Backtesting Value-at-Risk Models

Christophe Hurlin

University of Orléans

Séminaire Validation des Modèles Financiers. 29 Avril 2013
The Value-at-Risk (VaR) and more generally the Distortion Risk Measures (Expected Shortfall, etc.) are standard risk measures used in the current regulations introduced in Finance (Basel 2), or Insurance (Solvency 2) to fix the required capital (Pillar 1), or to monitor the risk by means of internal risk models (Pillar 2).
Let $\{r_t\}_{t=1}^T$ be a given P&L series. The daily (conditional) VaR for a nominal coverage rate $\alpha$ is defined as

$$\Pr[r_t < -\text{VaR}_{t|t-1}(\alpha) \mid \mathcal{F}_{t-1}] = \alpha$$

where $\mathcal{F}_{t-1}$ denotes the set of information available at time $t - 1$. 
<table>
<thead>
<tr>
<th>Who does use VaR?</th>
<th>What for?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank risk manager</td>
<td>Measure firm-level market, credit, op. risk</td>
</tr>
<tr>
<td>Bank executives</td>
<td>Set limits (management)</td>
</tr>
<tr>
<td>Banking regulators</td>
<td>Determine capital requirements</td>
</tr>
<tr>
<td>Exchanges</td>
<td>Compute margins</td>
</tr>
<tr>
<td>Regulators</td>
<td>Forecast systemic risk (CoVaR)</td>
</tr>
<tr>
<td>Industry</td>
<td>Ex: EDF, spot prices of electricity</td>
</tr>
</tbody>
</table>
"Disclosure of quantitative measures of market risk, such as value-at-risk, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance", Alan Greenspan (1996)
Definition

*Backtesting is a set of statistical procedures designed to check if the real losses are in line with VaR forecasts (Jorion, 2007).*
Whatever the type of use of VaR, the VaR forecasts are generated by an **internal risk model**.

This model is used to produce a sequence of pseudo out-of-sample VaR forecasts for a past period (typically one year).

The backtesting is based on the comparison of the observed P&L to these VaR forecasts.
Outlines

1. How to test the validity of a VaR model?
2. What are the backtesting strategies?
3. What are the good practices?
Backtesting Principles
Remark 1: Ex-post VaR is not observable, so it is impossible to compute traditional statistics or criteria such as MSFE.

Remark 2: There is no proxy for the VaR contrary to the volatility (realized volatility, Andersen and Bollerslev 1998)

Backtesting procedures are based on VaR exceptions

Definition
We denote $I_t(\alpha)$ the hit variable associated to the ex-post observation of an $\alpha\%$ VaR exception at time $t$:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } r_t < -\text{VaR}_{t|t-1}(\alpha) \\ 0 & \text{else} \end{cases}$$
Christoffersen (1998) : VaR forecasts are valid if and only if the violation process $I_t (\alpha)$ satisfies the following two assumptions:

1. The unconditional coverage (UC) hypothesis.
2. The independence (IND) hypothesis.

**Definition (unconditional coverage hypothesis)**

The unconditional probability of a violation must be equal to the \( \alpha \) coverage rate

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

- If \( \Pr [I_t(\alpha) = 1] > \alpha \), the risk is under-estimated
- If \( \Pr [I_t(\alpha) = 1] < \alpha \), the risk is over-estimated
**Definition (independence hypothesis)**

*VaR violations observed at two different dates must be independently distributed.*

\[ I_t(\alpha) \text{ and } I_s(\alpha) \text{ are independently distributed for } t \neq s \]
Figure: Illustration: violations’ cluster

\begin{figure}
\centering
\includegraphics[width=\textwidth]{violations_cluster}
\caption{Illustration: violations’ cluster}
\end{figure}
Introduction
Backtesting Principles
Testing strategies
Recommendations

Backtesting Principles

Figure: Illustration: violations' cluster

![Graph showing VaR(95%) and P&L over time. The graph indicates a cluster of violations.]
Definition (conditional coverage hypothesis)

The violation process satisfies a difference martingale assumption.

\[ \mathbb{E} [ l_t(\alpha) \mid \mathcal{F}_{t-1}] = \alpha \]
Remark: These assumptions can be expressed as distributional assumptions.

Under the UC assumption, each variable $I_t(\alpha)$ has a Bernoulli distribution with a probability $\alpha$.

$$I_{tt}(\alpha) \sim Bernoulli(\alpha)$$

Under the IND assumption, these variables are independent, and the number of violations has a Binomial distribution.

$$\sum_{t=1}^{T} I_t(\alpha) \sim B(T, \alpha)$$
Testing strategies

What are the backtesting strategies?
Let us consider a sequence of daily VaR out-of-sample forecasts \( \{ VaR_{t|t-1}(\alpha) \}^T_{t=1} \) and the corresponding observed P&L.

How to test the validity of the internal risk model?

