ISSN 2074-9317

The Economics of Weather and Climate Risks Working Paper Series

Working Paper No. 9/2009

On Mathematical Tools for Weather Risks

Dominik Kortschak[3]

---

[1] Wegener Zentrum für Klima und globalen Wandel, Universität Graz
[2] Institut für Technologie- und Regionalpolitik, Joanneum Research Graz
[3] Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences
[4] Zentralanstalt für Meteorologie und Geodynamik (ZAMG)
[5] Institut für Mathematik, TU Graz

Contents

List of Figures

On mathematical tools for weather risks

Abstract

In this article we survey some of the mathematical tools that can be useful in quantifying and managing weather risks. We give an overview of mathematical tools like utility theory and no arbitrage pricing, where we especially focus on the use of weather derivatives as a tool of weather risk management. Further we will give a brief overview of statistical models that can be used for weather risk, with an emphasis on the modelling of dependence of random variables.

1 Introduction

In this paper we want to give an overview of mathematical questions and models that arise when dealing with weather risk. Risk is inherently about uncertainty, and a general mathematical framework to deal with uncertain events is probability theory. In the following we want to concentrate on measurable quantities like weather indices (e.g. daily temperature, wind speed, precipitation indices, etc.) and business indices (e.g. profit, turnover, number of guests etc.). In this context we state the following definition: A weather index W is a random vector, with distribution function F_W. A business index U is a random vector, with distribution function F_U.

Since the random variable U will often depend on the decisions of a manager of a company, it is sometimes convenient to describe the random variable $U := U(\pi)$ as a function of a decision rule $\pi \in \Pi$, where Π is the set of all possible decision rules. We will assume that π is a vector. The manager then wants to find the best possible decision rule $\pi^\star$ such that $U(\pi^\star)$ is most favorable for this business. In this paper we want to study the connection between the business index U and W.

In Section 2 we will concentrate on the case when we have specific models for U and W. In Section 3 we will concentrate on statistical methods for finding models for the dependence between U and W.

We will assume that we are concerned with the planning of the next business period (e.g. next year, next season). In particular this means that we are looking at a time period where no accurate weather forecasts exists.

2 Dealing with weather risks

Assume that we are interested in a business index U which depends on a weather index W. Further assume that we know how the weather index influences the index U (i.e. we know the common distribution of the vector). How can we make advantage of knowing the dependence of weather and business index? At first, let us look at an example.

Example 2.1. *Consider a restaurant in the mountain area, where we know that the number of guests depends on the weather (rain, sunshine, temperature). If the weather is fine* *we expect many guests, whereas when it is raining the number of guests will dramatically decrease. If we know the weather we can (and will) try to adapt the capacity of the restaurant to the expected number of guests and hence optimize our profit.*

As Example 2.1 shows, if one is aware of the influence of the weather to the company and reacts to the weather forecast one can gain wealth from it. Another thing that one sees from the example is that if the number of days with bad weather is too large then we can still get into troubles. One possibility is to use weather derivatives (if they exist) to hedge against the risk of bad weather. See Leggio (2007) for an example of the use of weather derivatives for golf courses.
Before we go on in analysing weather derivatives we want to give a definition of weather derivatives:

Definition 2.1. *Following Jewson & Brix (2005), a weather derivative $D(W)$ is a pay-off function of a given weather situation, especially a weather derivative contract consists of (see Jewson & Brix (2005))*

- *the contract period: a start date and end date;*
- *a measurement station;*
- *a weather variable, measured at the measurement station, over the contract period;*
- *an index which aggregates the weather variable over the contract period in some way;*
- *a pay-off function, which converts the index into the cashflow that settles the derivative shortly after the end of the contract period;*
- *for some kinds of contracts, a premium paid from the buyer to seller at the start of the contract.*

In the following we want to find possible answers to the following questions that arise from the definition of weather:

1. Who should buy or sell a weather derivative?
2. What is a fair price P^D for $D(W)$?
3. What is the highest price a company should except for a given weather derivative?
4. What is the optimal weather derivative for the company?

2.1 The use of derivatives as an insurance product and the implied price

In Example 2.1 the idea is to use weather derivatives as an insurance against the impact of a unfortunate weather situation. At first we want to give the pros and cons of using a weather derivative instead of a business insurance.

+ Moral hazard: When the loss is taken by the insurance then there is only little motivation to minimize this loss.

+ Asymmetric information: The insurance knows less about the risk than the insured business ⇒ high security premiums.

- Basis Risk: e.g. the loss is higher than the payment of the derivative, the weather at the measurement station is different then in the needed area. There is a loss but the weather is good.

- There is no reliable weather station that is relevant to the business.

If we think about a weather derivative as an insurance product we should think about the way an insurance would price the derivative. In insurance mathematics the net premium, i.e. $P^D := \mathbb{E}[D(W)]$. On a first sight this looks like a simple formula but when one takes a closer look, actually that this simple formula causes some major troubles. At first one should note that as time goes by, our knowledge of the random variable and hence P^D will change. Let T be the time point when the derivative is evaluated. Denote with P_t^D the price of the derivative at time $0 \le t \le T$. To evaluate P_t^D we have to model our information up to time t. In mathematical terms this corresponds to introduce a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ which codes our information. The price can then be evaluated by:

$$P_t^D = \mathbb{E}[D(W)|\mathcal{F}_t], \tag{1}$$

and P_t^D is the (in some sense) best possible prediction of $D(W)$ given our available information at time t.

Remark 2.1. *Note that the best prediction of $D(W)$ heavily depends on the used underlying model.*

Hence, using $P^d = \mathbb{E}[D(W)]$ is only possible if all market participants that buy or sell weather derivatives have the same model of $D(W)$ and the same information. Never the less if we have a model for W in which we trust, then $\mathbb{E}[D(W)]$ gives a hint where P^D should lie.
One should note that an insurance will not only charge the net premium for an insurance product, but needs to add a risk premium to the net premium. Such security premiums often are linear functions of the expected value, the standard deviation or the variance of $D(W)$. But also other pricing principles are possible.

2.2 An approach via utility maximization

Let us consider the problem of finding the fair price from the viewpoint of Example 2.1 where we want to use weather derivatives to hedge against bad weather. In this case the main question is not the "fair" price of the derivative, but what price is acceptable. In this case we will use utility maximitation.

Definition 2.2. *A utility function u is a monotone increasing and concave function that assigns to every value x a utility $u(x)$. Someone who uses a utility function of this type is called risk-averse.*

If $U(\pi)$ is the return by given strategy π then we want to maximize the expected utility

$$\max_{\pi} \mathbb{E}\left[u\left(U(\pi)\right)\right]$$

Now assume that we can buy or sell a weather derivative at a price P^D. If we maximize

$$\max_{\pi,n} \mathbb{E}\left[u\left(U(\pi) - nP^D + nD(W)\right)\right], \tag{2}$$

then n^* which maximizes (2) is the number of derivatives we should buy.

Remark 2.2. *Note that this method only works if we know for every π the distribution of the vector $(U(\pi), W)$.*

Remark 2.3. *Note that if in this model $P^D \neq \mathbb{E}\left[D(W)\right]$, then it is possible that the optimal strategy is to buy or sell the derivative even when U and W are independent.*

If we can only decide between buying or not buying a derivative, then we will buy the derivative if and only if

$$\max_{\pi} \mathbb{E}\left[u\left(U(\pi)\right)\right] < \max_{\pi} \mathbb{E}\left[u\left(U(\pi) - P^D + D(W)\right)\right],$$

i.e our expected utility is higher when we buy the derivative.

