A new approach to backtesting and risk model selection

Jacopo Corbetta (École des Ponts - ParisTech)

Joint work with: Ilaria Peri (University of Greenwich)

June 18, 2016

Outline

1. Theoretical framework

2. Model validation

3. Model selection

An internal point of view

> VaR is only as good as its backtest. When someone shows me a VaR number, I don't ask how it is computed, I ask to see the backtest.
>
> Aaron Brown - Risk manager of the year 2012

An internal point of view

> VaR is only as good as its backtest. When someone shows me a VaR number, I don't ask how it is computed, I ask to see the backtest.
>
> Aaron Brown - Risk manager of the year 2012

> Many banks that have adopted an internal model-based approach to market risk measurement routinely compare daily profits and losses with model-generated risk measures to gauge the quality and accuracy of their risk measurement systems. This process, known as "backtesting", has been found useful by many institutions as they have developed and introduced their risk measurement models.
>
> Basel Committee (1996)

Open Problems

1. Methodologies for the backtesting of numerous alternative risk measures are less straight-forward (approximations or Monte-Carlo simulations).
2. Rigorous definition of backtesting and backtestability is still missing.
3. Model selection among risk models is meaningful only for elicitable estimator and only different forecasts of the same risk estimator can be compared.

Outline

Some notation

$(\Omega, \mathcal{F}, \mathbb{P})$, probability space.

- $\mathcal{M}_1$ set of all probability measures on the real line.
- L^0 set of all almost surely finite real valued random measures.
- $\mathcal{R}$ set of law invariant risk measure.

Some notation

$(\Omega, \mathcal{F}, \mathbb{P})$, probability space.

- $\mathcal{M}_1$ set of all probability measures on the real line.
- L^0 set of all almost surely finite real valued random measures.
- $\mathcal{R}$ set of law invariant risk measure.
- $X \in L^0$ represent the returns, its distribution is given by $\mathrm{P}(x) = \mathbb{P}(X < x)$.
- $\varrho : \mathcal{P}_\varrho \to \mathbb{R}^+$ is a law invariant risk measure.
- (P, ϱ) will be called a risk procedure.

Idea behind our framework

> The fraction actually covered can then be compared with the *intended level of coverage* to gauge the performance of the bank's risk model
>
> Basel Committee (1996)

Idea behind our framework

> The fraction actually covered can then be compared with the *intended level of coverage* to gauge the performance of the bank's risk model
>
> Basel Committee (1996)

Main question: What is the actual level of coverage provided by the risk model?

- For *VaR* the level of coverage can be naturally associated with its confidence level λ. For other risk measures alternative this is not straightforward;
- The choice of the confidence level is a critical issue pointed out by both practitioners and academics (see Kerkhof and Melenberg (2004) for detail explanation).

Level of coverage

Definition: Level of coverage

Given a $\varrho : \mathcal{P}_\varrho \subset \mathcal{M}_1 \to \mathbb{R}$ and a probability measure P, the level of coverage associated to the risk measure ϱ applied to P is defined by:

$$\lambda_{\mathrm{P}}^{\varrho} := \mathrm{P}(-\varrho(\mathrm{P})) = \mathbb{P}(X < -\varrho(-\mathrm{P})) \in [0,1]$$

Level of coverage

Definition: Level of coverage

Given a $\varrho : \mathcal{P}_\varrho \subset \mathcal{M}_1 \to \mathbb{R}$ and a probability measure P, the level of coverage associated to the risk measure ϱ applied to P is defined by:

$$\lambda_{\mathrm{P}}^{\varrho} := \mathrm{P}(-\varrho(\mathrm{P})) = \mathbb{P}(X < -\varrho(-\mathrm{P})) \in [0,1]$$

The level of coverage $\lambda_{\mathrm{P}}^{\varrho}$ offered by the risk model is the probability of having a violation, where:

- the probability is provided by the model P for forecasting returns;
- the violation occurs in respect of the capital requirement $-\varrho(\mathrm{P})$.

Basic example:

$$\lambda_{\mathrm{P}}^{Var_\varepsilon} = \varepsilon.$$

A new definition of Backtestability

Definition

A risk measure $\varrho : \mathcal{P}_\varrho \subset \mathcal{M}_1 \mapsto \mathbb{R}$ is backtestable over a set $\mathcal{P} \subseteq \mathcal{M}_1$ if there exists a map $\mathscr{P}_\varrho$ that associates to every $\mathrm{P} \in \mathcal{P}$ the coverage level given by ϱ applied to P.