Testing strategies:

1. Frequency-based tests
2. Magnitude-based tests
3. Multivariate tests
4. Independence tests
5. Duration-based tests
### Testing strategies: frequency-based tests (1/5)

**Figure:** BIS "Traffic Light" System

<table>
<thead>
<tr>
<th>Zone</th>
<th>Number of exceptions</th>
<th>Increase in scaling factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>Yellow Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.40</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0.65</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.75</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0.85</td>
</tr>
<tr>
<td>Red Zone</td>
<td>10 or more</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: VaR(1%, 1 day), 250 daily observations
Christoffersen (1998) proposes a Likelihood Ratio statistic for UC defined as:

\[
LR_{UC} = -2 \ln \left[ \left(1 - \alpha \right)^{T-H} \alpha^H \right] \\
+ 2 \ln \left[ \left(1 - H/T \right)^{T-H} \left(H/T \right)^H \right] \xrightarrow{T \to \infty} \chi^2 (1)
\]

where \( H = \sum_{t=1}^{T} I_t (\alpha) \) denotes the total number of exceedances.

- For a nominal risk of 5%, the null of UC can not be rejected if and only if \( H < 7 \) for \( T = 250 \) and \( \alpha = 1\% \).
Testing strategies: frequency-based tests (1/5)

Example

Berkowitz and O-Brien (2002) consider the VaR forecasts of six US commercial banks

Testing strategies: frequency-based tests (1/5)

Figure: Bank Daily VaR Models
Testing strategies: frequency-based tests (1/5)

**Figure:** Violations of Banks’ 99% VaR
### Testing strategies: frequency-based tests (1/5)

Table 4. Backtests of Large-Scale VaR Models

<table>
<thead>
<tr>
<th>Bank</th>
<th>Violation Rate</th>
<th>Coverage</th>
<th>Conditional Coverage</th>
<th>Independence</th>
<th>Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank 1</td>
<td>0.005</td>
<td>1.54</td>
<td>1.57</td>
<td>.0321</td>
<td>-.00533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.214]</td>
<td>[.455]</td>
<td>[.858]</td>
<td>[.885]</td>
</tr>
<tr>
<td>Bank 2</td>
<td>0.010</td>
<td>.00693</td>
<td>4.01</td>
<td>4.00*</td>
<td>.158*</td>
</tr>
<tr>
<td>Bank 3</td>
<td>0.005</td>
<td>1.70</td>
<td>8.81*</td>
<td>7.11**</td>
<td>.330**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.193]</td>
<td>[.012]</td>
<td>[.008]</td>
<td>[.001]</td>
</tr>
<tr>
<td>Bank 4</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank 5</td>
<td>0.001</td>
<td>8.92**</td>
<td>8.93*</td>
<td>.00271</td>
<td>-.00137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.003]</td>
<td>[.012]</td>
<td>[.959]</td>
<td>[.995]</td>
</tr>
<tr>
<td>Bank 6</td>
<td>0.005</td>
<td>1.71</td>
<td>1.74</td>
<td>.0312</td>
<td>-.00511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.191]</td>
<td>[.419]</td>
<td>[.860]</td>
<td>[.885]</td>
</tr>
</tbody>
</table>
## Testing strategies: frequency-based tests (1/5)

Table 4. Backtests of Large-Scale VaR Models

<table>
<thead>
<tr>
<th></th>
<th>Violation Rate</th>
<th>Coverage</th>
<th>Conditional Coverage</th>
<th>Independence</th>
<th>Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank 1</td>
<td>0.005</td>
<td>1.54</td>
<td>1.57</td>
<td>.0321</td>
<td>-.00533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.214]</td>
<td>[.455]</td>
<td>[.858]</td>
<td>[.885]</td>
</tr>
<tr>
<td>Bank 2</td>
<td>0.010</td>
<td>.00693</td>
<td>4.01</td>
<td>4.00*</td>
<td>.158*</td>
</tr>
<tr>
<td>Bank 3</td>
<td>0.005</td>
<td>1.70</td>
<td>8.81*</td>
<td>7.11**</td>
<td>.330**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.193]</td>
<td>[.012]</td>
<td>[.008]</td>
<td>[.001]</td>
</tr>
<tr>
<td>Bank 4</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bank 5</td>
<td>0.001</td>
<td>8.92**</td>
<td>8.93*</td>
<td>.00271</td>
<td>-.00137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.003]</td>
<td>[.012]</td>
<td>[.959]</td>
<td>[.995]</td>
</tr>
<tr>
<td>Bank 6</td>
<td>0.005</td>
<td>1.71</td>
<td>1.74</td>
<td>.0312</td>
<td>-.00511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.191]</td>
<td>[.419]</td>
<td>[.860]</td>
<td>[.885]</td>
</tr>
</tbody>
</table>

*Significant at 5% level
**Significant at 1% level

---

This table presents the backtests of large-scale VaR models for five different banks. The table includes the violation rate, coverage, conditional coverage, independence, and serial correlation. The results show varying degrees of model performance across the different banks, with Bank 5 and Bank 6 having particularly strong performance in terms of violation rates and coverage.
Testing strategies:

1. Frequency-based tests
2. Magnitude-based tests
3. Multivariate tests
4. Independence tests
5. Duration-based tests
Testing strategies: magnitude-based tests (2/5)

All these tests do not take into account the **magnitude of the losses beyond the VaR**

**Example**

Consider two banks that both have a one-day 1%-VaR of $100 million. Assume each bank reports three VaR exceptions, but the average VaR exceedance is $1 million for bank A and $500 million for bank B.

