ROBUST BACKTESTING TESTS FOR VALUE-AT-RISK MODELS

J. Carlos Escanciano*
Indiana University, Bloomington, IN, USA

Jose Olmo
City University, London, UK

November 2008

Abstract

Backtesting methods are statistical tests designed to uncover excessive risk-taking from financial institutions. We show in this paper that these methods are subject to the presence of model risk produced by a wrong specification of the conditional VaR model, and derive its effect on the asymptotic distribution of the relevant out-of-sample tests. We also show that in the absence of estimation risk, the unconditional backtest is affected by model misspecification but the independence test is not. Our solution for the general case consists on proposing robust subsampling techniques to approximate the true sampling distribution of these tests. We carry out a Monte Carlo study to see the importance of these effects in finite samples for location-scale models that are wrongly specified but correct on "average ". An application to Dow-Jones Index shows the impact of correcting for model risk on backtesting procedures for different dynamic VaR models measuring risk exposure.

Keywords and Phrases: Backtesting; Basel Accord; Conditional Quantile; Forecast evaluation; Model Risk; Risk management; Value at Risk.

*Corresponding Address: Dept. Economics, Indiana University, Bloomington, IN, USA. J. Carlos Escanciano, E-mail:jescanci@indiana.edu Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

1 Introduction

One of the implications of the creation of Basel Committee on Banking Supervision was the implementation of Value-at-Risk (VaR) as the standard tool for measuring market risk and of out-of-sample backtesting for banking risk monitoring. As a result of this agreement financial institutions have to report their VaR, defined as a conditional quantile with coverage probability α of the distribution of returns on their trading portfolio. To test the performance of this and alternative VaR measures the Basel Accord (1996) set a statistical device denoted backtesting that consisted of out-of-sample comparisons between the actual trading results with internally model-generated risk measures. The magnitude and sign of the difference between the model-generated measure and actual returns indicate whether the VaR model reported by an institution is correct for forecasting the underlying market risk and if this is not so, whether the departures are due to over- or under-risk exposure of the institution. The implications of over- or under- risk exposure being diametrically different: either extra penalties on the level of capital requirements or bad management of the outstanding equity by the institution.

These backtesting techniques are usually interpreted as statistical parametric tests for the coverage probability α defining the conditional quantile VaR measure. More formally, denote the real-valued time series of portfolio returns or Profit and Losses (P&L) account by Y_t, and assume that at time $t-1$ the agent's information set is given by a d_w-dimensional random vector W_{t-1}, which may contain past values of Y_t and other relevant economic and financial variables, *i.e.*, $W_{t-1} = \{Y_s, Z_s'\}_{s=t-h}^{t-1}$, $h < \infty$. Henceforth, A' denotes the transpose matrix of A. Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $\{Y_s, Z_s'\}_{s=-\infty}^{t-1}$. Assuming that the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is continuous, we say that $m_\alpha(W_{t-1}, \theta)$ is a correctly specified α-th conditional VaR of Y_t given W_{t-1} if and only if

$$P(Y_t \leq m_\alpha(W_{t-1}, \theta_0) \mid \mathcal{F}_{t-1}) = \alpha, \text{ almost surely (a.s.)}, \alpha \in (0,1), \ \forall t \in \mathbb{Z}, \tag{1}$$

for some unknown parameter θ_0 that belongs to Θ, with Θ a compact set in an Euclidean space $\mathbb{R}^p$. Inferences for models defined through the conditional moment restriction in (1) can be found extensively in the literature; see *e.g.* Engle and Manganelli (2004), Koenker and Xiao (2006) and Gourieroux and Jasiak (2006), among many others.

In particular, (1) implies the so-called joint hypothesis (cf. Christoffersen (1998)),

$$\{I_{t,\alpha}(\theta_0)\} \text{ are } iid \text{ Ber}(\alpha) \text{ random variables } (r.v.) \text{ for some } \theta_0 \in \Theta, \qquad (2)$$

where Ber(α) stands for a Bernoulli *r.v.* with parameter α and $I_{t,\alpha}(\theta) := 1(Y_t \leq m_\alpha(W_{t-1}, \theta))$ are the so called *hits* or *exceedances* associated to the model $m_\alpha(W_{t-1}, \theta_0)$. Henceforth, $1(A)$ is the indicator function, i.e. $1(A) = 1$ if the event A occurs and 0 otherwise. In the VaR literature, the satisfaction of condition (2) has been taken as the "loss function" for the out-of-sample evaluation of VaR forecasts, leading to the so-called *unconditional backtesting* (*i.e.* tests for $E[I_{t,\alpha}(\theta_0)] = \alpha$) and *tests of independence* (*i.e.* tests for $\{I_{t,\alpha}(\theta_0)\}$ being *iid*).

Existing Backtesting methodologies seem to assume that (1) and (2) are equivalent. Note, however, that, as has been pointed out in the literature, see examples in Engle and Manganelli (2004) or Kuester, Mittnik and Paolella (2006), condition (2) is a necessary but not sufficient condition of (1). The gap between these two conditions comprises a large class of misspecified models that lead however to *iid* exceedances with correct unconditional coverage probability α. The effect of this misspecification in the asymptotic distribution of the backtesting tests is what we will denote in this paper as *model risk*. Whereas other authors have studied the effect of estimation risk on backtesting methods, see Figlewski (2003), Christoffersen and Gonçalves (2005) or Escanciano and Olmo (2008, EO henceforth) the effect of model risk has not been studied yet. Our aim in this paper is therefore to propose, for the first time in the literature, robust unconditional, independence, and joint tests for testing (2) under possible misspecification of the conditional VaR model.

We show in this paper that model risk has a direct effect on the unconditional backtest, but just and indirect effect on the test of independence (through the estimation risk). This indirect effect is also present in the unconditional backtest. Some simulations in this paper, and a comparison with EO, suggest that model risk is higher than estimation risk for conventional choices of the ratio of out-of-sample size relative to in-sample size. We stress that estimation risk can be annihilated in the unconditional and independence tests by choosing a large in-sample size relative to the out of sample size. Model risk is, however, ubiquitous in the unconditional test, unless the independence hypothesis holds.

Our paper draws on the literature on out-of-sample predictive ability in *e.g.* West (1996), and more closely in McCracken (2000), adapting their general methods to the specific case of backtesting tests. These authors also acknowledge the presence of uncertainty due to parameter estimation and model risk in out-of-sample forecast inference, but they do not consider the problem we deal with here. In a similar spirit to our out-of-sample testing framework Corradi and Swanson (2007) propose a block-bootstrap procedure robust to estimation and model risk to approximate the asymptotic critical values of out-of-sample forecast tests. Unfortunately, their results are not applicable to the present backtesting framework, since both the objective function of the underlying M-estimators and the loss function in their prediction exercise have to be smooth in the parameters. Backtesting tests correspond to non-smooth loss functions, with estimators possibly from a non-smooth M-estimator objective function (*e.g.* quantile regression estimators).

We use subsampling rather than block bootstrap in this paper for several reasons. First, with subsampling one does not need to estimate any score or influence function, whereas in block bootstrap such estimation is generally needed, as shown in Corradi and Swanson (2007) with the recursive scheme. Notice that if the estimator is defined through a non-smooth objective function, e.g. quantile regression, the corresponding score involves (conditional) density functions which are difficult to estimate if the dimension of W_{t-1}, d_w, is large. Therefore, our subsampling approximation leads to substantially simpler tests. Second, subsampling is a general resampling method that works, i.e. it is consistent, under a minimal set of assumptions, including cases where the block bootstrap is inconsistent; see Politis, Romano and Wolf (1999) for a comprehensive monograph on subsampling and comparisons with other bootstrap methods. This robustness of subsampling is consistent with the main theme of this paper, that is, to provide robust backtests for market risk evaluation.

The rest of the paper is structured as follows. Section 2 introduces the unconditional and independence backtests robust to estimation and model risks. Section 3 discusses different subsampling schemes to approximate the asymptotic critical values of the previously considered backtesting tests. Section 4 evaluates the finite-sample performance of the subsampling approximation through a Monte Carlo analysis. Section 5 contains an application of these backtesting procedures for assessing the risk exposure of a passive portfolio tracking the Dow-Jones Industrial Average Index. This application illustrates the differences between robust and non-robust backtesting inferences. Finally, Section 6 concludes. Mathematical proofs, tables and figures are gathered in an Appendix.

2 Backtesting Techniques Robust to Model and Estimation Risks

2.1 Unconditional Backtesting

To introduce the section and motivate the existence and importance of model risk effects in risk management exercises we study in what follows different types of misspecification of the conditional quantile model under study. The section continues with the discussion of estimation risk effects, as studied in EO, and finalizes with a theorem that proposes valid asymptotic backtesting tests robust to estimation and model risk.

Due to backtesting requirements risk management standard methods for computing VaR are concerned with reporting the correct unconditional coverage probability of market risk failure, and with filtering the serial dependence of exceedances of VaR. As discussed in the introduction these two conditions do not preclude the choice of risk models to describe the conditional VaR model that are not correctly specified. The following three examples illustrate different missmatches between conditions (1) and (2). These are given for the test of unconditional coverage probability, for the test of independence and for the joint test, respectively. For simplicity in the exposition we will assume that the relevant parameters and distribution functions in these examples are known, and therefore there is no estimation stage involved.