A new definition of Backtestability

Definition

A risk measure $\varrho : \mathcal{P}_\varrho \subset \mathcal{M}_1 \mapsto \mathbb{R}$ is backtestable over a set $\mathcal{P} \subseteq \mathcal{M}_1$ if there exists a map $\mathscr{P}_\varrho$ that associates to every $\mathrm{P} \in \mathcal{P}$ the coverage level given by ϱ applied to P.

- the backtestability of a risk measure does not depend by any theoretical property, such as the elicitability or consistency;
- the backtestability only depends by the identification of the level of coverage provided by the risk model.

Universal Backtesting

Theorem

Every law invariant risk measure ϱ is backtestable over its domain of definition $\mathcal{P}_\varrho$ in the sense of Definition (9)

Universal Backtesting

Theorem

Every law invariant risk measure ϱ is backtestable over its domain of definition $\mathcal{P}_\varrho$ in the sense of Definition (9)

Observations:

- determining the capital requirement by a risk measure and its backtesting are two separate issues;
- the objective of the backtesting should be to verify if the coverage has been adequate and this prescinds from the process of generation of the capital requirement.

Outline

The setting

$(\Omega, \{\mathcal{F}_t\}_{t=1,\dots,T}, \mathbb{P})$ filtered space. X_t returns at time t.

- $F_t(x) = \mathbb{P}(X_t < x)$ real unknown probability distributions of the returns.

The setting

$(\Omega, \{\mathcal{F}_t\}_{t=1,\dots,T}, \mathbb{P})$ filtered space. X_t returns at time t.

- $F_t(x) = \mathbb{P}(X_t < x)$ real unknown probability distributions of the returns.
- $\mathrm{P}_t(x) = \mathbb{P}(X_t < x|\mathcal{F}_{t-1})$ forecast probability distributions of the returns.

The violations

Let ϱ be a risk measure. We define the violations as

$$I_t = \begin{cases} 1 & \text{if } X_t < -\varrho(\mathrm{P}_t) \\ 0 & \text{otherwise} \end{cases} .$$

The violations

Let ϱ be a risk measure. We define the violations as

$$I_t = \begin{cases} 1 & \text{if } X_t < -\varrho(\mathrm{P}_t) \\ 0 & \text{otherwise} \end{cases} .$$

Remark

The violations follow a Bernulli distribution with parameter $\lambda_t^0 = \lambda_{\mathrm{P}_t}^{\varrho}$ under the model probability, and with parameter $\lambda_t = F_t(X_t < -\varrho(\mathrm{P}_t))$ under the real probability

The violations

Let ϱ be a risk measure. We define the violations as

$$I_t = \begin{cases} 1 & \text{if } X_t < -\varrho(\mathrm{P}_t) \\ 0 & \text{otherwise} \end{cases} .$$

Remark

The violations follow a Bernulli distribution with parameter $\lambda_t^0 = \lambda_{\mathrm{P}_t}^{\varrho}$ under the model probability, and with parameter $\lambda_t = F_t(X_t < -\varrho(\mathrm{P}_t))$ under the real probability

Key Assumption

The violations are independent.

Test 1: unilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \geq \lambda_t^0$ for every t, with strict inequality for some t.

Test 1: unilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \geq \lambda_t^0$ for every t, with strict inequality for some t.

Test statistic

We define the test statistic

$$Z_1 := \sum_{t=1}^{T} I_t$$

Test 1: unilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \geq \lambda_t^0$ for every t, with strict inequality for some t.

Test statistic

We define the test statistic

$$Z_1 := \sum_{t=1}^{T} I_t$$

- Under H_0, Z_1 follows a Binomial Poisson distribution with parameters $\{\lambda_t^0\}_t$.
- For the significance level α, the rejection region is $C_{Z_1} = \{z_1 : P_{Z_1}(z_1) > 1 - \alpha\}$.
- Generalizes the traffic light approach (Basel, 1996) but only two regions.

Test 2: asymptotic bilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \neq \lambda_t^0$ for some t.

Test 2: asymptotic bilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \neq \lambda_t^0$ for some t.

Test statistic

We define the test statistic:

$$Z_2 := \frac{\sum_{t=1}^{T}(I_t - \lambda_t^0)}{\sqrt{\sum_{t=1}^{T}\lambda_t^0(1-\lambda_t^0)}}$$

Test 2: asymptotic bilateral coverage test

We set the null and alternative hypothesis as follows:

- H_0: $\lambda_t^0 = \mathrm{P}_t(-\varrho(\mathrm{P}_t)) = \lambda_t = F_t(-\varrho(\mathrm{P}_t))$ for every t
- H_1: $\lambda_t \neq \lambda_t^0$ for some t.