In this case, **standard backtesting methodologies** would indicate that the **performance of both models is equal and acceptable**.
Testing strategies: magnitude-based tests (2/5)

Figure: Daily VaR and P/L for SocGen 2008
Testing strategies: magnitude-based tests (2/5)

Figure: Daily VaR and P/L for SocGen 2008
Testing strategies: magnitude-based tests (2/5)

The Risk Map

We propose a VaR backtesting methodology based on the number and the severity of VaR exceptions: this approach exploits the concept of "super exception".

**Definition**

We define a super exception using a VaR with a much smaller coverage probability $\alpha'$, with $\alpha' < \alpha$. In this case, a super exception is defined as a loss greater than $\text{VaR}_t(\alpha')$. 

---

Christophe Hurlin | Backtesting
Testing strategies: magnitude-based tests (2/5)

Figure: VaR Exception vs. VaR Super Exception
Solution

Given VaR exceptions $I_t(\alpha)$ and VaR super exception $I_t(\alpha')$, we define a Risk Map that jointly accounts for the number and the magnitude of the VaR exceptions.

Let us consider a given UC test with a statistic $Z(\alpha)$ based on the violations sequence $\{I_t(\alpha)\}_{t=1}^T$.

$$H_0 : \mathbb{E}[I_t(\alpha)] = \alpha$$

$$H_1 : \mathbb{E}[I_t(\alpha)] \neq \alpha.$$
Testing strategies: magnitude-based tests (2/5)

Number of VaR Exceptions ($N$)

Non-rejection area for test on VaR exceptions
Based on the same UC test, it is possible to test for the **magnitude of VaR exceptions**, via the **VaR super exceptions** \( \{ l_t(\alpha') \} \)_{t=1}^{T}:

\[
H_0 : \mathbb{E} \left[ l_t(\alpha') \right] = \alpha' \\
H_1 : \mathbb{E} \left[ l_t(\alpha') \right] \neq \alpha'
\]
Testing strategies: magnitude-based tests (2/5)

Non-rejection area for test on VaR super exceptions

Number of VaR Super Exceptions (N’)

Christophe Hurlin
Backtesting
We can also **test jointly** for both magnitude and frequency of VaR exceptions:

\[ H_0 : \mathbb{E} [ l_t (\alpha) ] = \alpha \text{ and } \mathbb{E} [ l_t (\alpha') ] = \alpha' \]

*Multivariate approach*

Testing strategies: magnitude-based tests (2/5)

![Diagram showing number of VaR Super Exceptions (N') vs. number of VaR Exceptions (N) for nominal risk 5% and 1%]

- Frequency-based tests
- Magnitude-based tests
- Multivariate tests
- Independence tests
- Duration-based tests

Christophe Hurlin
Backtesting
Testing strategies: magnitude-based tests (2/5)

Figure: Backtesting Bank VaR: La Caixa (2007-2008)
Testing strategies: magnitude-based tests (2/5)
Testing strategies:

1. Frequency-based tests
2. Magnitude-based tests
3. **Multivariate tests**
4. Independence tests
5. Duration-based tests
Testing strategies: multivariate tests (3/5)

**Intuition:** Testing the validity of the VaR model for $M$ coverage rates, with $M > 1$.


Perignon and Smith (2008) consider the null:

\[ H_{0,MUC} : \mathbb{E} \left[ I_t (\alpha) \right] = \alpha \quad \text{and} \quad \mathbb{E} \left[ I_t (\alpha') \right] = \alpha'. \]

Let us denote:

\[
\begin{align*}
J_{0,t} &= 1 - J_{1,t} - J_{2,t} \\
J_{1,t} &= \begin{cases} 
1 & \text{if } -\text{VaR}_{t|t-1}(\alpha') < r_t < -\text{VaR}_{t|t-1}(\alpha) \\
0 & \text{otherwise}
\end{cases} \\
J_{2,t} &= \begin{cases} 
1 & \text{if } r_t < -\text{VaR}_{t|t-1}(\alpha') \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]
Definition (Perignon and Smith, 2008)

The multivariate unconditional coverage test is a LR test given by:

\[
LR_{MUC} = -2 \ln \left[ (1 - \alpha)^{H_0} (\alpha - \alpha')^{H_1} (\alpha')^{H_2} \right] + 2 \ln \left[ \left(1 - \frac{H_0}{T}\right)^{H_0} \left(\frac{H_0}{T} - \frac{H_1}{T}\right)^{H_1} \left(\frac{H_2}{T}\right)^{H_2} \right].
\]

where \( H_i = \sum_{t=1}^{T} J_{i,t} \), for \( i = 0, 1, 2 \), denote the count variable associated with each of the Bernoulli variables.
Hurlin and Tokpavi (2006):