Example 1. (Unconditional coverage test): Suppose that $\{Y_t\}$ is a stationary stochastic process following an unconditional strictly increasing distribution function F. The researcher proposes as VaR model the unconditional quantile $\widetilde{m}_\alpha(W_{t-1}, \theta_0) = F^{-1}(\alpha)$, with $F^{-1}(\cdot)$ the inverse function of the cumulative distribution function F. Whereas this function determines the unconditional quantile of the process $\{Y_t\}$, the true conditional VaR process is given by $m_\alpha(W_{t-1}, \theta_0) = F^{-1}_{W_{t-1}}(\alpha)$, with $F_{W_{t-1}}(x) := P\{Y_t \leq x | \Im_{t-1}\}$. The unconditional quantile is misspecified to assess the conditional quantile unless $F^{-1}_{W_{t-1}}(\alpha) = F^{-1}(\alpha)$ a.s. Note, however, that even if $F^{-1}_{W_{t-1}}(\alpha) \neq F^{-1}(\alpha)$ with positive probability, we still have

$$E[1(Y_t \leq \widetilde{m}_\alpha(W_{t-1}, \theta))] = F(F^{-1}(\alpha)) = \alpha.$$

The popular Historical simulation (HS) VaR falls in this group of unconditional risk models. Note that in the HS case the unconditional VaR quantile is estimated nonparametrically from the empirical distribution function.

We will show later in the paper that for the independence test there is no model risk. Note however that this does not imply that the conditional VaR model is correctly specified. The following example illustrates this case.

EXAMPLE 2. (INDEPENDENCE TEST): Suppose that $\{Y_t\}$ is an autoregressive model given by

$$Y_t = \rho_1 Y_{t-1} + \varepsilon_t, \tag{3}$$

with $|\rho_1| < 1$, and ε_t an error term serially independent that follows a distribution function $F_\varepsilon(\cdot)$. The corresponding conditional VaR that satisfies (2) is given by

$$m_\alpha(W_{t-1}, \theta_0) = \rho_1 Y_{t-1} + F_\varepsilon^{-1}(\alpha),$$

with $F_\varepsilon^{-1}(\alpha)$ the $\alpha-$quantile of F_ε. Suppose however that the researcher assumes incorrectly an alternative autoregressive process where the error term follows a Gaussian distribution function $(\Phi(\cdot))$; in this case the conditional (misspecified) quantile process is

$$\widetilde{m}_\alpha(W_{t-1}, \theta_0) = \rho_1 Y_{t-1} + \Phi_\varepsilon^{-1}(\alpha),$$

with $\Phi_\varepsilon^{-1}(\alpha) \neq F_\varepsilon^{-1}(\alpha)$.

Note that although the risk model is wrong the sequence of hits associated to $\widetilde{m}_\alpha(W_{t-1}, \theta_0)$ are *iid* since

$$P\{\varepsilon_t \leq \Phi_\varepsilon^{-1}(\alpha), \varepsilon_{t-1} \leq \Phi_\varepsilon^{-1}(\alpha)\} = P\{\varepsilon_t \leq \Phi_\varepsilon^{-1}(\alpha)\} P\{\varepsilon_{t-1} \leq \Phi_\varepsilon^{-1}(\alpha)\},$$

by independence of the error term.

The last example illustrates a misspecified location-scale model for the joint hypothesis (1).

EXAMPLE 3. (JOINT TEST): Suppose that $\{Y_t\}$, conditional on a sequence of exogenous *iid* variables $\{Z_t\}_{t=1}^n$, are independent, with Y_t given $\{Z_t\}_{t=1}^n$ distributed as $F_0(y \mid Z_t)$, which is a strictly increasing function of y. Therefore, the corresponding VaR model is $m_\alpha(W_{t-1}) = F_0^{-1}(\alpha \mid Z_t)$. Let F_Y be the unconditional cdf of Y_t, and suppose the researcher incorrectly specifies $\widetilde{m}_\alpha(W_{t-1}, \theta_0) \equiv \theta_0 = F_Y^{-1}(\alpha)$. Hence, by definition of $F_Y^{-1}(\alpha)$, $E[I_{t,\alpha}(\theta_0)] = \alpha$. Moreover, for $j \geq 1$,

$$\begin{aligned} E[I_{t,\alpha}(\theta_0)I_{t-j,\alpha}(\theta_0)] &= E\left[E[I_{t,\alpha}(\theta_0)I_{t-j,\alpha}(\theta_0) \mid \{Z_t\}_{t=1}^n]\right] \\ &= E\left[E[I_{t,\alpha}(\theta_0) \mid Z_t]E[I_{t-j,\alpha}(\theta_0) \mid Z_{t-j}]\right] \\ &= E\left[E[I_{t,\alpha}(\theta_0) \mid Z_t]\right]E[[I_{t-j,\alpha}(\theta_0) \mid Z_{t-j}]] \\ &= E[I_{t,\alpha}(\theta_0)]E[I_{t-j,\alpha}(\theta_0)]. \end{aligned}$$

Therefore, $\{I_{t,\alpha}(\theta_0)\}$ satisfy (2), but clearly (1) does not hold.

These examples illustrate the difference between hypotheses (1) and (2). The simulation section shows in more detail the effect of these different types of misspecification in the asymptotic backtesting procedures. Another source of perturbation in the practical implementation of these tests stems from the estimation of the unknown parameters of the conditional quantile process. This implies that the test statistic introduced by Kupiec (1995) and used for the unconditional coverage hypothesis is in practice

$$S_P \equiv S(P,R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} (I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha),$$

with $\widehat{\theta}_{t-1}$ an estimator of θ_0 satisfying certain regularity conditions (cf. A4 below), instead of

$$K_P \equiv K(P,R) := \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} (I_{t,\alpha}(\theta_0) - \alpha), \tag{4}$$

with R being the in-sample size and P the corresponding out-of-sample testing period.

The estimator $\widehat{\theta}_{t-1}$ will be computed according to the forecasting scheme used to create the forecasts. For the sake of completeness, and following *e.g.* West (1996) and McCraken (2000), we discuss three different forecasting schemes, namely, the recursive, rolling and fixed forecasting schemes. They differ in how the parameter θ_0 is estimated. In the recursive scheme, the estimator $\widehat{\theta}_t$ is 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 $\widehat{\theta}_t$, that is, $\widehat{\theta}_t$ is constructed from the sample $s = t - R + 1, ..., t$. Finally, in the fixed scheme the parameter is not updated when new observations become available, i.e., $\widehat{\theta}_t = \widehat{\theta}_R$, for all t, $R \leq t \leq n$.

EO showed that inference procedures computed from S_P but using the critical values of the asymptotic normal distribution followed by K_P may be misleading under very general circumstances. It should be noted that this inference procedure is the standard practice in empirical applications. More concretely, EO proved that under general conditions $(\alpha(1-\alpha))^{-1/2} S_P$ converges to a non-standard zero mean Gaussian random variable, with variance that depends on the model, the estimator $\widehat{\theta}_{t-1}$ and the forecast scheme. They assumed in their analysis, however, that the model was correctly specified, that is, (1) holds. In the present paper we relax such a strong assumption and prove that this introduces asymptotically an extra term in the, still normal, limiting distribution, changing the resulting asymptotic variance of S_P.

For sake of clarity we also discuss here Theorem 1 in EO, and the decomposition of the test statistic S_P in terms of estimation and model risk components. Under some regularity conditions, excluding (2) and (1), EO showed that

$$\begin{aligned} S_P &= \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[I_{t,\alpha}(\theta_0) - F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0)) \right] \\ & \underbrace{+E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0)) \right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} H(t-1)}_{\text{Estimation Risk}} \\ & \underbrace{+\frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0)) - \alpha \right]}_{\text{Model Risk}} + o_P(1), \end{aligned} \tag{5}$$

where $g_\alpha(W_{t-1},\theta)$ is the derivative of $m_\alpha(W_{t-1},\theta)$ with respect to θ, and $H(t-1)$ is defined in A4 below.

For their analysis EO assumed that $F_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0)) = \alpha$ *a.s.* If we relax this assumption it follows from (5) and a suitable Central Limit Theorem (CLT) applied to its third term that model risk has a non-negligible effect on the unconditional test, thereby invalidating the results in EO. Under suitable mixing conditions and following the assumptions in EO we can study simultaneously the effect of both, estimation and model risks. For a different set of assumptions see McCracken (2000).

We remark that in general the estimation risk will be affected by model risk, through a different asymptotic theory for the estimator. Estimators that under correct specification had a martingale difference influence function generally lose this property when model risk is present, see e.g. the quantile regression estimator analyzed in Kim and White (2003).

Define the α-mixing coefficients as

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

where the σ-fields $\mathcal{F}_n$ and $\mathcal{P}_n$ are $\mathcal{F}_n := \sigma(X_t, t \leq n)$ and $\mathcal{P}_n := \sigma(X_t, t \geq n)$, respectively, with $X_t = (Y_t, Z_t')'$.

Assumption A1: $\{Y_t, Z_t'\}_{t\in\mathbb{Z}}$ is strictly stationary and strong mixing process with mixing coefficients satisfying $\sum_{j=1}^{\infty} (\alpha(j))^{1-2/d} < \infty$, with $d > 2$ as in A4.

Assumption A2: The family of distributions functions $\{F_x, x \in \mathbb{R}^{d_w}\}$ has Lebesgue densities $\{f_x, x \in \mathbb{R}^{d_w}\}$ that are uniformly bounded $\left(\sup_{x\in\mathbb{R}^{d_w}, y\in\mathbb{R}} |f_x(y)| \leq C\right)$ and equicontinuous: for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\sup_{x\in\mathbb{R}^{d_w}, |y-z|\leq\delta} |f_x(y) - f_x(z)| \leq \epsilon$.