2.3 No-arbitrage and hedging arguments

No-arbitrage arguments In financial markets the pricing of derivatives is an important task. The main idea of pricing financial derivatives is via no-arbitrage arguments. At first a little example:

Example 2.2. *Assume that we have an asset S which is only traded at the times $\{0, 1\}$. The price of the asset at time 0 is S_0. At time 1 the price of the asset S_1 can have the values S_u and S_d. Assume that $\mathbb{P}(S_1 = S_u) = p$. Further assume that there exists a derivative $D(S)$ which pays $D(S_u)$ if $S_1 = S_u$ and $D(S_d)$ if $S_1 = S_d$. Further we assume that we can borrow money without any interest rates and that we can buy and sell any fraction of the asset as we want. We want to evaluate the price P^D of D at time 0. See Figure 1 for a visualization of the prices. The idea to evaluate the price P^D is to find a portfolio consisting of θ_0 cash and θ_1 assets which replicates the derivative. The price of the portfolio at time 0 should be the same as P^D. θ_0 and θ_1 can be evaluated with the equations:*

$$\theta_0 S_u + \theta_1 = D(S_u) \qquad \theta_0 = \frac{D(S_u) - D(S_d)}{S_u - S_d}$$

$$\theta_0 S_d + \theta_1 = D(S_d) \qquad \theta_1 = \frac{S_u D(S_d) - S_d D(S_u)}{S_u - S_d}$$

Figure 1: The price process in a two state model

For the price of the portfolio (and hence the price of D) we get

$$P^D = \theta_0 S_0 + \theta_1 = p^* D(S_u) - (1 - p^*) D(S_d),$$

where $p^* = (S_0 - S_d)/(S_u - S_d)$. Note that if $S_d < S_0 < S_u$ then $0 < p^* < 1$.

Remark 2.4. *We see that the price P^D can be written as $\mathbb{E}_{p^*}[D(S_1)]$ with respect to a measure. This measure is called a risk-neutral measure (note that $\mathbb{E}_{p^*}[S_1] = S_0$). Note that the risk-neutral measure is independent from the physical measure.*

Remark 2.5.

- *The idea of no-arbitrage also exists in more involved models.*
- *In general models the no-arbitrage-price is not unique ("incomplete market").*
- *In incomplete models it is still possible to evaluate the price as the mean with respect to a (possible not unique) risk neutral measure.*

The idea of the no-arbitrage pricing is to find a trading strategy π of a portfolio such that the value of the portfolio P_t^π at the terminating time T is the same as the value of the derivative D_T. If it is not possible to construct a portfolio that perfectly replicates D_T, then one might find trading strategies such that the value of P_T^π is greater or smaller than D_T independent of the performance of the asset. The value of P_t^π then serves as an upper or lower bound for D_t.
In the case of weather derivatives, the no-arbitrage arguments have one important drawback, namely we can not invest into a weather index. In (Jewson & Brix 2005, (chapter 11)) this problem is solved in assuming that there exists a derivative D^0 that is traded liquidly on the market. In this model the derivative D^0 is used as an underlying and no-arbitrage arguments are used to price other derivatives on the same index.

Utility pricing of derivatives Assume that we have an underlying S and we can invest in this underlying. Denote with Π the set of all possible trading strategies (denote with P_t^π the value of the portfolio a time t when trading with strategy $\pi \in \Pi$), and assume that there exists a derivative $D(S_T)$ with price P^D. Further assume that we have a utility function u and we want to maximize our expected utility. The question is: What is the maximal price we would pay for the derivative? We will buy the derivative if and only if the expected utility with the derivative is greater than the expected utility without the derivative i.e. if

$$\sup_{\pi\in\Pi}\mathbb{E}\left[u(P_T^\pi)\right] \le \sup_{\pi\in\Pi}\mathbb{E}\left[u\left(P_T^\pi - P^D + D(S_T)\right)\right].$$

Hence the maximal price $\overline{P^D}$ we would accept to pay for the derivative is given as the solution of

$$\sup_{\pi\in\Pi}\mathbb{E}\left[u(P_T^\pi)\right] = \sup_{\pi\in\Pi}\mathbb{E}\left[u\left(P_T^\pi - \overline{P^D} + D(S_T)\right)\right].$$

Similarly if we want to sell a derivate, then the minimal price $\underline{P}^D$ we would ask for the derivative is given as the solution of

$$\sup_{\pi\in\Pi}\mathbb{E}\left[u(P_T^\pi)\right] = \sup_{\pi\in\Pi}\mathbb{E}\left[u\left(P_T^\pi + \underline{P}^D - D(S_T)\right)\right].$$

We now want to give two examples with a derivative on a non-traded underlying Y_t and there exists a traded asset S_t that is correlated to Y_t.

Example 2.3. *In Monoyios (2004) two assets $(S,Y) := (S_t,Y_t)_{0\le t\le T}$ are considered which follow a log-normal diffusion*

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}\omega_t, \\ \mathrm{d}Y_t &= \mu_0 Y_t \mathrm{d}t + \sigma_0 Y_t \mathrm{d}\omega_t^0, \end{aligned}$$

where (ω,ω^0) are correlated Brownian motions with correlation ρ and $\mu,\sigma,\mu_0,\sigma_0,\rho$ are constants.

Example 2.4. *In Carmona & Diko (2005) the derivative is written on a precipitation index Y_t (the precipitation intensity at time t), modelled as a certain Markov process. Further, there exists a traded asset S_t which satisfies the stochastic differential equation*

$$\mathrm{d}S_t = \mu(Y_t) S_t \mathrm{d}t + \sigma(Y_t) S_t \mathrm{d}\omega_t,$$

where ω_t is a Brownian motion independent of Y_t and $\mu(\cdot)$ and $\sigma(\cdot)$ are functions. Note that in this model the drift and the volatility depend on the weather index.

Remark 2.6. *In Examples 2.3 and 2.4 one can use stochastic control theory to get rather explicit results for the price P^D of special derivatives. One should note that if one uses more general models for S_t and Y_t, then it will be very likely that explicit expressions can not be obtained.*

2.4 The optimal derivative

Assume that we are in the fortunate situation that we know the distribution of the business index U (which possibly depends on a strategy $\pi \in \Pi$) and the weather index W. Further we assume that we have a pricing mechanism $P^{(\cdot)}$ which assigns every admissible derivative $D \in \mathcal{D}$ a price P^D. If we have a utility function u and we want to maximize the expected utility, then we have to solve the optimization problem

$$\sup_{\pi \in \Pi, D \in \mathcal{D}} \mathbb{E}\left[U^\pi - P^D + D(W)\right].$$

See Barrieu & Karoui (2002) for a more involved example of finding the optimal derivative.

3 Models for weather and business indices

In Section 2 we have seen that it is of importance to have a reliable model for the considered business and weather index. In this section we want to discuss the problem of finding models for these quantities. Inspired by Carmona & Diko (2005) it is in principle possible to classify the models into different categories:

1. Physical models that try to use the current state of the atmosphere (or weather system) and use a physical model to calculate the future state of the atmosphere (or weather system).
2. Stochastic models that describe the weather index and the business index as a random variable that use only a few physically or economically meaningful parameters to describe the distribution of the business and weather index.
3. Pure statistical models that fit a distribution to the data, which has no physical or economical meaning.
4. No model, where one tries to get information from the raw data without any assumptions on the random variables.

If we look at the physical model then a problem is that the current state of the atmosphere is not known sufficiently well. Further the accuracy is limited by the computing power. In Jewson & Brix (2005) it is stated that for weather forecasts several simulations with different starting points are made instead of a single forecast. From a probabilistic point of view this can be seen as generating realizations of the random variable weather and this can be used to evaluate connected indices using Monte Carlo methods.

