemps, C. and N. Meddahi (2006), "Testing distributional assumptions: A GMM approach", *Working Paper.*

Campbell, S. D. (2007), "A review of backtesting and backtesting procedures", *Journal of Risk,* 9(2), pp 1-18.

Christoffersen, P. F. (1998), "Evaluating interval forecasts", *International Economic Review,* 39, pp. 841-862.

Christoffersen, P. F. and D. Pelletier (2004), "Backtesting Value-at-Risk: A duration-based approach", *Journal of Financial Econometrics,* 2, 1, pp. 84-108.

Dufour, J.-M. (2006), "Monte Carlo tests with nuisance parameters: a general approach to finite sample inference and nonstandard asymptotics", *Journal of Econometrics*, vol. 127(2), pp. 443-477.

Engle, R. F., and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value-at-Risk by regression quantiles", *Journal of Business and Economic Statistics,* 22, pp. 367-381.

Escanciano, J. C. and Olmo J. (2007), "Estimation risk effects on backtesting for parametric Value-at-Risk models", *Center for Applied Economics and Policy Research,* Working Paper*,* 05.

Escanciano, J. C. and Olmo J. (2008), "Robust Backtesting Tests for Value-at-Risk Models", Working Paper, *Dept. Economics, Indiana University.*

Escanciano, J. C. and Olmo J. (2009), "Backtesting Parametric Value-at-Risk with Estimation risk", *Center for Applied Economics and Policy Research,* Working Paper, 2007-05. *Journal of Business and Economic Statistics*, forthcoming.

Jorion, P. (2007), *Value-at-Risk*, Third edition, McGraw-Hill.

Haas, M. (2005), "Improved duration-based backtesting of Value-at-Risk", *Journal of Risk*, 8(2), pp. 17-36.

Hansen L.P., (1982), "Large sample properties of Generalized Method of Moments estimators", *Econometrica,* 50, pp. 1029–1054.

Kupiec, P.. (1995), "Techniques for verifying the accuracy of risk measurement models", *Journal of Derivatives,* 3, pp. 73-84.

Nakagawa, T., and Osaki, S. (1975), "The discrete Weibull distribution", *IEEE Transactions on Reliability*, R-24, pp 300–301.

Politis, D. N., Romano, J. P. and M. Wolf (1999), *Subsampling, Springer-Verlag*, New-York.

Table 1. Empirical size of 5% asymptotic CC tests

			Backtesting 5% VaR			
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{Markov}
$T = 250$	0.0467	0.0448	0.0369	0.0323	0.0866	0.0901
$T = 500$	0.0448	0.0474	0.0413	0.0342	0.0717	0.0878
$T = 750$	0.0473	0.0481	0.0405	0.0343	0.0725	0.1029
$T = 1000$	0.0533	0.0500	0.0440	0.0373	0.0828	0.1125
$T = 1500$	0.0496	0.0491	0.0439	0.0345	0.0929	0.1132
			Backtesting 5% VaR with censored durations			
Sample size	J_{UC}^{cens}	$J_{CC}^{cens}\,(2)$	$J_{CC}^{cens}\,(3)$	$J_{CC}^{cens}\,(5)$		
$T = 250$	0.0482	0.0435	0.0378	0.0316	—	—
$T = 500$	0.0512	0.0486	0.0419	0.0362	—	—
$T = 750$	0.0430	0.0475	0.0398	0.0329	—	—
$T = 1000$	0.0536	0.0508	0.0445	0.0377	—	—
$T = 1500$	0.0524	0.0503	0.0440	0.0347	—	—
			Backtesting 1% VaR			
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{Markov}
$T = 250$	0.0042	0.0053	0.0140	0.0401	0.0615	0.0281
$T = 500$	0.0098	0.0069	0.0076	0.0077	0.0849	0.0188
$T = 750$	0.0353	0.0263	0.0217	0.0185	0.1000	0.0285
$T = 1000$	0.0417	0.0350	0.0306	0.0282	0.0883	0.0363
$T = 1500$	0.0439	0.0445	0.0368	0.0346	0.0706	0.0439
			Backtesting 1% VaR with censored durations			
Sample size	J_{UC}^{cens}	$J_{CC}^{cens}\,(2)$	$J_{CC}^{cens}\,(3)$	$J_{CC}^{cens}\,(5)$		
$T = 250$	0.0053	0.0060	0.0054	0.0019	—	—
$T = 500$	0.0124	0.0058	0.0065	0.0031	—	—
$T = 750$	0.0188	0.0282	0.0235	0.0217	—	—
$T = 1000$	0.0447	0.0422	0.0379	0.0318	—	—
$T = 1500$	0.0543	0.0465	0.0375	0.0339	—	—

Notes: Under the null, the hit data are i.i.d. from a Bernoulli distribution. The results are based on 10,000 replications. For each sample, we provide the percentage of rejection at a 5% level. Jcc(p) denotes the GMM-based conditional coverage test with p moment conditions. Juc denotes the unconditional coverage test obtained for p=1. LRcc denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2009), and LRmarkov corresponds to the Christoffersen (1998) CC test based on a Markov chain approach.