Assumption A3: The model $m_\alpha(W_{t-1}, \theta)$ is continuously differentiable in θ (a.s.) with derivative $g_\alpha(W_{t-1}, \theta)$ such that $E\left[\sup_{\theta\in\Theta_0} |g_\alpha(W_{t-1}, \theta)|^2\right] < C$, for a neighborhood Θ_0 of θ_0.

Assumption A4: The parameter space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ. The estimator $\widehat{\theta}_t$ satisfies the asymptotic expansion $\widehat{\theta}_t - \theta_0 = H(t) + o_P(1)$, where $H(t)$ is a $p \times 1$ vector such that $H(t) = t^{-1}\sum_{s=1}^{t} l(Y_s, W_{s-1}, \theta_0)$, $R^{-1}\sum_{s=t-R+1}^{t} l(Y_s, W_{s-1}, \theta_0)$ and $R^{-1}\sum_{s=1}^{R} l(Y_s, W_{s-1}, \theta_0)$ for the recursive, rolling and fixed schemes, respectively. Moreover, $l(Y_t, W_{t-1}, \theta)$ is continuous (a.s.) in θ in Θ_0 and $E\left[|l(Y_t, W_{t-1}, \theta_0)|^d\right] < \infty$ for some $d > 2$.

Assumption A5: $R, P \to \infty$ as $n \to \infty$, and $\lim_{n\to\infty} P/R = \pi$, $0 \leq \pi < \infty$.

Assumptions A1 to A5 are standard in inferences on nonlinear quantile regression. Assumptions A4 and A5 are assumed in West (1996) and McCraken (2000), see e.g. the discussion in McCraken (2000, p. 200). Further, to simplify notation define $i_t := I_{t,\alpha}(\theta_0) - F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))$, $d_t := F_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) - \alpha$, $l_t := l(Y_t, W_{t-1}, \theta_0)$, $a_t := i_t + d_t$, and $A := E\left[g_\alpha'(W_{t-1}, \theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))\right]$. Note that d_t is bounded and strong mixing, therefore for weak convergence of the model risk it suffices that $\sum_{j=1}^{\infty} \alpha(j) < \infty$. The stronger assumption in A1 is needed for the weak convergence of the estimation risk. Define $\Gamma_{al}(j) = E[a_t l_{t-j}]$, $\Gamma_{ll}(j) = E[l_t l_{t-j}]$, $\Gamma_{aa}(j) = E[a_t a_{t-j}]$, $S_{al} = \sum_{j=-\infty}^{\infty} \Gamma_{al}(j)$, $S_{ll} = \sum_{j=-\infty}^{\infty} \Gamma_{ll}(j)$, and $S_{aa} = \sum_{j=-\infty}^{\infty} \Gamma_{aa}(j)$. Our assumptions imply that the previous long-run variances exist.

With this notation in place and the decomposition in (5) we can derive the right asymptotic distribution of the backtesting test statistic. Thus, next theorem provides the asymptotic distribution of S_P under the null hypothesis of correct unconditional coverage, and making allowance for the presence of estimation and model risk.

THEOREM 1: *Under Assumptions A1-A5 and* $E[I_{t,\alpha}(\theta_0)] = \alpha$,

$$S_P \xrightarrow{d} N(0, \sigma_a^2),$$

with $\sigma_a^2 = S_{aa} + \lambda_{al}(AS_{al} + S'_{al}A') + \lambda_{ll}AS_{ll}A'$, *where*

Scheme	λ_{al}	λ_{ll}
Recursive	$1 - \pi^{-1}\ln(1+\pi)$	$2\left[1 - \pi^{-1}\ln(1+\pi)\right]$
Rolling, $\pi \leq 1$	$\pi/2$	$\pi - \pi^2/3$
Rolling, $1 < \pi < \infty$	$1 - (2\pi)^{-1}$	$1 - (3\pi)^{-1}$
Fixed	0	π

(6)

In general, σ_a^2 may be greater, equal or smaller than $\alpha(1-\alpha)$. Note that if R is arbitrarily large relative to P, *i.e.* $\pi = 0$, there is "infinite" information contained in $\widehat{\theta}_{t-1}$ about θ_0 relative to S_P, and as a result the estimation risk component asymptotically vanishes. Model risk will be present even if $\pi = 0$, whenever the hits are correlated.

There are different approaches to implement the test above. McCraken (2000) discusses non-parametric estimation of σ_a^2, whereas Corradi and Swanson (2007) propose block-bootstrap procedures. In this paper we do not follow either of these methodologies, instead we use subsampling approximations. The subsampling has several advantages over the asymptotic theory or the block bootstrap. Most notably, with subsampling one does not need to estimate any score or influence function, whereas in asymptotic theory or in block bootstrap such estimation is generally needed, as shown in e.g. Corradi and Swanson (2007) with the recursive scheme. Therefore, subsampling avoides the cumbersome task of estimating quantities such as the conditional density $f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))$. The subsampling approximation is discussed in Section 3 in detail.

2.2 Independence and Joint Tests

This section is devoted to the hypothesis of independence in (2). To complement backtesting exercises one can be interested in testing the following

$$\{I_{t,\alpha}(\theta_0)\}_{t=R+1}^{n} \text{ are } \textit{iid}. \tag{7}$$

Since for Bernoulli random variables serial independence is equivalent to serial uncorrelation, it is natural to build a test for (7) based on the autocovariances

$$\gamma_j = Cov(I_{t,\alpha}(\theta_0), I_{t-j,\alpha}(\theta_0)) \qquad j \geq 1. \tag{8}$$

They can be consistently estimated (under $E[I_{t,\alpha}(\theta_0)] = \alpha$) by

$$\gamma_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\theta_0) - \alpha)(I_{t-j,\alpha}(\theta_0) - \alpha) \qquad j \geq 1.$$

Berkowitz, Christoffersen and Pelletier (2006) discuss Portmanteau tests in the spirit of those proposed by Box and Pierce (1970) and Ljung and Box (1978) that make use of the sequence of sample autocovariances $\{\gamma_{P,j}\}$. Note that tests based on $\{\gamma_{P,j}\}$ are actually joint tests of the *iid* and the unconditional hypothesis, since they explicitly use the fact that $E[I_{t,\alpha}(\theta_0)] = \alpha$. Instead, a proper marginal test of independence should be based on

$$\zeta_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\theta_0) - E_n[I_{t,\alpha}(\theta_0)])(I_{t-j,\alpha}(\theta_0) - E_n[I_{t-j,\alpha}(\theta_0)]),$$

where, for $\theta \in \Theta$,

$$E_n[I_{t,\alpha}(\theta)] = \frac{1}{P-j} \left\{ \sum_{t=R+j+1}^{n} I_{t,\alpha}(\theta) \right\}.$$

In practice, however, joint or marginal tests for (7) need to be based on estimates of the relevant parameters, such as

$$\widehat{\gamma}_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\widehat{\theta}_{t-1}) - \alpha)(I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - \alpha)$$

or

$$\widehat{\zeta}_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} (I_{t,\alpha}(\widehat{\theta}_{t-1}) - E_n[I_{t,\alpha}(\widehat{\theta}_{t-1})])(I_{t-j,\alpha}(\widehat{\theta}_{t-j-1}) - E_n[I_{t-j,\alpha}(\widehat{\theta}_{t-j-1})]).$$

Next theorem is the equivalent of Theorem 1 for the joint and independence backtesting tests, but before introducing it we need the following decompositions corresponding to $\widehat{\gamma}_{P,j}$ and $\widehat{\zeta}_{P,j}$;

$$\widehat{\gamma}_{P,j} = \gamma_{P,j} + \frac{B}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1), \tag{9}$$

with $B = E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\{I_{t-j,\alpha}(\theta_0)+\alpha\}\right]$, see Theorem 2 in EO for a detailed proof, and

$$\widehat{\zeta}_{P,j} = \zeta_{P,j} + \frac{C}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1), \tag{10}$$

with $C = E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\{I_{t-j,\alpha}(\theta_0)+E[I_{t,\alpha}(\theta_0)]\}\right]$.

Note that this statistic is different from the test statistics discussed in EO. The decomposition, nevertheless, follows immediately from (9). Now, define $b_t(\theta) := (I_{t,\alpha}(\theta)-\alpha)(I_{t-j,\alpha}(\theta)-\alpha)$, $b_t = b_t(\theta_0)$, and similarly $c_t(\theta) := (I_{t,\alpha}(\theta)-E[I_{t,\alpha}(\theta)])(I_{t-j,\alpha}(\theta)-E[I_{t-j,\alpha}(\theta)])$, $c_t = c_t(\theta_0)$. Define also $\Gamma_{bl}(j) = E[b_t l_{t-j}]$, $\Gamma_{cl}(j) = E[c_t l_{t-j}]$, $\Gamma_{bb} = E[b_t^2]$, $\Gamma_{cc} = E[c_t^2]$, $S_{bl} = \sum_{j=-\infty}^{\infty}\Gamma_{bl}(j)$ and $S_{cl} = \sum_{j=-\infty}^{\infty}\Gamma_{cl}(j)$.

Theorem 2: *Under Assumptions A1-A5 and (2), for any* $j \geq 1$,

1. $\widehat{\gamma}_{P,j} \xrightarrow{d} N(0,\sigma_b^2)$, *where* $\sigma_b^2 = \Gamma_{bb} + \lambda_{al}(BS_{bl} + S'_{bl}B') + \lambda_{ll}BS_{ll}B'$.
2. *If instead of (2), only* (7) *holds, then*

$$\widehat{\zeta}_{P,j} \xrightarrow{d} N(0,\sigma_c^2),$$

where $\sigma_c^2 = \Gamma_{cc} + \lambda_{al}(CS_{cl} + S'_{cl}C') + \lambda_{ll}CS_{ll}C'$.