3.1 The univariate case

We have chosen a weather or business index and we want to find out more about its behavior. At first we will assume that the index X consists of only one real number. Note that most of the statements in this section remain true in more complex situations (like multi-dimensional random variables). We will assume that the index X is random. We want to evaluate quantities like $\mathbb{E}[X]$, $\mathbb{V}ar[X]$ or $\mathbb{P}(X \in A)$ for some set A. In principle, we have to decide between two possibilities: We can use a parametric model where we assume that the random variable X has a specific distribution and we are only left with the problem of estimating the parameters of the distribution. Or we use a non-parametric model where we try to get information out of the observations without any assumptions anything on the distribution of X. Of course it is possible to combine these two methods.

3.1.1 Parametric models

If we have decided to use a parametric method, we have in general to go through the following steps.

1. Choose an (appropriate) parametric model for X.
2. Estimate the parameters of the model.
3. Check the plausibility of the model with a reasonable goodness-of-fit measure.
4. Evaluate the desired quantities with respect to the now fully specified distribution. If it is possible, one should use confidence intervals instead of one value.

We will concentrate on the first step. Further, one should note that the other steps heavily depend on the first step. At first we will provide some distributions that can be used for modelling random variables.

The Normal distribution Assume that the index $X := \sum_{i=1}^{n} Y_i$ where the Y_i are independent identical distributed random variables with $\mu = \mathbb{E}[Y_1]$ and $\sigma^2 = \mathbb{V}ar[Y_1] < \infty$. In this case we get from the Central Limit Theorem (e.g. Embrechts et al. (1997)) that (as $n \to \infty$)

$$\frac{\sum_{i=1}^{n}(Y_i - \mu)}{\sqrt{n\sigma^2}} \xrightarrow{d} \bar{Y}$$

where $\bar{Y}$ is normally distributed with $\mathbb{E}\left[\bar{Y}\right] = 0$ and $\mathbb{V}ar\left[\bar{Y}\right] = 1$. This implies that we could try to approximate (model) X with a normal distribution. An estimate of the error of the approximation is given by the Berry-Esséen Theorem. Denote with $\Phi(x) := \mathbb{P}(\bar{Y} \le x)$ and with $F_n(x) := \mathbb{P}(\sum_{i=1}^{n}(Y_i - \mu)/\sqrt{n\sigma^2} \le x)$. If the third absolute moment $\rho = \mathbb{E}\left[|Y_1|^3\right]$ exists then for all $x > 0$.

$$|\Phi(x) - F_n(x)| \le \frac{C\rho}{\sigma^3\sqrt{n}},$$

where one can take $C = 0.7056$ (see Shevtsova (2007)). If the third absolute moment does not exist then a similar result can be found in e.g. Feller (1971). Since we are using a limit relation we need to ensure that the number of summands n is large enough that the

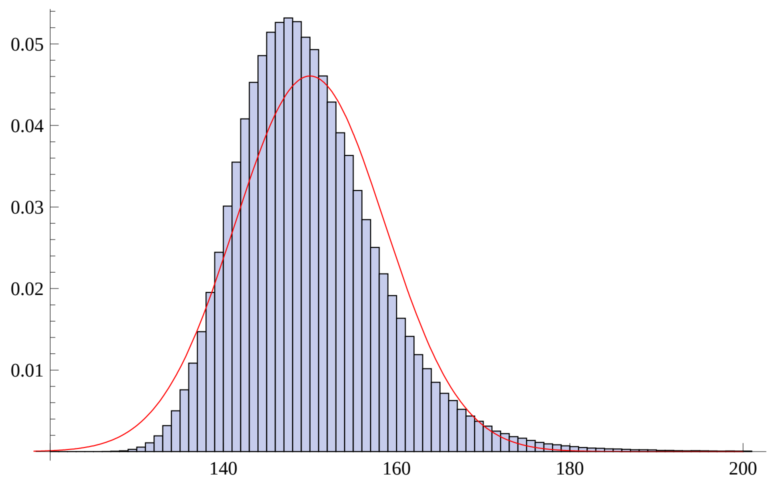

Figure 2: Histogram of $\sum_{i=1}^{100} Y_i$ where $\mathbb{P}(Y_i > x) = x^{-3}$, compared to normal distribution.

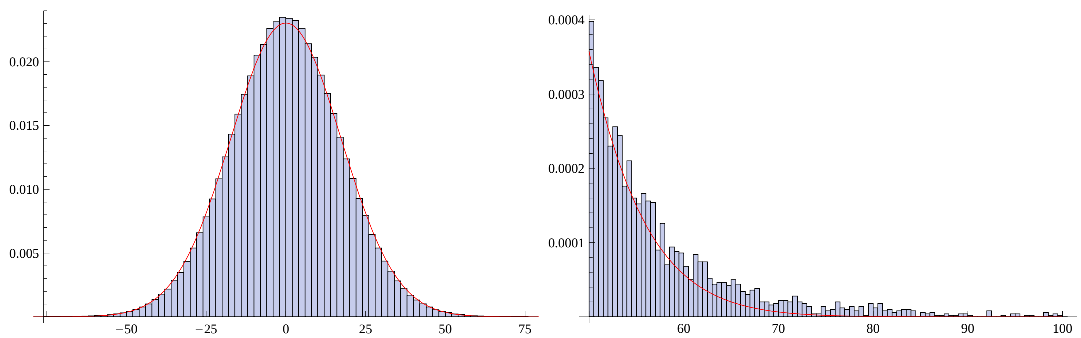

Figure 3: Histogram of $\sum_{i=1}^{100} S_i Y_i$ where $\mathbb{P}(Y_i > x) = x^{-3}$ and $S_i \in \{-1, 1\}$, compared to normal distribution.

limit relation holds. In the following we give some examples of the approximation of the sum by a normal distribution.

Example 3.1. *Let* $X = \sum_{i=1}^{100} Y_i$ *where the* Y_i *are independent identical distributed with density* $f(x) = 3x^{-4}$ *due to the asymmetry of the random variables* Y_i *we see in Figure 2 that the normal distribution is not a good approximation to* X*.*

Example 3.2. *Let* $X = \sum_{i=1}^{100} S_i Y_i$ *where the* Y_i *are independent identical distributed with density* $f(x) = 3x^{-4}$ *and* S_i *is uniform distributed on the set* $\{-1, 1\}$*. In Figure 3 we see that the approximation of* X *by the normal distribution is acceptable. On the other hand we see that if we are also interested in the tails of the distribution, then the approximation by the normal distribution may be too crude.*

Max-stable distributions Now assume that $X = \max(Y_1, \dots, Y_n)$ where the Y_i are independent identical distributed random variables. Similar to the case of sums, one can use a limit theorem to model the distribution of X. At first we define the possible limit distributions.

Definition 3.1 (Extreme value distribution)**.** *A distribution H is called extreme value distribution, if it is of one of the following three types*

$$\begin{aligned}\Phi_\alpha(x) =& e^{-x^{-\alpha}}, \quad x > 0, \alpha > 0.\\ \Psi_\alpha(x) =& e^{-(-x)^\alpha}, \quad x < 0, \alpha > 0.\\ \Lambda(x) =& e^{-e^{-x}}, \quad x \in \mathbb{R},\end{aligned}$$

where two distributions H_1 and H_2 are of the same type if there exists constants a and b such that $H_1(ax + b) = H_2(x)$.

If one aims to estimate the parameter of a distribution, then it is of advantage if the underlying family has only one parameter.

