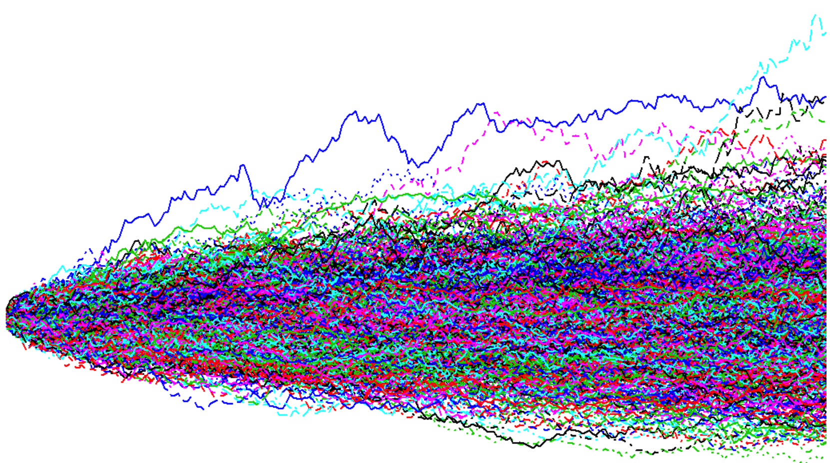
Quantitative Risk Management

MATH 510



Nicholas Hayek

Taught by Prof. Johanna Neslehova

CONTENTS

This course is based on the ETH course and textbook of the same name, by McNeil, Fray, and Embrechts. The present document is based on this textbook and Prof. Neshlehov's lectures. In order, we will focus on: stochastic volatility, extreme-value theory, multivariate models, risk aggregation, and backtesting.

Cover: index simulation following the stochastic volatility process

We will assume good working knowledge of probability and statistics. In addition, we remind ourselves of the *survivor function* $\overline{F}(x) := \mathbb{P}(X > x) = 1 - F(x)$, as well as the *$\alpha$-quantile*, where $\alpha \in (0, 1)$, defined to be

DEF 0.1

DEF 0.2

$$q_\alpha = F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$$

In this course, we would like to quantify one's exposure to bad consequences. The likelihood of loss or less-than-expected gains is called *risk*. The following are types of risk:

DEF 0.3

Risk is necessary: taking this course incurs a risk of a poor grade

____ Risk	Description
Credit Risk	Odds a debtor defaults on payment
Market Risk	Exposure to price fluctuations of bonds, stocks, or derivatives
Operational Risk	Risk relating to circumstantial adverse events (e.g. institutional fraud)
Liquidity Risk	Risk of damage from not having sufficient assets to pay off debts
Model Risk	Risk associated with financial model inaccuracies; closely related to operational risk
Underwriting Risk	Odds that an insured makes a claim on their policy

The above types of risk interact with each-other. Quantitative risk management aims to model these interactions and hedge against risk.

I Loss

RISK FACTORS

DEF 1.1

DEF 1.2

A *portfolio* is a collection of assets or liabilities. Denote by V_t the value of the portfolio at time t. We denote by Δt a *time horizon*, i.e. a duration of time. Assuming that V_t is known, that the composition of the portfolio remains constant over the time horizon, and that there are no payments made, we denote by V_{t+1} the value of the portfolio at time $t + \Delta t$.

Portfolios may include stocks, bonds, derivatives, risky loans, or insurance contracts, for example.

We write $\Delta V_{t+1} = V_{t+1} - V_t$. This is a random variable which takes a negative value on losses and a positive value on profits. We say that it follows the *profit-and-loss distribution*. Since we care about managing risk in this course, we prefer that losses are positively indicated. Define *loss* as

DEF 1.3

DEF 1.4

Hence, V_{t+k} is the value of the portfolio at time $t + k\Delta t$

$$L_{t+1} = -\Delta V_{t+1} = \begin{cases} V_t - V_{t+1} & \Delta t \text{ is a short horizon} \\ V_t - \frac{V_{t+1}}{1+r_{t,1}} & \Delta t \text{ is a long horizon} \end{cases}$$

Note that, for long time horizons, we must account for the time value of money. If $r_{t,1}$ is the risk-free interest rate, applied over Δt, then $\alpha(1 + r_{t,1})$ at time $t + 1$ is the equivalent

This is akin to the concept of opportunity cost

in value to α at time t. Hence, V_{t+1} must be adjusted to $V_{t+1}/(1+r_{r,1})$ to keep L_{t+1}'s time-value of money consistent. Working in "$t+1$ dollars," one could equivalently write $L_{t+1} = V_t(1+r_{t,1}) - V_{t+1}$ for long time intervals.

If one's portfolio consists of insurance policies, $Z_{t,4}$ might be the probability that a claim is made.

V_t is a function of multiple *risk factors*, denoted by $\mathbf{Z}_t = \langle Z_{t,1}, ..., Z_{t,d} \rangle$. Thus, we write $V_t = f(t, \mathbf{Z}_t)$ with $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, called the *risk function*. When $\mathbf{Z}_t$ is known, we write $\mathbf{Z}_t = \mathbf{z}_t$. In this case, $f(t, \mathbf{z}_t)$ is called the *realized value* of V_t.

