Robust Backtesting Tests for Value-at-Risk Models

Jose Olmo

City University London

(joint work with Juan Carlos Escanciano, Indiana University)
Outline

• Differences between a specification test for the quantile process and backtesting.

• Model risk in unconditional and independence Backtesting Exercises

• A Subsampling solution

• Monte Carlo analysis (Asymptotic theory vs Block bootstrap and Subsampling)

• Application to risk management in financial data

• Conclusions
Differences between Specification Tests and Backtesting

Value-at-Risk: We say that $m_\alpha(W_{t-1}, \theta)$ is a correctly specified $\alpha$-th conditional VaR of $Y_t$ given $W_{t-1} = \{Y_s, Z'_s\}_{s=t-h}^{t-1}$, $h < \infty$, if and only if

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

(1)

for some unknown parameter $\theta_0 \in \Theta$, with $\Theta$ a compact set in an Euclidean space $\mathbb{R}^p$ and $\mathcal{F}_{t-1}$ denoting the set of available information.

This condition is equivalent to the following in terms of conditional expectations:

$$E[I_{t,\alpha}(\theta_0) \mid \mathcal{F}_{t-1}] = \alpha, \text{ almost surely (a.s.), } \alpha \in (0, 1), \forall t \in \mathbb{Z},$$

(2)

with $I_{t,\alpha}(\theta) := 1(Y_t \leq m_\alpha(W_{t-1}, \theta))$ the so called hits or exceedances associated to the model $m_\alpha(W_{t-1}, \theta_0)$. 
This condition implies the so-called joint hypothesis (cf. Christoffersen (1998)),

\[ \{I_{t,\alpha}(\theta_0)\} \text{ are iid } \text{Ber}(\alpha) \text{ random variables (r.v.) for some } \theta_0 \in \Theta, \tag{3} \]

where \text{Ber}(\alpha) stands for a Bernoulli r.v. with parameter \( \alpha \).

In the risk management literature \textbf{Backtesting} is identified with this joint hypothesis:

\begin{itemize}
    \item \textbf{Unconditional Backtesting test:}
    \[ E[I_{t,\alpha}(\theta_0)] = \alpha, \tag{4} \]
    \item \textbf{Independence Backtesting test:}
    \[ \{I_{t,\alpha}(\theta_0)\} \text{ being iid.} \tag{5} \]
\end{itemize}

\textbf{Motivation:} The joint test \( E[I_{t,\alpha}(\theta_0) \mid \mathcal{F}_{t-1}] = \alpha \) and the two tests above are not equivalent. In fact, the latter two tests do not imply necessarily the former.
Example 1. (Unconditional coverage test):

Suppose that the true model for conditional VaR is

\[ m_\alpha(W_{t-1}, \theta_0) = F_{W_{t-1}}^{-1}(\alpha), \]  

with \( F_{W_{t-1}}(x) := P\{Y_t \leq x | \mathcal{F}_{t-1}\} \).

The researcher proposes, instead, an unconditional quantile:

\[ \tilde{m}_\alpha(W_{t-1}, \theta_0) = F^{-1}(\alpha), \quad \text{for all } t, \]  

with \( F^{-1}(\cdot) \) the inverse function of the unconditional distribution function \( F \).

Note from these equations that

\[ E[1(Y_t \leq \tilde{m}_\alpha(W_{t-1}, \theta))] = F(F^{-1}(\alpha)) = \alpha. \]
However, the unconditional quantile is misspecified to assess the conditional quantile unless $F_{W_{t-1}}^{-1}(\alpha) = F^{-1}(\alpha)$ a.s.

**Note:** 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 $\hat{F}_n$:

$$\tilde{m}_\alpha(W_{t-1}, \theta_0) = \hat{F}_n^{-1}(\alpha), \quad \text{for all } t,$$

and therefore implies that this method exhibits model risk effects.