A natural test of the CC is the univariate Ljung-Box test of

\[ H_{0,CC} : r_1 = \ldots = r_K = 0 \]

where \( r_k \) denotes the \( k^{th} \) autocorrelation:

\[
LB(K) = T(T + 2) \sum_{k=1}^{K} \frac{\hat{r}_k^2}{T - k} \xrightarrow{d} \chi^2(K)
\]
Testing strategies: multivariate tests (3/5)

Definition (Hurlin and Tokpavi, 2006)

Let $\Theta = \{\theta_1, \ldots, \theta_m\}$ be a discrete set of $m$ different coverage rates and $Hit_t = [Hit_t(\theta_1) : Hit_t(\theta_2) : \ldots : Hit_t(\theta_m)]'$

$$Hit_t(\theta_i) = \begin{cases} 1 - \theta_i & \text{if } r_t < -\text{VaR}_{t|t-1}(\theta_i) \\ -\theta_i & \text{else} \end{cases}$$

Under the null of CC (martingale difference):

$$H_{0, CC} : E \left( [Hit_t Hit'_{t-k}] \right) = 0_m \quad \forall k = 1, \ldots, K$$
Testing strategies:

1. Frequency-based tests
2. Magnitude-based tests
3. Multivariate tests
4. Independence tests
5. Duration-based tests
Testing strategies: independence tests (4/5)

**LR tests**

Christoffersen (1998) assumes that the violation process \( l_t(\alpha) \) can be represented as a Markov chain with two states:

\[
\Pi = \begin{pmatrix}
1 - \pi_{01} & \pi_{01} \\
1 - \pi_{11} & \pi_{11}
\end{pmatrix}
\]

\[
\pi_{ij} = \Pr [ l_t(\alpha) = j \mid l_{t-1}(\alpha) = i ]
\]

**Definition**

The null of CC can be defined as follows:

\[
H_{0,CC} : \, \Pi = \Pi_\alpha = \begin{pmatrix}
1 - \alpha & \alpha \\
1 - \alpha & \alpha
\end{pmatrix}
\]
LR tests

Christoffersen (1998) assumes that the violation process $l_t(\alpha)$ can be represented as a Markov chain with two states:

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

$$\pi_{ij} = \Pr [ l_t(\alpha) = j \mid l_{t-1}(\alpha) = i ]$$

**Definition**

The null of IND can be defined as follows:

$$H_{0,IND} : \Pi = \Pi_\beta = \begin{pmatrix} 1 - \beta & \beta \\ 1 - \beta & \beta \end{pmatrix}$$
Testing strategies: frequency-based tests (4/5)

The corresponding LR statistics are defined by:

\[ LR_{IND} = -2 \ln \left( 1 - \frac{H}{T} \right)^{T-H} \left( \frac{H}{T} \right)^{H} \]
\[ + 2 \ln \left[ \left( 1 - \hat{\pi}_{01} \right)^{n_{00}} \hat{\pi}_{01} \left( 1 - \hat{\pi}_{11} \right)^{n_{10}} \hat{\pi}_{11} \right] \overset{d}{\underset{T \to \infty}{\rightarrow}} \chi^2 (1) \]

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

By definition:

\[ LR_{CC} = LR_{UC} + LR_{IND} \]
### Testing strategies: independence tests (4/5)

Table 4. Backtests of Large-Scale VaR Models

<table>
<thead>
<tr>
<th>Bank</th>
<th>Violation Rate</th>
<th>Coverage</th>
<th>Conditional Coverage</th>
<th>Independence</th>
<th>Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank 1</td>
<td>0.005</td>
<td>1.54</td>
<td>1.57</td>
<td>.0321</td>
<td>-.00533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.214]</td>
<td>[.455]</td>
<td>[.858]</td>
<td>[.885]</td>
</tr>
<tr>
<td>Bank 2</td>
<td>0.010</td>
<td>.00693</td>
<td>4.01</td>
<td>4.00*</td>
<td>.158*</td>
</tr>
<tr>
<td>Bank 3</td>
<td>0.005</td>
<td>1.70</td>
<td>8.81*</td>
<td>7.11**</td>
<td>.330**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.193]</td>
<td>[.012]</td>
<td>[.008]</td>
<td>[.001]</td>
</tr>
<tr>
<td>Bank 4</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bank 5</td>
<td>0.001</td>
<td>8.92**</td>
<td>8.93*</td>
<td>.00271</td>
<td>-.00137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.003]</td>
<td>[.012]</td>
<td>[.959]</td>
<td>[.995]</td>
</tr>
<tr>
<td>Bank 6</td>
<td>0.005</td>
<td>1.71</td>
<td>1.74</td>
<td>.0312</td>
<td>-.00511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.191]</td>
<td>[.419]</td>
<td>[.860]</td>
<td>[.885]</td>
</tr>
</tbody>
</table>
## Testing strategies: independence tests (4/5)