In contrast to the unconditional case there is no model risk produced by the wrong specification of the VaR model in these tests. This can be observed from the expressions above for Γ_{bb} and Γ_{cc} that do not depend on autocorrelation terms. Note, however, that there will be an indirect model risk effect through the estimation risk. Nevertheless, as for the unconditional test, we use subsampling to develop robust inference for the joint and marginal independence tests. The subsampling approximation is described in the following section.

3 Subsampling Approximation

In this section we approximate the critical values of the aforementioned tests with the assistance of the subsampling methodology. Resampling methods from *iid* sequences have been used extensively in the literature on quantile regression models, see, *e.g.*, Hahn (1995), Horowitz (1998), Bilias, Chen and Ying (2000), Sakov and Bickel (2000) or He and Hu (2002). Under the presence of serial dependence the bootstrap approximation becomes more challenging. On the other hand the subsampling method is a powerful resampling scheme that allows an asymptotically valid inference under very general conditions on the data generating process, see Politis, Romano and Wolf (1999) for a general treatise on subsampling, and Chernozhukov (2002) and Whang (2004) for subsampling approximations in linear quantile regression models.

For completeness, we discuss two different subsampling approximations, depending on the existence or not of estimation risk.

3.1 Subsampling approximaton with known parameters

In our framework we apply first the subsampling methodology to approximate the critical values of tests based on K_P, $\gamma_{P,j}$ and $\zeta_{P,j}$. Henceforth, we use $T_P(\theta_0)$ to denote any of these test statistics. Also, with an abuse of notation we write the test statistics as a function of the data $\{X_t = (Y_t, W'_{t-1})' : t = 0, 1, 2, ...\}$, $T_P(\theta_0) = T_P(X_{R+1}, ..., X_n; \theta_0)$, finally, $G_P(w)$ denotes the cumulative distribution function of the test statistic,

$$G_P(w) = P(T_P(\theta_0) \leq w).$$

There are two approaches in standard subsampling. Subsampling for *iid* data where the subsamples are taken randomly, and subsampling from consecutive subsamples of observations for the dependent case. Since our interest is in defining a testing framework for (2) that is robust to possible misspecifications of condition (1) we will implement the second class of subsampling that preserves the dependence structure in the re-sampling exercise. We stress that if there is no estimation involved another possibility is subsampling from the sequence of hits rather than from the original sequence X_t. We also remark that for the independence test without estimation risk, it is possible to apply subsampling for *iid* data, by resampling from the hits.

Let $T_{b,i}(\theta_0) = T_b(X_i, ..., X_{i+b-1}; \theta_0)$, $i = 1, \ldots, n-b+1$, be the test statistic computed with the subsample $(X_i, ..., X_{i+b-1})$ of size b. We note that each subsample of size b (taken without replacement from the original data) is indeed a sample of size b from the true data generating process. Hence, one can approximate the sampling distribution $G_P(w)$ using the distribution of the values of $T_{b,i}(\theta_0)$ computed over the $n-b+1$ different consecutive subsamples of size b. This subsampling distribution is defined by

$$G_{P,b}(w) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1(T_{b,i}(\theta_0) \leq w), \qquad w \in [0, \infty),$$

and let $c_{P,1-\tau,b}$ be the $(1-\tau)$-th sample quantile of $G_{P,b}(w)$, *i.e.*, $c_{P,1-\tau,b} = \inf\{w : G_{P,b}(w) \geq 1-\tau\}$. The rejection region in this case is determined by $T_P(\theta_0) > c_{P,1-\tau,b}$. Of course, our subsampling approximation is also valid if we resample just from the out-of-sample observations $(X_i, ..., X_{i+b-1}; \theta_0)$, $i = P+1, \ldots, n-b+1$.

Note that the mixing assumption in A1 is sufficient but not necessary for the validity of the subsampling and of the empirical approximations of the asymptotic critical values, see Politis, Romano and Wolf (1999). The next result justifies theoretically the subsampling approximation.

THEOREM 3: *Assume A1-A5 and that $b/P \to 0$ with $b \to \infty$ as $P \to \infty$. Then, for $T_P(\theta_0)$ any of the tests statistics K_P, $\gamma_{P,j}$ or $\zeta_{P,j}$*, we have that

(i) under the joint null hypothesis (2), $c_{P,1-\tau,b} \xrightarrow{P} c_{1-\tau}$ *and* $P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow \tau$;

(ii) under the marginal null hypothesis of unconditional coverage probability, $c_{P,1-\tau,b} \xrightarrow{P} c_{1-\tau}$ *and* $P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow \tau$;

(iii) under any fixed alternative hypothesis, $P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow 1$.

Theorem 3 implies that the subsampling versions of the tests above have a correct asymptotic significance level and are consistent. It is worth remarking condition (iii); in contrast to bootstrap techniques this condition holds true even when $c_{P,1-\tau,b}$ is a critical value approximating the critical value of the asymptotic distribution of the test statistic under the alternative hypothesis. The consistency of the subsampling test is due to the fact that $T_P(\theta_0)$, based on P observations, converges to infinity as a faster rate than $T_{b,1}(\theta_0)$. This result will be very important in applications of backtesting methods to financial data in which the data generating process driving the conditional VaR process is not known. In this case one does not know whether the subsampling critical value converges to the asymptotic critical value of the null or alternative distributions. Another very appealing property of our subsampling tests compared to the asymptotic approximation is that they do not need estimation of the scores A, B and C.

3.2 Subsampling approximaton with unknown parameters

The previous subsampling approximation has to be modified when θ_0 is unknown. In this case the subsampling cumulative distribution function of the relevant test statistics, now S_P, $\widehat{\gamma}_{P,j}$ *or* $\widehat{\zeta}_{P,j}$, is

$$G^*_{P,b}(w) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1(\widehat{T}_{b,i} \leq w), \qquad w \in [0, \infty),$$

with $\widehat{T}_{b,i}$ the subsample version of the test statistic, computed with data $X_i, ..., X_{i+b-1}$. When estimation risk is present we divide each subsample $X_i, ..., X_{i+b-1}$ in in-sample and out-of-sample observations according to the ratio π used for computing the original test. That is, R_b observations are used for in-sample and P_b observations for out-of-sample, such that $R_b + P_b = b$ and $\lim_{n\to\infty} P_b/R_b = \pi$. The computation of $\widehat{T}_{b,i}$ is done as with the original test, say $\widehat{T}_P$, but replacing the original sample by $X_i, ..., X_{i+b-1}$ and R and P by R_b and P_b. Denote by $c^*_{P,1-\tau,b} = \inf\{w : G^*_{P,b}(w) \geq 1-\tau\}$ and $c^*_{1-\tau}$ the $(1-\tau)$-th sample and population quantile of $G^*_{P,b}(w)$ and $G^*_{\infty}(w)$, respectively, with $G^*_{\infty}(w)$ denoting the limit distribution of $\widehat{T}_P$.

The following theorem shows that the subsampling approximation $G^*_{P,b}(\cdot)$ of the asymptotic distribution incorporates estimation risk effects.

Theorem 4: *Assume A1-A5 and that* $b/P \to 0$ *with* $b \to \infty$ *as* $P \to \infty$. *Moreover, assume that* $\lim_{n\to\infty} P_b/R_b = \lim_{n\to\infty} P/R = \pi$. *Then, for* $\widehat{T}_P$ *any of the tests statistics* S_P, $\widehat{\gamma}_{P,j}$ *or* $\widehat{\zeta}_{P,j}$, *and, we have that*

(i) under the joint null hypothesis (2), $c^*_{P,1-\tau,b} \xrightarrow{P} c^*_{1-\tau}$ *and* $P(\widehat{T}_P > c^*_{P,1-\tau,b}) \longrightarrow \tau$;

(ii) under the marginal null hypothesis of unconditional coverage probability, $c^*_{P,1-\tau,b} \xrightarrow{P} c^*_{1-\tau}$ *and* $P(\widehat{T}_P > c^*_{P,1-\tau,b}) \longrightarrow \tau$;

(iii) under any fixed alternative hypothesis, $P(\widehat{T}_P > c^*_{P,1-\tau,b}) \longrightarrow 1$.

In practice it is well known that the empirical size and power of the tests depend on the choice of the parameter b. For appropriate choices the reader is referred to Politis, Romano and

Wolf (1999) or Sakov and Bickel (2000). In the present article, for simplicity in the computations we follow the suggestion of Sakov and Bickel (2000) and choose $b = \lfloor kP^{2/5} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, which yields the optimal minimax accuracy under certain conditions. Section 4 shows that this resampling procedure provides good approximations in finite samples for a variety of values of k.

4 Monte Carlo Experiment

This section investigates the effect of model risk in the standard unconditional coverage and independence backtesting methods for assessing VaR measures. To simplify the computation and stress the effects of model risk over estimation risk we assume that there is no estimation involved in the simulations. For a study of estimation effects from parametric VaR models on backtesting procedures the reader can refer to EO.