Definition 3.2. *The generalized extreme value distribution (GEV) is defined by*

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(e^{-x}\right), & \xi = 0, \end{cases}$$

where $1 + \xi x > 0$.

Note that $\lim_{\xi \to 0} H_\xi(x) = H_0(x)$. Now we can state the Fisher-Tippett theorem (e.g (Embrechts et al. 1997, p. 121)

Theorem 3.1 (Fisher-Tippett theorem)**.** *Let $X_1, X_2, \dots$ be independent identical distributed random variables with distribution F. Denote with $M_n = \max_{i=1,n} X_i$. If there exists a non-degenerated distribution H and constants a_n and b_n such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H,$$

then H is of the type of one of the extreme value distributions.

In view of the Fisher-Tippett theorem, we can try to model X by an extreme value distribution. Since the approximation is of similar type as in the case of sums, we will not go into details of it.

Stationary distribution of time series Assume that we have a time series X_t, $t \in \mathbb{N}$ or $t \in \mathbb{R}$ and for a fixed time T, $X = X_T$. If the limit distribution $\lim_{t\to\infty} X_t$ exists and T is sufficiently large, we can approximate X_T by the limit distribution.

3.2 Non-parametric methods

Assume that we have n *independent* realizations $x_1 \ldots, x_n$ of the random variable X and assume that we want to evaluate the index R_X (e.g. $\mathbb{E}[X]$, $\mathbb{V}ar[X]$, ...). The idea of non-parametric estimation is to find a statistic g, i.e. (for every n) a function $g : \mathbb{R}^n \to \mathbb{R}$ such that for independent identical distributed random variables $X_1 \ldots, X_n$ with the same distribution as X

$$R_X = \mathbb{E}[g(X_1, \ldots, X_n)] \quad \text{or at least} \quad R_X = \lim_{n\to\infty} \mathbb{E}[g(X_1, \ldots, X_n)].$$

In many cases the asymptotic distribution of $g(X_1, \ldots, X_n)$ (as $n \to \infty$) and hence one can obtain asymptotically valid confidence intervals. A simple and well known example is $g(X_1, \ldots, X_n) := 1/n \sum_{i=1}^n X_i$ which can be used to estimate $\mathbb{E}[X_1]$. In the context of weather derivate pricing, this method is also known as burn analysis.

3.3 A comparison between parametric and non-parametric models

We now want to give an example of the use of parametric models compared to non-parametric models. Let $X, X_1, \ldots, X_n$ be independent identical distributed random variables with common distribution $F(x) = 1 - (1 + x)^{-1.5}$. Assume that we want to evaluate $\mathbb{E}[X]$ and $\mathbb{P}(X > 20)$. As a model we will use two parametric models M_1 and M_2 and one non-parametric model NP. In M_1 we will use the "correct" model and assume that $F(x) = 1-(1+x)^{-\alpha}$ and estimate α with the maximum likelihood method. As M_2 we use an "incorrect" model and assume that $F(x) = 1-e^{-\lambda x}$. As in M_1, we estimate λ with the maximum likelihood method. As parametric estimates we just used the empirical mean and the relative frequency of realization laying over 20.
At first considere $n = 30$. We simulated 30 independent identical distributed realizations of X and evaluated $\mathbb{E}[X]$ and $\mathbb{P}(X > 20)$, for the three methods. We iterated this procedure 100 times. Figure 4 shows for each model these 100 estimates. At first one should note that the estimates of the model M_2 and NP for the mean are the same. This is due to the chosen estimation of the parameter λ. Further it is possible that the estimator of α in model M_1 can give a value less than 1, and hence $\mathbb{E}[X] = \infty$. The probability of this event is approximately 0.7% and hence not negligible. If we now compare the estimates with the correct values $\mathbb{E}[X] = 2$ and $\mathbb{P}(X > 20) \approx 0.0104$ we see that the non-parametric estimator tends to underestimate the real values while the model M_1 seems to behave better in this case. Further one can see that especially for $\mathbb{P}(X > 20)$ the estimates of M_2 can be quite far from the real value. This is an example of what can happen if one unreflectedly uses an incorrect model.

As a second example we used $n = 10000$ and performed the same procedure for the models M_1 and NP, the results can be found in Figure 5. In this case it seems that the model-based estimates are tighter around the correct values and hence should be preferred. Of course one should note that the probability that model M_1 evaluates $\mathbb{E}[X] = \infty$ is still positive, but in this situation negligibly small. At last we want to give some general

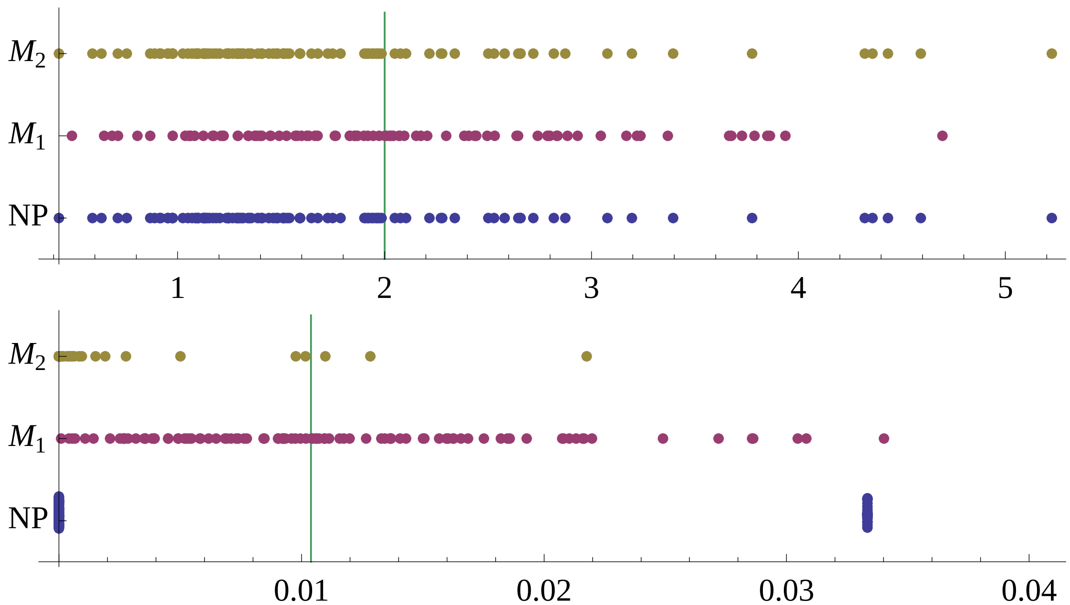

Figure 4: Estimates for $\mathbb{E}[X]$ and $\mathbb{P}(X > 20)$ for M_1, M_2 and NP with 30 datapoints.

