Unbiased estimation of risk measures

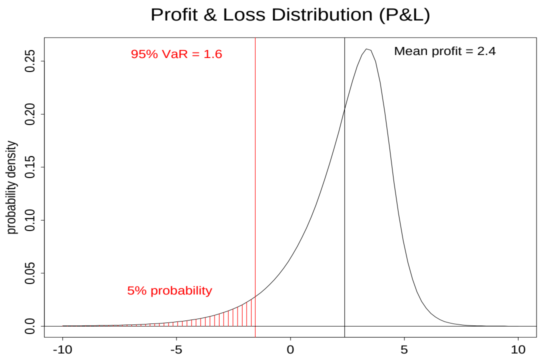
Albert-Ludwigs-Universität Freiburg

Thorsten Schmidt
joint work with Marcin Pitera
Fellow of FRIAS, Freiburg
Department for Mathematical Stochastics

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de
Lugano, 2017

Problem

- A standard risk measure is the **value-at-risk**. This risk measure is given by a quantile of the profit & loss distribution.



Density of X together with VaR at the level $\alpha = 0.05$. Source: McNeil, Frey, Embrechts **Quantitative Risk Management**.

- Assume that the **profit** of a portfolio, say X, is normally distributed with mean μ and variance σ^2. Then the value-at-risk at level α is given by

$$\mathrm{VaR}_\alpha(X) = -\Big(\mu + \sigma\Phi^{-1}(\alpha)\Big).$$

- But in practice, μ and σ are unknown and have to be estimated. In this regard, let us consider the simplest case: we have an i.i.d. sample $X_1, \dots, X_n =: \boldsymbol{X}$ at hand.
- **Efficient** estimators of μ and σ are at hand:

$$\hat\mu_n = \bar{\boldsymbol{X}}, \qquad \hat\sigma_n = \bar\sigma(\boldsymbol{X}) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{\boldsymbol{X}})^2}. \tag{1}$$

- **Common practice** is to use the plug-in estimator

$$\mathrm{VaR}_\alpha^{plugin} := -\Big(\hat\mu_n + \hat\sigma_n\Phi^{-1}(\alpha)\Big).$$

- Can this be efficient?

Motivation from Backtesting

- Let us perform a standard backtesting-procedure, i.e. we run several simulations, estimate the value-at-risk and check if the percentage of unsufficient capital does not exceed 5%.
- We show estimates of $\mathrm{VaR}_{0.05}$ with the plug-in procedure. The data is from NASDAQ100 and from a sample from normally distributed random variable with mean and variance fitted to the NASDAQ data ("Simulated", second column), both for 4.000 data points $\rightarrow$ we would expect **200** exceedances.
- **Exceeds** reports the number of exceptions in the sample, where the actual loss exceeded the risk estimate.

Estimator		NASDAQ		Simulated	
		exceeds	percentage	exceeds	percentage
Plug-in	$\hat{\mathrm{VaR}}_\alpha^{\text{plugin}}$	241	0.061	221	0.056

Motivation from Statistics

- In the normal case, for **known** σ, the likelihood-ratio test turns out to be the Gauss-test, or, equivalently, the confidence-interval is a normal distribution.
- If σ is **unknown**, one utilizes the t-distribution to obtain an efficient test: consider w.l.o.g. the test for $\mu = 0$ versus $\mu \neq 0$. The standardized test statistic is

$$T(X_1, \ldots, X_n) =: T(\boldsymbol{X}) = \frac{\sqrt{n}\bar{\boldsymbol{X}}}{\bar{\sigma}(\boldsymbol{X})}$$

and the test rejects the null hypothesis if

$$T(X) > t_n(1-\alpha).$$

- Shouldn't there be a similar adjustment towards the t-distribution in the estimator for VaR?

- Our findings suggest that the estimator is biased. In a statistical sense !
- Our goal is to analyse this problem and give a new notion of unbiasedness in an economic sense.

Measuring risk

We begin with well-known results on the measurement of risk, see McNeil et al. (2005).

