In this lecture blog I will briefly and possibly incompletely summarize what we did in each lecture. I welcome feedback if this is useful for you!

Lecture 1 (May 6)

After going over the course outline, we jumped right into Chapter 1: Risk in perspective. We saw attempts to define the term “risk” and concluded that for us, risk means a chance of loss, which is why we use probabilistic tools throughout the course. We also looked at five different types of risk: Market Risk, Credit Risk, Operational Risk as well as Liquidity Risk and Model Risk. We saw that Risk Management (RM) is about “ensuring resilience to future events” and saw some goals of RM. We saw a brief and incomplete history of risk management and what lead to current practices. We then talked about regulation (Basel Framework) and criticism thereof. We finally saw why we absolutely need QRM (not only in finance!) and discussed important and interesting challenges that the QRM community has to tackle.

Everything from Chapter 1 is relevant (including, but not limited to, definitions, examples, advantages/disadvantages,…) except for Section 1.2 (Slides 8 - 14): Even though it is useful to have a bit of the history that led to crisis, regulation and innovation in mind, there won’t be any exam questions about this section.

Lecture 2 (May 8)

We started Chapter 2: Basic Concepts in Risk Management. By looking at the balance sheet, we identified risks that are faced by a financial firm (market risk, credit risk, maturity mismatch and interest rate risk, inflation risk, longevity risk for insurers). We looked at three different notions of capital (Equity capital, Regulatory capital, Economic capital).

We then started with Section 2.2: Risk factors and Loss Distributions. We defined a probabilistic framework to model losses. We have Risk factors ($\mathbf{Z}_t$: a random vector), risk factor changes ($\mathbf{X}_t$) which are mapped, combined with a time variable t, to the value of a portfolio, denoted by V_t, that is $V_t = f(t, \mathbf{Z}_t)$. V_t is a random variable that can be observed at time t. Using a loss operator L we were then able to define L_{t+1}, the one-period-ahead loss, as a random variable whose randomness arises from $\mathbf{X}_{t+1}$. The rv L_{t+1} is typically not linear in $\mathbf{X}_{t+1}$, so it can be hard to work with it. We therefore considered the linearized loss, L^{Δ}_{t+1}, which is a first order Taylor approximation to L_{t+1}. We then drew a “concept map” briefly summarizing the notation introduced and looked at a concrete Example: A stock portfolio with d stocks. Here we defined the mapping f and the risk factors $\mathbf{Z}_t$ and then derived the loss L_{t+1}, the loss operator L and the linearized loss L^{Δ}_{t+1}. We finally saw how L^{Δ}_{t+1} can be useful: It allowed us to easily approximate the mean and variance of the random variable L_{t+1}. As always, we need to be careful with approximations. That’s also where we stopped: We finished at Slide 15.

Lecture 3 (May 13)

We continued Chapter 2 on Basic Concepts in RM. We looked at another concrete example illustrating how we might model our loss in a specific case (here: a portfolio consisting of one option) as a function of risk factor (changes): This is example 2.2 on slide 16. After the example, we looked into *loss distributions* and identified three key tasks that QRM needs to tackle (finding a model for $\mathbf{X}_{t+1}$, approximate the cdf $F_{L_{t+1}}$ of L_{t+1}, compute a risk measure). To achieve this, there are three general methods (Analytical method, Historical Simulation, Monte Carlo). We looked at each of these methods sepearatley and discussed advantages and drawbacks of each method.
Having defined a mapping f, risk factor changes $\mathbf{X}_{t+1}$, the loss L_{t+1} and having some way of finding the distribution of the random variable L_{t+1} (eg via its cdf $F_{L_{t+1}}$), we next need to *measure risk*. A big part of Chapter 2 is actually related to risk measurement, as it is an incredibly important task. We saw why one would want to measure risks in practice (Determining amount of capital to hold as buffer, limitting traders, determining riskiness of insurance contracts. But there are many more!) and saw the mathematical definition of a risk measure: A function ϱ mapping random variables to real numbers. There are three general methods to measure risk (notional amount approach, scenario-based risk measures and risk measures based on loss distributions): We discussed all three of them, again with advantages and limitations. The rest of the chapter will deal with risk measures based on loss distributions, which are risk measures defined as some *characteristic* of the loss distribution (eg mean, Value-at-risk (later), variance,…).
We stopped at Slide 32 with the definition of a *generalized inverse*, which we will review next time.

Lecture 4 (May 15)

We are still in the middle of Section 2.3: Risk Measurement. We started on Slide 32 and reviewed the definiton of a generalized inverse and sketched a typical cdf along with its generalized inverse. We were then finally able to define the *Value-at-risk*, that is, for a loss L (a rv!)

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

This is a very important risk measure with both advantages and limitations. We calculated the value-at-risk for a normal and t distribution; those distributions were quickly reviewed as well and we saw that the t distribution may be better suited than the normal for modeling extreme losses. We then saw typical parameter choices for Δt and α as well as the daily risk capital formula in Basel II, which in fact uses the VaR_α. As always, we discussed advantages and disadvantages of VaR as a risk measure. It is always important to be critical!

We then defined another important risk-measure, namely the expected shortfall: For a loss rv L with $E(L_+) < \infty$,

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

We saw a nice interpretation of that risk measure (Prop. 2.1) and calculated the expected shortfall for a $N(\mu, \sigma^2)$ distribution. The $t(\mu, \sigma^2)$ case should be done as an exercise.