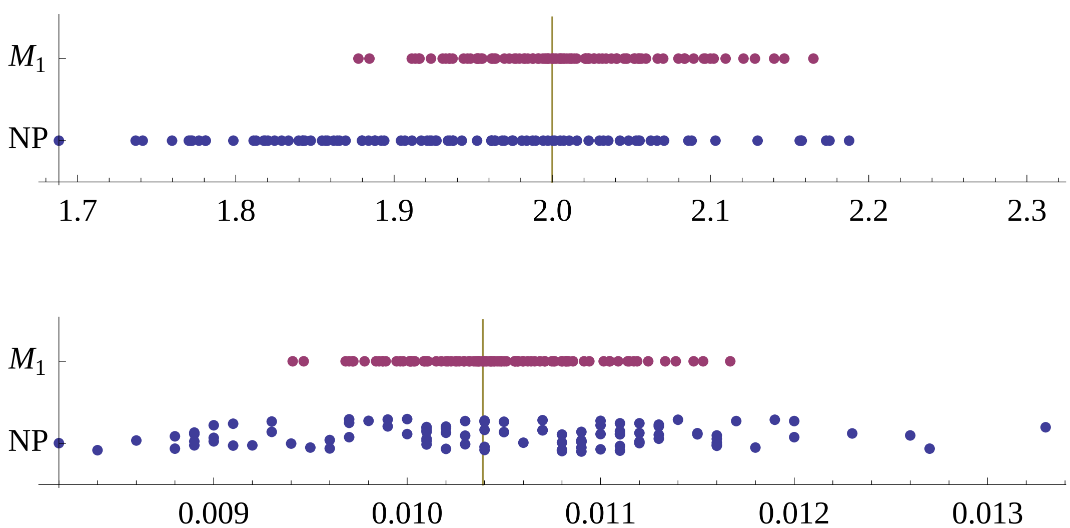

Figure 5: Estimates for $\mathbb{E}[X]$ and $\mathbb{P}(X > 20)$ for M_1, M_2 and NP with 10000 datapoints.

remarks on the advantages and disadvantages of parametric models.

- Parametric results are often of less variation than non-parametric results.
- Especially in the case of few datapoints, parametric models can improve the estimation quality.
- The chosen model has impact on the result (this can especially be seen in the tails of the distribution).
- When one uses a model there is always the danger that the model does not describe the reality sufficiently well.
- When one is only interested in special properties of the distribution (e.g. tail behavior, mean value) then a model which is not usable for other properties can still provide good approximations.
- Non-parametric methods only use acquired experiences. Especially in the case of rare events (like big floods), this can lead to an underestimation of the risk. For example if we have 30 data points (according to 30 years), then the worst event we can expect is a once-in-30 year event.

3.4 Extremes and heavy tails, a semi-parametric approach

Assume that $x_1, \dots, x_n$ are independent realizations of a random variable X. We are interested only in the right tail of the distribution, i.e. we want to evaluate $\mathbb{P}(X > u)$ for a high threshold u. If we do not want to assume that X has a specific distribution, we can try the following semi-parametric approach. For the following definition see e.g. (Embrechts et al. 1997, p. 162)

Definition 3.3 (The generalized Pareto Distribution (GDP))**.** *The GDP is defined by the distribution function*

$$G_\xi(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-x}, & \xi = 0, \end{cases}$$

where $x \geq 0$ *if* $\xi \geq 0$ *and* $0 \leq x \leq -1/\xi$ *if* $\xi < 0$.

Now the usefulness of the GDP is the following Theorem (e.g. (Embrechts et al. 1997, p. 165))

Theorem 3.2. *Denote with* x_F *the right endpoint of the distribution* F *of* X *(i.e.* $x_F = \inf\{x : \mathbb{P}(X \leq x) = 1$*. For every* $\xi \in \mathbb{R}$ $F \in MDA(H_\xi)$ *there exists a function* $b(u)$ *with*

$$\lim_{u \to x_F} \sup_{0 \leq x \leq x_F - u} \left| \mathbb{P}(X - u \leq x | X > u) - G_\xi\left(\frac{x}{\beta(u)}\right)\right| = 0$$

Remark 3.1. *At first one should note that if we want to create a sample of the random variable X_u with distribution $F_u(x) = \mathbb{P}(X - u \leq x | X > u)$, we can take all elements of a sample of X exceeding u and substract u from these elements.*

Remark 3.2. *From Theorem 3.2 it follows that we can approximate the distribution of F_u of X_u, for sufficiently large u by*

$$F_u(x) \approx G_\xi\left(\frac{x}{\beta(u)}\right) = 1 - \left(1 + \xi\frac{x}{\beta(u)}\right)^{-1/\xi}. \tag{3}$$

The main idea of the Peak-over-threshold methods (POT) is to replace F_u by $G_\xi(x/\beta)$ and estimate the parameters ξ and β with a sample of X_u. Here we are facing the following problems:

- *We do not know if $F \in MDA(H_\xi)$.*
- *If we take u that is too "small", then the approximation of $F_u(x)$ by $G_\xi(x/\beta)$ will not be good.*
- *If we choose u too "large", then the sample of X_u is too small to get reliable estimates for ξ and β.*
- *We do not know if there exists a u that is neither too "small" nor too "large".*

We now want to estimate $\mathbb{P}(X > u)$ for a high threshold u. Following (e.g. (Embrechts et al. 1997, p. 352, (chapter 6.5))) we can make the following procedure: Choose a $u_0 < u$ such that the approximation

$$\mathbb{P}(X > u_0) \approx \frac{N_u}{n},$$

where $N_u = \sum_{i=1}^n I_{\{x_i > u_0\}}$, is reliable. As mentioned in Remark 3.2 we can now estimate the parameters ξ and β and approximate

$$\mathbb{P}(X > u) = \mathbb{P}(X > u_0)\mathbb{P}(X - u_0 > u - u_0 | X > u_0) \approx \frac{N_u}{n} G_{\hat{\xi}}((u - u_0)/\hat{\beta}),$$

where $\hat{\xi}$ and $\hat{\beta}$ are estimators of ξ and β respectively. If we neglect the error in the estimation of the parameters (which one should not do practice), then the absolute value of the approximation error is given by

$$\mathbb{P}(X > u_0)\left|\mathbb{P}(X - u > u - u_0 | X > u_0) - G_\xi\left(\frac{u - u_0}{\beta(u_0)}\right)\right|.$$

Now Theorem 3.2 shows that for every $\epsilon > 0$ there exists an u_0 such that for every $u > u_0$

$$\left|\mathbb{P}(X - u > u - u_0 | X > u_0) - G_\xi\left(\frac{u - u_0}{\beta(u_0)}\right)\right| \leq \epsilon.$$

One should note that if $\mathbb{P}(X > u) \leq \epsilon\mathbb{P}(X > u_0)$, then the possible error is of the same size as the probability that we are estimating.

3.4.1 Estimation of the parameter ξ

This section follows de Haan & Ferreira (n.d.). Assume that $F \in \mathrm{MDA}(H_\xi)$, then ξ is connected with the tail behavior of F. For example, if $\xi < 0$, then F has a finite right end-point, or if $\xi > 0$, then F is regularly varying, i.e. for all $x > 0$, $\lim_{u\to\infty} \overline{F}(xu)/F(u) = x^{-\alpha}$ for an $\alpha > 0$. When $\xi = 0$, then the range of distribution which satisfy $F \in \mathrm{MDA}(H_\xi)$ is quite large. It goes from quite heavy distributions like the lognormal or the Weibull distribution to distributions like the exponential and the normal distribution. Further, there are also distributions with a finite right endpoint. Hence we see that a reliable estimation of ξ is of importance.

The Hill estimator. Assume that we know that $\xi > 0$. In this case we can use the Hill estimator for ξ. Denote with $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ the order statistic of $X_1, \dots, X_n$, then the Hill estimator is defined by

$$\hat{\xi}_H(k) := \hat{\xi}_H := \frac{1}{k} \sum_{i=0}^{k} \log(X_{n-i,n}) - \log(X_{n-k,n}).$$

Note that the appropriate choice of the value k is the main challenge of this estimator (as it is for all other estimators as well).

Theorem 3.3. *Let $X_1, X_2, \dots$ be independent identical distributed random variables with distribution F. If $F \in MDA(H_\xi)$ with $\xi > 0$ and $k = k(n)$, $\lim_{n\to\infty} k/n = 0$, then as $n \to \infty$,*

$$\hat{\xi}_H \xrightarrow{P} \xi,$$

where $\xrightarrow{P}$ denotes convergence in distribution.