Test statistic

We define the test statistic:

$$Z_2 := \frac{\sum_{t=1}^{T}(I_t - \lambda_t^0)}{\sqrt{\sum_{t=1}^{T}\lambda_t^0(1-\lambda_t^0)}}$$

- Under H_0, by Lyapunov Central Limit Theorem Z_2 converges to a Standard Normal: $Z_2 \xrightarrow{d} N(0,1)$.
- For the significance level α, the rejection region is $C_{Z_2} := \left\{z_2 : z_2(x) < q_N\left(\frac{\alpha}{2}\right)\right\} \cup \left\{z_2 : z_2(x) > q_N\left(1-\frac{\alpha}{2}\right)\right\}$.

Outline

1. Theoretical framework
2. Model validation
3. Model selection

We validated, and now?

- The sole study of the number of exception does not tell if the model is good or not. One should study also the **magnitude** of violations
- Model selection can be conducted by minimizing the average backtesting error (loss) (Lopez 1998)
- However, the common practice of using **loss** functions that are not consistent with the forecasting estimator leads to meaningless inferences (Gneiting 2011).
- As a consequence, model selection seems to be possible only for **elicitable** estimator.

Our proposal: the magnitude of exceptions

We propose a method for comparing forecasting outcomes of every risk estimator.

Definition

Let (P, ϱ) be a risk measure procedure with coverage level $\lambda_t^0 \in (0,1)$. The magnitude of the exceptions over the backtesting period T is:

$$M((\mathrm{P}, \varrho)) = \sum_{t=1}^{T} (x_t - (-\varrho(\mathrm{P}_t)))^+ g(\mathrm{P}_t(-\varrho(\mathrm{P}_t))) + (x_t - (-\varrho(\mathrm{P}_t)))^- g(1 - \mathrm{P}_t(-\varrho(\mathrm{P}_t)))$$

A new order

Consider the family $\{(\mathrm{P}, \varrho)^i\}_{i \in I} = \{(\{\mathrm{P}^i_t\}_t, \varrho^i)\}_{i \in I}$.

We say that $(\mathrm{P}, \varrho)^i$ is preferable to $(\mathrm{P}, \varrho)^j$ for the magnitude order, and we write

$$(\mathrm{P}, \varrho)^i \preceq_M (\mathrm{P}, \varrho)^j$$

if and only if

$$M((\mathrm{P}, \varrho)^i) \leq M((\mathrm{P}, \varrho)^j)$$

A new order

Consider the family $\{(\mathrm{P},\varrho)^i\}_{i\in I} = \{(\{\mathrm{P}_t^i\}_t, \varrho^i)\}_{i\in I}$.

We say that $(\mathrm{P},\varrho)^i$ is preferable to $(\mathrm{P},\varrho)^j$ for the magnitude order, and we write

$$(\mathrm{P},\varrho)^i \preceq_M (\mathrm{P},\varrho)^j$$

if and only if

$$M((\mathrm{P},\varrho)^i) \leq M((\mathrm{P},\varrho)^j)$$

- $\preceq_M$ is a total preorder.

Best performing risk measure procedure

We can choose the best performing procedure as

$$(\hat{\mathrm{P}}, \hat{\varrho}) = \arg\min_{i \in I} M((\mathrm{P}, \varrho)^i)$$

Best performing risk measure procedure

We can choose the best performing procedure as

$$(\hat{\mathrm{P}}, \hat{\varrho}) = \arg\min_{i \in I} M((\mathrm{P}, \varrho)^i)$$

- This method can be used for selecting the best estimator among **all** the risk models previously validated.
- Our aim is **not** to propose an alternative property for identifying "superior" statistical functionals (as done by the elicitability or consistency)

Conclusions

- We propose a practical approach to backtesting.
- Our framework as a natural interpretation: we provide 2 straightforward test.
- We propose a simple way to compare performance of different risk procedures.

Conclusions

- We propose a practical approach to backtesting.
- Our framework as a natural interpretation: we provide 2 straightforward test.
- We propose a simple way to compare performance of different risk procedures.

Thank you for your attention

References

- Acerbi, C., Székely, B., 2015. Backtesting Expected Shortfall.
- Campbell, S., 2005. A review of backtesting and backtesting procedures.
- Christoffersen, P., 2010. Encyclopedia of Quantitative Finance - Backtesting.
- Embrechts, P., Hofert, M., 2014. Statistics and Quantitative Risk Management for Banking and Insurance.
- Gneiting, T., 2011. Making and evaluating point forecasts.
- Kerkhof, J., Melenberg, B., 2004. Backtesting for Risk-Based Regulatory Capital.
- Lopez, J.A., 1998. Methods for Evaluating Value-at-Risk Estimates.