Lecture 5 (May 22)

We started by revisiting the definition of the expected shortfall and proved Proposition 2.1 which states that for a loss with *continuous* cdf, the expected shortfall is the conditional expectation of the loss given that the loss exceeds the value-at-risk. We also briefly saw how this proposition could have been used to show Example 2.4.
We compared the expected shortfall and the value at risk using the limiting shortfall-to-quantile ratio and worked through Example 2.5, in which we saw that only for large α, the normal model is "riskier" than the t-model. We then saw Lemma 2.2 which gives a nice interpretation of $\mathrm{ES}_\alpha(L)$ *and* an estimation procedure. Finally, we discussed advantages and disadvantes of the expected shortfall as a risk measure.

Moving forward, we started looking into *coherent and convex* risk measures. The goal here is to define a set of axioms that a "good" risk measure should satisfy. We defined what it means for a risk measure ϱ to be monotone, translation invariant, subadditive and positive homogeneous. We also interpreted those axioms and discussed criticisms. Finally, we said that a risk measure ϱ is coherent if it is monotone, translation invariant, subadditive and positive homogeneous.
At the end of the lecture, we looked at three risk measures that we already saw: Value-at-risk, expected shortfall and the generalized scenario risk measure. Value-at-risk is not coherent (Assignment 1), expected shortfall is coherent (proof next time) and the generalized scenario risk measure is coherent as well (proof next time). In fact, any coherent risk measure can be represented as generalized scenarios (Remark 2.3) - this gives nice mathematical structure. We stopped at slide 59.

Lecture 6 (May 27)

We continued our discussion of coherent risk measures. A coherent risk measure is monotone, translation invariant, subadditive and positive homogeneous. We proved subadditivity of ES_α using Lemma 2.2 - a pretty neat proof actually! You can show yourself that ES_α satisfies the other three axioms as well. This finishes the proof of Theorem 2.1 which states that ES_α is a coherent risk measure. We also showed that the generalized scenario risk measure $\Psi_{X,w}$ is subadditive - again, the other three axioms can be shown easily as well so that overall we proved Theorem 2.2 which states that the generalized scenario risk measure is coherent. In Assignment 1, you are verifying that the risk measure VaR_α is monotone, translation invarant and positive homogeneous. However, VaR_α is *not* subadditive, as we saw in two examples: One used two independent losses which have a skewed distribution, the other one used two independent losses having a heavy tailed distribution. Therefore, VaR_α cannot be subadditive.

Next, we defined *convex* risk measures as an alternative to coherent risk measures. Recall that a coherent risk measure is positive homogeneous, but positive homogeneity is sometimes critized as it does not consider liquidity risk (cf Slide 56). Convex risk measures on the other hand satisfy a convexity axiom, monotonicity and translation invariance. They do not satisfy positive homogeneity! If they did, they would actually be convex (Tutorial 3).

This concludes this Chapter and I gave a very brief overview of what we did in Chapter 2.

At the end of the lecture, we started Chapter 3: Empirical Properties of Financial Data. The goal here is to investigate empirical properties of financial risk factor change data (eg log-return data). We look at so called *stylized facts* which are a collection of empirical observations that apply to many time series of risk factor changes. Those are crucial to understand and keep in mind! We looked at weekly log-returns of `SP500` and saw that this data-set shows extreme events and volatility clustering. If we fit a normal distribution to the data and simulate from the fitted distribution, neither extreme events nor volatility clustering is present. Doing the same with a t distribution we saw that the latter does show extreme events, but no volatility clustering - which is clear, since we simulated iid data. We stopped at slide 6.

Lecture 7 (May 29)

We continued Chapter 3 about empirical properties of financial data. We looked at weekly log-returns for `SP500` and saw that the data show extreme events and volatility clustering. We then defined what it means for a process $(Z_t)_{t\in\mathbb{Z}}$ to be *covariance stationary* and defined the autocorrelation function

$$\mathrm{ACF}(h) = \mathrm{corr}(X_0, X_h).$$

An investigation of the estimated ACF for both the real data-set and the simulated data revealed no serial correlation. However, moving to *absolute* returns, we saw that there is profound serial correlation for *absolute returns* in the data, indicating that there is a tendency for large/small absolute returns to be followed by large/small absolute returns; in other words, returns cluster and are not iid. This is not true for the simulated data-set, since the simulated data are iid. We looked at the largest losses and then reviewed formal and graphical tests for testing wether data come from a certain distribution with cdf F. We summarized six univariate stylized facts and then moved to multivariate stylized facts. We looked at correlation between and within series as well as *tail dependence*, a concept that will be more properly defined in Chapter 7. We then summarized multivariate stylized facts which concludes Chapter 3.

We then tackled Chapter 4: Financial Time Series. This chapter shall just give a very rough introduction into time series modelling and how that can be used for risk management purposes. We defined an $\mathrm{ARMA}(p_1, q_1)$ process with $\mathrm{GARCH}(p_2, q_2)$ errors and saw the very rough idea of how time dependent risk measures can be estimated in a time series setting. Since we only covered the notation very superficially, Chapter 4 is not relevant for either exam - I still hope that it convinced you that the stylized facts we saw in Chapter 3 can indeed be modelled and that those models can be used in QRM.