Table 2. Fraction of samples where tests are feasible

	Size Simulations							
	1% VaR				5% VaR			
Sample Size	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}
$T = 250$	0.715	0.920	0.630	0.920	1.000	1.000	1.000	1.000
$T = 500$	0.959	0.993	0.934	0.993	1.000	1.000	1.000	1.000
$T = 750$	0.994	0.999	0.989	0.999	1.000	1.000	1.000	1.000
$T = 1000$	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	Power Simulations ($Te = 250$)							
	1% VaR				5% VaR			
Sample Size	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}
$T = 250$	0.775	0.901	0.742	0.901	0.992	0.997	0.990	0.997
$T = 500$	0.988	0.997	0.981	0.997	1.000	1.000	1.000	1.000
$T = 750$	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000
$T = 1000$	1.000	1.000	0.742	1.000	1.000	1.000	1.000	1.000
	Power Simulations ($Te = 500$)							
	1% VaR				5% VaR			
Sample Size	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}	J_{CC}	J_{CC}^{cens}	LR_{CC}	LR_{CC}^{M}
$T = 250$	0.628	0.793	0.598	0.793	0.974	0.990	0.967	0.990
$T = 500$	0.907	0.966	0.886	0.966	0.999	1.000	0.999	1.000
$T = 750$	0.990	0.998	0.985	0.998	0.999	1.000	0.999	1.000
$T = 1000$	0.999	0.999	0.998	0.999	1.000	1.000	1.000	1.000

Notes: The results are based on 10,000 replications. For each sample and for each test, we provide the percentage of samples for which the statistic can be computed. Jcc denotes the GMM-based (un)conditional coverage test based only on uncensored data. Jcc(cens) denotes the J-statistic based on both uncensored and censored durations. For the J test, note that the feasible ratios are independent of the number p of moments used. LRcc denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2009), and LR^{M} corresponds to the Christoffersen (1998) CC test based on the Markov chain approach. Te denotes the rolling window length.

Table 3. Power of 5% finite sample tests

Backtesting 1% VaR						
	Length of rolling estimation window $Te = 250$					
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$T = 250$	0.3868	0.4150	0.3669	0.3090	0.1791	0.2788
$T = 500$	0.3592	0.4202	0.4516	0.5024	0.2404	0.2994
$T = 750$	0.3238	0.4239	0.5062	0.5743	0.3341	0.3505
$T = 1000$	0.3276	0.4684	0.5603	0.6365	0.4557	0.3891
$T = 1500$	0.4045	0.5462	0.6632	0.7451	0.6593	0.4968
	Length of rolling estimation window $Te = 500$					
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$T = 250$	0.4034	0.4425	0.4011	0.3539	0.2262	0.3154
$T = 500$	0.3971	0.4557	0.4949	0.5387	0.3240	0.3207
$T = 750$	0.3333	0.4556	0.5197	0.5823	0.4033	0.3205
$T = 1000$	0.3068	0.4971	0.5836	0.6437	0.5248	0.3546
$T = 1500$	0.2969	0.5887	0 7078	0.7579	0.6997	0.4414
Backtesting 5% VaR						
	Length of rolling estimation window $Te = 250$					
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$T = 250$	0.3175	0.4241	0.4577	0.4527	0.2616	0.2561
$T = 500$	0.2300	0.6113	0.6730	0.6600	0.3927	0.2803
$T = 750$	0.1796	0.7515	0.8132	0.7976	0.5266	0.3255
$T = 1000$	0.1811	0.8524	0.8977	0.8873	0.6472	0.3861
$T = 1500$	0.1850	0.9511	0.9737	0.9675	0.8149	0.5099
	Length of rolling estimation window $Te = 500$					
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$T = 250$	0.3271	0.4350	0.4759	0.4688	0.3398	0.3134
$T = 500$	0.3426	0.6877	0.7370	0.7241	0.5006	0.3878
$T = 750$	0.2748	0.8083	0.8584	0.8523	0.6170	0.4280
$T = 1000$	0.2193	0.8889	0.9230	0.9145	0.7178	0.4359
$T = 1500$	0.1711	0.9629	0.9801	0.9773	0.8532	0.5416

Notes: The results are based on 10,000 replications and the MC procedure of Dufour (2006) with ns=9,999. The nominal size is 5%. Jcc(p) denotes the GMM-based CC test with p moment conditions. Juc denotes the UC test obtained for p=1. LRcc denotes the Weibull CC test proposed by Berkowitz et al. (2009), and LRcc(Markov) corresponds to the Christoffersen (1998) CC test.