- Let $(\Omega, \mathscr{A})$ be a measurable space and $(P_\theta : \theta \in \Theta)$ be a family of probability measures.
- For simplicity, we assume that the measures P_θ are equivalent, such that their null-sets coincide.
- For the estimation, we assume that we have a sample $X_1, X_2, \dots, X_n$ of observations at hand.
- A risk measure ρ is a mapping from L^0 to $\mathbb{R} \cup \{+\infty\}$.
- The value $\rho(X)$ is a quantification of risk for a future position: it is the amount of money one has to add to the position X such that the position becomes acceptable.

A priori, the definition of a risk measure is formulated without any relation to the underlying probability. However, in most practical applications one typically considers law-invariant risk-measures. Denote by $\mathscr{D}$ the convex space of cumulative distribution functions of real-valued random variables.

Definition

The family of risk-measures $(\rho_\theta)_{\theta\in\Theta}$ is called **law-invariant**, if there exists a function $R : \mathscr{D} \to \mathbb{R} \cup \{+\infty\}$ such that for all $\theta \in \Theta$ and $X \in L^0$

$$\rho_\theta(X) = R(F_X(\theta)), \tag{2}$$

$F_X(\theta) = P_\theta(X \le \cdot)$ denoting the cumulative distribution function of X under the parameter θ.

Estimation

We aim at estimating the risk of the future position when $\theta \in \Theta$ is unknown and needs to be estimated from a data sample $x_1, \dots, x_n$.

Definition

An **estimator** of a risk measure is a Borel function $\hat{\rho}_n : \mathbb{R}^n \to \boldsymbol{R} \cup \{+\infty\}$.

Sometimes we will call $\hat{\rho}_n$ also risk estimator.

The following definition introduces an economically motivated formulation of unbiasedness.

Definition

The estimator $\hat{\rho}_n$ is called **unbiased** for $\rho(X)$, if for all $\theta \in \Theta$,

$$\rho_\theta(X + \hat{\rho}_n) = 0. \tag{3}$$

- If the estimator is unbiased, adding the estimated amount of risk capital $\hat{\rho}_n$ to the position X makes the position $X + \hat{\rho}_n$ acceptable under all possible scenarios $\theta \in \Theta$.
- Requiring equality in Equation (3) ensures that the estimated capital is not too high.
- Except for the i.i.d. case, the distribution of $X + \hat{\rho}_n$ does also depends on the dependence structure of $X, X_1, \ldots, X_n$ and not only on the (marginal) laws.

Relation to the statistical definition of unbiasedness

Our Definition differs from unbiasedness in the statistical sense!

- The estimator $\hat{\rho}_n$ is called **statistically unbiased**, if

$$E_\theta[\hat{\rho}_n] = \rho_\theta(X), \qquad \text{for all } \theta \in \Theta, \tag{4}$$

- One point why the statistical unbiasedness is not reasonable here is that it does not behave well in various backtesting or stress-testing procedures.

Relation to level adjustment

- A further alternative is to adjust the level α, see Frank (2016) and Francioni and Herzog (2012).
- An existing estimator depending continuously on α can always be trimmed to match exactly the unbiased estimator. However, the adjusted α will typically depend on n and the sample (Example to follow)!

Unbiased estimation of value-at-risk under normality

- Let $X \sim \mathcal{N}(\theta_1, \theta_2^2)$ and denote $\theta = (\theta_1, \theta_1) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$.
- The value-at-risk is

$$\rho_\theta(X) = \inf\{x \in \boldsymbol{R} \colon P_\theta[X + x < 0] \leq \alpha\}, \qquad \theta \in \Theta, \tag{5}$$

- Unbiasedness as defined in Equation (3) is equivalent to

$$P_\theta[X + \hat{\rho} < 0] = \alpha, \qquad \text{for all } \theta \in \Theta. \tag{6}$$

- We define estimator $\hat{\rho}$, as

$$\hat{\rho}(x_1, \ldots, x_n) = -\bar{x} - \bar{\sigma}(\boldsymbol{x})\sqrt{\frac{n+1}{n}}\, t_{n-1}^{-1}(\alpha), \tag{7}$$

This estimator is **unbiased**: first, note that

$$X + \hat{\rho} = X - \bar{X} - \bar{\sigma}(\boldsymbol{X})\sqrt{\frac{n+1}{n}}t_{n-1}^{-1}(\alpha) \leq 0$$
$$\Leftrightarrow \qquad \sqrt{\frac{n}{n+1}} \cdot \frac{X - \bar{X}}{\bar{\sigma}(\boldsymbol{X})} \leq t_{n-1}^{-1}(\alpha).$$