Lecture 8 (June 3)

We looked at two `R` scripts, namely `R_Standard_methods_for_market_risk.R` and `R_Nonparametric_VaR_and_ES_estimators.R` (both can be found on Learn).

The first script puts us in the setting of Example 2.1. We have a portfolio of two stocks (BMW and SIEMENS) for which we have real stock data and we want to estimate VaR_α and ES_α for the next period, based on the past data, using different methods, namely

- Historic simulation
- Variance-covariance method
- Monte Carlo (based on Normal distribution)
- Monte Carlo (based on t distribution)
- POT (Peaks over threshold, will come in Chapter 5)

We saw how those methods can be implemented in `R` and how they can be used to estimate risk measures of a one period loss.

The main goal of the second script was to investigate the performance of non-parametric VaR_α and ES_α estimators as a function of α. We also constructed *confidence intervals* using *Bootstrap*, a method which can be used to approximate the distribution of estimators.
We saw that for large α, is is hard to (non-parametrically) estimate VaR_α and ES_α, the latter being even harder than the former. This is natural given that with n data points, the empirical cdf is a step function with steps of size $1/n$. It also highlights that one should not "blindly" believe in point-estimates!

Lecture 9 (June 5)

Today, we started Chapter 5: Extreme Value Theory. The motivation here is that QRM is concerned with *extreme* and *maximal* losses, as those are the ones that can have a severe impact on the solvency of a (financial) company.

We are interested in asymptotic properties of the *block maximum*, defined by $M_n = \max\{X_1, \dots, X_n\}$ for a sequence of iid rv's X_i with cdf F. We reviewed convergence concepts and asymptotic properties of the sample mean. For the block maximum, we saw that $M_n \xrightarrow{as} x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}$ and then wanted to see if we can get a "CLT like" theorem for M_n. To this end, we defined the maximum domain of attraction: IF we find sequences $(c_n) > 0$ and (d_n) such that $\frac{M_n - d_n}{c_n}$ converges in dist'n to a non degenerate rv with cdf H, then F (the cdf of the X_i) is said to be in the maximum domain of attraction of H, i.e. $F \in \mathrm{MDA}(H)$. In fact, we even know what H can look like: The Fisher-Tripett-Gnedenko theorem tells us that IF $\frac{M_n - d_n}{c_n}$ for some sequences $(c_n) > 0$ and (d_n) converges in distribution to a non-degenerate rv, the limit MUST be of GEV (generalized extreme value) type. We looked at this family of distributions, denoted by H_ξ, and the special cases $\xi > 0$, $\xi = 0$ and $\xi < 0$.

As a next step, we would like some easily verifiable conditions to show that a distribution F is in $\mathrm{MDA}(H_\xi)$, without having to find sequences (d_n) and (c_n). To this end, we defined *slowly* and *regularly varying* functions. At the end of the lecture, we saw Theorem 5.4: It states that $F \in \mathrm{MDA}(H_\xi)$ for $\xi > 0$ (Frechet case) if and only if $\bar{F}(x) = x^{-1/\xi} L(x)$ for a slowly varying function L. Not only does that give us a nice characterization of distributions in $\mathrm{MDA}(H_\xi)$ for $\xi > 0$ (this class contains distributions F that are heavy tailed (with a power-like tail) and $x_F = \infty$), the theorem even gives us the appropriate normalizing sequences (c_n) and (d_n).

Lecture 10 (June 10)

We picked up some of the remaing parts of Assignment 1 (Questions 1, 4, 5a-d). Questions 2, 3, 6 and 8 were discussed in the tutorial.

Lecture 11 (June 12)

We finished the discussion of Assignment 1. We then went back to Chapter 5: Extreme Value Theory. After reviewing what we had so far, we looked at Theorems 5.4, 5.5, 5.6. All those theorems are similar in structure and give handy necessary and sufficient conditions to check if a certain distribution belongs to $\mathrm{MDA}(H_\xi)$ for $\xi > 0$, $\xi = 0$ and $\xi < 0$. Not only do those Theorems give us conditions on a cdf F to belong to those classes, they also give us the form of the normalizing constants (c_n) and (d_n). Furthermore, we were able to characterize those classes: $\mathrm{MDA}(H_\xi)$ for $\xi > 0$ (Frechet class) contains heavy tailed distributions (tails decay like power functions) for which the existence of moments depends on the value of $1/\xi$. All distributions in this class have an infinite right endpoint. $\mathrm{MDA}(H_0)$ ($\xi = 0$, Gumbel class) contains a variety of distributions whose tails decay essentially like an exponential function. All moments exist for $F \in \mathrm{MDA}(H_0)$ and both, $x_F = \infty$ and $x_F < \infty$ is possible. The conditions to verify (Thm. 5.5) are more complicated here. Finally, the Weibull class, $\mathrm{MDA}(H_\xi)$ for $\xi > 0$ contains only short tailed (i.e. $x_F < \infty$) distributions. We summarized what we found so far and stopped after looking at an illustration of the Block Maxima Method (Slide 24).