Table 4. Backtests of Large-Scale VaR Models

<table>
<thead>
<tr>
<th>Bank</th>
<th>Violation Rate</th>
<th>Coverage</th>
<th>Conditional Coverage</th>
<th>Independence</th>
<th>Serial Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank 1</td>
<td>0.005</td>
<td>1.54</td>
<td>1.57</td>
<td>0.0321</td>
<td>-0.00533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.214]</td>
<td>[.455]</td>
<td>[.858]</td>
<td>[.885]</td>
</tr>
<tr>
<td>Bank 2</td>
<td>0.010</td>
<td>.00693</td>
<td>4.01</td>
<td>4.00*</td>
<td>.158*</td>
</tr>
<tr>
<td>Bank 3</td>
<td>0.005</td>
<td>1.70</td>
<td>8.81*</td>
<td>7.11**</td>
<td>.330**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.193]</td>
<td>[.012]</td>
<td>[.008]</td>
<td>[.001]</td>
</tr>
<tr>
<td>Bank 4</td>
<td>0.000</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Bank 5</td>
<td>0.001</td>
<td>8.92**</td>
<td>8.93*</td>
<td>0.00271</td>
<td>-0.00137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.003]</td>
<td>[.012]</td>
<td>[.959]</td>
<td>[.995]</td>
</tr>
<tr>
<td>Bank 6</td>
<td>0.005</td>
<td>1.71</td>
<td>1.74</td>
<td>0.0312</td>
<td>-0.00511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[.191]</td>
<td>[.419]</td>
<td>[.860]</td>
<td>[.885]</td>
</tr>
</tbody>
</table>
Testing strategies: frequency-based tests (1/5)

Figure: Violations of Banks’ 99% VaR
Testing strategies: independence tests (4/5)

Regression based tests

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

$$
\text{Hit}_t(\alpha) = \begin{cases} 
1 - \alpha & \text{if } r_t < -\text{VaR}_{t|t-1}(\alpha) \\
-\alpha & \text{otherwise}
\end{cases}
$$
Regression based tests

Consider the following linear regression model:

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

where the $z_{t-k}$ variables belong to the information set $\Omega_{t-1}$ (lagged P&L, squared past P&L, past margins, etc.)
Regression based tests

The null hypothesis test of CC corresponds to testing the joint nullity of all the regression coefficients:

\[ H_{0, CC} : \delta = \beta_k = \gamma_k = 0, \quad \forall k = 1, \ldots, K. \]

since under the null:

\[ \mathbb{E} [Hit_t (\alpha)] = \mathbb{E} [l_t (\alpha) - \alpha] = 0 \iff \Pr [l_t (\alpha) = 1] = \alpha \]
Testing strategies: independence tests (4/5)

Definition (Engle and Manganelli, 2004)

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

$$DQ_{CC} = \frac{\hat{\Psi}' Z' Z \hat{\Psi}}{\alpha (1 - \alpha)} \xrightarrow{\mathcal{D}} \chi^2 (2K + 1)$$

where $\hat{\Psi}$ is the OLS estimate of $\Psi$. 
Testing strategies: independence tests (4/5)

Regression based tests

Extension: A natural extension of the test of Engle and Manganelli (2004) consists in considering a (probit or logit) binary model linking current violations to past ones.

Dumitrescu E., Hurlin C. and Pham V. (2012), Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests, Finance
Definition (Dumitrescu et al., 2012)

We consider a dichotomic model:

\[
\Pr \left[ I_t(\alpha) = 1 \mid F_{t-1} \right] = F(\pi_t).
\]

where \( F(.) \) denotes a c.d.f. and the index \( \pi_t \) satisfies the following autoregressive representation:

\[
\pi_t = c + \sum_{j=1}^{q_1} \beta_j \pi_{t-j} + \sum_{j=1}^{q_2} \delta_j I_{t-j}(\alpha) + \sum_{j=1}^{q_3} \gamma_j x_{t-j},
\]

where \( I(.) \) is a function of a finite number of lagged values of observables, and \( x_t \) is a vector of explicative variables.
Testing strategies: independence tests (4/5)

Regression based tests

\[ H_0 : \beta = 0, \delta = 0, \gamma = 0 \text{ and } c = F^{-1}(\alpha). \]

since under the null of CC:

\[ \Pr(I_t = 1 \mid \mathcal{F}_{t-1}) = F(F^{-1}(\alpha)) = \alpha. \]

The Dynamic Binary (DB) LR test statistic is:

\[
DB_{LR_{CC}} = -2 \left\{ \ln L(0, F^{-1}(\alpha); I_t(\alpha), Z_t) - \ln L(\hat{\theta}, \hat{c}; I_t(\alpha), Z_t) \right\}
\]

\[
\xrightarrow{d \ T \to \infty} \chi^2 \left( \dim(Z_t) \right)
\]
Testing strategies:

1. Frequency-based tests
2. Magnitude-based tests
3. Multivariate tests
4. Independence tests
5. Duration-based tests
The UC, IND, and CC hypotheses also have some implications on the time between two consecutive VaR margin exceedances.