Table 4. Power of GMM duration-based tests

	Backtesting 1% VaR			
	Discrete distribution (geometric)			
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$
$T = 250$	0.3868	0.4150	0.3669	0.3090
$T = 500$	0.3592	0.4202	0.4516	0.5024
$T = 750$	0.3238	0.4239	0.5062	0.5743
$T = 1000$	0.3276	0.4684	0.5603	0.6365
$T = 1500$	0.4045	0.5462	0.6632	0.7451
	Continuous distribution (exponential)			
Sample size	$J_{UC}^{\exp}$	$J_{CC}^{\exp}\,(2)$	$J_{CC}^{\exp}\,(3)$	$J_{CC}^{\exp}\,(5)$
$T = 250$	0.3868	0.4218	0.4408	0.4576
$T = 500$	0.3591	0.4250	0.4784	0.5251
$T = 750$	0.3239	0.4368	0.5053	0.5670
$T = 1000$	0.3276	0.4660	0.5551	0.6269
$T = 1500$	0.4045	0.5437	0.6593	0.7369
	Backtesting 5% VaR			
	Discrete distribution (geometric)			
Sample size	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$
$T = 250$	0.3175	0.4241	0.4577	0.4527
$T = 500$	0.2300	0.6113	0.6730	0.6600
$T = 750$	0.1796	0.7515	0.8132	0.7976
$T = 1000$	0.1811	0.8524	0.8977	0.8873
$T = 1500$	0.1850	0.9511	0.9737	0.9675
	Continuous distribution (exponential)			
Sample size	$J_{UC}^{\exp}$	$J_{CC}^{\exp}\,(2)$	$J_{CC}^{\exp}\,(3)$	$J_{CC}^{\exp}\,(5)$
$T = 250$	0.3176	0.4175	0.4437	0.4309
$T = 500$	0.2300	0.5951	0.6439	0.6228
$T = 750$	0.1796	0.7311	0.7831	0.7574
$T = 1000$	0.1811	0.8314	0.8715	0.8535
$T = 1500$	0.1850	0.9396	0.9586	0.9498

Notes: The results are based on 10,000 replications and the MC procedure of Dufour (2006) with ns=9,999. The nominal size is 5%. Jcc(p) denotes the GMM-based CC test based on a geometric distribution. $J_{cc}^{\exp}$(p) denotes the GMM-based CC test based on an exponential distribution. In both cases, Juc denotes the UC test (p=1).

Table 5. Empirical size of 5% tests and estimation risk

Backtesting 5% VaR with estimation risk								
			Uncorrected sizes					
Sample size			J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$P = 1000$			0.2860	0.3080	0.2770	0.2480	0.3140	0.3400
			Sub-sampling Corrected sizes					
	N_b	P_b	J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$K = 65$	471	687	0.0920	0.0820	0.0820	0.0840	0.0890	0.0710
$K = 70$	392	740	0.0950	0.0850	0.0860	0.0870	0.1010	0.0850
$K = 75$	313	792	0.1170	0.1020	0.0940	0.0950	0.1160	0.0810
$K = 80$	324	845	0.1060	0.0960	0.0990	0.0890	0.1120	0.0930
Backtesting 1% VaR with estimation risk								
			Uncorrected sizes					
Sample size			J_{UC}	$J_{CC}\,(2)$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$P = 1000$			0.1520	0.1410	0.1150	0.0870	0.1950	0.1430
			Sub-sampling Corrected sizes					
	N_b	P_b	J_{UC}	$J_{CC(2)}$	$J_{CC}\,(3)$	$J_{CC}\,(5)$	LR_{CC}	LR_{CC}^{markov}
$K = 65$	471	687	0.0830	0.1100	0.1210	0.1380	0.0850	0.0640
$K = 70$	392	740	0.0850	0.1000	0.1170	0.1230	0.0930	0.0900
$K = 75$	313	792	0.0980	0.1090	0.1210	0.1310	0.1010	0.0880
$K = 80$	324	845	0.0930	0.1010	0.1070	0.1170	0.1170	0.1110

Notes: For each replication, the returns are simulated according to a t-GARCH. The t-GARVH is then estimated over R=500 periods, and the VaR forecasts are produced for P periods. Given these forecasts, the hits and durations are computed. The estimation risk affects the uncorrected sizes (nominal size is fixed at 5%) of the various backtesting tests. Juc denotes the unconditional coverage test obtained for p=1. LRcc denotes the Weibull conditional coverage test proposed by Berkowitz et al. (2009), and LRcc(Markov) corresponds to the Christoffersen (1998) CC test based on a Markov chain approach. Finally, the corrected rejection rates are reported for various values of K, Nb and Pb. In each case, the critical value is obtained through the ub-sampling procedure described in section 5. The results are based on 1,000 replications.