A method of choosing the value of k is using a so called Hill-plot. The Hill-plot depicts $\hat{\xi}_H(k)$ as a function of k. For large values of k, $\hat{\xi}_H(k)$ has a bias, for intermediate k it usually stabilize to a line, and for small k it will vary again. A rule of thumb is to take the first value of k, when the plot of $\hat{\xi}_H(k)$ behaves like a straight line.

The Pickands Estimator In the case $\xi \in \mathbb{R}$, the first proposed estimator is the Pickands Estimator:

$$\hat{\xi}_P(k) := \hat{\xi}_P := \frac{1}{\log(2)} \log\left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}}\right).$$

Theorem 3.4. *Let $X_1, X_2, \dots$ be independent identical distributed random variables with distribution F. If $F \in MDA(H_\xi)$ with $\xi \in \mathbb{R}$ and $k = k(n)$, $\lim_{n\to\infty} k/n = 0$, then as $n \to \infty$,*

$$\hat{\xi}_P \xrightarrow{P} \xi,$$

Maximum Likelihood Estimator The idea is to approximate the distribution of the k largest elements by (3) and estimate the parameters with the maximum liklihood estimators. For $k > 0$ define $u_0 = X_{n-k,n}$ and $Z_i = X_{n-i+1,n} - X_{n-k,n}$, $i = 1, \dots, k$. From Theorem 3.2 we get that the distribution of $Z_1, \dots, Z_k$ can be approximated by

$$F(x) = 1 - (1 + \frac{\xi}{\sigma} x)^{-1/\xi}.$$

We can estimate (ξ, σ) via the maximum likelihood method. Note that one has to restrict to values of $\xi > -1/2$, since the maximum likelihood estimator behaves irregularly for $\xi \leq 1/2$.

The Moment Estimator In case $\xi \in \mathbb{R}$, another estimator is the Moment Estimator. Define

$$M_n^{(j)} := \frac{1}{k} \sum_{i=1}^{k-1} (\log(X_{n-i,n} - X_{n-k,n})^j \,,$$

then the Moment Estimator is given by

$$\hat{\gamma}_M := M_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} \right)^{-1}$$

3.4.2 Detection of heavy tails and the mean excess function

If we are interested in tail probabilities, one of the main questions is the heavyness of the tails (i.e. is $\mathbb{E}\left[X^k\right] < \infty$ do exponential moments $\mathbb{E}\left[e^{sX}\right]$ exist, etc.). One idea is to use the mean excess function.

Definition 3.4. *The mean excess function of a random variable X is given by*

$$e(u) = \mathbb{E}\left[X - u | X > u\right]$$

If the distribution of the random variable is in the maximum domain of attraction of the Fréchet distribution (with parameter α), then the mean excess function behaves asymptotically like $e(u) \sim u/(\alpha - 1) + d$, where d is some constant. Loosely speaking, the faster the growth of the mean excess function, the heavier is the tail of the distribution. If one wants to get an idea of the heavyness of the tail of a distribution, one possibility is to look at the empirical mean excess function and compare it to the mean excess function of parametric distributions (see Figure 6).

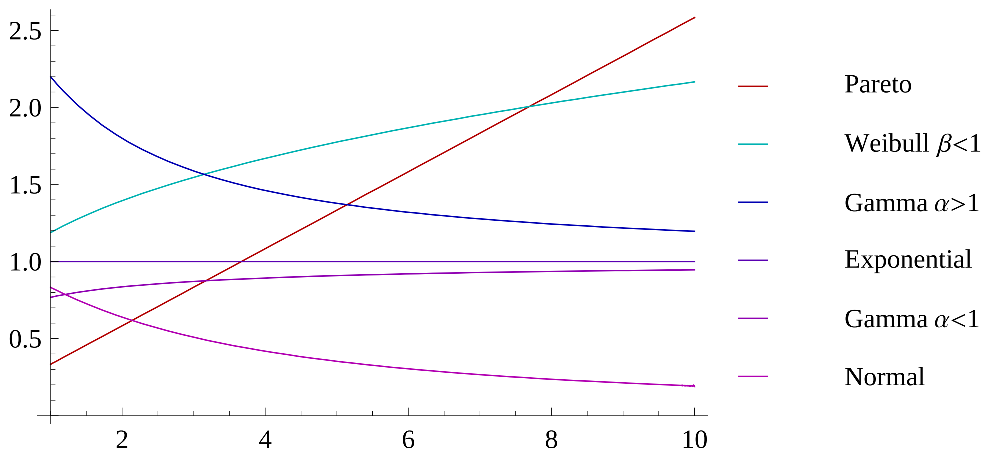

Figure 6: The mean excess function for different distributions

4 Dependence of random variables

We are dealing with a random vector $(X^1, \ldots, X^d)$ which contains the weather and the wealth. We now ask the question of the dependence among the components of the random vector. Since in many cases it is hard to estimate the complete dependence structure we first will give some dependence measures.

4.1 Dependence measures

4.1.1 Linear correlation coefficient

The correlation of two random variables X_1, X_2 is defined by

$$\rho_{X_1,X_2} = \frac{\mathbb{E}\left[(X_1 - \mathbb{E}\left[X_1\right])(X_2 - \mathbb{E}\left[X_2\right])\right]}{\sqrt{\mathbb{V}ar\left[X_1\right]\mathbb{V}ar\left[X_2\right]}}.$$

Note that $-1 \leq \rho_{X_1,X_2} \leq 1$ and (e.g. (Malevergne & Sornette 2006, p. 149))

$$\rho^2_{X_1,X_2} = \frac{\mathbb{V}ar\left[X_2\right] - \min_{a,b}\mathbb{E}\left[(X_2 - (aX_1 + b))^2\right]}{\mathbb{V}ar\left[X_2\right]},$$

hence the correlation can be seen as measure of how well X_2 can be approximated by a linear function of X_1. If one is only interested in the linear dependence of X_1 and X_2, the correlation might serve as a possible measure for dependence. In other cases the value of the correlation can be rather missleading.

Example 4.1. *Let $X_1 \sim N(0,1)$ and $X_2 = e^{\sigma X_1}$ then X_2 is a function of X_1 but the linear correlations given by:*

$$\rho_{X_1,X_2} = \frac{\sigma}{\sqrt{e^{\sigma^2}-1}}$$

For $\sigma = 1$ we have $\rho_{X_1,X_2} = 0.762874$ and for $\sigma = 4$ we have $\rho_{X_1,X_2} = 0.00134185$.