DEF 1.5
DEF 1.6
DEF 1.7

Eg. 1.1 Consider a portfolio of d stocks, and let λ_i denote the number of shares in stock i at time t. Let $S_{t,i}$ denote the price of the stock i. Write

$$Z_{t,i} = \log S_{t,i} : i \in [d]$$

The value of the portfolio is then

$$V_t = \sum_{i=1}^{d} \lambda_i S_{t,i} = \sum_{i=1}^{d} \lambda_i e^{Z_{t,i}}$$

This formula is called the *value with log prices*. The purpose of writing $\mathbf{Z}_t = \log(\mathbf{S}_t)$ versus $\mathbf{Z}_t = \mathbf{S}_t$ is purely numerical.

DEF 1.8

Risk factors may change over the time horizon. We call time-series changes in risk factors *risk factor changes*, denoted by $\mathbf{X}_{t+1}$. In particular,

DEF 1.9

$$\mathbf{X}_{t+1} = \mathbf{Z}_{t+1} - \mathbf{Z}_t$$

Rearranging Def 1.6 gives

$$L_{t+1} = f(t, \mathbf{Z}_t) - f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1})$$

"Observable" = "Known" = "Realized" = "Actualized"

When modeling loss, we assume that $\mathbf{z}_t$ is observable. Thus, we are forced only to consider $\mathbf{X}_{t+1}$ as a random variable. If $\mathbf{x}_{t+1}$ is observable, then L_{t+1} is called the *realized loss*. Realized loss is notated by $l_{[t]}(\mathbf{x}) = L_{t+1}$, especially when we wish to parameterize $\mathbf{x}_{t+1}$. For simplicity, we drop the subscript, i.e. $\mathbf{x} = \mathbf{x}_{t+1}$. We call this function the *loss operator*.

DEF 1.10
DEF 1.11

We say that L_{t+1} follows a *loss distribution*. The loss distribution is typically right-skewed and has a fat right tail: the probability of heavy profits is small, but it is easy to incur heavy losses.

DEF 1.12

We may use a linearization to obtain the *linearized loss*, denoted $L^{\Delta}_{t+1} \approx L_{t+1}$.

DEF 1.13

The linearized loss, assuming differentiability of f along t and $\mathbf{Z}_t$, is given by

PROP 1.1

$$L^{\Delta}_{t+1} = -f_t(t, \mathbf{z}_t) - \sum_{i=1}^{d} f_{z_{t,i}}(t, \mathbf{z}_t) X_{t+1,i}$$

PROOF. Recall $L_{t+1} = f(t, \mathbf{z}_t) - f(t+1, \mathbf{z}_{t+1})$. We will approximate $f(t+1, \mathbf{z}_{t+1})$:

$$f(t+1, \mathbf{z}_{t+1}) \approx f(t, \mathbf{z}_t) + \nabla f(t, \mathbf{z}_t) \bullet (\langle t+1, \mathbf{z}_{t+1}\rangle - \langle t, \mathbf{z}_t\rangle)$$

This is

$$f(t, \mathbf{z}_t) + f_t(t, \mathbf{z}_t) \cdot 1 + \sum_{i=1}^{d} f_{z_{t,i}}(t, \mathbf{z}_t) X_{t+1,i}$$

Applying to L_{t+1} gives

$$L^{\Delta}_{t+1} = f(t, \mathbf{z}_t) - \left[f(t, \mathbf{z}_t) + f_t(t, \mathbf{z}_t) + \sum_{i=1}^{d} f_{z_{t,i}}(t, \mathbf{z}_t) X_{t+1,i} \right] = -f_t(t, \mathbf{z}_t) - \sum_{i=1}^{d} f_{z_{t,i}}(t, \mathbf{z}_t) X_{t+1,i}$$

as desired. □

Linearized loss is convenient and easy to understand, but the assumption of its accuracy is a heavy one. An easy improvement would be to consider higher-order Taylor approximations, but the estimation of the necessary derivatives may not be numerically stable.

Eg. 1.2 Following Example 1.1, we compute

$$\mathbf{X}_{t+1} = \mathbf{Z}_{t+1} - \mathbf{Z}_t = \left\langle \log \frac{S_{t+1,1}}{S_{t,1}}, \ldots, \log \frac{S_{t+1,d}}{S_{t,d}} \right\rangle$$

DEF 1.14 which, when $\mathbf{Z}_t = \log(\mathbf{S}_t)$, we call *log returns*. Rearranging L_{t+1}:

$$L_{t+1} = \sum_{i=1}^{d} \lambda_i e^{Z_{t,i}} - \sum_{i=1}^{d} \lambda_i e^{X_{t+1,i} - Z_{t,i}} = -\sum_{i=1}^{d} \lambda_i e^{Z_{t,i}} (e^{X_{t+1,i}} - 1) = -V_t \sum_{i=1}^{d} w_{t,i} (e^{X_{t+1,i}} - 1)$$