**Message:** This method to quantify risk based on the unconditional $\alpha$—quantile can be successful in passing the backtesting unconditional coverage test, being however a wrong specification of the true conditional quantile process.
Example 2. (Independence test):

Suppose that $\{Y_t\}$ is an autoregressive model given by

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

(9)

with $|\rho_1| < 1$, and $\varepsilon_t$ an error term serially independent that follows a distribution function $F_\varepsilon(\cdot)$. Then

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

with $F_\varepsilon^{-1}(\alpha)$ the $\alpha-$quantile of $F_\varepsilon$.

Suppose now that the researcher assumes incorrectly an alternative autoregressive process where the error term follows a Gaussian distribution function ($\Phi(\cdot)$). The conditional (misspecified) quantile process is

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

with $\Phi_\varepsilon^{-1}(\alpha) \neq F_\varepsilon^{-1}(\alpha)$. 
Note that although the risk model is wrong:

\[ \tilde{m}_\alpha(W_{t-1}, \theta_0) \neq m_\alpha(W_{t-1}, \theta_0), \]

the sequence of hits associated to \( \tilde{m}_\alpha(W_{t-1}, \theta_0) \) are iid since

\[ P\{\varepsilon_t \leq \Phi^{-1}_\varepsilon(\alpha), \varepsilon_{t-1} \leq \Phi^{-1}_\varepsilon(\alpha)\} = P\{\varepsilon_t \leq \Phi^{-1}_\varepsilon(\alpha)\}P\{\varepsilon_{t-1} \leq \Phi^{-1}_\varepsilon(\alpha)\}, \]

by independence of the error term.
Implementation of Backtesting

The test statistic introduced by Kupiec (1995) for the unconditional coverage hypothesis is

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

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

In practice, however, the test statistic implemented is

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

(10)

with \( \hat{\theta}_{t-1} \) an estimator of \( \theta_0 \) satisfying certain regularity conditions (cf. A4 below).
Estimation and Model Risk I: Assumptions

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

Assumption A3: The model \( m_\alpha(W_{t-1}, \theta) \) is continuously differentiable in \( \theta \) (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 \( \Theta_0 \) of \( \theta_0 \).

Assumption A4: The parameter space \( \Theta \) is compact in \( \mathbb{R}^p \). The true parameter \( \theta_0 \) belongs to the interior of \( \Theta \). The estimator \( \hat{\theta}_t \) satisfies the asymptotic expansion \( \hat{\theta}_t - \theta_0 = H(t) + o_P(1) \), 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 \( \theta \) in \( \Theta_0 \) 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 \( \lim_{n \to \infty} P/R = \pi, \ 0 \leq \pi < \infty \).
Under the above assumptions Escanciano and Olmo (2008) showed that

\[ S_P = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[ I_{t,\alpha}(\theta_0) - F_{W_{t-1}}(m_{\alpha}(W_{t-1}, \theta_0)) \right] \]

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

\[ + \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left[ F_{W_{t-1}}(m_{\alpha}(W_{t-1}, \theta_0)) - \alpha \right] + o_P(1), \]

where \( g_{\alpha}(W_{t-1}, \theta) \) is the derivative of \( m_{\alpha}(W_{t-1}, \theta) \) with respect to \( \theta \).
**Estimation and Model Risk III: Theorem**

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

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

with $\sigma_a^2 = \sum_{j=-\infty}^{\infty} E[a_t a_{t-j}] + \lambda_{al}(A \sum_{j=-\infty}^{\infty} E[a_t l_{t-j}] + \sum_{j=-\infty}^{\infty} E[a_t l_{t-j}] A') + \lambda_{ll} A \sum_{j=-\infty}^{\infty} E[l_t l_{t-j}] A'$, where

