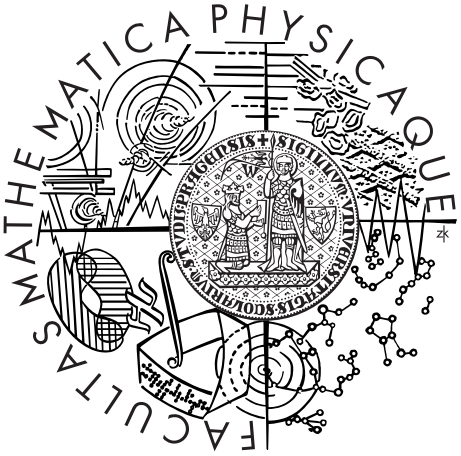
Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS



Mgr. Vadim Omelčenko

Problems of Stochastic Optimization under Uncertainty, Quantitative Methods, Simulations, Applications in Gas Storage Valuation

Department of Probability Theory and Mathematical Statistics

Supervisor of the doctoral thesis: RNDr. Vlasta Kaňková, CSc.

Study programme: Mathematics

Specialization: 4M5 - Econometrics and Operations Research

Prague 2016

Dedication. I dedicate my dissertation to my parents, and to many people who encouraged me and gave me valuable advice. A special feeling of gratitude goes towards my parents, Natalia and Sergey Omelchenko, my grandparents Nina and Mikhail Omelchenko, and my sister Marina. The encouragement of my closest relatives still rings in my ears, which I deeply appreciate. Among my teachers, there were those who either instilled intellectual curiosity in me, or provided inspiration for handling challenging tasks. They are: Nina Volchenko, Kira Zaytseva, Alexandra Shvydkaya, Natalia Tsaran, Tatiana Lebed, Milan Kněžour, Marta Pětová, Přemysl Šedivý, and Miroslav Novak. I send my gratitude to you all.

Acknowledgement. I am very grateful to my supervisor RNDr. Vlasta Kaňková, CSc., for her valuable advice and help. I also want to express my gratitude to Prof. Lev Klebanov, DrSc., RNDr. Martin Šmíd, PhD., Doc., RNDr., Ing. Miloš Kopa, PhD., and RNDr. Vaclav Kozmík, PhD. for their help and comments.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, August 30, 2016 Mgr. Vadim Omelčenko

Název práce: Problémy stochastické optimalizace za neurčitosti, kvantitativní metody, simulace, aplikace na ohodnocení plynových zásobníků.

Autor: Mgr. Vadim Omelčenko

Katedra: Katedra Pravděpodobnosti a Matematické Statistiky

Vedoucí disertační práce: RNDr. Vlasta Kaňková, CSc.

Abstrakt:

Tato dizertační práce se zabývá problematikou pravděpodobnostních rozdělení s těžkými chvosty a problematikou stochastické dominance v případě stabilních rozdělení. Pro stochastickou dominanci v případě stabilních rozdělení jsou dokázány nové výsledky, většinou založené na doméně atrakce tvořené stabilními rozděleními. Dále je v práci zkonstruována podrodina dvourozměrných stabilních rozdělení, které se snadno nasimulují a mohou být použity pro modelování závislých složek dvourozměrných dat (např. forwardových a spotových cen); příslušná marginální rozdělení jsou též stabilní a to obecně s rozdílnou hodnotou parametru alpha (který vyjadřuje tíži chvostu). Konečně v práci je prezentována metoda odhadu parametrů stabilních rozdělení. Dosažené teoretické výsledky jsou aplikovány pro hodnocení plynových zásobníků. V této části jsou využity metody stochastického dynamického programování pro ohodnocení plynových zásobníků a sestrojeno je několik algoritmů řešení.

Klíčová slova: Stabilní rozdělení, Bendersovy řezy, stochastické dynamické programování, stochastická dominance.

Title: Problems of Stochastic Optimization under Uncertainty, Quantitative Methods, Simulations, Applications in Gas Storage Valuation

Author: Mgr. Vadim Omelčenko

Department: Department of Probability Theory and Mathematical Statistics

Supervisor: RNDr. Vlasta Kaňková, CSc.

Abstract: This dissertation deals with heavy-tailed distributions and the problematics of stochastic dominance for stable distributions. In terms of stochastic dominance in the setup of stable distributions, we prove novel results which are mostly based on the domain of attraction of stable distributions. We introduce a bivariate sub-family of stable distributions, which can easily be simulated and used for the joint modelling of dependent data (such as spot and forward prices). The marginals of these bivariate distributions are stable and can have a different tail index. We also present our approach for parameter estimation of stable distributions. The theoretical results achieved are used for the valuation of gas storage units. In this part of the dissertation, we use stochastic dynamic programming to address this problem, and we present several algorithms.

Keywords: Stable distributions, Benders cuts, stochastic dynamic programming, stochastic dominance.

Contents

Introduction

This dissertation focuses on heavy-tailed distributions and stochastic dominance for stable distributions. We present novel results, such as conditions for parameters under which we can infer that one stable random variable first or second order dominates another stable random variable. In addition to this, we explore the second-order stochastic dominance for linear combinations of stable random variables, which relates to portfolio theory. We introduce a bivariate sub-family of stable distributions (denoted by $\mathcal{S}$), which enables us to relate stable random variables with different tail indices: this relationship is linear, if the tail indices coincide. We provide the basic properties of these distributions, and apply them to the joint fitting of prices of gas and electricity, which can be used for the valuation of spark-spread options, and for joint modelling of both spot and forward gas prices. We concentrate on these topics due to their actuality, and their applicability to practical problems.

The analysis and application of heavy tails is still a widely discussed topic [55, 29, 100, 69, 6], and part of this discussion is the question of whether financial data is heavy-tailed or not [47, 33]. Since prices of many commodities such as electricity and gas still exhibit behaviour which is not typical for the Gaussian assumption [100, 69], we consider alternatives to the Gaussian assumption, which enable a more jumpy behaviour of prices. One such alternative is stable distributions, because their application is justified by the Generalised Central Limit Theorem [96]. However, real financial data possesses lighter tails than seen in stable distributions [33]. Therefore, many a researcher has considered truncated and tempered stable distributions as alternatives. This topic is important, because heavy tails may be present in price models, and price models serve as inputs in many practical problems (such as asset valuation, risk management, trading, etc.).

Stochastic dominance enables us to compare different random variables e.g. different payoffs or different sources of risk and in addition to this, it is deeply related to the theory of utility functions [10]. In decision-making theory stochastic dominance takes into account the whole statistical distribution instead of its parts such as moments [10, 104]. Stochastic dominance of the first and second order has a clear economic interpretation, i.e. first-order stochastic dominance is a 'stochastically larger', and second order stochastic dominance a 'stochastically less volatile' or 'less risky' relationship [104]. This means that a 'larger' random variable is preferred by all agents who prefer higher payoffs and a 'less volatile' random variable is preferred by all agents who dislike risk [104]. Another advantage of stochastic dominance is that it can be used for incorporating probabilistic constraints into optimisation problems [94], which is important from many points of view, e.g. risk management. Additionally, it is also related to the value-at-risk and conditional value-at-risk [49]. In recent times, there has been much research on this topic, e.g. [4, 20, 27, 76, 77] to name a few. In this dissertation, we explore stochastic dominance for stable distributions and present novel results, and discuss ways in which we can incorporate probabilistic constraints into the

problem of gas storage valuation, where the underlying innovation of the price process is stable.