DEF 1.15 where $w_{t,i} = \frac{\lambda_i S_{t,i}}{V_t}$ is the *relative weight* of stock i at time t. We also consider $\rho_{t,i} = \lambda_i S_{t,i}$. In the language of operators:

$$l_{[t]}(\mathbf{x}) = -V_t \sum_{i=1}^{d} w_{t,i} (e^{x_i} - 1)$$

Following Def 1.13,

$$L^{\Delta}_{t+1} = -\sum_{i=1}^{d} \lambda_i e^{Z_{t,i}} X_{t+1,i} = -V_t \sum_{i=1}^{d} w_{t,i} X_{t+1,i} = -\boldsymbol{\rho_t} \bullet \mathbf{X}_{t+1}$$

Thus, if we assume that $\mathbf{X}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ follows the multivariate normal distribution,

$$\mathbb{E}[L^{\Delta}_{t+1}] = -V_t(\boldsymbol{w_t} \bullet \boldsymbol{\mu}) = -\boldsymbol{\rho_t} \bullet \boldsymbol{\mu}$$

with $\boldsymbol{w_t} = \langle w_{t,1}, ..., w_{t,d} \rangle$. Similarly,

$$\mathrm{Var}(L^{\Delta}_{t+1}) = V_t^2 \sum_{i=1}^{d} \mathrm{Var}(w_{t+1} X_{t+1,i}) = V_t^2 (\boldsymbol{w_t}^T \Sigma \boldsymbol{w_t}) = \boldsymbol{\rho_t}^T \Sigma \boldsymbol{\rho_t}$$

Note that Σ is a covariance matrix, so $\boldsymbol{w_t}^T \Sigma \boldsymbol{w_t}$ is indeed a scalar, as expected.

Eg. 1.3 Consider a portfolio with one standard European call on a stock S with maturity T and exercise price K. The value of European options is modeled by the Black-Scholes equation:

$$V_t = C^{BS}(t, S_t, r_t, \sigma_t)$$

where S_t and σ_t is the price and volatility of the underlying stock, respectively, and r_t is the risk-free interest rate. We write $\mathbf{Z}_t = \langle \log(S_t), r_t, \sigma_t \rangle$. The Black-Scholes equation satisfies

$$\frac{\partial C^{BS}}{\partial t} + \frac{\sigma^2}{2} S^2 \frac{\partial^2 C^{BS}}{\partial S^2} + rS \frac{\partial C^{BS}}{\partial S} - rC^{BS} = 0$$

Differentiating C^{BS} gives "the Greeks"

$$C_t^{BS}(\text{theta}) \quad C_r^{BS}(\text{rho}) \quad C_\sigma^{BS}(\text{vega}) \quad C_S^{BS}(\text{delta})$$

with $C_z^{BS} = S_t \times C_S^{BS}$. One can compute the linearized loss to be

$$L^{\Delta}_{t+1} = -C_t^{BS} - S_t C_S^{BS} X_{t+1,1} - C_r^{BS} X_{t+1,2} - C_\sigma^{BS} X_{t+1,3}$$

Note that

$$\mathbf{X}_{t+1} = \left\langle \log\left(\frac{S_{t+1}}{S_t}\right), r_{t+1} - r_t, \sigma_{t+1} - \sigma_t \right\rangle$$

Stylized Loan Portfolio

This is an example of a similar flavor to Example 1.2, but deserves a section of its own.

We will consider a stylized loan portfolio. Let the time horizon Δt be one year (with this horizon, we should account for the time value of money).

We lend to m obligors. For each obligor i, let k_i denote the amount to be paid at the end of the present time period, consisting of the loan principle plus interest. Since this payment is made at time $t+1$, $e_i = \frac{k_i}{1+r_{t,1}}$, called the exposure, is the present-day money value of this payment, where $r_{t,1}$ is the risk-free interest rate.

There is a possibility that an obligor defaults. Let $Y_{t,i}$ be an indicator variable, called the default state, that detects whether an obligor defaults by time t. We will assume $Y_{t,i} = 0$ and $\mathbb{E}[Y_{t+1,i}] = p_i$. In case of default, the lender may recover a portion of the loan, i.e. $(1-\delta_i)k_i$, where $\delta_i \in (0,1]$. The expected shortfall associated with obligor i is the difference between [the present-dollar value of the loan payment at time $t+1$] and [the present-dollar expected loan payment at time $t+1$]

$$\frac{k_i}{1+r_{t,i}} - \underbrace{\left[p_i(1-\delta_i)\frac{k_i}{1+r_{t,i}} + (1-p_i)\frac{k_i}{1+r_{t,i}}\right]}_{e_i(1-p_i\delta_i)} = p_i\delta_i\frac{k_i}{1+r_{t,i}} = p_i\delta_i e_i$$

The value of the loan itself, then, is the [present-dollar loan payment's value] minus the [expected shortfall], or $e_i - p_i\delta_i e_i$. Note that this is exactly [the present-dollar expected loan payment at time $t+1$], which we already computed. The value of whole the portfolio is given by

$$V_t = \sum_{i=1}^{m} e_i - p_i\delta_i e_i = \sum_{i=1}^{m} e_i(1-p_i\delta_i)$$