4.1.2 Concordance measures

This section follows Nelsen (1999). Let $(x_1^1, x_1^2), \ldots, (x_n^1, x_n^2)$ be a set of independent realizations of the random vector (X^1, X^2). A pair of observations (x_i^1, x_i^2) and (x_j^1, x_j^2) is called concordant if $(x_i^1 - x_j^1)(x_i^2 - x_j^2) > 0$ (i.e. if $x_i^1 > x_j^1$ then $x_i^2 > x_j^2$ and if $x_i^1 < x_j^1$ then $x_i^2 < x_j^2$). We call the pair discordant if $(x_i^1 - x_j^1)(x_i^2 - x_j^2) < 0$. Denote with c concordant pairs in the sample and d the number of discordant pairs. If a high value of X^1 would imply a high value of X^2, then we would assume that c is significantly lager than d. If X^1 and X^2 are independent, then we would assume that c and d are approximately the same. Kendall's tau is a measure for the difference of c and d. The sample version is defined by

$$t = \frac{c-d}{c+d}.$$

If (X_1^1, X_1^2) and (X_2^1, X_2^2) are two independent vectors with the distribution of (X^1, X^2), then Kendall's tau is defined by

$$\tau_{X^1,X^2} = \mathbb{P}\left((X_1^1 - X_2^1)(X_2^1 - X_2^2) > 0\right) - \mathbb{P}\left((X_1^1 - X_2^1)(X_2^1 - X_2^2) < 0\right).$$

Another concordance measure is Spearman's ρ. If (X_1^1, X_1^2), (X_2^1, X_2^2) and (X_3^1, X_3^2) are three independent copies of (X^1, X^2) then Spearman's ρ is defined through:

$$\rho^S_{X^1,X^2} = 3\left(\mathbb{P}\left((X_1^1 - X_2^1)(X_2^1 - X_2^3) > 0\right) - \mathbb{P}\left((X_1^1 - X_2^1)(X_2^1 - X_2^3) < 0\right)\right).$$

If we define the random variables $U = F_X^1(X^1)$ and $V = F_X^2(X^2)$ (where $F_X^i(x)$ is the marginal distribution function of X^i), then we get

$$\rho^S_{X^1,X^2} = \rho_{U,V} = \frac{\mathbb{E}\left[(U - \mathbb{E}[U])(V - \mathbb{E}[V])\right]}{\sqrt{\mathbb{V}ar[U]\,\mathbb{V}ar[V]}}$$

One should mention that these measures are invariant under monotone transformations. Hence we get for the example of the correlation:

Example 4.2. *Let $X^1 \sim N(0,1)$ and $X^2 = e^{\sigma X^1}$ then*

$$\tau_{X^1,X^2} = \rho^S_{X^1,X^2} = 1$$

One should also be careful when using these two dependence measures:

Example 4.3. *Let $X^1 \sim N(0,1)$ and P a random variable with $\mathbb{P}(P=1) = \mathbb{P}(P=-1) = \frac{1}{2}$. Let $X^2 = P \cdot X^1$, then $X^2 \sim N(0,1)$ and*

$$\rho_{X^1,X^2} = 0, \qquad \tau_{X^1,X^2} = \rho^S_{X^1,X^2} = 0$$

Note that X^1, X^2 are normally distributed with $\rho_{X,Y} = 0$, but X^1 and X^2 are not independent.

4.2 Conditional dependence measures

In some cases one is not interested in the dependence on the whole support, but one is interested in special areas of the support. For example, one can ask what is the influence of the weather to the wealth given an extreme weather situation (whatever extreme weather means). Since the presented dependence measures are of a global type, they are not well suited to describe the dependence structures in this situation. One possible solution to this problem is to use conditional versions of these dependence measures. For more details of these dependence measures we refer to Malevergne & Sornette (2006).

Definition 4.1 (Conditional Correlation)**.** *The conditional correlation of two random variables X^1 and X^2 with respect to a subset $\mathcal{A} \in \mathbb{R}$ with $\mathbb{P}((X^1, X^2) \in \mathcal{A}) > 0$ is defined through:*

$$\rho^{\mathcal{A}}_{X^1,X^2} = \frac{\mathbb{E}[(X^1 - \mathbb{E}[X^1|(X^1,X^2) \in \mathcal{A}])(X^2 - \mathbb{E}[X^2|(X^1,X^2) \in \mathcal{A}])|(X^1,X^2) \in \mathcal{A}]}{\sqrt{\mathbb{V}ar[X^1|(X^1,X^2) \in \mathcal{A}]\mathbb{V}ar[X^2|(X^1,X^2) \in \mathcal{A}]}}.$$

One can also define conditional versions of other dependence measures. In these cases one has to decide if one wants to keep the invariance under monotone transformations (for details see e.g. Malevergne & Sornette (2006).

5 Modeling the dependence structure

If we know that the random variable $X_1, \ldots, X_n$ are dependent, we can ask the question how one can model the dependence structure of these random variables. In this section we want to give some methods how to construct the dependence between the random variables.

5.1 The Copula approach

Assume that $X_1, \ldots, X_d$ are dependent real-valued random variables with marginal distribution function F_i (i.e. $F_i(x) := \mathbb{P}(X_i \leq x)$). We say that a function $C(u_1, \ldots, u_d)$ is a copula of the random variables $X_1, \ldots, X_d$, if the multivariate distribution function can be written as:

$$F(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

We should ask which are the requirements to C for being a copula. For $a_i \leq b_i$ ($i = 1, \ldots, d$), define a n-box $B = [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$. For a function H define

$$V_H(B) := \sum \mathrm{sgn}(c_1, \ldots, c_d) H(c_1, \ldots, c_d),$$

where the sum is taken over all $(c_1, \ldots, c_d)$ with $c_i \in \{a_i, b_i\}$ and

$$\mathrm{sgn}(c_1, \ldots, c_d) := \begin{cases} 1, & \text{if } c_k = a_k \text{ for an even number of k's,} \\ -1, & \text{if } c_k = a_k \text{ for an odd number of k's.} \end{cases}$$

Note that $V_F(B) = \mathbb{P}((X_1, \ldots, X_d) \in B)$. We now can define the general term of a copula:

Definition 5.1. *A copula C is a function from $[0,1]^d \to [0,1]$ such that*

- *if $(u_1, \ldots, u_d) \in [0,1]^d$ and at least one $u_i = 0$, then*
$$C(u_1, \ldots, u_d) = 0.$$
- *$(u_1, \ldots, u_d) \in [0,1]^d$ and $u_i = 1$ for all $i \neq k$, then*
$$C(u_1, \ldots, u_d) = u_k.$$
- *For every $0 \leq a_i \leq b_i \leq 1$ $(i = 1, \ldots, d)$ and $B = [a_1, b_1] \times \cdots \times [a_d, b_d]$,*
$$V_C(B) \geq 0.$$

Note that C is a copula if and only if there exists a random vector $(X_1, \ldots, X_d)$ with uniform marginal distributions (i.e. $F_i(x) = x$ for $0 \leq x \leq 1$) and C is the distribution function of the vector. The connection between copulas and arbitrary random vectors is given by Sklar's theorem.

Theorem 5.1 (Sklar's theorem)**.** *Let $X_1, \ldots, X_d$ be dependent real-valued random variables with distribution function F and marginal distribution functions F_i. Then there exists a copula C such that for all $x_i \in \mathbb{R} \cup \{-\infty, \infty\}$*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \quad (4)$$

If all F_i are continuous, then C is unique. Conversely if C is a copula and F_i are distributions, then there exists a random vector with distribution function defined by (4).

The use of copulas If we are in the situation that we have a random vector $X_1, \ldots, X_d$, where we know the marginal distributions F_i, we now want to find a model for the common distribution. In this situation one can try to fit a parametric copula to the data (see e.g. Joe (1997) or Nelsen (1999) for various families of parametric copulas). Of course one then should apply Goodness-of-fit test (a recent overview can be found in Genest et al. (2008)). Even if one can not be sure that one picks the right copula, it is at least possible to see the impact of dependence to the random variables by using different families of copulas.

5.2 Conditional distributions

If we want to model the common distribution of the weather W and the business index U, one method is to use conditional distributions. Since in nearly all cases the business index has no influence on the weather, we will concentrate on modeling the conditional distribution of U given W. If we denote with w a specific weather situation, then we want to evaluate (for every set A)

$$\mathbb{P}(Y \in A | W = w).$$

Note that this is just the problem of finding a univariate distribution for $Y|W = w$. In some situations, this will be a quite natural approach. Let us reconsider Example 2.1 and use as U the number of guests. In this case one might have an idea what the distribution of U is, given the weather condition.