The valuation of a gas storage unit is an important practical problem, and the by-product of solving it is a *decision policy* or a *decision rule* which produces a decision based on the available information. When we evaluate a gas storage unit, we also take into account that it can be connected to several markets and show numerically that this feature enables us both to earn more and to reduce risks. There has been recently much research on this topic [91, 39, 105] and there is still space for improvement in addressing this challenging task. We apply the aforementioned theory for the valuation of gas storage units under price uncertainty where we consider several innovations. One such innovation is the distribution $\mathcal{S}$ which we use in the valuation of gas storage within the problems which we denote as *Advanced Problem I* and *Advanced Problem II* where we take into account the connection of gas storage units to several markets. For the valuation of gas storage units, we use stochastic dynamic programming, present situations in which we face the *curse of dimensionality*, and suggest solutions for reducing it. We solve the problem of dynamic programming by linear interpolation and Benders cuts, and in our algorithms, which we denote as *Alg1* and *Alg2*, we incorporate heavy tails of prices into stochastic dynamic programming. Within the valuation of gas storage, we consider three problems:

- *Simple Problem* in which prices are a low-dimensional process, which is either the transition matrix inferred from the price history or the AR(1) process.

- *Advanced Problem I*, which takes into account that European gas storage units are usually connected to several markets; the feature which we have not yet found in the literature.

- *Advanced Problem II*, in which we try to incorporate as many features of real market prices as possible, but within Markovian restrictions. We conduct data analysis, compare different models in terms of different criteria, and then consider ways of dimensionality reduction (one of which is the empirical observation that the spreads of prices at different spot markets possess the mean reverting property).

The main theoretical contribution is as follows:
In the exploration of stochastic dominance in the framework of stable distributions, we found parametric regions where stochastic dominance property holds, and where it can never happen. We introduced a family of symmetric Pareto distributions, which belongs to the domain of attraction of stable distributions. We used this family and the limit properties of stable distributions for the proofs of theorems describing the stochastic dominance properties of stable distributions. We also present our two-phase method of parameter estimation for stable distributions based on Kagan's methodology [41]. This enables us to overcome the limitation of stable distributions not having an explicit form of density and distribution functions.

This dissertation has the following structure:

1. In Chapter 1, we provide the definition of stable distributions: the properties within this which are relevant for our analysis; parameter estimation of stable distribution, and the aforementioned sub-family of stable distributions. We also consider the issues of truncating stable distributions, because their tail might be substantially heavier than demonstrated in real data. Finally, using stable truncated distributions, we compare the fit of gas and electricity prices against alternatives such as hyperbolic, NIG, Student-t, Normal, and stable. When considering joint electricity and gas models, we compare the fit of $\mathcal{S}$ versus sub-Gaussian and normal distributions.

2. In Chapter 2, we explore stochastic dominance for stable distributions and stable portfolios, and briefly mention the Wasserstein metric.

3. In Chapter 3, we describe the problem of gas storage valuation: describe algorithms *Alg1* and *Alg2*; introduce *Simple Problem*, *Advanced Problem I*, and *Advanced Problem II*. We provide a heuristic method which enables us to accelerate calculations at the cost that the solution will be sub-optimal. In Table 3.8 and Table 3.9, we demonstrate how it is important to incorporate the fact that gas storage is connected to several markets, because ignoring this fact leads to underestimated values of the storage. At the end of this chapter, we discuss stochastic dominance constraints and briefly mention forward and spot forward strategies for gas storage valuation.

From a practical point of view, the ultimate goal of this work is to provide a tool which can be useful for problems of asset valuation. This logic can be used not only for the valuation of gas storage, but also other assets (such as hydro storage or power plants) which cannot be valued via classical methods of financial analysis (due to technical parameters).

1. Heavy-Tailed Distributions

Classical models in financial risk management and portfolio optimization are based on normality, but normal models are known to have a number of shortcomings and there is overwhelming empirical evidence that the normality assumption must be rejected [24, 28, 54, 63, 80]. On the other hand, the use of normality is theoretically justifiable because random effects that influence data are caused by a range of micro-effects, which add up and hence allow us to use the central limit theorem [45, 89]. The assumption of normality enables us to make many problems tractable, while replacing normality by another distribution capable of capturing more features of the time series of prices may lead to useless or even dangerous models unless the modeller can properly handle the arising complexity entailed by introducing a non-normal innovation [60]. Hence, the question is: what is the best compromise? How can we transform the model without the need to rule out normality and without dropping common-sense assumptions which entail normality? If we are to replace the normal assumption with a stable one, we will also be able to use the central limit theorem and convolution properties of the normal distribution will be also preserved [89]. However we will need to be able to estimate the parameters of the stable distributions properly, which is a challenging task because stable distributions in general do not have an explicit form of neither the density nor the distribution function except for a few cases [45, 89]. In this paper, we take up this challenge. In the case of density functions of exponential types such as normal or exponential densities, we can easily obtain explicit estimators of parameters such as mean or variance. In such cases, we can easily obtain estimates of parameters for large numbers of observations. The main goal of this work is to develop a method based on maximum likelihood capable of handling large amounts of data and producing decent precision as in the case of exponential type densities. This method consists of two phases: in the first phase, we calculate the means of the harmonic functions of observations and in the second phase, we conduct the main procedure using those average values as inputs. The first phase enables us to deal with large amounts of data and the second phase enables us to attain high precision and is independent of the sample size of the data [2]. This two-phase algorithm is called a method of projections which converges to the maximum likelihood methodology [41, 2]. We apply this to univariate stable and multivariate sub-Gaussian distributions. We denote our methodology with *MLP* in order to emphasise that our algorithm is based on maximum likelihood estimation and projections. Apart from stable distributions, we also consider geometric stable distributions introduced by Klebanov [45], hyperbolic and NIG, whose definition and applications can be found in [13, 8, 9, 42], and Student-t distribution. In section 1.7, we define a family of distributions which we call *dependent stable distributions* and denote as $\mathcal{S}$.

The Chapter's Structure

This chapter is organised as follows. In the introduction, we give a brief description to stable distributions and challenges that arise in estimating their parameters. In the following sections, we give a complete definition of stable distributions, provide their basic properties and discuss contributions of other

authors to estimating their parameters. Then, we provide the description of our method of the estimation of the tail index and dispersion matrix of sub-Gaussian distributions. We introduce our subfamily of multivariate stable distributions, where the tail index might change across the marginals. After this we demonstrate how to fit the financial data by stable and Gaussian models and then move to conclusions.

1.1 Definition of Stable Distributions

There are four equivalent definitions of univariate stable distributions that concern their different statistical properties: two concern convolution properties, one concerns limit properties and the last definition concerns the form of the characteristic function [89].