Lecture 12 (June 17)

Midterm

Lecture 13 (June 19)

We continued studying the Block Maxima Method (BMM), where a dataset is split into m blocks of size n. If n is big enough, assuming that the data come from $F \in \mathrm{MDA}(H_\xi)$, the block maxima (i.e. the maxima in each block) can be assumed to come from $H_{\xi,\mu,\sigma}$. We can then maximize the log-likelihood of the block maxima and obtain estimates $\hat{\xi}$, $\hat{\mu}$ and $\hat{\sigma}$. There is a bias-variance trade-off when choosing n and m that needs to be kept in mind. We then defined the *return level* $r_{n,k}$ and *return period* $k_{n,u}$. As an example, we looked at SP500 data up to Black Monday and saw that we didn't have enough data to forsee an extreme event as it happened on Black Monday.

A big disadvantage of the BMM method is that it is fairly wasteful of data (only the m block maxima are used). This motivated the peak-over-threshold method (POT), where one looks only at those data above a certain high threshold u and models the excess above u. To this end, we defined the generalized Pareto distribution, saw some properties and noted that $G_{\xi,\beta} \in \mathrm{MDA}(H_\xi)$. We defined the *excess distribution over the threshold* u to be $F_u(x) = P(X - u \leq x \mid X > u)$ which will be crucial in the rest of the section. In fact, Pickans-Bamkema-de Haan Theorem tells us that there is a positive, measurable function $\beta(\cdot)$ so that

$$\lim_{u \to x_F} \sup_{0 \leq x < x_f - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$$

IFF $F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$. Thus, if $F \in \mathrm{MDA}(H_\xi)$, in the limit for large u, we can approximate F_u as $G_{\xi,\beta}$ - no matter what F is, as long as $F \in \mathrm{MDA}(H_\xi)$. This motivates the POT method, where the log-likelihood of the excesses (those observations above u and then u subtracted) is maximized in order to estimate the GPD parameters ξ and β. We stopped at slide 42.

Lecture 14 (June 24)

We continued and finished Chapter 5.2. We reviewed the Peaks-over-threshold method where given some threshold u, the excesses above u are used to fit the GPD parameters β and ξ. One question arises naturally: How should we pick the threshold u? We know that if $F \in \mathrm{MDA}(H_\xi)$, then by Pickans-Bamkema-de Haan, $\lim_{u \to x_F} \sup_{0 \leq x < x_f - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$. That is: Only for large u are we justified to approximate F_u by the GPD distribution $G_{\xi,\beta}$. In order to derive a graphical method of finding u, we saw in Lemma 5.1 that if $F_u = G_{\xi,\beta}$, $F_v = G_{\xi,\beta+\xi(v-u)}$ for any $v \geq u$ and the mean excess function is then

$$e(v) = E(G_{\xi,\beta+\xi(v-u)}) = \frac{\xi}{1-\xi} v + \frac{\beta - \xi u}{1-\xi}$$

which is linear in $v \in [u, x_F)$. Thus: IF $F_u = G_{\xi,\beta}$, then $e(v)$ is linear for $v \geq u$. This forms the basis for a graphical method of choosing u. We construct a sample mean excess plot ($e_n(v)$ is a non-parametric estimator for $e(v)$) and choose the threshold u as the smallest point where $e_n(v)$ becomes linear. The idea is that if the data support a GPD model over a high threshold, Lemma 5.1 suggests that $e(v)$ (estimated by $e_n(v)$) should become increasingly linear.
It should be emphasized that the sample mean excess plot is rarely perfectly linear and that choosing u is not easy: Again, we face a bias-variance tradeoff. Choosing u is as difficult as choosing the block size in the BMM method.
We then analyzed a concrete data-set (Danish Fire Loss Data) as an example. We found a threshold u and then estimated the GPD parameter using excesses over u. The resulting estimates can then be used to model F_u and F_v for any $v \geq u$ (using Lemma 5.1). But maybe we want to do more: We also want to estimate the unconditional distribution function and using that, we want to estimate risk measures such as VaR_α and ES_α. To this end, we saw on Slide 53 that if $F_u = G_{\xi,\beta}$, we get

$$\bar{F}(x) = \bar{F}(u)\left(1 + \xi \frac{x-u}{\beta}\right)^{-1/\xi}$$

for $x \geq u$. This can be used to derive formulas for VaR_α and ES_α. If we estimate $\bar{F}(u)$ by N_u/n, we immediately get estimates for $\bar{F}(x)$ (for $x \geq u$) as well as estimates for VaR_α and ES_α for $\alpha \geq F(u)$ (since then $\mathrm{VaR}_\alpha \geq u$).

Lecture 15 (June 26)

I first gave a high-level overview over Chapter 5: Extreme Value Theory (see 'Additional Material' on Learn). We then started Chapter 6 about multivariate models and reviewed basics of multivariate modelling.
Recall that we are interested in modelling losses and in Chapter 2 we saw that we can model

$$L_{t+1} = L(\mathbf{X}_{t+1})$$

where L is the loss operator and $\mathbf{X}_{t+1}$ is a d dimensional *random vector* of risk factor changes. In Chapters 5 and 6, we look in more detail into random vectors and how those can be used for modelling in risk management. To this end, we reviewed basic concepts such as joint/marginal distribution/density functions, the notion of independence, the definition of mean vector and covariance/cross-covariance matrix as well as properties thereof.