Using the fact that X, $\bar{X}$ and $\bar{\sigma}(\boldsymbol{X})$ are independent for any $\theta \in \Theta$, we obtain

$$T := \sqrt{\frac{n}{n+1}} \cdot \frac{X - \bar{X}}{\bar{\sigma}(\boldsymbol{X})} = \frac{X - \bar{X}}{\sqrt{\frac{n+1}{n}}\theta_2} \cdot \sqrt{\frac{n-1}{\sum_{i=1}^{n}(\frac{X_i - \bar{X}}{\theta_2})^2}} \sim t_{n-1}.$$

Thus, the random variable T is a pivotal quantity and

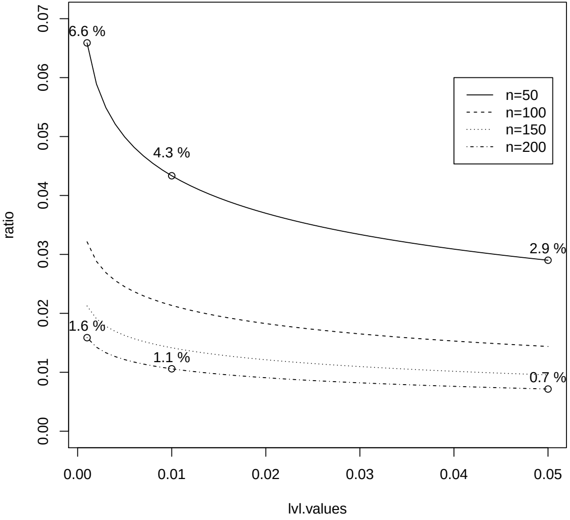
$$P_\theta[X + \hat{\rho} < 0] = P_\theta[T < q_{t_{n-1}}(\alpha)] = \alpha.$$

Let us elaborate a little bit on the difference between the plug-in and the unbiased estimator.

$$\hat{\text{VaR}}^{\text{u}}_\alpha = -\bar{x} - \bar{\sigma}(\boldsymbol{x})\sqrt{\frac{n+1}{n}}t^{-1}_{n-1}(\alpha)$$
$$\hat{\text{VaR}}^{\text{plugin}}_\alpha = -\bar{x} - \bar{\sigma}(\boldsymbol{x})\Phi^{-1}(\alpha)$$

The percentage of additional capital over the mean needed for the unbiased estimator is given by

$$\sqrt{\frac{n+1}{n}}\frac{t^{-1}_{n-1}(\alpha)}{\Phi^{-1}(\alpha)}. \tag{8}$$



The percentage of additional capital over the mean from Equation (8).

- Assume that
$$S_t = S_0 \exp\left(\mu t + \sigma\sqrt{t}\xi\right),$$
with $\xi \sim \mathcal{N}(0,1)$. (Note that this is the portfolio position).
- The estimators are (assuming i.i.d. data to S_t and using log-differences)
$$\mathrm{VaR}_\alpha^{\text{plugin}} = -S_0\left(e^{\bar{X}+\bar{\sigma}\Phi^{-1}(\alpha)} - 1\right)$$
$$\mathrm{VaR}_\alpha^{\text{plugin}} = -S_0\left(e^{\bar{X}+\bar{\sigma}\sqrt{n+1/n}\,t_{n-1}^{-1}(\alpha)} - 1\right)$$
- This time, the percentage of additional capital needed will also depend on mean and variance!

Unbiased estimation of expected shortfall under normality

- We continue in the previous setting,
- The expected shortfall at level α under a continuous distribution is

$$\rho_\theta(X) = E_\theta[-X|X \leq q_X(\theta,\alpha)],$$

where $q_X(\theta,\alpha)$ is α-quantile of X under P_θ.

- We consider estimators of the form

$$\hat{\rho}(x_1,\dots,x_n) = -\bar{x} - \bar{\sigma}(x)a_n, \tag{9}$$

for some $(a_n)_{n\in\mathbf{N}}$, where $a_n \in \mathbb{R}$.