Definition 1 (Stable distribution 1.)**.** *A random variable X is stable if for X_1 and X_2 independent copies of X and any positive constants a and b,*

$$aX_1 + bX_2 =_d cX + d \tag{1.1}$$

holds for some positive c and some $d \in \mathbb{R}$. The random variable is strictly stable or stable in the narrow sense if (1.1) holds with $d = 0$ for all choices of a and b. A random variable is symmetric stable if it is stable and symmetrically distributed around 0, i.e. $X =_d -X$

Definition 2 (Stable distribution 2.)**.** *A non-degenerate random variable X is stable if and only if for all $n > 1$, there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$X_1 + ... + X_n =_d c_n X + d_n,$$

where $X_1, ..., X_n$ are independent, identical copies of X. X is strictly stable if and only if $d_n = 0$ for all n.

Definition 3 (Stable distribution 3. (Domain of Attraction))**.** *A random variable X is said to have a stable distribution if it has a domain of attraction, i.e., if there is sequence of i.i.d. random variables $Y_1, Y_2, ..., Y_n$ and two sequences of constants d_n and a_n such that*

$$\frac{Y_1 + Y_2 + ... + Y_n}{d_n} + a_n \longrightarrow_d X,$$

where $\longrightarrow_d$ denotes convergence in distribution.

Definition 3 is the reason why stable distribution are popular, because the sums of all random variables which are fat-tailed converge to a stable distribution which is stated in the generalised central limit theorem [96].

Definition 4 (Stable distribution 4.)**.** *A random variable X has a univariate stable distribution if its characteristic function is of the form:*

$$\psi(u) = \begin{cases} \exp(i\mu u) \exp\left(-\sigma^\alpha |u|^\alpha \left(1 - i\beta \cdot \text{sign}(u) \tan\frac{\pi\alpha}{2}\right)\right), & \alpha \neq 1 \\ \exp(i\mu u) \exp\left(-\sigma|u| \left(1 + i\beta \cdot \text{sign}(u) \ln|u| \frac{2}{\pi}\right)\right), & \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$, $\mu \in (-\infty, \infty)$, $\beta \in [-1, 1]$ and $\sigma > 0$.

Note that if $\alpha = 2$ we have a characteristic function of the normal distribution. If $\alpha < 2$ then any moment EX^a with $a \geq \alpha$ is infinite. If $a < \alpha$ then EX^a is finite [45, 89]. Hence if $\alpha < 2$, then the variance of X is infinite. A general univariate stable distribution is denoted as $S_\alpha(\sigma, \beta, \mu)$, where α, σ, β and μ are the tail index, the scale parameter, the skewness parameter and the location parameter, respectively.
Nolan [66] introduced two types of parametrisations for stable distributions: 1-parametrisation which corresponds to Definition 4, and 0-parametrisation for stable distribution which differs in the location parameter from 1-parametrisation, and is defined as follows:

Definition 5 (0-Parametrisation of Stable Distributions [66])**.** *A random variable X is $S_\alpha(\sigma, \beta, \mu; 0)$, i.e. X has a stable distributions with parameters α, σ, β, and μ in 0-parametrisation, if*

$$X =_d \begin{cases} \sigma\left(Z - \beta \tan\frac{\pi\alpha}{2}\right) + \mu, & \alpha \neq 1 \\ \\ \sigma Z + \mu, & \alpha = 1, \end{cases}$$

where Z is a random variable with the characteristic function

$$\mathbb{E}\exp(iuZ) = \begin{cases} \exp\left(-|u|^\alpha \left(1 - i\beta \cdot \text{sign}(u) \tan\frac{\pi\alpha}{2}\right)\right), & \alpha \neq 1 \\ \\ \exp\left(-|u|\left(1 + i\beta \cdot \text{sign}(u) \ln|u|\frac{2}{\pi}\right)\right), & \alpha = 1, \end{cases}$$

i.e. $Z \sim S_\alpha(1, \beta, 0)$. X has the characteristic function

$$\psi(u) = \begin{cases} \exp(i\mu u)\exp\left(-\sigma^\alpha|u|^\alpha\left(1 + i\beta \cdot \text{sign}(u)\left(|\sigma u|^{1-\alpha} - 1\right)\tan\frac{\pi\alpha}{2}\right)\right), & \alpha \neq 1 \\ \\ \exp(i\mu u)\exp\left(-\sigma|u|\left(1 + i\beta \cdot \text{sign}(u)\ln(\sigma|u|)\frac{2}{\pi}\right)\right), & \alpha = 1, \end{cases}$$

Theorem 1 (Generalised Central Limit Theorem [96])**.** *Let $X_1, X_2, ..., X_n$ be iid random variables with the distribution function $F_X(\cdot)$ satisfying the following conditions*

$$1 - F_X(x) \sim \frac{c}{x^\epsilon} \qquad x \longrightarrow \infty, \tag{1.2}$$

$$F_X(x) \sim \frac{d}{|x|^\epsilon} \qquad x \longrightarrow -\infty, \tag{1.3}$$

with $\epsilon > 0$. Then there exist sequences a_n and $b_n > 0$ such that the distribution of the centred and normalised sum

$$Z_n = \frac{1}{b_n}\left(\sum_{i=1}^{n} X_i - a_n\right) \tag{1.4}$$

converges in distribution to the stable random variable with parameters

$$\alpha = \begin{cases} \epsilon & , \epsilon \leq 2 \\ \\ 2 & , \epsilon > 2. \end{cases} \tag{1.5}$$

$$\beta = \frac{c - d}{c + d} \tag{1.6}$$

as $n \longrightarrow \infty$.

Proof. See [96]. □

From Theorem 1, it follows that not only do stable distributions have similar convolution properties to the Gaussian distribution, but also have similar limit properties. In other words, if we assume that the randomness is given as the sum of independent heavy-tailed random effects, then stable distribution is the natural candidate.

The Density Function of Stable Distributions

Let us note that neither the density nor the distribution functions of stable distributions can be expressed in terms of elementary functions in general. The exceptions are as follows:

1. The Gaussian distribution $S_2(\sigma, \beta, \mu) = S_2(\sigma, 0, \mu) = N(\mu, \sigma^2/2)$ [89],
2. The Cauchy distribution $S_1(\sigma, 0, \mu)$ [89],
3. The Lévy distribution $S_{1/2}(\sigma, 1, \mu)$ [89],
4. The Holtsmark distribution $S_{1.5}(\sigma, 0, \mu)$, which can be expressed in terms of hyperbolic functions [53, 65].
5. The Landau distribution $S_1(\sigma, 1, \mu)$ whose density function $\phi(\cdot)$ for $\sigma = 1$ and $\mu = 0$ can be approximated as follows [51, 52]:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x + e^{-x})\right\}$$

The asymptotic properties of the densities of stable distributions $S_\alpha(\sigma, \beta, \cdot)$ with $\alpha < 2$ are given in the following form [65]

$$\phi(x) \sim \frac{\sigma^\alpha (1 + \operatorname{sgn}(x)\beta) \sin(\pi\alpha/2)\Gamma(\alpha + 1)/\pi}{|x|^{1+\alpha}} \text{ as } |x| \to \infty. \tag{1.7}$$

Only the characteristic function of stable distributions $\psi(\cdot)$ has an explicit form so the straightforward way of calculating the density of a stable distribution is the inverse Fourier transform of ψ, i.e.