Let us denote by $d_v$ the duration between two consecutive VaR margin violations:

$$d_v = t_v - t_{v-1}$$

where $t_v$ denotes the date of the $v^{th}$ exceedance.
Under CC hypothesis, the duration process $d_v$ has a geometric distribution:

$$\Pr [d_v = k] = \alpha (1 - \alpha)^{k-1} \quad k \in \mathbb{N}^*.$$ 

This distribution characterizes the memory-free property of the violation process $l_t(\alpha)$ with $\mathbb{E} (d_v) = 1/\alpha$. 
Christoffersen and Pelletier (2004) use under the null hypothesis the exponential distribution:

$$g(d_v; \alpha) = \alpha \exp(-\alpha d_v).$$

Under the alternative hypothesis, they postulate a Weibull distribution for the duration variable:

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

$$H_{0,IND} : b = 1 \quad H_{0,CC} : b = 1, \quad a = \alpha$$
Testing strategies: duration-based tests (5/5)

**Drawback:** we have to postulate a distribution for the duration under the alternative (misspecification of the VaR model).

**Solution:** Candelon et al. (2001) propose a J-test based on orthonormal polynomials associated to the geometric distribution.

Candelon et al. (2001)

In the case of continuous distributions, the Pearson distributions (Normal, Student, etc.) are associated to some particular orthonormal polynomials whose expectation is equal to zero.

- These polynomials can be used as special moments to test for a distributional assumption (see. Bontemps and Meddahi, *Journal of Econometrics*, 2005).

- In the discrete case, orthonormal polynomials are defined for distributions belonging to the Ord’s family (Poisson, Pascal, hypergeometric, etc.).
Testing strategies: duration-based tests (5/5)

Candelon et al. (2001)

**Definition**

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

\[ M_{j+1} (d; \beta) = \frac{(1 - \beta) (2j + 1) + \beta (j - d + 1)}{(j + 1) \sqrt{1 - \beta}} M_j (d; \beta) \]

\[ - \left( \frac{j}{j + 1} \right) M_{j-1} (d; \beta), \]

for any order $j \in \mathbb{N}$, with $M_{-1} (d; \beta) = 0$ and $M_0 (d; \beta) = 1$ and:

\[ \mathbb{E} [M_j (d; \beta)] = 0 \quad \forall j \in \mathbb{N}^*, \forall d \in \mathbb{N}^*. \]
Candelon et al. (2001)

Example

We can show that if $d$ follows a geometric distribution of parameter $\beta$, then:

$$M_1 (d; \beta) = (1 - \beta d) / \sqrt{1 - \beta}$$

with

$$\mathbb{E} [M_1 (d; \beta)] = 0$$
Testing strategies: duration-based tests (5/5)

Candelon et al. (2001)

Our duration-based backtest procedure exploits these moment conditions.

- More precisely, let us define \( \{d_1; \ldots; d_N\} \) a sequence of \( N \) durations between VaR violations, computed from the sequence of the hit variables \( \{I_t(\alpha)\}_{t=1}^T \).

- Under the CC assumption, the durations \( d_i, i = 1, \ldots, N \), are i.i.d. geometric(\( \alpha \)). Hence, the null of CC can be expressed as follows:

\[
H_{0,CC} : \mathbb{E}[M_j(d_i; \alpha)] = 0, \quad j = \{1, \ldots, p\},
\]

where \( p \) denotes the number of moment conditions.
Testing strategies: duration-based tests (5/5)

Candelon et al. (2001)

**Definition**

The null hypothesis of CC can be expressed as

\[ H_{0, CC} : E \left[ M (d_i; \alpha) \right] = 0, \]

where \( M (d_i; \alpha) \) denotes a \((p, 1)\) vector whose components are the orthonormal polynomials \( M_j (d_i; \alpha) \), for \( j = 1, \ldots, p \). Under some regularity conditions:

\[
J_{CC} (p) = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M (d_i; \alpha) \right)^T \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M (d_i; \alpha) \right) \xrightarrow{d} N \to \infty \chi^2 (p)
\]
Candelon et al. (2001)

**Definition**

Under UC, the mean of durations between two violations is equal to $1/\alpha$, and the null hypothesis is

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

with a test statistic equal to

$$J_{UC} = \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_1 (d_i; \alpha) \right)^2 \xrightarrow{d} \chi^2(1)$$
Recommandations
Recommandation 1: Test, test and test
Recommandation 2: Check the P&L data
Recommandation 3: The power of your tests may be low..
Recommandation 4: Take into account the estimation risk
Recommandation 1: Test, test and test

- Each type of test (frequency, severity, independence, conditional coverage, multivariate test etc..) captures one type of potential misspecification of the VaR model.