<table>
<thead>
<tr>
<th>Scheme</th>
<th>$\lambda_{al}$</th>
<th>$\lambda_{ll}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recursive</td>
<td>$1 - \pi^{-1} \ln(1 + \pi)$</td>
<td>$2 [1 - \pi^{-1} \ln(1 + \pi)]$</td>
</tr>
<tr>
<td>Rolling, $\pi \leq 1$</td>
<td>$\pi/2$</td>
<td>$\pi - \pi^2/3$</td>
</tr>
<tr>
<td>Rolling, $1 &lt; \pi &lt; \infty$</td>
<td>$1 - (2\pi)^{-1}$</td>
<td>$1 - (3\pi)^{-1}$</td>
</tr>
<tr>
<td>Fixed</td>
<td>$0$</td>
<td>$\pi$</td>
</tr>
</tbody>
</table>

and $A = E \left[ g'_\alpha(W_{t-1}, \theta_0) f_{W_{t-1}}(m_\alpha(W_{t-1}, \theta_0)) \right] \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} H(t - 1)$ and $l(Y_s, W_{s-1}, \theta_0)$ is the score function.
Independence Tests I: test statistics

To complement backtesting exercises one can be interested in testing the following

\[ \{I_{t,\alpha(\theta_0)}\}_{t=R+1}^n \text{ are iid.} \] (13)

Since for Bernoulli random variables serial independence is equivalent to serial uncorrelation, it is natural to build a test for (13) based on the autocovariances

\[ \gamma_j = Cov(I_{t,\alpha(\theta_0)}, I_{t-j,\alpha(\theta_0)}) \quad j \geq 1. \] (14)

They can be consistently estimated (under \( E[I_{t,\alpha(\theta_0)}] = \alpha \)) by either of these tests:

- (Joint Test)

\[ \gamma_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^n (I_{t,\alpha(\theta_0)} - \alpha)(I_{t-j,\alpha(\theta_0)} - \alpha) \quad j \geq 1. \]
• (Marginal Test)

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

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 (13) need to be based on estimates of the relevant parameters, such as

\[
\hat{\gamma}_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} \left( I_{t,\alpha}(\hat{\theta}_{t-1}) - \alpha \right) \left( I_{t-j,\alpha}(\hat{\theta}_{t-j-1}) - \alpha \right)
\]

or

\[
\hat{\zeta}_{P,j} = \frac{1}{\sqrt{P-j}} \sum_{t=R+j+1}^{n} \left( I_{t,\alpha}(\hat{\theta}_{t-1}) - E_n[I_{t,\alpha}(\hat{\theta}_{t-1})] \right) \left( I_{t-j,\alpha}(\hat{\theta}_{t-j-1}) - E_n[I_{t-j,\alpha}(\hat{\theta}_{t-j-1})] \right).
\]
Independence Tests II: Theorem

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

\[
\hat{\gamma}_{P,j} \xrightarrow{d} N(0, \sigma^2_b),
\]

where

\[
\sigma^2_b = E[b_t^2] + \lambda_{al}(B \sum_{j=-\infty}^{\infty} E[b_{t_l-t-j}] + \sum_{j=-\infty}^{\infty} E[b_{t_l-t-j}]'B') + \lambda_{ll}B \sum_{j=-\infty}^{\infty} E[l_{t_l-t-j}]B',
\]

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

\[
b_t = (I_{t,\alpha}(\theta) - \alpha)(I_{t-j,\alpha}(\theta) - \alpha).
\]

If instead of the joint hypothesis, only (13) holds, then

\[
\hat{\zeta}_{P,j} \xrightarrow{d} N(0, \sigma^2_c),
\]

where

\[
\sigma^2_c = E[c_t^2] + \lambda_{al}(C \sum_{j=-\infty}^{\infty} E[c_{t_l-t-j}] + \sum_{j=-\infty}^{\infty} E[c_{t_l-t-j}]'C') + \lambda_{ll}C \sum_{j=-\infty}^{\infty} E[l_{t_l-t-j}]C',
\]

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

\[
c_t = (I_{t,\alpha}(\theta) - E[I_{t,\alpha}(\theta)])(I_{t-j,\alpha}(\theta) - E[I_{t-j,\alpha}(\theta)]).
Subsampling approximation: Known parameters

Let \( T_P(\theta_0) = T_P(X_{R+1}, \ldots, X_n; \theta_0) \), be the relevant test statistic and \( G_P(w) \) the corresponding cumulative distribution function: \( G_P(w) = P(T_P(\theta_0) \leq w) \).