Lecture 16 (July 2)

We continued studying multivariate models. We first finished the review section by looking into standard estimators of μ, Σ and P and discussing characteristic functions. We then looked in more detail into the multivariate normal distribution, which has shown up quite a few times so far! We defined the multivariate normal distribution via its stochastic representation, but also saw several equivalent definitions. We derived a lot of useful and handy properties of the multivariate normal distribution: Margins are normal, sums (or in general: linear combinations) of components are normal, components of a multivariate normal dist'n are independent if and only if they are uncorrelated (this does not hold in general!), conditional distributions are normal, convolutions are normal and the distribution of the quadratic form $(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu)$ is a χ^2_d, which is useful for GOF testing. We then discussed how we can statistically test wether or not a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is drawn from the multivariate normal distribution, both formally (via Mardia's test) and graphically (via QQ Plots). We stopped at Slide 32.

Lecture 17 (July 3)

We summarized what we learned about the multivariate normal (MVN) distribution and discussed advantages and disadvantages of the MVN distribution. The main drawbacks of the MVN for RM purposes are that univariate and joint tails are too thin, the MVN generates too few extreme events (jointly and marginally). Another drawback is the radial symmetry.

In order to adress the shortcomings, while still keeping some of the desirable properties of the MVN, we want to construct *classes* of multivariate distributions. One class of distributions that generalizes the MVN are so called normal variance mixtures. Those are obtained by randomizing Σ with a non-negative random variable W, ie

$$\mathbf{X} = \mu + \sqrt{W} A \mathbf{Z}$$

where $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$ is independent of $W \geq 0$. Note that

$$\mathbf{X} \mid W \sim N_d(\mu, W\Sigma)$$

This property is very useful in proving properties of normal variance mixtures. We studied these distributions in more detail and calculated $E(\mathbf{X})$, $\mathrm{cov}(\mathbf{X})$ as well as the characteristic funtion $\varphi_{\mathbf{X}}$ and the density of $\mathbf{X}$, provided Σ has full rank (ow does not exist). We saw that if $\mathbf{X}$ is an uncorrelated normal variance mixture, components are independent if and only if W is constant (in which case we obtain a MVN). We also saw a sampling algorithm (which immediately follows from the stochastic representation). We saw that normal variance mixtures are closed under linear transformations, which is very useful for RM purposes. We concluded this section by considering some examples, such as the multivariate t dist'n (very important example). We briefly defined normal mean-variance mixtures as well, but didn't go into much detail.

We can even generalize multivariate normal variance mixture distributions: Those belong to the wider class of elliptical distributions, as will be seen later. In order to construct elliptical distributions, we need the notion of spherical distributions. A random vector $\mathbf{Y}$ is called spherical if $\mathbf{Y} \overset{d}{=} U\mathbf{Y}$ for all orthogonal matrices U (ie U such that $UU^T = I_d$). There are many different ways of characterizing these distributions, as seen in Theorem 6.1: Either via the characteristic function ($\varphi_{\mathbf{Y}}(\mathbf{t}) = \psi(||\mathbf{t}||^2)$ for some characteristic generator $\psi : [0, \infty) \to \mathbb{R}$) or via $\mathbf{a}^T\mathbf{Y} \overset{d}{=} ||\mathbf{a}||Y_1$ (meaning that linear combinations are of "the same type"). Note that the characteristic function (as well as the density, if it exists) is constant on spheres, hence spherical distributions. We also saw that $\mathbf{Y} \sim S_d(\psi)$ (ie a spherical dist'n in dimension d with characteristic generator ψ) can be represented as

$$\mathbf{Y} = R\mathbf{S}$$

where $\mathbf{S}$ is uniformly distributed on the sphere and $R \geq 0$ is independent of $\mathbf{S}$. This result is extremely useful as it implies (Cor. 6.1) that $(||Y||, Y/||Y||) \overset{d}{=} (R, \mathbf{S})$. Note that this implies that any random vector $\mathbf{Y}$ (as long as $\mathbf{Y} \sim S_d(\psi)$) can be used to sample from the sphere. We stopped at Lemma 6.3 where we calculate $E(\mathbf{S})$, $\mathrm{cov}(\mathbf{S})$ as well as $\mathrm{cov}(\mathbf{Y})$ and $\mathrm{corr}(\mathbf{Y})$.

Lecture 18 (July 8)

We continued studying spherical and elliptical distributions. We defined elliptical distributions to be a multivariate affine transformations of spherical distributions: That is, $\mathbf{X}$ follows an elliptical distribution if, in distribution,

$$\mathbf{X} = \mu + A\mathbf{Y}$$

where $Y \sim S_d(\psi)$ is a spherical random vector. A lot of properties about elliptical distributions actually follow from the properties of the (underlying) spherical distribution: Stochastic representation, characteristic function, distribution of the quadratic form $(\mathbf{X} - \mu)^T\Sigma^{-1}(\mathbf{X} - \mu)$, density (if Σ positive definite), linear combinations, convolutions, marginal distribution functions and conditional distributions. Ultimately, many (not all) nice and desirable properties of $N_d(\mu, \Sigma)$ still hold for elliptical distributions while providing more flexible models with different joint/marginal tail behaviours. We then proved Proposition 6.7 which shows that the risk measure value-at-risk is subadditive in elliptical models.