$$\phi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\psi(t)dt$$

1.2 Basic Properties of Stable Distributions

Theorem 2. *If $X \sim S_\alpha(\sigma, \beta, \mu)$, then $X + a \sim S_\alpha(\sigma, \beta, \mu + a)$.*

Proof. See [89]. □

Theorem 3. *If $X \sim S_\alpha(\sigma, \beta, \mu)$ and $a \in \mathbb{R}$, then*

$$\begin{aligned} a \cdot X &\sim S_\alpha(|a|\sigma, \operatorname{sign}(a)\beta, a \cdot \mu), && \alpha \neq 1 \\ a \cdot X &\sim S_\alpha(|a|\sigma, \operatorname{sign}(a)\beta, a \cdot \mu - \tfrac{2}{\pi} a \ln |a|\sigma\beta), && \alpha = 1 \end{aligned} \tag{1.8}$$

Proof. See [89]. □

Theorem 4. *If $X \sim S_\alpha(\sigma, \beta, \mu)$, then $-X \sim S_\alpha(\sigma, -\beta, -\mu)$.*

Proof. See [89]. □

Theorem 5 (Convolution)**.** *If $X \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$, $Y \sim S_\alpha(\sigma_2, \beta_2, \mu_2)$, $\alpha \neq 1$, and $a, b \in \mathbb{R}$, then*

$$\begin{aligned} &Z = a \cdot X + b \cdot Y \sim S_\alpha(\sigma, \beta, \mu), \\ &\textit{where} \\ &\mu = a\mu_1 + b\mu_2, \\ &\sigma = \sqrt[\alpha]{|a\sigma_1|^\alpha + |b\sigma_2|^\alpha}, \\ &\beta = \frac{\beta_1 |a\sigma_1|^\alpha + \beta_2 |b\sigma_2|^\alpha}{\sigma^\alpha}. \end{aligned}$$

If $X_1, ..., X_N \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$, then the scale parameter of $\sum_{i=1}^N b_i X_i$, with $b_i \in \mathbb{R}$, $i = 1, ..., N$, is

$$\sqrt[\alpha]{\sum_{i=1}^{N} |b_i|^\alpha \sigma_i^\alpha}.$$

Proof. See [89]. □

Corollary

If $X_1, X_2, ..., X_{2n} \sim S_\alpha(\sigma, \beta, \mu)$ then

$$X_1 - X_2, X_3 - X_4, ..., X_{2n-1} - X_{2n} \sim S_\alpha(2^{1/\alpha}\sigma, 0, 0).$$

Since neither the density nor the distribution functions are given in any explicit form, we propose to calculate the former by means of the inverse Fourier transform of the characteristic function, and the latter by means of the Gil-Pilaez Theorem.

Theorem 6 (Gil-Pilaez (1959)[31])**.** *If X is a continuous random variable with the characteristic function $\psi(\cdot)$, where $\mathbb{E}|X| < \infty$, and for some ϵ and c*

$$\psi(u) < c \cdot u^{-\epsilon} \quad \forall u \in \mathbb{R},$$

then

$$\mathbb{P}(X < x) = \frac{1}{2} - \int_{-\infty}^{\infty} \mathrm{Im}\left(\frac{\psi(u)\exp(-iux)}{2\pi u}\right) du \textit{ for all } x \in \mathbb{R}.$$

Proof. See [31]. □

Corollary

If $X \sim S_\alpha(\sigma, 0, 0)$ and $\alpha > 1$, then

$$\mathbb{P}(X < x) = \frac{1}{2} + \int_{-\infty}^{\infty} \frac{\exp(-\sigma^\alpha |u|^\alpha) \cdot \sin(ux)}{2\pi u} du.$$

1.3 Sub-Gaussian Distributions

Definition 6 (Sub-Gaussian Distributions)**.** *The random vector* $\mathbf{X} \in \mathbb{R}^d$ *with* $d \in \mathbb{N}$*, has a sub-Gaussian Distribution with the location parameter* $\boldsymbol{\mu}$*, the tail index* $\alpha \in (0, 2]$*, and the positively semi-definite dispersion matrix* $\mathbf{Q}$ *if its characteristic function is of the form:*

$$\psi(\mathbf{u}) = \exp(i \cdot \mathbf{u}^T \boldsymbol{\mu}) \exp\left(-\left|\mathbf{u}^T \mathbf{Q} \mathbf{u}\right|^{\alpha/2}\right).$$

This vector is called sub-Gaussian.

Here $\boldsymbol{\mu}$ is the location parameter and $\mathbf{Q}$ is the matrix that determines the dependence structure between the marginals. (Another name for these distributions is *multivariate elliptical stable distributions* [1]) $\mathbf{Q}$ is a positively definite matrix and in the case of α equal to 2 we get the multivariate normal distribution whose covariance matrix is $\mathbf{Q}$ [67]. If $\alpha > 1$ then $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$. If $\mathbf{X}_1$, $\mathbf{X}_2$,...,$\mathbf{X}_n$, $\mathbf{X}$ are i.i.d. sub-Gaussian distributions $\mathbf{X}_i \sim \psi(\mathbf{u}) = \exp(-\left|\mathbf{u}^T \mathbf{Q} \mathbf{u}\right|^{\alpha/2})$, $i = 1, 2, ...$, i.e. $\boldsymbol{\mu} = \mathbf{0}$ then

$$\mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_n =_d \frac{1}{n^{1/\alpha}} \mathbf{X},$$

[63, 67]. Sub-Gaussian distributions generalise the multivariate Gaussian distribution as an elliptical sub-family.

1.3.1 Simulating of Sub-Gaussian Distributions

Theorem 7 ([23, 67])**.**

1. *If* $X \sim S_\alpha(1, \beta, 0)$*,* $W \sim \exp(1)$*,* $V \sim U(-\pi/2, \pi/2)$*,* V *and* W *are independent, then:*

$$X =_d \begin{cases} S_{\alpha,\beta} \cdot \frac{\sin(\alpha(V+B_{\alpha,\beta}))}{\cos(V)^{1/\alpha}} \left[\frac{\cos(V-\alpha(V+B_{\alpha,\beta}))}{W}\right]^{\frac{1-\alpha}{\alpha}}, & \alpha \neq 1, \\ \frac{2}{\pi}\left\{\left(\frac{\pi}{2} + \beta V\right)\tan V - \beta \ln\left(\frac{\frac{\pi}{2} W \cos V}{\frac{\pi}{2} + \beta V}\right)\right\}, & \alpha = 1, \end{cases} \tag{1.9}$$

   where

$$B_{\alpha,\beta} = \frac{\arctan\left(\beta \tan\frac{\pi\alpha}{2}\right)}{\alpha},$$

$$S_{\alpha,\beta} = \left[1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right]^{1/2\alpha},$$

   If $X \sim S_\alpha(1, \beta, 0)$ *then for all* $\sigma > 0$ *and* $\mu \in \mathbb{R}$ *we have* $Y = \sigma X + \mu \sim S_\alpha(\sigma, \beta, \mu)$*.*

2. *If* $\mathbf{Z}$ *is a random vector with a characteristic function* $\psi(\mathbf{u}) = \exp(-\{\mathbf{u}^T\mathbf{Q}\mathbf{u}\}^{\alpha/2})$ *then*

$$\mathbf{Z} = \sqrt{s} \cdot \mathbf{G}, \tag{1.10}$$

where $s \sim S_{\alpha/2}\left(\left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{2/\alpha}, 1, 0\right)$ *and* $\mathbf{G} \sim N(\mathbf{0}, \mathbf{Q})$. $\mathbf{G} = \mathbf{C}^T\mathbf{Y}$, $\mathbf{C} \cdot \mathbf{C}^T = \mathbf{Q}$, $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)^T$, $Y_i \sim N(0, 1)$, $i = 1, 2, .., n$ *and* Y_i, $i = 1, 2, .., n$ *are i.i.d. and finally,* s *and* $\mathbf{G}$ *are independent.*

Proof. See [23, 67]. □

Expressions (1.10) and (1.9) enable us to simulate any sub-Gaussian distribution. The simulation of a general multivariate stable distribution is outside of the scope of this thesis.