Table 6. Backtesting tests of 5% VaR forecasts for Nasdaq index

		VaR forecasting methods		
Backtesting Tests	Statistic	GARCH-t(d)	HS	CAViaR
Unconditional Coverage	Hits Freq.	0.036	0.036	0.028
	J_{UC}	1.1977 (0.261) (0.446)	1.1977 (0.261) (0.140)	4.0662 (0.036) (0.097)
Independence Tests	J_{IND} (2)	0.2489 (0.646) (0.713)	0.1866 (0.719) (0.839)	0.0801 (0.867) (0.924)
	J_{IND} (4)	0.3839 (0.779) (0.838)	4.6521 (0.036) (0.070)	0.5323 (0.663) (0.787)
	J_{IND} (6)	0.3856 (0.907) (0.929)	7.8571 (0.016) (0.042)	2.1271 (0.299) (0.445)
	LR_{IND}	0.5121 (0.532) (0.710)	1.7725 (0.236) (0.529)	0.0584 (0.823) (0.819)
Conditional Coverage	J_{CC} (2)	1.5228 (0.315) (0.477)	2.7080 (0.157) (0.067)	4.4663 (0.071) (0.023)
	J_{CC} (4)	1.5456 (0.522) (0.610)	11.147 (0.025) ($<$0.001)	9.8744 (0.030) (0.078)
	J_{CC} (6)	1.5867 (0.646) (0.752)	11.891 (0.030) (0.014)	12.050 (0.031) (0.075)
	LR_{CC}	2.3356 (0.364) (0.550)	3.5960 (0.213) (0.449)	4.1989 (0.168) (0.314)

Notes: The hit empirical frequency is the ratio of VaR violations to the sample size (T=250) observed for the Nasdaq between June 22, 2005 and June 20, 2006. Three methods of VaR forecasting are used: the GARCH with Student conditional, distribution, historical simulation (HS) and a CAViaR (Engle and Manganelli, 2004. For the VaR method, Juc denotes the unconditional coverage test statistic obtained for p=1. Jind(p), and Jcc(p) denotes the GMM-based independence and conditional coverage tests base on p moments conditions. The number of moments is fixed at 2, 4 or 6. LRind and LRcc respectively denote the Weibull independence and conditional coverage tests proposed by Christoffersen and Pelletier (2004) and Berkowitz et al. (2005). For all of these tests, the two numbers in parentheses respectively denote the p-values corrected by Dufour's Monte Carlo procedure (Dufour, 2006) and the p-value obtained using the sub-sampling approach (with K=15, Rb=Pb=68 and 365 sub-samples)

Figure 1: GARCH-t(v) Simulated Returns with 1% and 5% VaR from HS ($Te = 250$).

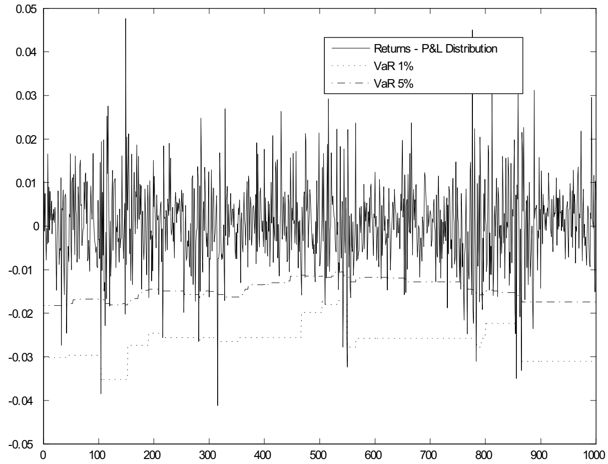


Figure 2: Empirical Power: Sensitivity Analysis to the choice of p

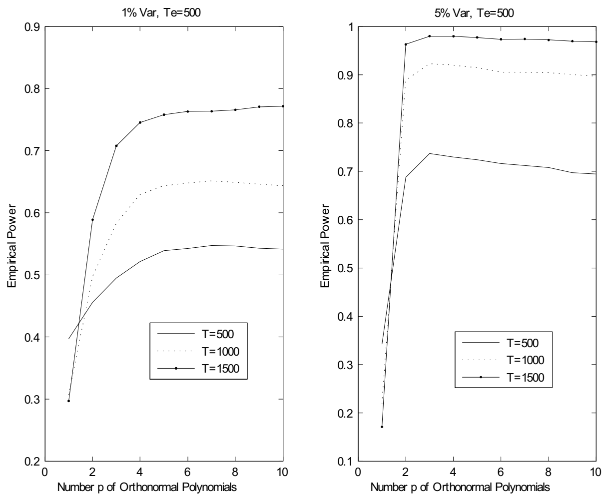


Figure 3: Historical Returns and 5% VaR Forecasts. Nasdaq (June 2005- June 2006)