We implement these two tests for the following homoscedastic AR(1) data generating process defined by

$$Y_t = \rho_1 Y_{t-1} + \varepsilon_t, \tag{11}$$

and where ε_t follows a standard Gaussian cumulative distribution function. The true conditional VaR process derived from this model is given by

$$m_\alpha(W_{t-1}, \theta_0) = \rho_1 Y_{t-1} + \Phi^{-1}(\alpha). \tag{12}$$

Let us assume for our simulation experiment that the researcher chooses instead a misspecified VaR process, that however satisfies the unconditional coverage hypothesis in (2). This is given by

$$\widetilde{m}_\alpha(W_{t-1}, \theta_0) = \frac{1}{1-\rho_1^2}\Phi^{-1}(\alpha). \tag{13}$$

The Value at Risk of these models is calculated at 10%, 5% and 1% to gauge the effect of different coverage probabilities. Following Sakov and Bikel (2000) we have used a subsample size $b = \lfloor kP^{2/5} \rfloor$, that have implemented for $k = 1, 2, \ldots, 20$, and $P = 500$, 1000. In these simulations we observe a better performance of the empirical size as k increases that stabilizes after $k = 5$ and decreases after $k = 15$. Hence, we have chosen just to report $k = 8$ and $k = 15$ in the tables below. The relevant test statistics are S_n for the correct test statistic obtained from (12), S_n^m for the misspecified test statistic derived from (13) and using the asymptotic distribution, and S_n^s for the misspecified test statistic derived from (13) and using the subsampling robust approximation. Finally note that for simplicity and to stress model risk effects, the parameters of the model are considered known, in this case the choice of out-of-sample forecasting method is irrelevant for the simulations. It should be also noted that S_n is not feasible in practice, since we do not know the true VaR model.

The conclusions from the simulation experiment reported in tables 1 and 2 shed some light about the importance of model risk in unconditional backtesting exercises. The comparison of the first and second row for any of the reported tables shows no doubt about the magnitude of model risk effects for different significance levels and coverage probabilities. Also, we observe for $\alpha = 0.10$ and 0.05 coverage probabilities and values of k greater than five, that the subsampling methods offer a reliable approximation of the asymptotic critical values. Finally note that the empirical size is closer to the nominal size as P increases. Interestingly, for $\alpha = 0.01$ the misspecified model seems to provide good results and the subsampling, however, fails to approximate the true critical values of the test. This puzzling result may be due to the lack of significant data at this extreme coverage probability that hinders the subsampling implementation for reasonable sample sizes.

To study the consistency of the different tests under alternative hypotheses of (2) we report in tables 3 to 6 the empirical power for the processes

$$m_{\alpha}(W_{t-1}, \theta_0) = \rho_1 Y_{t-1} + \Phi^{-1}(\widetilde{\alpha}), \tag{14}$$

with $\widetilde{\alpha} \neq \alpha$, and

$$m_{\alpha}(W_{t-1}, \theta_0) = \rho_2 Y_{t-1} + \Phi^{-1}(\alpha), \tag{15}$$

with $\rho_2 \neq \rho_1$, when we fit model (12).

The set of simulations corresponding to (14), see tables 3 and 4, shows that the subsampling framework has strong power to reject the VaR specification of wrong risk models. We find a slight decrease of power of the subsampling method compared to the asymptotic distribution in table 3. The results in terms of power are similar across k values. On the other hand there is a significant decrease of power for the subsampling robust approximation when the alternative hypothesis is determined by a different autoregressive parameter (tables 5 and 6). Nevertheless, the subsampling method still has reasonable power to detect wrong VaR models. Note that, as discussed previously, the comparison between S_n with S_n^s is somewhat unfair, since the former is not feasible in practice. Finally note that the method implemented with $k = 15$ yields marginal better power results for $\alpha = 0.05, 0.01$ coverage probabilities. For $\alpha = 0.10$, it seems better to use $k = 8$ for the choice of subsampling block size.

In the previous section we studied the test statistics $\gamma_{P,j}$ and $\zeta_{P,j}$ with j denoting the order of autocorrelation. Note that whereas the first test statistic assumes the correct unconditional VaR coverage of the process, the second one does not, being therefore a marginal test specific to the independence hypothesis. For illustration purposes we will study the finite-sample size of $\gamma_{P,1}$ using the asymptotic normal critical values; and also, of $\zeta_{P,1}$ using the asymptotic normal critical values and the subsampling approximation. For the three test statistics we assume a risk model that is misspecified to report VaR but satisfies the independence hypothesis,

$$\widetilde{m}_{\alpha}(W_{t-1}, \theta_0) = \rho_1 Y_{t-1} + \Phi^{-1}(\widetilde{\alpha}), \tag{16}$$

with $\widetilde{\alpha} \neq \alpha$.

In the block of simulations in tables 7 and 8 we consider the following cases: $\alpha = 0.10, 0.05, 0.01$ and $\widetilde{\alpha} = 0.05$. Obviously, for $\alpha = 0.05$ the above model is correctly specified. We observe from the tables, as mentioned in Section 2.2, that model risk is absent for this type of tests. Moreover, whereas the test statistic $\gamma_{P,1}$ for the joint hypothesis (2) leads to clear rejections of the model, the two versions (asymptotic and subsampling) of the marginal independence test, $\zeta_{P,1}$, report the correct size. It is also remarkable the good finite-sample performance of the subsampling approximation.

Finally, in order to study the consistency of both tests under alternative hypotheses we also perform a small simulation experiment for an alternative model given by the following AR(1)-GARCH(1,1) process;

$$m_{\alpha}(W_{t-1}, \theta_0) = \rho_2 Y_{t-1} + \sigma_t \Phi^{-1}(\widetilde{\alpha}), \tag{17}$$

with $\rho_2 \neq \rho_1$, $\sigma_t^2 = \beta_0 + \beta_1 a_{t-1}^2 + \beta_2 \sigma_{t-1}^2$, $a_t = \sigma_t \varepsilon_t$ the innovation sequence and ε_t the error sequence that follows a standard Gaussian cumulative distribution function.

The results of tables 9 and 10 show the outperformance of the asymptotic marginal test over the subsampling version. As a by-product, we also observe the marginal gain in power from an alternative model (17) given by different parameter values and a coverage probability $\widetilde{\alpha} \neq \alpha$

compared to an alternative determined by changes only on the parameter values. This can be seen by comparing the left and right columns of tables 9 and 10 against the estimates in the middle column corresponding to $\widetilde{\alpha} = \alpha = 0.05$.

5 Application: Backtesting performance of VaR forecast models

This section analyzes the effect of model risk in backtesting for a battery of econometric models widely employed for forecasting out-of-sample VaR. The models under study are the hybrid method (historical simulation applied to the residuals of a GARCH model), defined by

$$m_{\alpha,1}(W_{t-1}, \widehat{\theta}_{t-1}) = \widehat{\sigma}_t F^{-1}_{R,\widehat{\varepsilon}}(\alpha), \tag{18}$$

with $F_{R,\widehat{\varepsilon}}$ the empirical distribution function constructed from an in-sample size of R observations, and $\widehat{\sigma}^2_t = \widehat{\beta}_0 + \widehat{\beta}_1 Y^2_{t-1} + \widehat{\beta}_2 \widehat{\sigma}^2_{t-1}$, the estimated GARCH process, where $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$ is the vector of estimates of the parameter vector $(\beta_0, \beta_1, \beta_2)$ and $\{\widehat{\varepsilon}_t\}$ is the residual sequence. The second method is a parametric GARCH model with residual sequence assumed to be normal;

$$m_{\alpha,2}(W_{t-1}, \widehat{\theta}_{t-1}) = \widehat{\sigma}_t \Phi^{-1}(\alpha). \tag{19}$$

The third risk model extends the previous one, allowing for heavier than normal tails. It is given by

$$m_{\alpha,3}(W_{t-1}, \widehat{\theta}_{t-1}) = \widehat{\sigma}_t t^{-1}_{\widehat{\nu}}(\alpha), \tag{20}$$

with $\widehat{\nu}$ the degrees of freedom of a Student-t estimated by Quasi-Maximum likelihood. Finally, the last method considered was proposed by Morgan (1995). This method is similar in spirit to the GARCH methodology but the volatility filtering is an exponential smoothing ($\beta_0 = 0$) where the parameters are fixed, in our case to $\beta_1 = 0.04$ and $\beta_2 = 1 - \beta_1$. The risk model is as follows

$$m_{\alpha,4}(W_{t-1}, \widehat{\theta}_{t-1}) = \widehat{\sigma}^{Rm}_t \Phi^{-1}(\alpha), \tag{21}$$

with $(\widehat{\sigma}^{Rm}_t)^2 = \beta_1 Y^2_{t-1} + (1 - \beta_1)(\widehat{\sigma}^{Rm}_{t-1})^2$.

The experiment consists on carrying out a backtesting exercise for daily returns on Dow Jones Industrial Average Index over the period 02/01/1998 until 23/05/2008 containing 2500 observations. For ease of computation we have implemented a fixed forecasting scheme for estimating the relevant empirical distribution functions F_R and the GARCH parameters, and where the in-sample size $R = 1000$ is considerably greater than the out-of-sample size $P = 500$. The choice of this sample size is a compromise between absence of estimation risk effects ($P/R < 1$) and meaningful results of the subsampling and asymptotic tests (P sufficiently large). We repeat this experiment using all the information of the time series available considering rolling windows of 250 observations and giving a total of five different periods where computing backtesting tests.

The rejection regions at 5% considered for the unconditional coverage and independence tests are those determined by the corresponding asymptotic normal distributions and the alternative subsampling approximations. For sake of space we only report the plots computed via subsampling with $k = 8$ and coverage probability $\alpha = 0.05$.

The conclusions from the sequence of plots in figure 1 are threefold. First, we observe the differences between the subsampling rejection regions and the asymptotic ones for the first three methods based on GARCH estimates, and the similarities of the Riskmetrics approach with the asymptotic intervals. This fact implies that whereas we reject the VaR forecasting methods obtained from using GARCH models we do not do it if Riskmetrics is employed. Second, the choice of the asymptotic critical values is misleading most of the times since one can accept VaR models that are wrong if no attention is paid to the misspecification effects. Finally, we observe an increase in the uncertainty in all of the unconditional backtesting tests that use subsampling.