1.4 General Multivariate Stable Distributions

As it was noted, sub-Gaussian distributions represent a special case of multivariate stable distributions and the latter are defined as follows.

Theorem 8 ([45, 89])**.** *Let* $0 < \alpha \leq 2$. *Then* $\mathbf{X} = (X_1, X_2, ..., X_n)$ *is a stable random vector with the tail index* α *iff there exists a finite measure* Γ *on the unit hypersphere* $S_n = \{\mathbf{s} \in \mathbb{R}^n | \quad \|s\| = 1\}$ *and a vector* $\boldsymbol{\mu} \in \mathbb{R}^n$ *such that for* $\alpha > 0$, *the characteristic function takes the form:*

$$\mathbf{\Psi}_\alpha(\mathbf{u}) = \exp\left\{-\int_{S^d}\left(1 - i \cdot \text{sign}((\mathbf{u}, \mathbf{s}))\tan\left(\frac{\pi\alpha}{2}\right)\right)\Gamma(d\mathbf{s}) + i \cdot (\mathbf{u}, \boldsymbol{\mu})\right\}.$$

The pair $(\boldsymbol{\mu}, \Gamma)$ *is unique.*
The measure Γ*, called the spectral measure of the stable random vector* $\boldsymbol{X}$*, specifies the dependence structure between its marginal distributions.*

Remark.

Note that α is the same for all such marginal distributions.

1.5 Geometric Stable Distributions

Definition 7 ([45])**.** *A random variable* Y *is said to be geometric stable with respect to the summation scheme (in short, geometric stable, or GS) if there exists a sequence of iid. random variables* $X_1, X_2, ...$, *a geometric random variable* ν_p *independent of all* X_i, *and constants* $a = a(p) > 0$ *and* $b = b(p) \in \mathbb{R}$, *such that:*

$$a(p)\sum_{i=1}^{\nu_p}(X_i + b(p)) \longrightarrow_d Y, \text{ as } p \to 0.$$

Geometric stable distributions are, like stable distributions, heavy-tailed [45], and are also characterised by four parameters and denoted $GS_\alpha(\sigma, \beta, \mu)$. These distributions, like stable distributions, do not have an explicit form of either density or the distribution function, but their characteristic function does have an explicit form [45]. The case when $\alpha = 2$ corresponds to the Laplace distributions [45].

Definition 8 (The Definition of $GS_\alpha(\sigma, \beta, \mu)$ [45])**.** *A random variable Y is said to have a geometric stable distribution if its characteristic function ψ has the following form:*

$$\psi(t) = (1 + \sigma^\alpha |t|^\alpha \omega(t, \alpha, \beta) - i\mu t)^{-1},$$

where

$$\omega(t, \alpha, \beta) = \begin{cases} 1 - i\beta \tan(\pi\alpha/2)\text{sign}(t) & , \text{ if } \alpha \neq 1 \\ 1 + i\beta(2/\pi) \ln |t| \text{sign}(t) & , \alpha = 2, \end{cases} \tag{1.11}$$

and

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad -\infty < \mu < \infty, \quad \sigma \geq 0. \tag{1.12}$$

The parameters α, β, σ, and μ are unique (β is irrelevant if $\alpha = 2$).

When $\beta = \mu = 0$ the characteristic function ψ has the following form [45]

$$\psi(t) = \frac{1}{1 + \sigma^\alpha |t|^\alpha} \tag{1.13}$$

By using the shape of ψ, we can prove that $Y \sim GS_\alpha(\sigma, 0, 0)$ belongs to the domain of attraction of $S_\alpha(\sigma, 0, 0)$.
This can be proved as follows: if $Y_1, Y_2, ..$ are iid random variable which have $GS_\alpha(\sigma, 0, 0)$, then $\frac{Y_1 + ... + Y_n}{n^{1/\alpha}}$ has the following characteristic function: Since $Y_i/n^{1/\alpha}$ has the characteristic function

$$\psi_{Y_i/n^{1/\alpha}}(t) = \frac{1}{1 + \frac{\sigma^\alpha |t|^\alpha}{n}},$$

we get that the characteristic function of $(Y_1 + ... + Y_n)/n^{1/\alpha}$ is

$$\prod_{i=1}^{n} \mathbb{E} \exp\left(\frac{itY_i}{n^{1/\alpha}}\right) = \prod_{i=1}^{n} \frac{1}{1 + \frac{\sigma^\alpha |t|^\alpha}{n}}$$

and

$$\prod_{i=1}^{n} \frac{1}{1 + \frac{\sigma^\alpha |t|^\alpha}{n}} = \frac{1}{\left(1 + \frac{\sigma^\alpha |t|^\alpha}{n}\right)^n} \longrightarrow e^{-\sigma^\alpha |t|^\alpha} \text{ as } n \longrightarrow \infty \tag{1.14}$$

In effect, all of geometric stable distributions belong to the domain of attraction of the stable distribution [45] which implies that they also have Paretian tails. The difference between the stable and geometric stable distributions consists in the fact that the latter are given by the sum of a random number of influences which might take place on real markets, therefore this distribution can be an alternative to the stable one if the former provides a fit of the data with an unsatisfying precision [45].

1.6 Truncated and Tempered Stable Distributions

According to many a researcher, stable models are an appropriate choice for modelling financial and commodity data. Rachev et al. (2004) [83] fit Gaussian, hyperbolic, normal inverse Gaussian (NIG), and stable laws to the first differences of the deseasonalised (2001-2003) daily prices at EEX (European Energy Exchange) market, and concluded that the stable laws yielded the best fit, closely followed by the NIG law. Weron (2005) [102] analysed the distributional properties of EEX (2001-2003) and Nord Pool (1997-2000) deseasonalised daily returns and found stable law to perform better than NIG, hyperbolic, and Gaussian distributions, especially for the Nordic market data. Mugele et al. (2005) [62] fit stable and Gaussian laws to Nord Pool (1997-2002), EEX (2000-2002), and PoIPX (five months of 2002) raw and deseasonalised daily spot price differences. They found the stable law to yield a better fit in terms of the goodness-of-fit statistics in all cases except raw PoIPX differences which might be due to the small size of the dataset [16]. In [16, 101], there are analogous results in favour of the stable laws. However Grabchak and Samorodnitsky (2012) [33] when solving the problem of whether the returns of financial data have infinite variance, came to the conclusion that they mostly come from tempered heavy tails (although they might appear to be heavy-tailed). They also demonstrate that the distribution of the sum of many independent iid random variables with tempered heavy tails can be well-approximated by an infinite variance α-stable distribution even though these random variables are not in the domain of attraction of the stable distribution. This implies that we have to be cautious even though the corresponding tests of goodness-of-fit such as Kolmogorov-Smirnov or Anderson-Darling suggest that the data comes from the heavy-tailed family. Weron (2013) made comparisons of models with different innovations, but this time there is no unanimous winner in fitting: he compares the stable distribution with NIG and non-parametric innovations. However, a semi-parametric ARX-NP model generally yields the best performance. For risk management purposes, requiring accurate approximation of the upper quantiles, the heavy-tailed models, i.e. stable and NIG, may be preferential.