At time $t+1$, we assume that all obligors have paid off the principle. Hence, V_{t+1} encapsulates the actualized payoff made at time $t+1$. However, if the loan payments continued into a second year, we would *add on* to V_{t+1} an estimation of the expected loan payment at time $t+2$. In the former case, we have

$$V_{t+1} = \sum_{i=1}^{m} Y_{t,i}(1-\delta_i)k_i + (1-Y_{t,i})k_i = k_i(1-Y_{t,i}\delta_i)$$

We adjust for the time value of money in the next calculation:

$$L_{t+1} = V_t - \frac{V_{t+1}}{1+r_{t,1}} = \sum_{i=1}^{m} e_i(1-\delta_i p_i) - e_i(1-Y_{t,i}\delta_i) = \sum_{i=1}^{m} e_i\delta_i(Y_{t,i}-p_i)$$

Determining Loss Distributions

In order to determine the loss distribution (Def 1.12), one must model risk factor changes $\mathbf{X}_{t+1}$, given a known mapping $f(t,\mathbf{Z}_t) = V_t$. We distinguish between:

1. The conditional distribution of risk factor changes, as a function of all information up to time t. We call the resulting loss distribution is called the *conditional loss disrtibution*. Note that an estimation of risk factor changes which relies on historical data is *not* necessarily a conditional distribution. DEF 1.16

2. The stationary distribution of risk factor changes. The resulting loss distribution is called the *unconditional loss distribution*. DEF 1.17

VARIANCE-COVARIANCE METHOD

We will first consider an analytical method for the unconditional distribution of L_{t+1}. Assume that $\mathbf{X}_{t+1}$ has a multivariate normal distribution, and write $\mathbf{X}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Suppose also that the linearized loss, written as

$$L_{t+1}^{\Delta} = -(c_t + \mathbf{b}_t^T \mathbf{X}_{t+1})$$

This is not new: c_t denotes the partial of the realized value with respect to t, and $\mathbf{b}_t$ denotes the partials with respect to $\mathbf{z}_t$.

is sufficiently accurate, i.e. $L^{\Delta}_{t+1} = L_{t+1}$. Consequently,

$$L^{\Delta}_{t+1} \sim \mathcal{N}(-c_t - \mathbf{b}_t^T \boldsymbol{\mu}, \mathbf{b}_t^T \Sigma \mathbf{b}_t)$$

In this case, the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ are estimated from past data. Hence, inference about the loss distribution are made using these estimates:

$$L^{\Delta}_{t+1} \sim \mathcal{N}(-c_t - \mathbf{b}_t^T \hat{\boldsymbol{\mu}}, \mathbf{b}_t^T \hat{\Sigma} \mathbf{b}_t)$$

The estimation of L_{t+1} in this manner is called the *Variance-Covariance Method*. Of its advantages: it is easy to implement and understand, and it is a closed solution, which eases compute requirements. However, we reply on heavy assumptions. In particular, linearized loss is crude, and the normality assumption may seriously underestimate the tail of the loss distribution. DEF 1.18

Granted, we can use the t-distribution to achieve heavy tailed-ness

Note: this method is identical to the brief analysis we include at the end of Example 1.2.

HISTORICAL SIMULATION METHOD

This also estimates the *unconditional* loss distribution

Alternatively, we use an empirical distribution based on historical data $\mathbf{X}_{t-n+1}, ..., \mathbf{X}_t$. To do so, we construct the historically simulated loss data:

$$\tilde{L}_s = f(t, \mathbf{z}_t) - f(t+1, \mathbf{z}_t + \mathbf{X}_s) : s \in [t-n+1, t]$$

Then, we average these estimates to yield a final estimation of the loss distribution:

$$\mathbb{P}(L_{t+1} \leq x) \approx \frac{1}{n} \sum_{s=1}^{n} \mathbb{1}(\tilde{L}_{t-n+s} \leq x)$$

If $\mathbf{X}_s$ are IID, the convergence of the empirical distribution to the true distribution is ensured. (Real-life risk factor changes are not IID.) Note that this method can only estimate the unconditional loss distribution.

Past performance does not indicate future gains. The historical simulation method is akin to driving by only looking through the rear view mirror

MONTE CARLO METHOD

We create a model for risk factor changes, based on data $\mathbf{X}_{t-n+1}, ..., \mathbf{X}_t$, to simulate m new data points $\tilde{\mathbf{X}}^{(1)}_{t+1}, ..., \tilde{\mathbf{X}}^{(m)}_{t+1}$. Using this data, we construct simulated future loss data:

$$\tilde{L}_s = f(t, \mathbf{z}_t) - f(t+1, \mathbf{z}_t + \mathbf{X}^{(m)}_{t+1}) : m \in [M]$$

As before, we average these estimates into a final loss distribution

$$\mathbb{P}(L_{t+1} \leq x) \approx \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}(\tilde{L}^{(m)}_{t+1} \leq x)$$

This method can potential circumvent a lack of historical data with a better model, more simulations with the model (i.e. $M \ggg 0$), or more models. However, this method is potentially expensive, and it relies on training a suitably accurate model.