This finishes Chapter 6 and I concluded Chapter 6 by giving a brief, high level overview thereof.

We then started our last chapter, Chapter 7, about copulas. The main motivation will become clearer once we saw Sklar's theorem, but the main idea is that we would like to study dependence independent of margins. That will become more clear next time! We defined the term *copula*, which is a $d-$dimensional distribution function with standard uniform univariate margins. That is: Every one-dimensional margin (!) is $U(0,1)$ distributed. We saw three examples of copulas and Proposition 7.1, which gives a mathematical characterization of copulas: A function $C : [0,1]^d \to [0,1]$ is a copula if and only if C is grounded, has standard uniform univariate margins and is $d-$increasing. $d-$increasingness ensures that the probability of a random vector $\mathbf{U} \sim C$ falling into any rectangle $(\mathbf{a}, \mathbf{b}]$ is always non-negative. $d-$ increasingness can be nicely illustrated in $d = 2$, which we did using a sketch.

Lecture 19 (July 10)

We continued our discussion of copulas. We reviewed the probability transform ($F(X) \sim U(0,1)$ for $X \sim F$, F continuous) and the quantile transformation ($F^{\leftarrow}(U) \sim F$ for $U \sim U(0,1)$ and any F) - those are key to all applications involving copulas!

We then discussed and proved Sklar's Theorem: Given any multivariate df F and margins $F_1, \dots, F_d$ there is a copula C such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

where C is uniquely defined on $\prod_{j=1}^{d} \mathrm{ran}(F_j)$ and (there) given by

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)).$$

Conversely, given any copula C and univaraite dfs $F_1, \dots, F_d$, above defined F is indeed a df with margins $F_1, \dots, F_d$.

Copulas allow us thus to study dependence (via C) independent of the margins; furthermore, Sklar's theorem gives a bottom-up apprach for building multivariate models.
We then discussed Lemma 7.2: If $X_j \sim F_j$, F_j continuous for $j \in \{1, \dots, d\}$, then

$$\mathbf{X} \text{ has copula } C \Leftrightarrow (F_1(X_1), \dots, F_d(X_d)) \sim C$$

That is, we can study dependence via the margin-free $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$ which has the same copula as $\mathbf{X}$.
Theorem 7.2 (invariance principle) statest that if $\mathbf{X} \sim F$ with continuous margins $F_1, \dots F_d$ and copula C, T_j are strictly increasing functions on $ran(X_j)$ for $j \in \{1, \dots d\}$, then $(T_1(X_1), \dots, T_d(X_d))$ also has copula C. Hence, applying strictly increasing functions on margins doesn't change the copula. We illustrated this with a few examples.

We then saw Theorem 7.3 about the Frechet Hoeffding bounds: Any copula C is bounded by M and W. Note that we saw the copulas M and W (remember that M is only a copula in $d = 2$) before, so this Theorem tells us why they are "special".
As examples of copulas we then discussed fundamental copulas and their properties (Π, M and W) as well as implicit copulas, notably Gauss and t– copulas as the probably most prominent examples. We stopped at Slide 31.

Lecture 20 (July 15)

We continued looking at examples of copulas. Having introduced fundamental and implicit copulas last time, we introduced Archimedean copulas as examples of *explicit copulas*. A copula C is called *Archimedean* if it admits the form

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$$

for $\mathbf{u} \in [0,1]^d$ where $\psi \in \Psi$ is an Archimedean generator (ie a continuous, decreasing function $\psi : [0, \infty) \to [0,1]$ with $\psi(0) = 0$ and $\lim_{x \to \infty} \psi(x) = 0$ which is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$). All properties of C are related to the univariate generator ψ. We will see those copulas in more details later; for now, we looked at two examples (Gumbel and Clayton) and looked at some advantages and disadvantages of using Archimedean copulas.

After looking at examples of copulas, we discussed how copulas and meta distributions can be simulated. This is very important: Often, one would specify a certain model on the risk factor changes $\mathbf{X}$ (via a copula C and margins $F_1, \dots, F_d$, for instance) and then simulate from $\mathbf{X}$ to estimate certain quantities of interest (such as ES_α, for instance). In order to simulate from $\mathbf{X}$, it suffices to simulate from the underlying copula C and then one can apply $F_i^{\leftarrow}$ marginally to transform the margins (without changing the copula by the invariance principle).

A general algorithm to simulate from any copula C is the *conditional distribution method*, which we saw in Theorem 5.4. In order to obtain a realization $\mathbf{U} = (U_1, \dots, U_d) \sim C$, one simulates $U_1 \sim U(0,1)$, then one simulates from U_2 conditional on U_1, from U_3 conditional on U_1 and U_2 and so on. That is, starting with $\mathbf{U}' \sim U(0,1)^d$, an observation $\mathbf{U} \sim C$ is obtained by setting $U_1 = U_1'$ and
$U_j = C_{j|1,\dots,j-1}^{\leftarrow}(U_j' \mid U_1, \dots, U_{i-1})$ for $j = 2, \dots, d$. The conditional distribution functions

$$C_{j|1,\dots,j-1}(u_j \mid u_1, \dots, u_{j-1}) = \mathbb{P}(U_j \le u_j \mid U_1 = u_1, \dots, U_{j-1} = u_{j-1})$$

can, for sufficiently smooth (differentiable) copulas C be obtained using Theorem 7.5.