- It is important to use a variety of tests
Backtesting

Marginal Backtesting
By Christophe Hurlin, and Christophe Perignon
University of Orleans, HEC Paris (2011)

This code allows users to implement various backtesting methods for a series of VaR margins or VaR forecasts. The statistical tests used are the following: Z-statistic (Jorion, 2007), LR tests (Kupiec, 1995 and Christoffersen, 1998), DG tests (Engle and Manganelli, 2004), LR duration-based tests (Christoffersen and Pelletier, 2004), and the Risk Map (Colletaz, Hurlin and Perignon, 2011).

Christophe Hurlin
University of Orleans
France
CODER PAGE

Christophe Perignon
HEC Paris
France
CODER PAGE
Backtesting

Introduction
Backtesting Principles
Testing strategies
Recommendations

Test, test and test
Check the P&L data
The power of your tests may be low...
Estimation risk

1. Load your data
2. RunMyCode
3. Receive your results

Choose the test that you want to implement
- LR tests
- All the available tests
- Dynamic Quantile (DQ) tests
- LR duration based tests
- LR tests
- The Risk Map
- z-statistic and Kupiec’s LR test

Coverage rate

P&L
Value-at-Risk

Required input
Type: choice list
Recommandation 2: Check the P&L data


1. A large fraction of US and international banks validate their market risk model using P&L data that include fees and commissions and intraday trading revenues.

2. Distinction between dirty P/L and hypothetical P/L (JP. Morgan, Romain Berry 2011).
Recommandation 3: The power of your tests may be low..

Definition

The power of a backtesting test corresponds to its capacity to detect misspecified VaR model.

\[ \Pr [ \text{Rejection } H_0 \mid H_1] \]

Example

### Table 5: Power of 10% Finite Sample CC Tests on 5% VaR in Four Business Lines

<table>
<thead>
<tr>
<th>Sample</th>
<th>LB(1)</th>
<th>LB(5)</th>
<th>Markov</th>
<th>Weibull</th>
<th>Geometric</th>
<th>Caviar</th>
<th>KS</th>
<th>CVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.2964</td>
<td>0.3852</td>
<td>0.2048</td>
<td>0.1613</td>
<td>0.3190</td>
<td>0.4240</td>
<td>0.3485</td>
<td>0.3442</td>
</tr>
<tr>
<td>500</td>
<td>0.3912</td>
<td>0.5275</td>
<td>0.2139</td>
<td>0.1825</td>
<td>0.4466</td>
<td>0.5100</td>
<td>0.4429</td>
<td>0.4638</td>
</tr>
<tr>
<td>750</td>
<td>0.4356</td>
<td>0.6334</td>
<td>0.2257</td>
<td>0.2305</td>
<td>0.5684</td>
<td>0.6240</td>
<td>0.5316</td>
<td>0.5534</td>
</tr>
<tr>
<td>1000</td>
<td>0.4836</td>
<td>0.6957</td>
<td>0.2511</td>
<td>0.2698</td>
<td>0.6794</td>
<td>0.7290</td>
<td>0.5858</td>
<td>0.6068</td>
</tr>
<tr>
<td>1250</td>
<td>0.5431</td>
<td>0.7621</td>
<td>0.2935</td>
<td>0.3246</td>
<td>0.7560</td>
<td>0.7940</td>
<td>0.6654</td>
<td>0.6745</td>
</tr>
<tr>
<td>1500</td>
<td>0.5925</td>
<td>0.8146</td>
<td>0.3279</td>
<td>0.3790</td>
<td>0.8188</td>
<td>0.8590</td>
<td>0.7199</td>
<td>0.7220</td>
</tr>
</tbody>
</table>
Table 5: Power of 10% Finite Sample CC Tests on 5% VaR in Four Business Lines

<table>
<thead>
<tr>
<th>Sample</th>
<th>LB(1)</th>
<th>LB(5)</th>
<th>Markov</th>
<th>Weibull</th>
<th>Geometric</th>
<th>Caviar</th>
<th>KS</th>
<th>CVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.2964</td>
<td>0.3852</td>
<td>0.2048</td>
<td>0.1613</td>
<td>0.3190</td>
<td>0.4240</td>
<td>0.3485</td>
<td>0.3442</td>
</tr>
<tr>
<td>500</td>
<td>0.3912</td>
<td>0.5275</td>
<td>0.2139</td>
<td>0.1825</td>
<td>0.4466</td>
<td>0.5100</td>
<td>0.4429</td>
<td>0.4638</td>
</tr>
<tr>
<td>750</td>
<td>0.4356</td>
<td>0.6334</td>
<td>0.2257</td>
<td>0.2305</td>
<td>0.5684</td>
<td>0.6240</td>
<td>0.5316</td>
<td>0.5534</td>
</tr>
<tr>
<td>1000</td>
<td>0.4836</td>
<td>0.6957</td>
<td>0.2511</td>
<td>0.2698</td>
<td>0.6794</td>
<td>0.7290</td>
<td>0.5858</td>
<td>0.6068</td>
</tr>
<tr>
<td>1250</td>
<td>0.5431</td>
<td>0.7621</td>
<td>0.2935</td>
<td>0.3246</td>
<td>0.7560</td>
<td>0.7940</td>
<td>0.6654</td>
<td>0.6745</td>
</tr>
<tr>
<td>1500</td>
<td>0.5925</td>
<td>0.8146</td>
<td>0.3279</td>
<td>0.3790</td>
<td>0.8188</td>
<td>0.8590</td>
<td>0.7199</td>
<td>0.7220</td>
</tr>
</tbody>
</table>
### Table 5: Power of 10% Finite Sample CC Tests on 5% VaR in Four Business Lines