At last, note that the common distribution function of U and W is given by

$$\mathbb{P}(U \leq y, W \leq w) = \mathbb{P}(Y \leq y | W \leq w)\mathbb{P}(W \leq w) = \int_{-\infty}^{w} \mathbb{P}(Y \leq y | W = w) \mathrm{d}W(w).$$

5.3 Factor models

In a factor model the dependence between two (or more) random variables U and V is modeled with a common factor, i.e. there exists independent random variables $X_1, \ldots, X_n$ and functions f_u and f_w such that $U = f_u(X_1, \ldots, X_n)$ and $W = f_w(X_1, \ldots, X_n)$. For example, let $n = 3$ and

$$U = X_1 + \alpha X_3 \quad \text{and} \quad W = X_2 + \beta X_3,$$

where $\alpha, \beta \in \mathbb{R}$. In this example U and W have the common factor X_3. Since U and W are linear functions of the factors this is also called a linear factor model.

Instead of modeling the distributions of U and W one has to model the distributions of the factors $X_1, \ldots, X_n$ and the functions f_u and f_w. Note that every factor model implies a special copula for the random variables U and W.

5.4 Time series

A time series $(X_t)_{t \in \mathbb{N}}$ is a series of random variables. For example, X_t can be the mean temperature at day t. Since the analysis of time series is a classical topic in statistics as well as in econometrics and meteorology, we do not want to give an overview of existing models of time series. We only want to give a brief example of using time series in the connection of weather risk.

Let us go back to Example 2.1. Assume that we have opened the restaurant in the summer and we want to hedge against the risk of bad weather. Assume that we know for a single day t the connection between U_t (the profit we make on that day) and the weather $W_t := (T_t, P_t)$ (the mean temperature T_t and the precipitation P_t). A priori we do not want to exclude that this connection depends on the parameter t (e.g. if t corresponds to a weekend day or working day). Assume that we are planning for a period of N days and that we can buy a derivate on the cumulative weather index $\hat{W} := (1/N \sum_{t=1}^{N} T_t, 1/N \sum_{t=1}^{N} P_t)$. Our goal is to stabilize the cumulative profit $\hat{U} = 1/N \sum_{t=1}^{N} U_t$. (e.g. we want to maximize our expected utility). To fulfill our task, we have to find the dependence structure of

$(\hat{U}, \hat{T}, \hat{P})$. If we only look at the cumulative values, it can be a quite demanding task to find a good model for the dependence structure. In this case it might be easier to find a model for the time series (U_t, T_t, P_t). See Jewson & Brix (2005) for more details of the pro and cons of these two approaches (daily v.s. index modeling).

5.5 Stochastic processes

A stochastic process is a collection of random variables $(X_t)_{t\geq 0}$, where $t \in \mathbb{R}^+$. Hence it can be seen as a random function of t. On the other hand, one can see the stochastic process as a continuous version of a time series. Now one can ask what are the advantages to go from a discrete model to a continuous model. One possible answer is that it is sometimes easier to get "explicit" solutions when one uses a more involved model (see e.g. Examples 2.3 and 2.3) or in some situations it is more natural to use a continuous model than a discrete model. Since a overview of the theory of stochastic process would be far beyond the aim of this paper we only want to mention some stochastic processes.

- The Brownian motion W_t is a stochastic process with
  - The paths of B_t are continuous.
  - $B_0 = 0$
  - The increments $B_{t_1} - B_0, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent $N(0, t_i - t_{i-1})$ random variables.
- The fractional Brownian motion with Hurst index H is a stochastic process with
  - The paths of B_t^H are continuous.
  - $B_0^H = 0$
  - $B^H(t)$ is a $N(0, t^{2H})$ random variable and $\mathbb{E}[B^H(t)B^H(s)] = 1/2(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$.
- An Ornstein-Uhlenbeck process X_t is a stochastic process that fullfills the stochastic differential equations:
  $$\mathrm{d}X_t = -\theta(X_t - \mu)\mathrm{d}t + \sigma \mathrm{d}B_t,$$
  where B_t is a Brownian motion, σ is the constant volatility, μ is the long term mean, which the process attempts to reach and θ is the rate of convergence to μ.
- A Levy processes L_t is a process with independent and stationary increments. For $t_1 \leq t_2, \cdots \leq t_n$, this means
  $$L_{t_1}, L_{t_1} - L_{t_2}, \ldots, L_{t_n} - T_{t_{n-1}}$$
  are independent and for every $h > 0$, $L_{t_1} - L_{t_1+h}$ has the same distribution as $L_{t_2} - L_{t_2+h}$.

A The Monte Carlo Method

We have seen that in some cases we have to evaluate $m := \mathbb{E}[f(X_1, \ldots, X_d)]$, where f is a function and $X_1, \ldots, X_d$ are random variables. In principle, this problem is equivalent to evaluating a d-dimensional integral. Numerical integration of high-dimensional integrands is not an easy task. An alternative to classical integration rules is the Monte Carlo method. To evaluate m, one generates n independent realization $\mathbf{x}_i$ of the vector $\mathbf{X} = (X_1, \ldots, X_d)$ and then approximates

$$m \approx \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i) =: m_n.$$

If $m < \infty$, then we get by the strong low of large numbers that $\lim_{n\to\infty} m_n = m$. Note that m_n is the realization of the random variable $M_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i)$. If $\sigma^2 = \mathbb{V}ar[f(X_1, \ldots, X_d)] < \infty$ then $\mathbb{V}ar[M_n] = \sigma^2/n$ and from the central limit theorem we get that

$$\lim_{n\to\infty} \sqrt{n}(M_n - m) \xrightarrow{d} N(0, \sigma^2).$$

This means that the length of an 95% confidence interval for m tends to zero like $\mathcal{O}(\sqrt{n})$. Hence we get a probabilistic error bound of order $\mathcal{O}(\sqrt{n})$. Note that this bound does not depend on the dimension d of the integral.

References

Barrieu, P. & Karoui, N. E. (2002), 'Optimal design of weather derivatives', *Algo Research Quaterly* **5**, 79–92.

Carmona, R. & Diko, P. (2005), 'Pricing precipitation based derivatives', *Int. J. Theor. Appl. Finance* **8**(7), 959–988.

de Haan, L. & Ferreira, A. (n.d.), *Extreme value theory*, Springer Series in Operations Research and Financial Engineering.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events, for insurance and finance*, Vol. 33 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin.

Feller, W. (1971), *An Introduction to Probability Theory and its Applications, II (2nd ed.)*, John Wiley & Sons Inc., New York.

Genest, C., Rémillard, B. & Beaudoin, D. (2008), 'Goodness-of-fit tests for copulas: A review and a power study', *Insurance: Mathematics and Economics* **To appear**.

Jewson, S. & Brix, A. (2005), *Weather Derivative Valuation*, Cambridge University Press, Cambridge.

Joe, H. (1997), *Multivariate models and dependence concepts*, Vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Leggio, K. B. (2007), 'Using weather derivatives to hedge precipitation exposure', *Managerial Finance* **33**(4), 246–252.

Malevergne, Y. & Sornette, D. (2006), *Extreme financial risks. From dependence to risk management*, Springer-Verlag, Berlin.

Monoyios, M. (2004), 'Performance of utility-based strategies for hedging basis risk', *Quant. Finance* **4**(3), 245–255.

Nelsen, R. B. (1999), *An introduction to copulas*, Vol. 139 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Shevtsova, I. G. (2007), 'Sharpening of the upper bound of the absolute constant in the berry–esseen inequality', *Theory of Probability and its Applications* **51**(3), 549–553.