While the conditional distribution method is general in that it is applicable for any copula C, faster sampling algorithms can be obtained for certain copula classes. For instance, we saw that sampling implicit copulas (eg Gauss, t) is straigtforward: Sample from the multivariate distribtution, then apply the corresponding marginal univariate dfs.
Many Archimedean copulas can be sampled efficiently as well: We will see more about that in Section 7.4.
Having discussed how copulas can be sampled from, sampling from meta-C models is a straightforward application of the quantile transform (cf Algorithm 7.2).
We finished Section 7.1 by introducing a few more concepts such as survival copulas, radian symmetry, exchangeability and copula densities.

Lecture 21 (July 17)

We started Section 7.2 about dependence concepts and measures. We first studied the counter- and comonotonicity copulas M and W - they model perfect positive dependence and perfect negative dependence as seen in Prop. 7.2 (note that showing "$\Rightarrow$" is an easy exercise). In terms of subadditivity of the risk measure VaR_α this implies that if $X_1, \dots, X_d$ are comonotone (ie their copula is M),
$VaR_\alpha(\sum_{j=1}^d X_j) = \sum_{j=1}^d VaR_\alpha(X_j)$ so that VaR_α is *comonotone additive*. We proved this in Prop. 7.2 and also discussed implications of that result from a RM perspective.
We then studied the linear (or Pearson) correlation coefficient ϱ. A few main points that we made is that ϱ does not exist for all pairs of random variables (eg not defined for the multivariate t with df $0 < \nu \le 2$), it depends on the marginal distributions and not only on the underlying copula, it is invariant under strictly increasing linear transformations (good exercise to prove this!) but not invariant under strictly increasing functions in general and independence implies uncorrelatedness. Note also that $\varrho \in [-1.1]$ and $\varrho(X_1, X_2) = \pm 1$ if and only if $X_2 = aX_1 + b$ almost surely for $a, b \in \mathbb{R}$, $a \neq 0$.
We moved on by looking at three correlation fallacies: The first one is that F_1, F_2, ϱ uniquely determine F - this is in general wrong. The second fallacy argues that given marginal dfs F_1, F_2, any $\varrho \in [-1, 1]$ should be attainable - again, this is wrong in general (but true for elliptical models). Using Hoeffdings formula and Sklar's theorem, by considering the copulas M and W, one can actually determine bounds on the correlation coefficient: That is, one can calculate numbers ϱ_{min} and ϱ_{max} so that given margins F_1, F_2, $\varrho \in [\varrho_{min}, \varrho_{max}]$. This attainable range can be quite small, as we saw in Example 7.8. The third fallacy was concerned with the value-at-risk of a sum: If ϱ is maximal, then $VaR_\alpha(X_1 + X_2)$ should be maximal. While this is true for elliptical models (cf Prop. 6.7 and Prop. 7.3) it is wrong in general: We saw plenty of counterexamples in Chapter 2.
All those fallacies demonstrate that the correlation coefficient has some undesirable properties. As alternatives we studied rank correlation

coefficients: *Kendall's tau* ϱ_τ which is the probability of concordance minus the probability of discordance and *Spearman's rho*, which is defined as $\varrho_S = \varrho(F_1(X_1), F_2(X_2))$ (this is just the linear correlation of the margin free $\mathbf{U} = (F_1(X_1), F_2(X_2))$). We gave integral formulas for both measures. Note that both, ϱ_s and ϱ_τ, only depend on the underlying copula, are always defined and can be easily estimated.

Lecture 22 (July 22)

We continued our discussion of depdence concepts and measures. We reviewed what we learned last time about Spearman's rho and Kendall's tau and finished the discussion of rank correlation coefficients.
As a next step, we want to study *tail dependence*. We have (loosely) used the word tail dependence quite a few times so far, and now we formally define it. The upper tail dependence coefficients λ_u of a bivariate random vector (X_1, X_2) with continous margins is defined as the limiting probability that X_2 exceeds its u quantile given that X_1 exceeds its u quantile for u growing to 1, ie

$$\lambda_u = \lim_{u \uparrow 1} \mathbb{P}(X_2 > F_2^{\leftarrow}(u) \mid X_1 > F_1^{\leftarrow}(u)).$$

Similarly one defines the lower tail dependence coefficient λ_l as

$$\lambda_l = \lim_{u \downarrow 0} \mathbb{P}(X_2 \leq F_2^{\leftarrow}(u) \mid X_1 \leq F_1^{\leftarrow}(u)).$$

Note that λ_l and λ_u, being probabilities, are between 0 and 1. We speak of lower (upper) tail dependence, if λ_l (λ_u) is positive, otherwise we speak of lower (upper) tail independence. We showed in Lemma 7.3 how these quantities can be calculated and also noted the important fact that tail dependence is a copula property, it does not depend on the marginal distributions. There are extensions to $d > 2$, but we won't study them.
This finished the discussion of Section 7.2 and we moved on to talk in a bit more detail about normal mixture copulas, which are implicit copulas constructed from normal variance mixture distributions. That is, we start from a normal variance mixture

$$\mathbf{X} \overset{d}{=} \sqrt{W} A Z$$

where $AA^\top = P$ is a correlation matrix, $W \geq 0$ is a rv independent of $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$ and then extract the underlying copula via Sklar's theorem. In these models, there is an easy formula for Kendall's tau as a function of the linear correlation ϱ. Such a formula has not (yet?) been found for Spearman's rho. This allows us to transform the correlation matrix to a matrix consisting of pairwise Kendall's tau and vice versa.
In terms of tail dependence, using results we derived in the end of Section 7.2, upper and lower tail dependence are equal in normal variance mixture copulas (by radial symmetry!) and can be calculated using Formula (3) on Slide 87.