- We can show that there exists a_n which makes $\hat{\rho}$ unbiased. This a_n can easily be computed numerically.

Empirical study

- It is the aim of this section to analyse the performance of selected estimators on various sets of real market data (Market) as well as on simulated data (Simulated). Our focus is on the practically most relevant risk measures, VaR and ES.
- The market data we use are returns from the data library Fama and French (2015), containing returns of 25 portfolios formed on book-to-market and operating profitability in the period from 27.01.2005 to 01.01.2015. We obtain exactly 2500 observations for each portfolio.
- The sample is split into 50 separate subsets, each consisting of 50 consecutive trading days. For $i = 1, 2, \dots, 49$, we estimate the risk measure using the i-th subset and test it's adequacy on $(i+1)$-th subset.
- The simulation study uses i.i.d. normally distributed random variables whose mean and variance was fitted to each of the 25 portfolios. The sample size was set to 2500 for each set of parameters. In this way we are able to exclude difficulties due to dependencies in the data or bad model fit.

Backtesting VaR

- We considered the unbiased estimator $\hat{\mathrm{VaR}}^{\mathrm{u}}_\alpha$, the empirical sample quantile $\hat{\mathrm{VaR}}^{\mathrm{emp}}_\alpha$, the modified Cornish-Fisher estimator $\hat{\mathrm{VaR}}^{\mathrm{CF}}_\alpha$, the plug-in estimator $\hat{\mathrm{VaR}}^{\mathrm{norm}}_\alpha$ and the GPD plug-in estimator[1] $\hat{\mathrm{VaR}}^{\mathrm{GPD}}_\alpha$.

$$\hat{\mathrm{VaR}}^{\mathrm{emp}}_\alpha(x) := -\left(x_{(\lfloor h\rfloor)} + (h-\lfloor h\rfloor)(x_{(\lfloor h+1\rfloor)} - x_{(\lfloor h\rfloor)})\right),$$

$$\hat{\mathrm{VaR}}^{\mathrm{CF}}_\alpha(x) := -\left(\bar{x} + \bar{\sigma}(x)\bar{Z}^\alpha_{CF}(x)\right),$$

$$\hat{\mathrm{VaR}}^{\mathrm{norm}}_\alpha(x) := -\left(\bar{x} + \bar{\sigma}(x)\Phi^{-1}(\alpha)\right),$$

$$\hat{\mathrm{VaR}}^{\mathrm{GPD}}_\alpha := -u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{\alpha n}{k}\right)^{-\hat{\xi}} - 1\right),$$

$$\hat{\mathrm{VaR}}^{\mathrm{u}}_\alpha(x_1,\dots,x_n) := -\left(\bar{x} + \bar{\sigma}(x)\sqrt{\frac{n+1}{n}}\,t^{-1}_{n-1}(\alpha)\right),$$

where $x_{(k)}$ is the k-th order statistic of $x = (x_1,\dots,x_n)$, the value $\lfloor z\rfloor$ denotes the integer part of $z \in \boldsymbol{R}$, $h = \alpha(n-1)+1$, Φ denotes the cumulative distribution function of the standard normal distribution and $\bar{Z}^\alpha_{CF}$ is a standard Cornish-Fisher α-quantile estimator.

---

[1] For each portfolio, we set the threshold value u to match the 0.7-empirical quantile of the corresponding sample.

Tabelle: Top: the results for portfolios in the period from 27.01.2005 to 01.01.2015 from the Fama & French dataset. Bottom: the results on simulated Gaussian data. We perform the standard backtest, splitting into intervals of length 50 and computing average rate of exceptions.