The Monte Carlo Method is akin to driving by using the rear view mirror to view a front-facing funny mirror. Slightly better.

RISK MEASURES

In order to simplify our view of a portfolio's exposure to risk, we employ mappings from risk distributions to $\mathbb{R}$, called risk measures. Historically, risk measurement has been comprised of the following methods:

Notational Risk is simply derived from the sum total of the values of the portfolio's securities.

Scenario-Based One predicts (e.g. via Markov chains) future scenarios, and measures the expected maximum future loss of the portfolio against these scenarios.

Loss-Based We provide statistical descriptions of the loss distribution, or an estimate of it.

Building on the previous section, we will primarily concern ourselves with the last method. However, we consider some examples for each.

Eg. 1.4 Using the notational method, we weigh assets by their riskiness. For instance, when measuring operational risk, we might use

$$\alpha \frac{\sum_{i=1}^{3} \max(G^{t-i}, 0)}{\sum_{i=1}^{3} \mathbb{1}(G^{t-i} > 0)}$$

where $\alpha \approx 0.15$ and G is gross annual income. This measure tracks the average gross income over the past 3 years, excluding years where incoming is 0 or negative. Heuristically, our operational risk (i.e. risk of loss by internal mismanagement) is proportional to cash flows processed by the institution.

Eg. 1.5 Using the scenario-based method, we consider some fixed risk factor changes $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$, with associated weights $\{w_1, ..., w_n\}$ describing their likelihood of occurring. A scenario-based risk assessment is

$$\psi(L) = \max\{w_i L(\mathbf{x}_i)\}_{i \in [n]}$$

where L is a loss function.

Eg. 1.6 The oldest loss-based measure simply used the variance of the $P\&L$ distribution as a risk measure. However, this approach requires the existence of a second moment.

DEF 1.19 A *coherent risk measure* is a function $\rho : \mathcal{M} \to \mathbb{R} : L \mapsto \rho(L)$ on the space of random variables $\mathcal{M}$ representing losses. We view $\rho(L)$ has the total amount of equity capital required to back a position with loss function L. It satisfies the following:

Monotonicity For all $L_1, L_2 \in \mathcal{M}$ with $L_1 \geq L_2$ almost surely,

$$\rho(L_1) \geq \rho(L_2)$$

Recall the notion of almost surely: $\mathbb{P}(L_1 \geq L_2) = 1$

Invariance For any $L \in \mathcal{M}$ and $\ell \in \mathbb{R}$,

$$\rho(L + \ell) = \rho(L) + \ell$$

Subadditivity For any $L_1, L_2 \in \mathcal{M}$, we have

$$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$$

Homogeneity For any $L \in \mathcal{M}$ and $\lambda \geq 0 \in \mathbb{R}$,

$$\rho(\lambda L) = \lambda \rho(L)$$

These axioms are informed by practice and financial intuition:

M. If we are confident that L_1 is a riskier position than L_2, then surely we would require more equity to back the first position

I.

S. A diversification of our portfolio should not create extraneous risk (if fact, this is usually considered a good thing). If this were not the case, an institution could reduce risk capital by splitting into subsidiaries.

H. $\rho(L + \cdots + L) \leq n\rho(L)$ by subadditivity. But we are not diversifying our portfolio, so it seems silly that we could reduce our risk by simply multiplying the size of our portfolio. Hence, this holds with equality. In fact, a "don't put your eggs in one basket" argument suggests $\rho(nL) > n\rho(L)$, but this is prevented by subadditivity. This last point has sparked much debate over the validity of the last two axioms.

A coherent risk measure is convex, i.e. for all $L_1, L_2 \in \mathcal{M}$ and $\lambda \in (0, 1)$ **PROP 1.2**

$$\rho(\lambda L_! + (1 - \lambda L_2)) \leq \lambda \rho(L_1) + (1 - \lambda)\rho(L_2)$$

However, the converse is not true.

PROOF. This follows immediately from subadditivity and homogeneity. □

Any function $\rho : \mathcal{M} \to \mathbb{R} : L \mapsto \rho(L)$, whether it satisfies the axioms of Def 1.19, is simply called a *risk measure*. DEF 1.20

Given a tolerance $\alpha \in (0, 1)$, *Value at Risk*, a primary risk measure, is defined to be DEF 1.21

$$\mathrm{VaR}_\alpha(L) = q_\alpha(F_L) = F_L^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$$

where F_L is the CDF of the loss distribution. In other words, $\mathrm{VaR}_\alpha(L)$ is the α-quantile of F_L. With probability at least α we would see losses in the range $(-\infty, \mathrm{VaR}_\alpha(L)]$.

Since $\mathrm{VaR}_\alpha(L)$ provides a boundary between "likely" and "unlikely" losses, we frequently put $\alpha = 0.95$.