Although heavy tails might remain for risk management, they seem to be inappropriate for the actual financial data. Menn and Rachev (2005) [59] proposed using smooth truncations of the density function. This methodology is described in the following subsection. Apart from smoothly truncated distributions there exist such modifications of stable distributions like tempered stable distributions [87, 88] and discrete stable distributions [33, 95]. Rachev (2015) [12] explored Ornstein-Uhlenbeck processes' features for different types of tempered stable laws such as *Classical Tempered Stable* (CTS), *Rapidly Increasing Tempered Stable* (RITS), *Tempered Infinitely Divisible Stable* (TIDS), etc. Klebanov and Slamová (2012) [95] elaborated discrete stable distributions which even have a compact support. They argue that financial data is per se discrete and finite, hence the heavy-tailed assumption is not appropriate. Additionally, they model financial returns with discrete stable laws. Klebanov and Volchenkova (2015) [47] provide empirical results in favour of rejecting the assumption that financial returns are heavy-tailed.

In Section 1.11, we provide our results of the analysis of electricity and gas prices by means of a plethora of models part of which are stable and stable truncated models.

Truncation

When truncating the tails, one approach is to do it 'non-smoothly' : i.e. we would choose an upper and a lower bounds b_u and b_l and if $X \sim S_\alpha(\sigma, \beta, \mu)$, then the truncated random variable X^* can be defined as follows

$$X^* = \begin{cases} b_u, & \text{if } X \geq b_u \\ X, & \text{if } X \in (b_l, b_u) \\ b_l, & \text{if } X \leq b_l. \end{cases}$$

Hence we get a random variable which not only has all of the moments, but also has a compact support. However its density function is not continuous. Another way of truncating the distribution which is proposed by Rachev and Menn (2005) [59] is as follows: we take the density function $\phi_\theta(\cdot)$, where $\theta = (\alpha, \sigma, \beta, \mu)$ and choose points at which we will truncate it, and at these points we replace the tails of the stable distribution with the normal tails in such a way that the value of the corresponding density functions are the same at these points and the squares from those points onward coincide with the initial densities, i.e.:
If we choose b_l and b_u, then

$$\phi^*_\theta(x) = \begin{cases} \phi_1(x), & \text{for } X \geq b_u \text{ where} \\ & \phi_1(b_u) = \phi_\theta(b_u) \text{ and } \int_{b_u}^{\infty}(\phi_\theta(x) - \phi_1(x))dx = 0 \\ \phi_\theta(x), & \text{for } X \in (b_l, b_u) \\ \phi_2(x), & \text{for } X \leq b_l \text{ where} \\ & \phi_2(b_l) = \phi_\theta(b_l) \text{ and } \int_{-\infty}^{b_l}(\phi_\theta(x) - \phi_2(x))dx = 0. \end{cases} \tag{1.15}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are normal densities whose mean and variance are calculated to match (1.15). In this work, we use only stable and truncated stable distributions for data analysis except for Section 1.11.

1.7 Dependent Stable Distributions

Stable distributions with dependent marginals can be simulated by means of Sub-Gaussian, multivariate stable, or operator-stable distributions. Operator-stable distributions represent a family of general multivariate stable distributions where the tail index of the marginals does not have to be the same [65]. The first and second alternatives contain a limitation which consists in the same tail index across all of the marginals. The second and third alternatives entail handling with a (possibly continuous) spectral measure, which brings intractability into many models [65, 40, 58]. In order to respond to the aforementioned limitation and challenge, we introduce our own sub-family of operator stable distributions (which we denote $\mathcal{S}$ and define in this section).

1.7.1 The Function stable$(\cdot,\cdot,\cdot,\cdot)$

We use (1.9) for simulating a stable random variable and following this equation's notation, we define the following function

$$\textbf{stable} : (0,2] \times [-1,1] \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times \mathbb{R}_+ \to \mathbb{R} :$$

$$\begin{aligned}
&\text{If } \alpha \neq 1, \text{then} \\
&\textbf{stable}(\alpha, \beta, V, W) = \\
&= S_{\alpha,\beta} \cdot \frac{\sin(\alpha(V + B_{\alpha,\beta}))}{\{\cos(V)\}^{1/\alpha}} \cdot \left[\frac{\cos(V - \alpha(V + B_{\alpha,\beta}))}{W}\right]^{(1-\alpha)/\alpha}. \\
&\text{If } \alpha = 1, \text{then} \\
&\textbf{stable}(1, \beta, V, W) = \\
&= \frac{2}{\pi}\left\{\left(\frac{\pi}{2} + \beta V\right)\tan V - \beta \ln\left(\frac{\frac{\pi}{2} W \cos V}{\frac{\pi}{2} + \beta V}\right)\right\}.
\end{aligned}$$

So if $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $V \sim U(-\pi/2, \pi 2)$, $W \sim \exp(1)$, and V and W are independent, then by (1.9), $\textbf{stable}(\alpha, \beta, V, W) \sim S_\alpha(1, \beta, 0)$. Additionally, if $X = \textbf{stable}(\alpha, \beta, V, W)$ then a stable distribution Y with arbitrary scale and location parameters can be defined as follows:

$$Y = \begin{cases} \mu + \sigma X & , \alpha \neq 1 \\ \\ \mu + \sigma X + \sigma\beta(2/\pi)\log(\sigma) & , \alpha = 1. \end{cases}$$

Hence, $Y \sim S_\alpha(\sigma, \beta, \mu)$.
Since the location and scale parameters can be changed by sheer linear transformation, we only concentrate on the tail index α and the skewness β in this section.