Type of data:	MARKET				
Portfolio	Estimator type				
	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{emp}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{norm}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{CF}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{GPD}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{u}}$
LoBM.LoOP	0.071	0.073	0.067	0.067	0.069
BM1.OP2	0.076	0.070	0.069	0.069	0.065
BM1.OP3	0.071	0.064	0.063	0.064	0.061
BM1.OP4	0.069	0.071	0.067	0.067	0.068
LoBM.HiOP	0.071	0.071	0.070	0.067	0.068
...	...	...	...	...	
mean	0.073	0.071	0.068	0.067	0.067

Type of data:	SIMULATED				
	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{emp}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{norm}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{CF}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{GPD}}$	$\hat{\mathrm{VaR}}_\alpha^{\mathrm{u}}$
LoBM.LoOP	0.065	0.057	0.055	0.056	**0.051**
BM1.OP2	0.064	0.053	0.053	0.053	**0.050**
BM1.OP3	0.069	0.058	0.058	0.060	**0.052**
BM1.OP4	0.069	0.057	0.058	0.062	**0.053**
LoBM.HiOP	0.060	0.054	0.053	0.056	**0.047**
...	...	...	...	...	
mean	0.066	0.057	0.057	0.058	**0.051**

Backtesting Expected Shortfall

In this example we will use the same dataset, but instead of VaR at level 5% we consider ES at level 10%. We consider

$$\hat{\mathrm{ES}}_\alpha^{\mathrm{emp}}(x) := -\left(\frac{\sum_{i=1}^n x_i \mathbb{1}_{\{x_i + \hat{\mathrm{VaR}}_\alpha^{\mathrm{emp}}(x) < 0\}}}{\sum_{i=1}^n \mathbb{1}_{\{x_i + \hat{\mathrm{VaR}}_\alpha^{\mathrm{emp}}(x) < 0\}}}\right),$$

$$\hat{\mathrm{ES}}_\alpha^{\mathrm{CF}}(x) := -\left(\bar{x} + \bar{\sigma}(x) C(\bar{Z}_{CF}^{\alpha}(x))\right),$$

$$\hat{\mathrm{ES}}_\alpha^{\mathrm{norm}}(x) := -\left(\bar{x} + \bar{\sigma}(x) \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\right),$$

$$\hat{\mathrm{ES}}_\alpha^{\mathrm{GPD}}(x) := \frac{\hat{\mathrm{VaR}}_\alpha^{\mathrm{emp}}(x)}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{1-\hat{\xi}},$$

Let the Gaussian unbiased Expected Shortfall estimator be

$$\hat{\mathrm{ES}}_\alpha^{\mathrm{u}}(x) := -\left(\bar{x} - \bar{\sigma}(x) a_n\right), \tag{10}$$

where a_n was computed numerically.

Note that the **non-elicitability of ES** is directly visible !

- For the backtest we follow *Test 2* suggested in Acerbi and Székely (2014) utilizing the 50 separate subsets of our data denoted by $(x_1^i, \dots, x_{50}^i)$.
- The test statistic for the backtest is given by

$$Z := \frac{1}{49}\sum_{i=1}^{49}\left(\frac{1}{50}\sum_{j=1}^{50}\frac{x_j^{i+1}\,\mathbb{1}_{\{x_j^{i+1}+\hat{\mathrm{VaR}}_\alpha^i<0\}}}{\alpha\,\hat{\mathrm{ES}}_\alpha^i}\right)+1. \tag{11}$$

- The results of our backtest are presented in Table 2.
- While 0 would be optimal, negative values of the test statistic Z correspond to **underestimtion** of risk of the considered estimator.

Tabelle: The results of our backtest. Note that while 0 would be optimal, negative values of the test statistic Z correspond to **underestimtion** of risk.

Type of data:	MARKET				
Portfolio	Estimator type				
	ES_α^{emp}	ES_α^{norm}	ES_α^{CF}	ES_α^{GPD}	ES_α^{u}
LoBM.LoOP	-0.357	-0.393	-0.325	-0.302	-0.331
BM1.OP2	-0.428	-0.303	-0.338	-0.335	-0.235
BM1.OP3	-0.327	-0.322	-0.336	-0.295	-0.254
BM1.OP4	-0.326	-0.354	-0.348	-0.282	-0.272
LoBM.HiOP	-0.424	-0.421	-0.371	-0.335	-0.331
mean	-0.374	-0.363	-0.339	-0.308	-0.290

Type of data:	SIMULATED				
	ES_α^{emp}	ES_α^{norm}	ES_α^{CF}	ES_α^{GPD}	ES_α^{u}
LoBM.LoOP	-0.177	-0.073	-0.077	-0.104	**-0.005**
BM1.OP2	-0.143	-0.083	-0.069	-0.074	**-0.014**
BM1.OP3	-0.220	-0.084	-0.100	-0.157	**-0.019**
BM1.OP4	-0.224	-0.086	-0.101	-0.150	**-0.012**
LoBM.HiOP	-0.183	-0.082	-0.072	-0.098	**-0.016**
mean	-0.174	-0.101	-0.103	-0.109	**-0.030**

Elicitability

The remarkable article Gneiting (2011) critically reviews the evaluation of point forecasts: a good performance in backtesting might not necessarily imply that a given estimator is good.