<table>
<thead>
<tr>
<th>Business Line 1</th>
<th>Sample</th>
<th>LB(1)</th>
<th>LB(5)</th>
<th>Markov</th>
<th>Weibull</th>
<th>Geometric</th>
<th>Caviar</th>
<th>KS</th>
<th>CVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>250</td>
<td>0.2964</td>
<td>0.3852</td>
<td>0.2048</td>
<td>0.1613</td>
<td>0.3190</td>
<td>0.4240</td>
<td>0.3485</td>
<td>0.3442</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.3912</td>
<td>0.5275</td>
<td>0.2139</td>
<td>0.1825</td>
<td>0.4466</td>
<td>0.5100</td>
<td>0.4429</td>
<td>0.4638</td>
</tr>
<tr>
<td></td>
<td>750</td>
<td>0.4356</td>
<td>0.6334</td>
<td>0.2257</td>
<td>0.2305</td>
<td>0.5684</td>
<td>0.6240</td>
<td>0.5316</td>
<td>0.5534</td>
</tr>
<tr>
<td><strong>1000</strong></td>
<td><strong>0.4836</strong></td>
<td><strong>0.6957</strong></td>
<td><strong>0.2511</strong></td>
<td><strong>0.2698</strong></td>
<td><strong>0.6794</strong></td>
<td><strong>0.7290</strong></td>
<td><strong>0.5858</strong></td>
<td><strong>0.6068</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1250</td>
<td>0.5431</td>
<td>0.7621</td>
<td>0.2935</td>
<td>0.3246</td>
<td>0.7560</td>
<td>0.7940</td>
<td>0.6654</td>
<td>0.6745</td>
</tr>
<tr>
<td></td>
<td>1500</td>
<td>0.5925</td>
<td>0.8146</td>
<td>0.3279</td>
<td>0.3790</td>
<td>0.8188</td>
<td>0.8590</td>
<td>0.7199</td>
<td>0.7220</td>
</tr>
</tbody>
</table>
Backtesting

Hurlin C. et Tokpavi S. (2008), "Une Evaluation des Procédures de Backtesting : Tout va pour le Mieux dans le Meilleur des Mondes", Finance

**Idea:** we use 6 different methods (GARCH, RiskMetrics, HS, CaviaR, Hybride, Delta Normale) to forecast a VaR(5%) on the same asset (GM, Nasdaq), and we apply the backtests (LR, DQ, Duration based tests) on a set of 500 samples (rolling window) of $T = 250$ forecasts.
Example

$LR_{CC}$ tests: for 47% of the samples, we don’t reject (at 5%) the null for any of the six VaR forecasts. In 71% of the samples, we reject at the most one VaR.

Example

$DQ_{CC}$ tests: for 20% of the samples, we don’t reject (at 5%) the null for any of the six VaR forecasts. In 51% of the samples, we reject at the most one VaR.
The power of a **consistent test** tends to 1 when the sample size tends to infinity.

**Recommandation:** increase at the maximum the sample size of your backtest.. ($T = 500, 750$ or more.)
Recommandation 4: Take into account the estimation risk

- The risk dynamic is usually represented by a parametric or semi-parametric model, which has to be estimated in a preliminary step. However, the estimated counterparts of risk measures are subject to estimation uncertainty.

- Replacing, in the theoretical formulas, the true parameter value by an estimator induces a bias in the coverage probabilities.
Backtesting

**VaR model. Parameters:** $\beta$

**Ex:** VaR-GARCH

$$VaR = -\Phi^{-1}(\alpha)\sigma_t$$

$$\sigma_t^2 = \beta_0 + \beta_1 \tau_{t-1}^2 + \beta_2 \sigma_{t-1}^2$$

**Estimation (ML)**

$$\hat{\beta}$$

**VaR Forecasts**

**BACKTESTING**
Escanciano and Olmo (2010, 2011) studied the effects of estimation risk on backtesting procedures. They showed how to correct the critical values in standard tests used to assess VaR models.


Backtesting

Estimation Adjusted VaR

Gouriéroux and Zakoian (2013) a method to directly adjust the VaR to estimation risk ensuring the right conditional coverage probability at order $1/T$:

$$
\Pr \left[ r_t < -EVaR_{t|t-1} (\alpha) \right] = \alpha + o_P \left( 1/T \right)
$$

Thank you for your attention