Project 5: Risk quantification in a low-dimensional portfolio

MATH-516 Applied Statistics

Linda Mhalla

2025-04-14

Dataset

Each student has to use a specific dataset corresponding to their ID (see last slide)

Datasets are available on the course Moodle webpage (`dataset_xx.Rdata`).

Each dataset includes:

- the calibration date range (`dates.calib`)
- the testing date range (`dates.test`)
- the portfolio weights (`weights`)
- the price series of three assets (`prices`)

Analysis: overview

The goal is to compute Value at Risk (VaR) estimates using a univariate approach and assess its performance via backtesting

- The Value at Risk can be viewed mathematically as an extreme quantile of a distribution, i.e., it is the value v_α such that

$$1 - \alpha = \bar{F}(v_\alpha), \quad \text{with } \alpha \text{ a large quantile level}$$

- The portfolio consists of three stocks from the Dow Jones Industrial Average index
- The portfolio is rebalanced daily: weights $w = (w_1, w_2, w_3)$ are fixed each day
- The approach involves:
  1. Model selection on the calibration period
  2. Fitting and evaluating VaR predictions during the testing period

Univariate modelling

Let P_t denote the price of an asset at time t. Then, the logarithmic return r_t and the logarithmic loss ℓ_t of the asset at time t are defined as:

$$r_t = \log(P_t) - \log(P_{t-1}), \quad \ell_t = -r_t = \log(P_{t-1}) - \log(P_t)$$

The portfolio's logarithmic loss $\ell_{p,t}$ at time t can be approximated as the weighted sum of the logarithmic losses of each asset:

$$\ell_{p,t} \approx \sum_{i=1}^{d} w_i \ell_{i,t}$$

where w_i and $\ell_{i,t}$ denote, respectively, the weight and logarithmic loss at time t of asset i

Univariate modelling

In this approach, the log-losses of the rebalanced portfolio are modeled as a univariate time series

- The method focuses on filtered log-losses
- VaR estimates are obtained via a semi-parametric approach involving extreme value theory

step A.1: exploratory analysis

Perform an exploratory analysis of the calibration-period log-losses:

- Summary statistics
- Correlograms
- Portmanteau tests (e.g., Ljung–Box to test independence in time series)
- Discussion of stylized facts of financial time series (returns show low serial correlation and absolute returns high + volatility clustering)

Correlogram

Given time series data $X_1, \dots, X_n$ we calculate the sample autocovariances

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} \left(X_t - \bar{X}\right)\left(X_{t+h} - \bar{X}\right) \quad \text{where} \quad \bar{X} = \sum_{t=1}^{n} X_t / n$$

- The sample autocorrelations are given by

$$\hat{\rho}(h) := \hat{\gamma}(h)/\hat{\gamma}(0), h = 0, 1, 2, \dots .$$

- The correlogram is the plot $\{(h, \hat{\rho}(h)), h = 0, 1, 2, ...\}$

step A.2: model fitting

Financial time series usually exhibit heteroskedasticity (consequence of volatility clustering)

An appropriate model for such settings is the $AR(k)$–$GARCH(p,q)$ model, suitable for the logarithmic losses of the portfolio:

$$
\begin{aligned}
X_t &= \mu_t + \sigma_t Z_t \\
\mu_t &= \mu + \sum_{i=1}^{k} \phi_i (X_{t-i} - \mu) \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2
\end{aligned}
$$

Here, $(Z_t)_{t\in\mathbb{Z}}$ is a white noise process with zero mean and unit variance, but its distribution F is unknown

Note: A white noise process is a time series process with no serial autocorrelation, i.e., $\rho(h) = \mathrm{Cov}(Z_h, Z_0)/\mathrm{Var}(Z_0) = 0$, for $h \neq 0$ and $\rho(0) = 1$

step A.2: model fitting

Based on results from A.1, fit a collection of univariate time series models to the portfolio's log-losses