Let \( T_{b,i}(\theta_0) = T_b(X_i, \ldots, X_{i+b-1}; \theta_0), i = 1, \ldots, n-b+1 \), be the test statistic computed with the subsample \((X_i, \ldots, X_{i+b-1})\) of size \( b \).

**Subsampling:** 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

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

and let \( c_{P,1-\tau,b} \) be the \((1-\tau)\)-th sample quantile of \( G_{P,b}(w) \), i.e.,

\[
c_{P,1-\tau,b} = \inf\{w : G_{P,b}(w) \geq 1-\tau\}.
\]
Subsampling approximation: Theorem

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, $c_{P,1-\tau,b} \xrightarrow{P} c_{1-\tau}$ and

$$P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow \tau;$$

(ii) under the **marginal** null hypothesis of unconditional coverage probability,

$$c_{P,1-\tau,b} \xrightarrow{P} c_{1-\tau} \quad \text{and}$$

$$P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow \tau;$$

(iii) under any fixed alternative hypothesis,

$$P(T_P(\theta_0) > c_{P,1-\tau,b}) \longrightarrow 1.$$
Subsampling approximation: Unknown parameters

The subsampling distribution of the relevant test statistics, now $S_P$, $\hat{\gamma}_{P,j}$ or $\hat{\zeta}_{P,j}$, is

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

with $\hat{T}_{b,i}$ the subsample version of the test statistic, computed with data $X_i, \ldots, X_{i+b-1}$.

Under the presence of estimation risk we divide each subsample $X_i, \ldots, X_{i+b-1}$ in in-sample and out-of-sample observations according to the ratio $\pi$ used for computing the original test: $R_b$ observations are used for in-sample and $P_b$ observations for out-of-sample, such that $R_b + P_b = b$ and $\lim_{b \to \infty} P_b/R_b = \pi$.

The computation of $\hat{T}_{b,i}$ is done as with the original test, say $\hat{T}_P$, but replacing the original sample by $X_i, \ldots, X_{i+b-1}$ and $R$ and $P$ by $R_b$ and $P_b$. 
Theorem 4: Assume A1-A5 and that \( b/P \to 0 \) with \( b \to \infty \) as \( P \to \infty \). Moreover, assume that \( \lim_{b \to \infty} P_b/R_b = \lim_{n \to \infty} P/R = \pi \). Then, for \( \hat{T}_P \) any of the tests statistics \( S_P, \hat{\gamma}_{P,j} \) or \( \hat{\zeta}_{P,j} \), and, we have that

(i) under the joint null hypothesis,

\[
\begin{align*}
    c_{P,1-\tau,b}^* & \xrightarrow{P} c_{1-\tau}^* \quad \text{and} \\
    P(\hat{T}_P > c_{P,1-\tau,b}^*) & \to \tau;
\end{align*}
\]

(ii) under the marginal null hypothesis of unconditional coverage probability,

\[
\begin{align*}
    c_{P,1-\tau,b}^* & \xrightarrow{P} c_{1-\tau}^* \quad \text{and} \\
    P(\hat{T}_P > c_{P,1-\tau,b}^*) & \to \tau;
\end{align*}
\]

(iii) under any fixed alternative hypothesis,

\[
P(\hat{T}_P > c_{P,1-\tau,b}^*) \to 1.
\]
Monte Carlo Experiment: Unconditional Backtesting test

Data generating process: \( Y_t = \sigma_t \varepsilon_t \), with \( \varepsilon_t \sim N(0, 1) \equiv \Phi(\cdot) \), and

\[
\sigma_t^2 = \beta_0 + \beta_1 Y_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \tag{15}
\]

where \( \beta_0 = 0.05 \), \( \beta_1 = 0.10 \) and \( \beta_2 = 0.85 \).