The set of plots in figure 2 reports the results of the independence test. In contrast to the unconditional test the four subsampling methods yield very similar rejection regions. These are very different from the asymptotic ones. Note that in contrast to the conclusions drawn from focusing on the asymptotic tests there are no grounds to reject the null hypothesis of independence. The four methods are valid filters of the underlying serial dependence in the volatility process.

6 Conclusion

Backtesting techniques are of paramount importance for risk managers and regulators concerned with assessing the risk exposure of a financial institution to market risk. These methods are implemented as statistical tests designed, specially, to uncover an excessive risk-taking from financial institutions and measured by the number of exceedances of the VaR model under scrutiny.

It is also well known in the forecasting literature that econometric methods that are well specified in-sample for describing the dynamics of extreme quantiles are not necessarily those that best forecast their future dynamics. Therefore, financial institutions can choose risk models for forecasting conditional VaR that although badly specified they succeed to satisfy unconditional and independence backtesting requirements.

We have shown in this paper that in order to implement correctly the standard backtesting procedures one needs to incorporate in the asymptotic theory certain components accounting for possible misspecifications of the risk model. Also, since these components can be difficult to derive analytically and take different forms depending on the correct specification of the conditional quantile process we have proposed instead subsampling methods to approximate the true sampling distribution of the relevant test statistics. The choice of subsampling surges as a natural simpler alternative to the asymptotic theory and block bootstrap techniques that overcomes technical problems related to estimation and inference.

Our simulations indicate that values of $b = \lfloor kP^{2/5} \rfloor$ determined by $k > 5$ provide valid approximations of the true sampling distributions and hence of the true rejection regions for backtesting tests. More sophisticated data-driven choices of b are possible, see Chapter 9 in Politis, Romano and Wolf (1999). Our simulations also showed that the subsampling approximation is not appropriate for $\alpha = 0.01$, at least for the conventional sample sizes used. Note that EO found that when estimation risk is present asymptotic theory for $\alpha = 0.01$ breaks down. Hence, the problem of constructing backtests procedures robust to estimation and model risk for values of $\alpha < 0.05$ remains an important unsolved practical problem. This problem is beyond the scope of this paper and deserves further research.

We have shown that although estimation risk can be diversified by choosing a large in-sample size relative to out-of-sample, model risk cannot. Model risk is pervasive in unconditional backtests. This has been confirmed by our simulations. Our theorical and empirical results suggest the use of robust techniques based on subsampling approximations to handle simultaneously model risk and estimation risk in general dynamic models.

Appendix: Mathematical Proofs

Proof of Theorem 1: We apply Lemma A1 in EO. To that end, we need to verify the following uniform tightness condition

$$\max_{R\le t\le n} \sqrt{t}(\widehat{\theta}_t - \theta_0) = O_P(1), \tag{22}$$

for the three forecasting schemes. But (22) follows from simple arguments using the mixing property, see McCracken (2000, pg. 221) for the proof of (22). By Lemma A1 in EO we conclude that

$$S_P = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} [I_{t,\alpha}(\theta_0) - \alpha] \tag{23}$$

$$+E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} H(t-1) + o_P(1). \tag{24}$$

Now, assumptions A1-A5 guarantee that the central limit theorem can be applied to the bivariate process

$$\left[\begin{array}{c} \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n} [I_{t,\alpha}(\theta_0) - \alpha] \\ \frac{1}{\sqrt{P}}\sum_{t=R+1}^{n} H(t-1) \end{array}\right].$$

The rest of the proof follows from McCracken (2000, Theorem 2.3.1). □

Proof of Theorem 2: The proof of this theorem is similar to the proof of theorem 1. Now, along with result (22), we also use the following decomposition introduced in (9) and shown in EO,

$$\widehat{\gamma}_{P,j} = \gamma_{P,j} + \frac{B}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1),$$

with $B = E\left[g'_\alpha(W_{t-1},\theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1},\theta_0))\{I_{t-j,\alpha}(\theta_0) + \alpha\}\right]$.

In this case the weak convergence of the joint and independence backtesting tests is guaranteed by the boundness and strong mixing property of the sequence of products of centered indicator functions. It is immediate to see that a central limit theorem result can be applied. Finally, in order to compute the expression for the asymptotic variance of $\widehat{\gamma}_{P,j}$ we use the following notation above introduced; $b_t(\theta) = (I_{t,\alpha}(\theta) - \alpha)(I_{t-j,\alpha}(\theta) - \alpha)$ and $b_t = b_t(\theta_0)$. Now, after simple variance calculations and using the results on asymptotic ratio of convergence between R and P

in McCracken (2000) we obtain

$$\sigma_b^2 = \Gamma_{bb} + \lambda_{al}(BS_{bl} + S_{bl}'B') + \lambda_{ll}BS_{ll}B',$$

where $\Gamma_{bl}(j) = E[b_t l_{t-j}]$, $\Gamma_{bb} = E[b_t^2]$, and $S_{bl} = \sum_{j=-\infty}^{\infty}\Gamma_{bl}(j)$. Then

$$\widehat{\gamma}_{P,j} \xrightarrow{d} N(0, \sigma_b^2).$$

Similarly, if instead of (2) only (7) holds, we have the following decomposition for $\widehat{\zeta}_{P,j}$;

$$\widehat{\zeta}_{P,j} = \zeta_{P,j} + \frac{C}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} H(t-j-1) + o_P(1),$$

with $C = E\left[g_\alpha'(W_{t-1}, \theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0))\{I_{t-j,\alpha}(\theta_0) + E[I_{t,\alpha}(\theta_0)]\}\right]$.

The corresponding asymptotic variance is

$$\sigma_c^2 = \Gamma_{cc} + \lambda_{al}(CS_{cl} + S_{cl}'C') + \lambda_{ll}CS_{ll}C',$$

with $c_t(\theta) = (I_{t,\alpha}(\theta) - E[I_{t,\alpha}(\theta)])(I_{t-j,\alpha}(\theta) - E[I_{t-j,\alpha}(\theta)])$, $c_t = c_t(\theta_0)$, $\Gamma_{cl}(j) = E[c_t l_{t-j}]$, $\Gamma_{cc} = E[c_t^2]$, and $S_{cl} = \sum_{j=-\infty}^{\infty}\Gamma_{cl}(j)$. Therefore

$$\widehat{\zeta}_{P,j} \xrightarrow{d} N(0, \sigma_c^2).$$

□

PROOF OF THEOREM 3: Let $T_P(\theta_0)$ be any of the backtests when θ_0 is known. As shown in Theorem 1 and Theorem 2 the limit distribution of $T_P(\theta_0)$ is a normal distribution, which is, of course, continuous. Since the mixing coefficients converge to zero, Theorem 3.5.1 in Politis, Romano and Wolf (1999) can be applied, and the proof trivially follows from there. □

PROOF OF THEOREM 4: It follows from the same arguments as in Theorem 3. □

TABLES

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.096	0.046	0.010	0.086	0.044	0.014	0.104	0.030	0.014
S_P^m	0.233	0.149	0.050	0.203	0.132	0.044	0.145	0.047	0.017
S_P^s	0.045	0.030	0.026	0.090	0.072	0.059	0.207	0.202	0.200
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.116	0.061	0.011	0.088	0.050	0.010	0.098	0.050	0.007
S_P^m	0.230	0.147	0.064	0.194	0.128	0.043	0.126	0.063	0.008
S_P^s	0.052	0.035	0.021	0.053	0.039	0.027	0.218	0.208	0.207

Table 1. Empirical size for unconditional backtesting test for models (12) and (13) with $\rho_1 = 0.5$. Out-of-sample size $P = 500$, 1000. Coverage probability $\alpha = 0.10$, 0.05, 0.01. $b = \lfloor kP^{2/5} \rfloor$ with $k = 8$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test, S_P^m the corresponding asymptotic test from the misspecified model and S_P^s the robust subsampling approximation.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.081	0.034	0.008	0.100	0.049	0.010	0.108	0.042	0.017
S_P^m	0.215	0.130	0.056	0.182	0.127	0.048	0.153	0.052	0.020
S_P^s	0.073	0.051	0.033	0.078	0.061	0.048	0.137	0.117	0.101
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.108	0.051	0.004	0.109	0.053	0.008	0.083	0.039	0.005
S_P^m	0.216	0.157	0.061	0.207	0.135	0.055	0.139	0.066	0.008
S_P^s	0.056	0.038	0.021	0.055	0.039	0.025	0.152	0.135	0.122

Table 2. Empirical size for unconditional backtesting test for models (12) and (13) with $\rho_1 = 0.5$. Out-of-sample size $P = 500$, 1000. Coverage probability $\alpha = 0.10$, 0.05, 0.01. $b = \lfloor kP^{2/5} \rfloor$ with $k = 15$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test, S_P^m the corresponding asymptotic test from the misspecified model and S_P^s the robust subsampling approximation.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.996	0.990	0.972	0.087	0.045	0.013	1.000	1.000	1.000
S_P^s	0.858	0.799	0.712	0.063	0.050	0.036	1.000	1.000	1.000
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	1.000	1.000	1.000	0.088	0.040	0.010	1.000	1.000	1.000
S_P^s	0.995	0.983	0.954	0.047	0.031	0.020	1.000	1.000	1.000