1.7.2 The Definition of Family $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$

Definition 9. *Let $V, V' \sim U(-\pi/2, \pi 2)$, $W, W' \sim \exp(1)$, and the random variables V, V', W, W' be independent of each other and let us assume that $\alpha_1, \alpha_2 \in (0, 2]$, $\beta_1, \beta_2, \beta'' \in [-1, 1]$, and $r \in [0, 1]$. Then we call the random vector $\mathbf{Y}$ defined as follows*

$$\begin{aligned}
&\mathbf{Y} = (Y_1, Y_2), \\
&\textit{where} \\
&Y_1 = \textbf{stable}(\alpha_1, \beta_1, V, W), \\
&Y_2' = \textbf{stable}(\alpha_2, \beta_1, V, W), \\
&Y_2'' = \textbf{stable}(\alpha_2, \beta'', V', W'), \\
&Y_2 = rY_2' + (1 - |r|^{\alpha_2})^{1/\alpha_2} Y_2'', \\
&\beta_2 = r^{\alpha_2}\beta_1 + (1 - |r|^{\alpha_2})\beta'', \qquad (1.16) \\
&-1 \leq \frac{\beta_2 - r^{\alpha_2}\beta_1}{1 - r^{\alpha_2}} \leq 1, \qquad (1.17) \\
&Y_1 \sim S_{\alpha_1}(1, \beta_1, 0) \textit{ and } Y_2 \sim S_{\alpha_2}(1, \beta_2, 0)
\end{aligned}$$

dependent stable *and denote it $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ or $\mathcal{S}$.*

Note that in (1.16), we used Theorem 5. Inequality (1.17) does not have always to hold and in connection with this feature, we introduce the property of *impelementability*.

Definition 10 (Implementability)**.** *We call a random vector* $(X, Y) \sim \mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *or a distribution* $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *implementable if*

$$-1 \leq \frac{\beta_2 - r^{\alpha_2}\beta_1}{1 - r^{\alpha_2}} \leq 1.$$

We denote the value $\frac{\beta_2 - r^{\alpha_2}\beta_1}{1 - r^{\alpha_2}}$ *by* $\zeta([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *or* ζ.

Theorem 9 (Implementability theorem)**.** *If* $\alpha_1 = \alpha_2 = \alpha$, $\beta_1, \beta_2 \in [0, 1]$, $r \in [0, 1]$,*and*

$$-1 \leq \zeta([\alpha, \beta_1], [\alpha, \beta_2], r) \leq 2,$$

then at least one of the following distributions

1. $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *and*
2. $\mathcal{S}([\alpha_2, \beta_2], [\alpha_1, \beta_1], r)$

is implementable.

Proof. Let us note that

$$\begin{aligned}&\zeta([\alpha, \beta_1], [\alpha, \beta_2], r) + \zeta([\alpha, \beta_2], [\alpha, \beta_1], r) = \\ &= \frac{\beta_2 - r^{\alpha}\beta_1}{1 - r^{\alpha}} + \frac{\beta_1 - r^{\alpha}\beta_2}{1 - r^{\alpha}} = \beta_1 + \beta_2 \leq 2\end{aligned}$$

So

$$\zeta([\alpha, \beta_2], [\alpha, \beta_1], r) = \beta_1 + \beta_2 - \zeta([\alpha, \beta_1], [\alpha, \beta_2], r). \tag{1.18}$$

If $\zeta([\alpha, \beta_1], [\alpha, \beta_2], r) > 1$, then since $\zeta([\alpha, \beta_1], [\alpha, \beta_2], r) + \zeta([\alpha, \beta_2], [\alpha, \beta_1], r) \leq 2$, we get that $\zeta([\alpha, \beta_2], [\alpha, \beta_1], r) < 1$. Otherwise if $\zeta([\alpha, \beta_1], [\alpha, \beta_2], r) \in [0, 1]$, then the conditions of the theorem are fulfilled. □

From Theorem 9, we can deduce that

Corollary 1. *If* $\alpha_1 = \alpha_2 = \alpha$, $\beta_1, \beta_2 \in [-1, 0]$, $r \in [0, 1]$,*and*

$$-2 \leq \zeta([\alpha, \beta_1], [\alpha, \beta_2], r) \leq 1,$$

then at least one of the following distributions

1. $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *and*
2. $\mathcal{S}([\alpha_2, \beta_2], [\alpha_1, \beta_1], r)$

is implementable.

Theorem 10 (Evident properties of implementability)**.** $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ *is implementable, if*

1. $r = 0$,
2. $\beta_1 = \beta_2$,
3. $\alpha_1 = \alpha_2 = 2$.

Proof. The first and second properties are evident from the shape of ζ. The third property corresponds to the Gaussian distribution which does not depend on β by Definition 4. □

1.7.3 The Characteristic Function of $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$

If $\alpha_1, \alpha_2 \in (0, 2]$, $\beta_1, \beta_2 \in [-1, 1]$, and $r \in [-1, 1]$, then we can easily simulate any random vector $\mathbf{Y} \in \mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$, but the characteristic function has an explicit form only for special cases of α_j, β_j and r, where $j = 1, 2$.

1. When $r = 0$ then Y_1 and Y_2 are independent and the characteristic function of $\mathbf{Y}$ can be determined from the univariate characteristic functions of the marginals.

2. If $\alpha_1 = \alpha_2 = \alpha$ and when β_1, β_2 and r satisfy the condition of the following theorem.

Theorem 11. *Let us assume that $\alpha \in (0, 2]$, $\beta_1, \beta_2 \in [-1, 1]$, and $r \in [0, 1)$. Then $\mathbf{Y} \sim \mathcal{S}([\alpha, \beta_1], [\alpha, \beta_2], r)$ has the characteristic function in the form:*

$$\psi(t_1, t_2) = \exp\left(-|t_1 + rt_2|^\alpha K_1 - |t_2|^\alpha (1 - r^\alpha) K_2\right), \tag{1.19}$$

with

$$K_j = \begin{cases} 1 - i\beta_j \text{sign}(\mathbf{t}_j) \tan\frac{\pi\alpha}{2}, & \alpha \neq 1 \\ 1 - i\beta_j \text{sign}(\mathbf{t}_j) \frac{2}{\pi} \ln|t_j|, & \alpha = 1 \end{cases}, \quad j = 1, 2,$$

and $\mathbf{t}_1 = t_1 + rt_2, \mathbf{t}_2 = t_2$,

if

$$-1 \leq \frac{\beta_2 - r^\alpha \beta_1}{1 - r^\alpha} \leq 1. \tag{1.20}$$

Proof. Let us note that in Definition 9, $Y_1 \equiv Y_2$ if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. So following the notation of Definition 9, we get $Y_2 = rY_2' + (1 - |r|^\alpha)^{1/\alpha} Y_2''$. By Theorem 5, the random variable $rY_2' + (1 - |r|^\alpha)^{1/\alpha} Y_2''$ has the following β parameter

$$\beta_1 r^\alpha + \beta''(1 - r^\alpha) = \beta_2,$$

where β_2 is reachable through β_1 and r only if (1.20) holds. So if (1.20) is satisfied, then taking into account that under the condition $\alpha_1 = \alpha_2$, the random variable Y_1 equals Y_2' a.s., we get the following characteristic function