Observe that $x_0 = \mathrm{VaR}_\alpha(L)$ if an only if $\mathbb{P}(L \leq x_0) \geq \alpha$ and $\mathbb{P}(L \leq x) < \alpha$ whenever $x < x_0$. **PROP 1.3**

$\mathrm{Var}_\alpha(L)$ is not coherent. In particular, it does not satisfy subadditivity, but it satisfies the other axioms of Def 1.19. **PROP 1.4**

PROOF. As proof of the fact that VaR does not satisfy subadditivity, we'll consider a worked (counter)example.

Consider a portfolio of 50 defaultable bonds with independent defaults. Default on each bond is IID $Y_i \sim \text{Ber}(0.02)$. The current price of bonds is 95, with a face value equal to 100. In **Portfolio A**, we buy 100 units of bond 1. In **Portfolio B**, we buy 2 units of each bond. The value of each portfolio is 9500.

Intuitively, we understand Portfolio A to be riskier. However, we'll see that VaR shows the opposite.

$$L_i = 95 - 100(1 - Y_i) \implies \mathbb{P}(L_i \le x) = \begin{cases} 1 & x \le 95 \\ 0 & x < -5 \\ 0.98 & x \in [-5, 95) \end{cases}$$

Then, with a tolerance of $\alpha = 0.95$, $\text{VaR}_\alpha(L_i) = -5$. We conclude that

$$\text{VaR}_\alpha(L^A) = \text{VaR}_\alpha(100L_1) = 100\text{VaR}_\alpha(L_1) = -500$$

Noting that $L^B = \sum_{i=1}^{50} 2L_i = -500 + 200\sum_{i=1}^{50} Y_i$, we have that

$$L^B \sim -500 + 200\text{Bin}(50, 0.02)$$

It follows that $\text{VaR}_\alpha(L^B) = -500 + 200\text{VaR}_\alpha(\text{Bin}(50, 0.02))$, which, one can calculate, is $-500 + 200 \cdot 3 = 100$. Hence, $\text{VaR}_\alpha(L^A) < \text{VaR}_\alpha(L^B)$, which contradicts subadditivity.

□

DEF 1.22 Alternatively, we use the *expected shortfall* as a risk measure, defined to be

$$\text{ES}_\alpha(L) = \mathbb{E}[L | L \ge \text{VaR}_\alpha(L)]$$

In other words, assuming that we incur losses above the α-quantile, how much should we expect to lose? This is akin to a bad-case scenario estimation. If $\alpha = 0.95$, $\text{ES}_\alpha(L)$ describes the expected loss among only extreme (probability < 0.05) losses.

PROP 1.5 If L is a continuous distribution, then $\text{ES}_\alpha(L)$ is a coherent risk measure.

PROOF. Axioms 1, 2, and 4 of Def 1.19 are inherited from $\text{VaR}_\alpha(L)$. For subadditivity, note that

$$\text{ES}_\alpha(L) = \frac{1}{1-\alpha}\mathbb{E}(L\mathbb{1}(L \ge \text{VaR}_\alpha(L)))$$

We will show

$$(1-\alpha)\left[\text{ES}_\alpha(L_1) + \text{ES}_\alpha(L_2) - ES_\alpha(L_1 + L_2)\right] \ge 0 \qquad \star$$

Let $I_i := \mathbb{1}(L_i \ge \text{VaR}_\alpha(L_i))$ for $i = 1, 2$, and $I_{12} = \mathbb{1}(L_1 + L_2 \ge \text{VaR}_\alpha(L_1 + L_2))$. Then

$$\begin{aligned} \star &\iff \mathbb{E}(L_1 I_1 + L_2 I_2 - (L_1 + L_2) I_{12}) \ge 0 \\ &\iff \mathbb{E}(L_1(I_1 - I_{12})) + L_2(I_2 - I_{12}) \ge 0 \end{aligned}$$

We show that

$$\begin{aligned}\mathbb{E}(L_1(I_1 - I_{12})) &\geq \mathbb{E}[\mathrm{VaR}_\alpha(L_1)(I_1 - I_{12})] = \mathrm{VaR}_\alpha(L_1)\mathbb{E}(I_1 - I_{12}) \\ &= \mathbb{E}[I_1] - \mathbb{E}[I_{12}] = \mathbb{P}(L_1 \geq \mathrm{VaR}(L_1)) - \mathbb{P}(L_1 + L_2 \geq \mathrm{VaR}_\alpha(L_1)) = 0\end{aligned}$$

The L_2 case is symmetric. If $I_1 = 1$, then $I_1 - I_{12} \geq 0$. Conversely, if $I_1 = 0$, then $I_1 - I_{12} \leq 0$. Hence

$$\mathbb{E}((L_1 - \mathrm{VaR}_\alpha(L_1))(I_1 - I_{12})) \geq 0 \qquad \square$$

PROP 1.6 If L has continuous density f, then

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_{\mathrm{VaR}_\alpha(L)}^{\infty} xf(x)dx$$