**Model 1:** True conditional VaR process (Correct specification).

\[
m_{\alpha,1}(W_{t-1}, \beta) = \sigma_t \Phi^{-1}(\alpha), \tag{16}
\]

with \( \Phi^{-1}(\alpha) \) the \( \alpha \)-quantile of the standard Gaussian distribution function.

**Comments:**

- The asymptotic distribution of the unconditional coverage test is \( N(0, \alpha(1 - \alpha)) \).
- Subsampling and block bootstrap methods are in this case valid alternatives to approximate the finite-sample distribution of the test statistic.
Model 2: Estimated version of the true VaR process.

\[ m_{\alpha,2}(W_{t-1}, \hat{\beta}) = \hat{\sigma}_t \Phi^{-1}(\alpha), \]  

(17)

with \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) \), and \( \hat{\sigma}^2_t = \hat{\beta}_0 + \hat{\beta}_1 Y^2_{t-1} + \hat{\beta}_2 \hat{\sigma}^2_{t-1} \) the estimated parameters obtained from the in-sample size (R=1000 in-sample size to estimate the process, and P=500 out-of-sample evaluation period).

Problem: Estimation effects. The asymptotic distribution is Gaussian but the variance is different from \( \alpha(1 - \alpha) \). It depends on a nonlinear manner on the asymptotic ratio \( P/R \) and the values of the parameter vector \( \beta \).

Solutions: Subsampling and block bootstrap versions of the test account for estimation effects.
Model 3: Historical Simulation method.

\[ m_{\alpha,3}(W_{t-1}, \hat{\beta}) = \hat{F}_R^{-1}(\alpha), \quad (18) \]

with \( \hat{F}_R \) the empirical distribution function of the \( R = 1000 \) in-sample data.

**Problem:** This method can potentially exhibit very strong estimation and model risk effects. Therefore, the asymptotic critical values of the unconditional backtesting test is not \( N(0, \alpha(1 - \alpha)) \). The size of the test using these critical values will be very distorted under general situations.

**Comment:** This method is widely used by practitioners in the industry due to its simplicity.

**Solutions:** Subsampling and block bootstrap versions of the test account for estimation and model risk effects and are, in theory, able to approximate the appropriate nominal sizes.
Simulations of Size

<table>
<thead>
<tr>
<th>$\Phi(\cdot)$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P = 500$/size</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>$K_P^{asy,1}$</td>
<td>0.092 0.052 0.016</td>
<td>0.102 0.048 0.010</td>
<td>0.087 0.040 0.010</td>
</tr>
<tr>
<td>$K_P^{bb,1}$</td>
<td>0.077 0.041 0.021</td>
<td>0.078 0.039 0.018</td>
<td>0.071 0.040 0.05</td>
</tr>
<tr>
<td>$K_P^{ss,1}$</td>
<td>0.060 0.032 0.020</td>
<td>0.038 0.026 0.014</td>
<td>0.1633 0.157 0.151</td>
</tr>
<tr>
<td>$P = 500$/size</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>$S_P^{asy,2}$</td>
<td>0.137 0.086 0.014</td>
<td>0.123 0.087 0.028</td>
<td>0.147 0.069 0.026</td>
</tr>
<tr>
<td>$S_P^{bb,2}$</td>
<td>0.085 0.032 0.002</td>
<td>0.077 0.022 0.002</td>
<td>0.046 0.020 0.002</td>
</tr>
<tr>
<td>$S_P^{ss,2}$</td>
<td>0.090 0.052 0.044</td>
<td>0.113 0.100 0.100</td>
<td>0.252 0.248 0.247</td>
</tr>
<tr>
<td>$P = 500$/size</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>$S_P^{asy,3}$</td>
<td>0.397 0.327 0.202</td>
<td>0.419 0.375 0.221</td>
<td>0.383 0.250 0.149</td>
</tr>
<tr>
<td>$S_P^{bb,3}$</td>
<td>0.195 0.119 0.044</td>
<td>0.181 0.105 0.030</td>
<td>0.119 0.060 0.020</td>
</tr>
<tr>
<td>$S_P^{ss,3}$</td>
<td>0.254 0.224 0.197</td>
<td>0.316 0.282 0.268</td>
<td>0.453 0.423 0.407</td>
</tr>
</tbody>
</table>