- Select the best model using model selection criteria (e.g., AIC)
- Assess the quality of the fit

A tutorial on fitting such models can be found here

step A.3: extreme value modeling

For the innovations of the selected model:

- Choose a threshold u using:
  - mean residual life plot
  - parameter stability plots
- Fit a GPD to the excesses over u

A semiparametric approximation to the distribution F of the innovations combines the empirical distribution function below u and a generalized Pareto fit above it and is given by

$$\hat{F}_u(x) = \begin{cases} n^{-1} \sum_{j=1}^{n} I\left(x_j \leq x\right), & x \leq u \\ 1 - \frac{n_u}{n} \left(1 + \hat{\xi} \frac{x-u}{\hat{\sigma}_u}\right)_+^{-1/\hat{\xi}}, & x > u, \end{cases}$$

where $n_u = \#\{j : x_j > u\}$

step A.4: quantiles

Innovation quantiles

Compute the $\alpha = 0.95, 0.99$ quantiles of the innovations using the estimate of F given by $\hat{F}_u$

If $\alpha > F^{-1}(u)$, then the $\text{VaR}_\alpha = v_\alpha$ of the innovations is such that

$$1 - \alpha = \bar{F}(v_\alpha) = \bar{F}(u)\left(1 + \xi \frac{v_\alpha - u}{\sigma}\right)^{-1/\xi}$$

so that

$$v_\alpha = u + \frac{\sigma}{\xi}\left\{\left(\frac{1-\alpha}{\bar{F}(u)}\right)^{-\xi} - 1\right\}$$

In-sample VaR

- Compute in-sample $\text{VaR}_{0.95}$ and $\text{VaR}_{0.99}$ for the losses (negative log-returns) on the calibration period

step A.6: VaR prediction for testing period

For each day t in the testing period:

1. Fit the model from A.2 to the most recent n days
2. Extract standardised innovations $z_{t-n+1}, \dots, z_t$
3. Predict:
   - conditional mean μ_{t+1}
   - conditional volatility σ_{t+1}
   - 0.95 and 0.99 quantiles of the innovations using semi-parametric method from A.4
4. Compute VaR predictions for day $t+1$

step A.7: backtesting

- Assess the quality of the VaR predictions using a backtesting approach

Backtesting approach: binomial test

To evaluate the accuracy of the VaR predictions, we use the binomial test to assess the number of VaR violations

- For a given confidence level α (e.g., 0.95 or 0.99), the expected proportion of violations is $1 - \alpha$
- Over a testing period of T days, if x violations are observed, we model this as a binomial random variable:

$$x \sim \text{Binomial}(T, 1 - \alpha)$$

- We thus test the null hypothesis: The observed violation rate equals the expected rate (model is accurate)

Assigned dataset

Name of student	Assigned dataset
Ahou Samuel	dataset_12.Rdata
Boissier Charles Louis Pierre Bogdan	dataset_02.Rdata
Bouhadra Kalil Brahim	dataset_05.Rdata
Caldarone Alex John	dataset_21.Rdata
Carron Léo Jérémy	dataset_01.Rdata
Chu Tianle	dataset_10.Rdata
Do Alexis	dataset_04.Rdata
Ferrera Alessandro	dataset_16.Rdata
Frasa Nina	dataset_14.Rdata
Garcia Averell Regina	dataset_13.Rdata
Giuli Daniele	dataset_07.Rdata
Hengl Stephan	dataset_06.Rdata
Khella Georg	dataset_03.Rdata
Kriem Zayed	dataset_08.Rdata
Loukaidis Andronikos	dataset_18.Rdata
Mailänder Lennart Wolfgang	dataset_20.Rdata
Olaye Whaboaman Joel Thierry E	dataset_15.Rdata
Pettersen Julie Sofie	dataset_17.Rdata
Pfander Mila	dataset_19.Rdata
Sigillo' Massara Vincenzo	dataset_09.Rdata
Zakharov Daniil	dataset_11.Rdata