PROOF. We can view $L|L \geq \mathrm{VaR}_\alpha(L)$ as a variable with CDF

$$\begin{aligned}G(x) = \mathbb{P}(L \leq x | L \geq \mathrm{VaR}_\alpha(L)) &= \begin{cases} 0 & x < \mathrm{VaR}_\alpha(L) \\ \frac{\mathbb{P}(L \in [\mathrm{VaR}_\alpha(L), x])}{\mathbb{P}(L \geq \mathrm{VaR}_\alpha(L))} & \end{cases} \\ &= \frac{F_L(x) - F_L(\mathrm{Var}_\alpha(L))}{1-\alpha} = \frac{F_L(x) - \alpha}{1-\alpha} : x \geq \mathrm{VaR}_\alpha(L)\end{aligned}$$

Hence, $L|L \geq \mathrm{VaR}_\alpha(L)$ will have PDF $\frac{f(x)}{1-\alpha}$ when $x \geq \mathrm{VaR}_\alpha(L)$, and 0 otherwise. $\square$

Eg. 1.7 Suppose $L \sim \mathcal{N}(\mu, \sigma^2)$. We know $L \overset{d}{=} \mu + \sigma L^*$, where $L^* \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned}\mathbb{P}(L \leq x) = F_L(x) &= \mathbb{P}(\mu + \sigma L^* \leq x) \\ &= \mathbb{P}(L^* \leq \frac{x-\mu}{\sigma}) = \Phi\left(\frac{x-\mu}{\sigma}\right) =: \alpha\end{aligned}$$

where Φ is the CDF of $\mathcal{N}(0,1)$. Then

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \alpha \iff \frac{x-\mu}{\sigma} = \Phi^{-1}(\alpha) \iff x = \mu + \sigma\Phi^{-1}(\alpha)$$

We conclude that $\mathrm{VaR}_\alpha(L) = \mu + \sigma \mathrm{VaR}_\alpha(L^*) = \mu + \sigma\Phi^{-1}(\alpha)$.

Eg. 1.8 Similarly, under the same assumptions, we compute $\mathrm{ES}_\alpha(L)$. This is

$$\begin{aligned}\mathrm{ES}_\alpha(L) &= \mathbb{E}(\mu + \sigma L^* | \mu + \sigma L^* \geq \mathrm{VaR}_\alpha(L)) = \mathbb{E}(\mu + \sigma L^* | \mu + \sigma L^* \geq \mu + \sigma \mathrm{VaR}_\alpha(L^*)) \\ &= \mathbb{E}(\mu + \sigma L^* | L^* \geq \mathrm{VaR}_\alpha(L^*)) = \mu + \sigma \mathbb{E}(L^* | L^* \geq \mathrm{VaR}_\alpha(L^*)) \\ &= \mu + \sigma \mathrm{ES}_\alpha(L^*)\end{aligned}$$

We must now calculate $\mathrm{ES}_\alpha(L^*)$ directly, using Prop 1.3:

$$\mathrm{ES}_\alpha(L^*) = \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} \phi(x)x\,dx$$

where $\mathcal{N}(0,1)$ follows the PDF ϕ. But this has the property that $\phi'(x) = -x\phi(x)$, so we conclude

$$\mathrm{ES}_\alpha(L^*) = \frac{1}{1-\alpha}\left[-\phi(x)\right]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha}$$

and hence $\mathrm{ES}_\alpha(L) = \mu + \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$

DEF 1.23 The *shortfall-to-quatile ratio* is

$$\lim_{\alpha\to1}\frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)}$$

PROP 1.7 When $L^* \sim \mathcal{N}(0,1)$, we have

$$\lim_{\alpha\to1}\frac{\mathrm{ES}_\alpha(L^*)}{\mathrm{VaR}_\alpha(L^*)} = \lim_{\alpha\to1}\frac{\frac{1}{1-\alpha}\phi(\Phi^{-1}(\alpha))}{\Phi^{-1}(\alpha)} = \lim_{x\to\infty}\frac{\frac{1}{1-\Phi(x)}\phi(x)}{x} = \lim_{x\to\infty}\frac{\frac{\phi(x)}{x}}{1-\Phi(x)}$$

Using L'Hopital's rule,

$$\lim_{x\to\infty}\frac{\frac{\phi(x)}{x}}{1-\Phi(x)} = \lim_{x\to\infty}\frac{1+x^2}{x^2} = 1$$

Hence, when $L \sim \mathcal{N}(\mu,\sigma^2)$, the shortfall-to-quantile ratio also goes to 1.