Table 1. Unconditional backtesting test for models (16), (17) and (18). $R = 1000$ and $P = 500$. Coverage probability $\alpha = 0.10, 0.05, 0.01$. $b = \left\lfloor kP^{2/5} \right\rfloor$ with $k = 8$. 500 Monte-Carlo replications.
Monte Carlo Experiment: Independence test

Data generating process: \( Y_t = \sigma_t \varepsilon_t \), with \( \varepsilon_t \sim \Phi(\cdot) \), and \( \sigma_t^2 = \beta_0 + \beta_1 Y_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \), where \( \beta_0 = 0.05 \), \( \beta_1 = 0.10 \) and \( \beta_2 = 0.85 \).

Note that

- \( \gamma_{P,j} \) is a joint test (unconditional and independence). If \( E[I_{t,\alpha}(\theta_0)] = \alpha \) this test does not exhibit model neither estimation risk, its asymptotic distribution is \( N(0, \alpha^2(1-\alpha)^2) \).

- \( \xi_{P,j} \), does not exhibit model risk, its asymptotic distribution is \( N(0, \alpha^2(1-\alpha)^2) \).

- The estimated version of these test statistics, \( \hat{\gamma}_{P,j} \) and \( \hat{\xi}_{P,j} \), however exhibit estimation risk.

  **Solution:** subsampling and block bootstrap methods.

In the simulations below we only study \( \gamma_{P,1} \) and \( \hat{\xi}_{P,1} \) in terms of size.
### Simulations of Size

<table>
<thead>
<tr>
<th>$\Phi(\cdot)$</th>
<th>( \alpha = 0.10 )</th>
<th>( \alpha = 0.05 )</th>
<th>( \alpha = 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P = 500/\text{size}$</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>$\gamma_{asy,1}^{P,1}$</td>
<td>0.072 0.040 0.016</td>
<td>0.070 0.036 0.011</td>
<td>0.133 0.105 0.000</td>
</tr>
<tr>
<td>$\gamma_{bb,1}^{P,1}$</td>
<td>0.042 0.019 0.004</td>
<td>0.042 0.008 0.000</td>
<td>0.041 0.019 0.004</td>
</tr>
<tr>
<td>$\gamma_{ss,1}^{P,1}$</td>
<td>0.053 0.030 0.010</td>
<td>0.093 0.072 0.040</td>
<td>0.306 0.190 0.107</td>
</tr>
<tr>
<td>$P = 500/\text{size}$</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
<td>0.10 0.05 0.01</td>
</tr>
<tr>
<td>$\xi_{asy,2}^{P,1}$</td>
<td>0.121 0.061 0.019</td>
<td>0.115 0.047 0.016</td>
<td>0.051 0.051 0.051</td>
</tr>
<tr>
<td>$\xi_{bb,2}^{P,1}$</td>
<td>0.070 0.025 0.004</td>
<td>0.068 0.023 0.002</td>
<td>0.051 0.023 0.004</td>
</tr>
<tr>
<td>$\xi_{ss,2}^{P,1}$</td>
<td>0.085 0.025 0.002</td>
<td>0.049 0.011 0.002</td>
<td>0.053 0.028 0.004</td>
</tr>
</tbody>
</table>

**Table 2.** Empirical size for independence backtesting test for models (16) and (17). $R = 1000$ and $P = 500$. Coverage probability $\alpha = 0.10, 0.05, 0.01$. $b = \left\lfloor kP^{2/5} \right\rfloor$ with $k = 8$. 500 Monte-Carlo replications.
Application: Backtesting performance of VaR forecast models