$$\lambda = \lambda_l = \lambda_u = 2 \lim_{x \downarrow -\infty} \mathbb{P}(X_2 \leq x \mid X_1 = x).$$

We used this formula to calculate λ for the Gaussian copula and saw that the Gaussian copula is tail-independent (unless $\varrho = 1$). The t–copula, as expected, has tail dependence for any $\varrho > -1$ and the tail dependence is increasing in ϱ and decreasing in ν. In general, it can be shown that in a normal variance mixture copula, tail dependence is present only when W has a power tail - keeping in mind that W affects all components of $\mathbf{X}$, this makes intuitive sense, as in those cases, the distribution of W is more heavy tailed. We finished by looking at joint quantile exceedence probabilities for the Gaussian and the t copula - we will start again on Slide 90.

Lecture 23 (July 24)

We started by looking at joint tail probabilities

$$\mathbb{P}(U_1 > u, \dots, U_d > u)$$

for the Gaussian and the t-copula, as a function of u and as a function of d for different combinations of ϱ and ν. Numerical results revealed that this probability is *much* bigger when a t-copula is used as oppsed to a Gaussian copula: This again highlights the importance of tail dependence.
We then reviewed advantages and disadvantages of Normal Mixture Copulas.
So far, we have mostly discussed implicit copulas (such as normal mixture copulas). As an example of explicit copulas - i.e. copulas that are explicitly constructed rather than extracted from known multivariate distributions - we already saw the basic definition of Archimedean copulas. In Section 7.4 we look at them a bit more closely. Recall that a copula C is called *Archimedean* if it admits the form

$$C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d))$$

for $\mathbf{u} \in [0,1]^d$ where $\psi \in \Psi$ is an Archimedean generator (ie a continuous, decreasing function $\psi : [0, \infty) \to [0,1]$ with $\psi(0) = 0$ and $\lim_{x \to \infty} \psi(x) = 0$ which is strictly decreasing on $[0, \inf\{t : \psi(t) = 0\}]$).
Not every $\psi \in \Psi$ generates a copula! Hence, the question *which* Archimedean generators actually generate a copula arises naturally. For the bivariate case we saw that for $\psi \in \Psi$

$$C(u) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2)) \text{ is a copula } \Leftrightarrow \psi \text{ is convex}$$

So any $\psi \in \Psi$ which is convex generates a bivariate copula - this is a nice characterization, as convexity is usally easy to show.
Moving forward, from the definition of an Archimedean copula it can be seen that properties of these copulas can be expressed in terms of the *univariate* function ψ. For instance, Prop. 7.7 gives a (relatively) simple formula for Kendall's tau ϱ_τ and Prop. 7.8, whose proof is straightforward, gives very easy formulas for the lower/upper tail dependence coefficients λ_l / λ_u. This is a big advantage of Archimedean copulas: If ψ is

reasonably tractable, most properties can be easily derived.
But what about multivariate (i.e $d > 2$) Archimedean copulas? The main result is as follows:

$C(\mathbf{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right)$ is a copula $\forall d \geq 2 \iff \psi$ completely monotone $\iff \Psi$ is the Laplace Transform of some rv $V \sim G$ where G is a df on

where we call ψ completely monotone if it is infinitely often differentiable and $(-1)^k \psi^{(k)}(t) \geq 0$ for all $t \in (0, \infty)$ and $k \in \mathbb{N}_0$: That is, the derivatives alternate in sign.
Thus: If we take ψ to be a Laplace Transform (LT) of some non-negative rv, ψ will generate a copula in any dimension. Conversely, if C is an Archimedean copula in any dimension, there must be a non-negative random variable V such that ψ is the LT of V.
The random variable V is crucial for sampling, as we saw in the Marshall Olkin algorithm: In order to simulate from the copula $C(\mathbf{u}) = \psi\left(\sum_{j=1}^{d} \psi^{-1}(u_j)\right)$, one needs to generate $E_1, \ldots, E_d \overset{ind}{\sim} Exp(1)$ indepedently of V (whose LT is ψ) and then return the vector $\mathbf{U} = (\psi(X_1), \ldots, \psi(X_d))$: Thus, V (characterized by its LT ψ) generates the dependence.
We finished by discussing some advantages and drawbacks of Archimedean copulas and looked at some more examples of copulas in Section 7.5.

This concludes the course ACTSC 445/845 and therefore also this blog. The tutorial on Friday and the last lecture on Monday will be used for review/ talking about the final/ answering questions...

I hope this lecture blog was helpful for you. I am happy about any feedback on this! I also hope you learned a lot of interesting and useful concepts in this course and got an appreciation of QRM overall and I wish you all nothing but the best for your future! :-)

Yours, Erik