Table 3. Empirical power for unconditional backtesting test for H_0 : model (12) and H_A : model (14) with $\widetilde{\alpha} = 0.05$. $\rho_1 = 0.5$ in both models. Out-of-sample size $P = 500$, 1000. True coverage probability $\alpha = 0.10, 0.05, 0.01$. $b = \lfloor kP^{2/5} \rfloor$ with $k = 8$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test and S_P^s the robust subsampling approximation from the misspecified model.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.994	0.988	0.950	0.087	0.045	0.013	1.000	1.000	1.000
S_P^s	0.940	0.924	0.900	0.077	0.051	0.035	0.990	0.982	0.966
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	1.000	1.000	1.000	0.088	0.040	0.010	1.000	1.000	1.000
S_P^s	0.999	0.998	0.996	0.044	0.028	0.018	0.847	0.790	0.740

Table 4. Empirical power for unconditional backtesting test for H_0 : model (12) and H_A : model (14) with $\widetilde{\alpha} = 0.05$. $\rho_1 = 0.5$ in both models. Out-of-sample size $P = 500$, 1000. True coverage $\alpha = 0.10$, 0.05, 0.01. $b = \lfloor kP^{2/5} \rfloor$ with $k = 15$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test and S_P^s the robust subsampling approximation from the misspecified model.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.732	0.598	0.431	0.755	0.672	0.504	0.648	0.522	0.425
S_P^s	0.198	0.137	0.084	0.226	0.160	0.108	0.145	0.100	0.067
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.897	0.846	0.723	0.926	0.883	0.777	0.809	0.759	0.634
S_P^s	0.433	0.311	0.201	0.457	0.332	0.223	0.257	0.177	0.124

Table 5. Empirical power for unconditional backtesting test for H_0 : model (12) with $\rho_1 = 0.5$ and H_A : model (15) with $\rho_2 = 0$. Out-of-sample size $P = 500$, 1000. Coverage probability $\alpha = 0.10, 0.05, 0.01$. $b = \lfloor kP^{2/5} \rfloor$ with $k = 8$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test and S_P^s the robust subsampling approximation from the misspecified model.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.732	0.598	0.431	0.755	0.672	0.504	0.648	0.522	0.425
S_P^s	0.217	0.160	0.112	0.248	0.183	0.134	0.177	0.140	0.098
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
S_P	0.897	0.846	0.723	0.926	0.883	0.777	0.809	0.759	0.634
S_P^s	0.387	0.293	0.195	0.435	0.354	0.260	0.295	0.218	0.159

Table 6. Empirical power for unconditional backtesting test for H_0 : model (12) with $\rho_1 = 0.5$ and H_A : model (15) with $\rho_2 = 0$. Out-of-sample size $P = 500$, 1000. Coverage probability $\alpha = 0.10, 0.05, 0.01$. $b = \lfloor kP^{2/5} \rfloor$ with $k = 15$. 1000 Monte-Carlo replications. S_P denotes the unconditional coverage backtesting asymptotic test and S_P^s the robust subsampling approximation from the misspecified model.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\gamma_{P,1}$	0.349	0.235	0.119	0.106	0.045	0.022	0.135	0.131	0.043
$\zeta^m_{P,1}$	0.104	0.046	0.019	0.104	0.046	0.019	0.080	0.039	0.014
$\zeta^s_{P,1}$	0.118	0.084	0.048	0.118	0.084	0.048	0.142	0.089	0.049
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\gamma_{P,1}$	0.479	0.370	0.184	0.091	0.047	0.015	0.247	0.225	0.111
$\zeta^m_{P,1}$	0.094	0.047	0.015	0.094	0.047	0.015	0.106	0.057	0.017
$\zeta^s_{P,1}$	0.090	0.062	0.037	0.090	0.062	0.037	0.099	0.068	0.037

Table 7. Empirical size of independence backtesting test for model (16) with $\rho_1 = 0.5$, and misspecified coverage probability $\widetilde{\alpha} = 0.05$. Out-of-sample size $P = 500$, 1000. True coverage probability $\alpha = 0.10$, 0.05, 0.01. $b = \lfloor kP^{2/5} \rfloor$ with $k = 8$. 1000 Monte-Carlo replications. $\gamma_{P,1}$ denotes the joint backtesting asymptotic test, $\zeta^m_{P,1}$ the corresponding marginal backtesting asymptotic test and $\zeta^s_{P,1}$ the robust subsampling approximation.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\gamma_{P,1}$	0.331	0.213	0.118	0.089	0.037	0.012	0.142	0.140	0.043
$\zeta^m_{P,1}$	0.089	0.037	0.012	0.089	0.037	0.012	0.094	0.037	0.015
$\zeta^s_{P,1}$	0.155	0.113	0.077	0.155	0.113	0.077	0.163	0.128	0.096
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\gamma_{P,1}$	0.495	0.386	0.203	0.095	0.043	0.012	0.227	0.204	0.102
$\zeta^m_{P,1}$	0.094	0.043	0.009	0.094	0.043	0.009	0.088	0.041	0.013
$\zeta^s_{P,1}$	0.116	0.087	0.066	0.116	0.087	0.066	0.102	0.082	0.054

Table 8. Empirical size of independence backtesting test for model (16) with $\rho_1 = 0.5$, and misspecified coverage probability $\widetilde{\alpha} = 0.05$. Out-of-sample size $P = 500$, 1000. True coverage probability $\alpha = 0.10$, 0.05, 0.01. $b = \lfloor kP^{2/5} \rfloor$ with $k = 15$. 1000 Monte-Carlo replications. $\gamma_{P,1}$ denotes the joint backtesting asymptotic test, $\zeta^m_{P,1}$ the corresponding marginal backtesting asymptotic test and $\zeta^s_{P,1}$ the robust subsampling approximation.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\zeta^m_{P,1}$	0.599	0.405	0.220	0.572	0.402	0.205	0.604	0.425	0.195
$\zeta^s_{P,1}$	0.302	0.225	0.130	0.287	0.194	0.126	0.308	0.215	0.139
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\zeta^m_{P,1}$	0.879	0.757	0.495	0.862	0.741	0.486	0.880	0.774	0.505
$\zeta^s_{P,1}$	0.564	0.425	0.289	0.562	0.427	0.288	0.547	0.424	0.267

Table 9. Empirical power for independence backtesting test for H_0 : model (16) with $\rho_1 = 0.5$ and H_A : model (17) with $\rho_2 = 0$, $\beta_0 = 0.1$, $\beta_1 = 0.05$, and $\beta_2 = 0.85$, and $\widetilde{\alpha} = 0.05$. Out-of-sample size $P = 500$, 1000. True coverage probabilities are $\alpha = 0.10, 0.05, 0.01$. $b = \lfloor kP^{2/5} \rfloor$ with $k = 8$. 1000 Monte-Carlo replications. $\zeta^m_{P,1}$ denotes the marginal backtesting asymptotic test and $\zeta^s_{P,1}$ the corresponding robust subsampling approximation.

$\Phi(\cdot)$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
$P = 500$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\zeta^m_{P,1}$	0.593	0.418	0.208	0.622	0.442	0.217	0.604	0.426	0.204
$\zeta^s_{P,1}$	0.293	0.236	0.178	0.325	0.264	0.202	0.336	0.266	0.196
$P = 1000$/size	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
$\zeta^m_{P,1}$	0.867	0.764	0.489	0.841	0.740	0.477	0.860	0.773	0.499
$\zeta^s_{P,1}$	0.553	0.462	0.353	0.556	0.461	0.370	0.551	0.475	0.364

Table 10. Empirical power for independence backtesting test for H_0 : model (16) with $\rho_1 = 0.5$ and H_A : model (17) with $\rho_2 = 0$, $\beta_0 = 0.1$, $\beta_1 = 0.05$, and $\beta_2 = 0.85$, and $\widetilde{\alpha} = 0.05$. Out-of-sample size $P = 500$, 1000. True coverage probabilities are $\alpha = 0.10, 0.05, 0.01$. $b = \lfloor kP^{2/5} \rfloor$ with $k = 15$. 1000 Monte-Carlo replications. $\zeta^m_{P,1}$ denotes the marginal backtesting asymptotic test and $\zeta^s_{P,1}$ the corresponding robust subsampling approximation.