The models under study are

- Hybrid method:

  \[ m_{\alpha,1}(W_{t-1}, \hat{\theta}_{t-1}) = \hat{\sigma}_t F_{R,\hat{\varepsilon}}^{-1}(\alpha) \]

  (19)

  with \( F_{R,\hat{\varepsilon}} \) the empirical distribution function constructed from an in-sample size of \( R \) observations, and \( \hat{\sigma}_t^2 = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1}^2 + \hat{\beta}_2 \hat{\sigma}_{t-1}^2 \), the estimated GARCH process, where \((\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)\) is the vector of estimates of the parameter vector \((\beta_0, \beta_1, \beta_2)\) and \(\{\hat{\varepsilon}_t\}\) is the residual sequence.
- Gaussian GARCH model:

\[ m_{\alpha,2}(W_{t-1}, \hat{\theta}_{t-1}) = \hat{\sigma}_t \Phi^{-1}(\alpha). \] (20)

- Student-t GARCH model:

\[ m_{\alpha,3}(W_{t-1}, \hat{\theta}_{t-1}) = \hat{\sigma}_t t_{\hat{\nu}}^{-1}(\alpha). \] (21)

- Riskmetrics model:

\[ m_{\alpha,4}(W_{t-1}, \hat{\theta}_{t-1}) = \hat{\sigma}_t Rm \Phi^{-1}(\alpha), \] (22)

with \( (\hat{\sigma}_t Rm)^2 = \beta_1 Y_{t-1}^2 + (1 - \beta_1)(\hat{\sigma}_{t-1})^2. \)
Design of Backtesting Exercise

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.

Use 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.
Unconditional test: Hybrid Method

Backtesting test $S_n$ at 5% for HS-GARCH, $k=8$. 

- $S_n$: Subsampling
- Asymptotic

Out-of-sample rolling period

Far East and South Asia Meeting of the Econometric Society 2009
Unconditional test: Gaussian GARCH(1,1) Method
Unconditional test: Student-t GARCH(1,1) Method
Unconditional test: Riskmetrics Method

Backtesting test $S_n$ at 5% for Riskmetrics. $k=8$. 

Far East and South Asia Meeting of the Econometric Society 2009
Comments on the plots

• 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.

• Therefore, whereas we reject the VaR forecasting methods obtained from using GARCH models we do not do it if Riskmetrics is employed.

• 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.

• We observe an increase in the uncertainty in all of the unconditional backtesting tests that use subsampling.
Independence test: Hybrid Method

Dependence backtesting test $\hat{\xi}_n$ at 5% for HS–GARCH, $k=8$.  

- Subsampling
- Asymptotic
Independence test: Gaussian GARCH(1,1) Method

Graph showing dependence backtesting test $\hat{\xi}_n$ at 5% for Gaussian-GARCH. $k=8$. Subsampling and asymptotic results are plotted.
Independence test: Student-t GARCH(1,1) Method

![Graph showing dependence backtesting test \(\hat{\xi}_n\) at 5% for Student-GARCH, \(k=8\).]
Independence test: Riskmetrics Method

![Graph showing dependence backtesting test \( \hat{\xi}_n \) at 5% for Riskmetrics.](image)

- **Out-of-sample rolling period**
- **Dependence backtesting test** \( \hat{\xi}_n \) at 5% for Riskmetrics. \( k=8 \).

- **Subsampling**
- **Asymptotic**
Conclusions

• Backtesting techniques are 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.

• 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.

• In order to implement correctly the standard backtesting procedures one needs to incorporate in the asymptotic theory certain components accounting for possible mis-specifications of the risk model. A feasible alternative to the complicated asymptotic theory is resampling methods: block bootstrap and subsampling.
• Our simulations indicate that values of $b = \left\lfloor k P^{2/5} \right\rfloor$ determined by $k > 5$ provide valid approximations of the true sampling distributions and hence of the true rejection regions for backtesting tests.

• 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.