Example (Perfect backtesting performance)

Assume a sample $(x_i)_{i=1,\dots,200}$ being centred with support $[-1, 1]$. Then, choosing 190 times the value 1 and 10 times the value -1 gives a perfect backtesting performance (for $\alpha = 0.05$) when measured only by the exceedance rate.

Elicitability will lead to a backtesting concept which will remedy this issue.

- A **scoring function** is a mapping which compares two values of risk measures: $S(x,y)$ measures the deviation from the forecast x to the realization y; the squared error $S(x,y) = (x-y)^2$ being a standard example.
- The scoring function S is called **consistent** for the law-invariant risk measure R relative to the class $\{F_X(\theta) : \theta \in \Theta\}$, if

$$E_\theta[S(R(F_X(\theta)), Y)] \leq E_\theta[S(r, Y)] \tag{12}$$

  for all $\theta \in \Theta$ and all $r \in \mathbb{R} \cup +\infty$; here E_θ denotes the expectation under which the random variable Y has distribution $F_X(\theta)$.
- The scoring function is called **strictly consistent** if it is consistent and equality in (12) implies that $r = R(F_X(\theta))$. For example, the squared error is strictly consistent relative to the class of probability measures of finite second moment.

Definition

The risk measure R is called **elicitable** relative to $\{F_X(\theta) : \theta \in \Theta\}$, if there exists a scoring function that is strictly consistent.

The prime example in our context is VaR_α (Value-at-Risk at level α). A possible specification of a scoring function is given by

$$S(x,y) = (\mathbb{1}_{\{x \geq y\}} - \alpha)(x - y). \tag{13}$$

Evaluating with the performance criterion

$$\bar{S} = \frac{1}{n} \sum_{i=1}^{n} S(x_i, y_i), \tag{14}$$

denoting by $x_1, \ldots, x_n$ the forecasts and by $y_1, \ldots, y_n$ the verifying observations, guarantees that the optimal point forecast outperforms all other estimators.
This in turn allows to identify flawed estimators like in Example 4.

Backtesting using scoring functions

- Running the backtests with scoring functions indeed shows that the unbiased estimators also outperform the other estimators with respect to this measures.
- For expected shortfall the situation is more complicated: it is only elicitable **jointly** with value-at-risk. Still, in this setting the unbiased estimators outperform the other estimators.

Conclusion

- We studied the estimation of risk, with a particular view on **unbiased** estimators and backtesting.
- The new notion of unbiasedness introduced is motivated from economic principles rather than from statistical reasoning, which links this concept to a better performance in backtesting.
- Some unbiased estimators, for example the unbiased estimator for value-at-risk in the Gaussian case, can be computed in closed form while for many other cases numerical methods are available.
- A small empirical analysis underlines the outperformance of the unbiased estimators with respect to standard backtesting measures.
- The extension to GPD-distributions is the next step.

The paper is available on SSRN: `https://ssrn.com/abstract=2890034`

Many thanks for your attention !

Acerbi, C. and Székely, B. (2014), 'Back-testing expected shortfall', **Risk magazine** (November).

Fama, E. F. and French, K. R. (2015), http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. accessed 20.10.2015.

Francioni, I. and Herzog, F. (2012), 'Probability-unbiased value-at-risk estimators', **Quantitative Finance 12**(5), 755–768.

Frank, D. (2016), 'Adjusting var to correct sample volatility bias', **Risk magazine** (October).

Gneiting, T. (2011), 'Making and evaluating point forecasts', **Journal of the American Statistical Association 106**(494), 746–762.

McNeil, A., Frey, R. and Embrechts, P. (2005), **Quantitative Risk Management: Concepts, Techniques and Tools**, Princeton University Press.