FIGURES

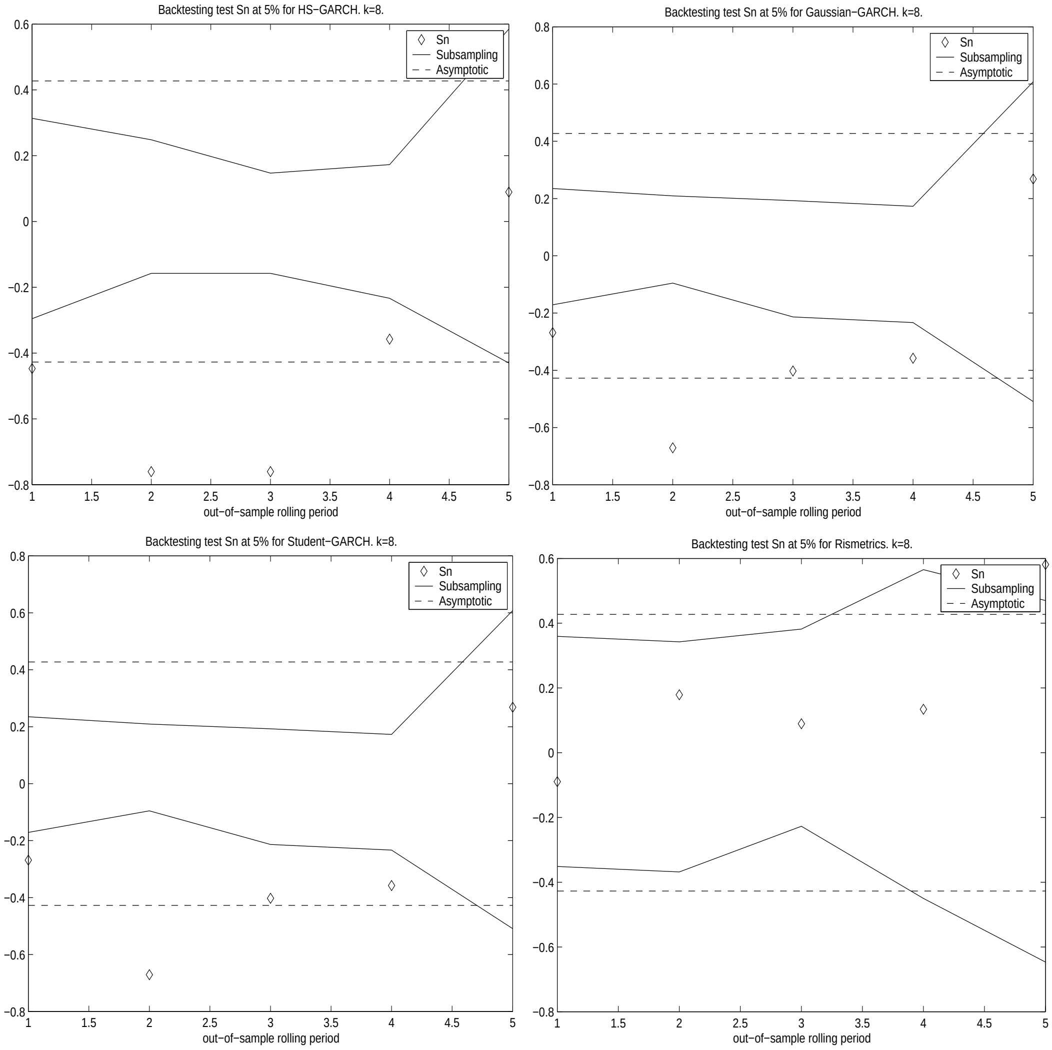

Figure 1. Unconditional backtesting for $\alpha = 0.05$. Left and right upper panels for models (18) and (19) respectively. Left and right lower panels for models (20) and (21) respectively. $\diamondsuit$ denotes S_P. Dashed lines for the asymptotic rejection regions and solid lines for the subsampling approximation. $R = 1000$, $P = 500$, rolling sample=250. $b = 8P^{2/5}$.

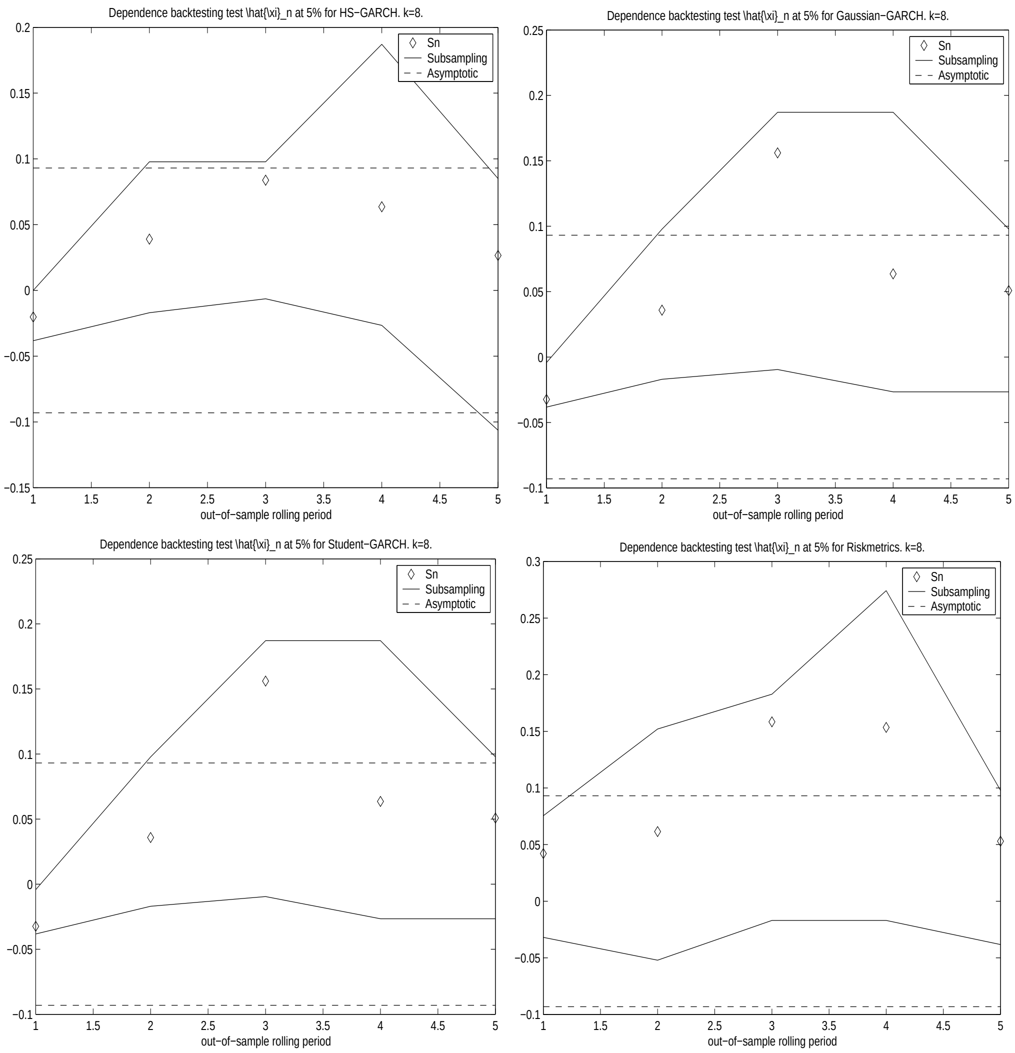

Figure 2. Independence backtesting for $\alpha = 0.05$. Left and right upper panels for models (18) and (19) respectively. Left and right lower panels for models (20) and (21) respectively. $\diamondsuit$ denotes $\widehat{\zeta}_{P,1}$. Dashed lines for the asymptotic rejection regions and solid lines for the subsampling approximation. $R = 1000$, $P = 500$, rolling sample=250. $b = 8P^{2/5}$.

REFERENCES

Basel Committee on Banking Supervision (1996): "Amendment to the Capital Accord to Incorporate Market Risks, Bank for International Settlements. "

Bilias, Y., Chen, S. and Ying, Z. (2000): "Simple resampling methods for the censored regression quantiles", *Journal of Econometrics*, 99, 373–386.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2006): "Evaluating Value-at-Risk Models with Desk-level Data. " *Working Paper Series, 010* Department of Economics, North Carolina State University.

Box, G. and Pierce, D. (1970): "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models." *Journal of American Statistical Association*, 65, 1509-1527.

Chernozhukov, V. (2002): "Inference on the quantile regression process, an alternative," Working article 02-12 (February), MIT (www.ssrn.com).

Christoffersen, P. (1998): "Evaluating Interval Forecasts. "*International Economic Review,* 39, 841-862.

Christoffersen, P., Hahn, J., and Inoue, A. (2001): "Testing and Comparing Value at Risk Measures "*Journal of Empirical Finance*, 8, 325-342.

Christoffersen, P., Gonçalves, S. (2005): "Estimation Risk in Financial Risk Management. "*Journal of Risk*, 7, 1-28.

Corradi, V., and Swanson, N.R. (2007): "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes", forthcoming in *International Economic Review.*

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

Escanciano, J. C. and Olmo, J. (2008): "Backtesting Parametric VaR with Estimation Risk. "Forthcoming in *Journal of Business and Economic Statistics.*

Figlewski, S. (2003): "Estimation Error in the Assessment of Financial Risk Exposure. "Available at SSRN: http://ssrn.com/abstract=424363

Goureioux, C., and Jasiak, J. (2006): "Dynamic Quantile Models. " Working Paper 2006-4, Department of Economics, York University.

Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators", *Econometric Theory*, 11, 105-121.

He, X., and Hu, F. (2002): "Markov chain marginal bootstrap", *Journal of the American Statistical Association*, 97, 783-795.

Horowitz, J.L. (1998): "Bootstrap methods for median regression models", *Econometrica*, 66, 1327-1351.

Kim,T.H. and White,H. (2003): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions", in T. Fomby and R.C. Hill, eds., Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. New York: Elsevier, 107-132.

Koenker, R. and Xiao, Z. (2006): "Quantile autoregression." *Journal of the American Statistical Association*, 101, 980-990.

Kuester, K., Mittnik, S., and Paolella, M.S. (2006): "Value-at-Risk Prediction: A Comparison of Alternative Strategies. "*Journal of Financial Econometrics*, 4, 1, 53-89.

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

Ljung, G. M. and Box, G. E. P. (1978): "A Measure of Lack of Fit in Time Series Models."*Biometrika,* 65, 297-303.

McCracken, (2000): "Robust Out-of-sample Inference." *Journal of Econometrics,* 99, 5, 195-223.

Morgan, J.P. (1995): *Riskmetrics Technical Manual,* 3^{rd} *edition.*

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

Sakov, A. and Bickel, P. (2000): "An Edgeworth expansion for the m out of n bootstrap median", *Statistical and Probability Letters*, 49, 217-223.

West, K.D. (1996): "Asymptotic Inference about Predictive Ability."*Econometrica,* 64, 5, 1067-1084.