Eg. 1.9 If L follows the Pareto distribution with parameter $\theta > 0$, i.e.

$$F_L(x) = 1 - x^{-\theta} : x \geq 1$$

From this, we derive the PDF $L(x) = \theta x^{-\theta-1} : x \geq 1$. Computing VaR_α:

$$1 - x^{-\theta} = \alpha \implies x = \mathrm{VaR}_\alpha(L) = (1-\alpha)^{-1/\theta}$$

Similarly,

$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_{(1-\alpha)^{-1/\alpha}}^{\infty} x\theta x^{-\theta-1}\,dx = \frac{1}{1-\alpha}\int_{(1-\alpha)^{-1/\alpha}}^{\infty}\theta x^{-\theta}\,dx$$

Note that this integral only converges when $\theta > 1$. Computing the integral gives

$$\mathrm{ES}_\alpha(L) = \frac{\theta}{\theta-1}(1-\alpha)^{-1/\theta} = \frac{\theta}{\theta-1}\mathrm{VaR}_\alpha(L)$$

Hence, $\frac{\mathrm{ES}_\alpha(L)}{\mathrm{VaR}_\alpha(L)} = \frac{\theta}{\theta-1}$

Suppose F_L is continuous and $\mathbb{E}[|L|] < \infty$. Then, for $\alpha \in (0,1)$ **PROP 1.8**

$$\mathrm{ES}_\alpha(L) = \int_\alpha^1 \mathrm{VaR}_u(L)du$$

II Financial Time Series

A *time series* is a collection $(X_t : t \in \mathbb{Z})$ of random variables. Some pertinent examples include log-returns, i.e. $\log\left(S_t/S_{t-1}\right)$, as well as *stylized facts*, which are derived from a collection of more elementary time series. DEF 2.1 DEF 2.2

We call a time series X_t *strictly stationary* if DEF 2.3

$$(X_{t_1}, \dots, X_{t_n}) \overset{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$$

for any choice of $\mathbf{t} \in \mathbb{Z}^n$ and $h \in \mathbb{Z}_+$, which we call *lag*. This is akin to the notion of time homogeneity of Markov chains. In particular, setting $n = 1$, X_t has the same distribution over all $t \in \mathbb{Z}$. In this case, we say that X_t follows a *stationary distribution*. DEF 2.4 DEF 2.5

Stationary $\neq$ IID. In particular, we may not (*probably do not*) have independence between time-steps.

For a time series X_t, with $\mathbb{E}[X_t^2] < \infty \; \forall t \in Z$, we define

$$\mu(t) = \mathbb{E}[X_t] \qquad \gamma(s,t) = \mathrm{cov}(X_t, X_s)$$

which we call the *mean function* and *autocovariance function*, respectively. DEF 2.6 DEF 2.7

For a series that satisfies $\mathbb{E}[X_t^2] < \infty \; \forall t \in Z$, we call it *covariance-stationary* if DEF 2.8

$$\mu(t) = \mu \; \forall t \in \mathbb{Z} \quad \text{and} \quad \gamma(s,t) = \gamma(s+h, t+h) \; \forall t, s, h \in \mathbb{Z}$$

Note that a strictly stationary time series which has a finite second moment is also covariance-stationary.

For a covariance-stationary series, $\gamma(0) = \mathrm{Var}(X_t) \; \forall t \in \mathbb{Z}$. **PROP 2.1**

PROOF.

$$\gamma(h) := \gamma(h, 0) = \mathrm{cov}(X_h, X_0) \; \forall h \in \mathbb{Z}$$

□

The *autocorrelation function*, also called serial correlation, is given by DEF 2.9

$$\rho(h) = \mathrm{cor}(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}$$

using the notation used in the proof above. In particular, $\rho(0) = 1$. Recall that

$$\mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

DEF 2.10 Let X have mean μ and variance σ^2. Suppose $\mathbb{E}[|Y|^3] < \infty$. Then the *skewness* of X is

$$\beta := \mathbb{E}\left[\left(\frac{Y-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}[(Y-\mu)^3]}{\mathbb{E}[(Y-\mu)^2]^{3/2}}$$

DEF 2.12 If $\beta < 0$, we say that X is *left skewed*. Similarly, when $\beta > 0$, we say that X is *right skewed*. A left skewed distribution typically has large mass on its right side, and similarly for a right skewed distribution. The naming convention describes how a left skewed variable's mean is to the left of both its median and mode.

DEF 2.13 Under the same conditions, suppose $\mathbb{E}[|Y|^4] < \infty$. Then the *kurtosis* of X is

$$\kappa := \mathbb{E}\left[\left(\frac{Y-\mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}[(X-\mu)^4]}{\mathbb{E}[(Y-\mu)^2]^2}$$

A large kurtosis indicates a higher concentration around the variable's mean.

PROP 2.2 If $Y \sim \mathcal{N}(\mu, \sigma^2)$, $\beta = 0$ and $\kappa = 3$.

We can easily find statistical estimates for β and κ as follows:

$$\beta_n := \frac{1/n \sum_{i=1}^n (X_i - \overline{X})^3}{\left[1/n \sum_{i=1}^n (X_i - \overline{X})^2\right]^{3/2}} \qquad \kappa_n := \frac{1/n \sum_{i=1}^n (X_i - \overline{X})^4}{\left[1/n \sum_{i=1}^n (X_i - \overline{X})^2\right]^2}$$

DEF 2.14 The *Jarque–Berra statistic* is

$$T_n = \frac{n}{6}\left(\beta_n^2 + \frac{1}{4}(\kappa_n - 3)^2\right)$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ are unknown, then $T_n \sim \chi_2^2$. Hence, we may use T_n to reject the null hypothesis that "X is normally distributed."

INDEX OF DEFINITIONS