$$\begin{aligned}
&\psi(t_1, t_2) = \\
&\mathbb{E}\exp\left(i(t_1 Y_1 + t_2 Y_2)\right) = \\
&\mathbb{E}\exp\left(i(t_1 Y_1 + t_2 r Y_1 + (1 - r^\alpha)^{1/\alpha} t_2 Y_2'')\right) = \\
&\mathbb{E}\exp\left(i(t_1 + t_2 r) Y_1 + i(1 - r^\alpha)^{1/\alpha} t_2 Y_2'')\right) = \\
&\mathbb{E}\exp\left(i(t_1 + t_2 r) Y_1\right) \cdot \exp\left(i(1 - r^\alpha)^{1/\alpha} t_2 Y_2'')\right) = \\
&\exp\left(-|t_1 + rt_2|^\alpha K_1\right) \cdot \exp\left(-|t_2|^\alpha (1 - r^\alpha) K_2\right) = \\
&\exp\left(-|t_1 + rt_2|^\alpha K_1 - |t_2|^\alpha (1 - r^\alpha) K_2\right), \\
&\text{with} \\
&K_j = \begin{cases} 1 - i\beta_j \text{sign}(\mathbf{t}_j) \tan\frac{\pi\alpha}{2}, & \alpha \neq 1 \\ 1 - i\beta_j \text{sign}(\mathbf{t}_j) \frac{2}{\pi} \ln|t_j|, & \alpha = 1 \end{cases}, \quad j = 1, 2, \\
&\text{and } \mathbf{t}_1 = t_1 + rt_2, \mathbf{t}_2 = t_2,
\end{aligned}$$

which proves the statement. □

Definition 11. *Let $V, V' \sim U(-\pi/2, \pi/2)$, $W, W' \sim \exp(1)$, and the random variables V, V', W, W' be independent of each other and let us assume that $\alpha_1, \alpha_2 \in (0, 2]$, $\beta_1, \beta_2, \beta'' \in [-1, 1]$, and $r \in [0, 1]$. Then we call the random vector $\mathbf{Y}$ defined as follows*

$$
\begin{aligned}
&\mathbf{Y} = (Y_1, Y_2),\\
&\textit{where}\\
&Y_1 = \mathbf{stable}(\alpha_1, \beta_1, V, W),\\
&Y_2' = \mathbf{stable}(\alpha_2, \beta', V, W),\\
&Y_2'' = \mathbf{stable}(\alpha_2, \beta'', V', W'),\\
&Y_2 = rY_2' + (1 - |r|^{\alpha_2})^{1/\alpha_2} Y_2'',\\
&\beta_2 = r^{\alpha_2}\beta' + (1 - |r|^{\alpha_2})\beta'',\\
&\beta', \beta'' \in [-1, 1],\\
&Y_1 \sim S_{\alpha_1}(1, \beta_1, 0) \textit{ and } Y_2 \sim S_{\alpha_2}(1, \beta_2, 0)
\end{aligned}
$$

generalised dependent stable *and denote it $\mathcal{T}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r, (\beta', \beta''))$.*

A Remark on $\mathcal{S}$ and $\mathcal{T}$

It is evident that for every $\alpha_1, \alpha_2 \in (0, 2], \beta_1, \beta_2 \in [-1, 1]$, and $r \in [0, 1]$ there exist such β' and β'' that $\mathcal{T}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r, (\beta', \beta''))$ is implementable.
$\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ is a special case of $\mathcal{T}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r, (\beta', \beta''))$ where $\beta' = \beta_1$ and $\beta'' = \zeta$. We prefer $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ because for $\alpha_1 = \alpha_2$, there is a linear dependence between two marginal random variables.
Let us assume that $\mathbf{X}_1$ and $\mathbf{X}_2$ ($(\mathbf{X}_1, \mathbf{X}_2) = (X_{i1}, X_{i2}), i = 1, 2, .., n$ where n is the sample size) are two normalised samples from a stable distribution $S_{\alpha_1}(1, \beta_1, 0)$ and $S_{\alpha_2}(1, \beta_2, 0)$, respectively.
When we fit two dimensional data $(\mathbf{X}_1, \mathbf{X}_2)$ by the $\mathcal{S}$ distribution, we conduct procedures:

1. We fit $(\mathbf{X}_1, \mathbf{X}_2)$ by $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$,
2. We fit $(\mathbf{X}_2, \mathbf{X}_1)$ by $\mathcal{S}([\alpha_2, \beta_2], [\alpha_1, \beta_1], r)$,

We choose such a distribution out of the two for which the p-value is higher and the Kolmogorov-Smirnov statistic is larger.

1.8 Parameter Estimation of Stable Distributions

In this section, we make a survey into methodologies of the estimation of the parameters of stable distributions, and provide the description of the method of projection which is based on the work of Kagan [41].

Approaches to Parameters Estimation of Stable Distributions

In the case of univariate stable distributions, we need to estimate four parameters: α, σ, β, and μ. The most challenging task is estimating the parameter α [45]. In the case of a general multivariate stable distribution, we estimate the parameter α, vector $\boldsymbol{\mu}$ and the *spectral measure* Γ, which bears all of the information about the values β and σ of their univariate marginals. It can easily be shown that the matrix $\mathbf{Q}$ is a special case of Γ [24, 45, 89]. Univariate marginals of sub-Gaussian distributions are symmetric, i.e. $\beta = 0$ for all of the marginals [89].

1.8.1 Parameters of Univariate Stable Distributions

For estimating α, there are many methods. There are estimators in a form of an explicit function of observations like the Hill estimator [37, 61]. The expression for the Hill estimator is:

$$\hat{\alpha} = \frac{1}{\frac{1}{k}\sum_{j=1}^{k} \log X_{n+1-j:n} - \log X_{n-k:n}},$$

where $X_{j:n}$ is the sample j-th order statistics and k is the *window* parameter. This estimator is easy to implement, but there are a number of shortcomings associated with it because we take into account only those observations that are in the tail of the distribution whose number is a small fraction of all observations. The definition of what is in the tail and what is not there is problematic if we do not know α. For this purpose, there exists the *window* parameter of the Hill estimator whose calibration requires from us to use sophisticated numerical methods and heavy dependency on the *window* size is the main problem [37, 61]. There are methods for estimating α based on quantiles. The main idea of McCulloch's methodology (1986) is to use differences in quantiles, properly normalised, in order to get rid of our dependence on location and scale parameters. Then, two functions of the stability index and the skewness are numerically calculated from the sample quantiles values and inverted to get the corresponding parameter estimates. This method is consistent and can be used for the estimation of all of the parameters of a univariate stable distribution, however it does not work well for certain choices of the parameters' values [56]. There are methods of estimating α and the rest of the parameters by using the empirical characteristic function. The main idea of this methodology consists in minimizing the distance between the characteristic function (CF) and the *empirical characteristic function* (ECF) in an appropriate norm [21, 22, 30, 48]. Let us denote by ψ the characteristic

function of a stable distribution and by

$$\hat{\psi}(u) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(iuX_j\right)$$

its *empirical characteristic function*. Since $|\hat{\psi}(u)|$ is bounded, all moments of $\hat{\psi}(u)$ are finite for any fixed u. By the Law of Large Numbers $\hat{\psi}(u)$ is a consistent estimator of $\psi(u)$.
The method finds

$$\theta = \text{argmin}_{\theta\in\Theta}\|\hat{\psi} - \psi\|,$$

where θ is a point in the parametric space Θ and $\|\cdot\|$ is a norm usually L^∞ or an L^r weighted norm with $r > 0$. The last type of the norm is more useful for implementation and it can be written as

$$h(\theta) = \int_{-\infty}^{\infty} |\hat{\psi}(u) - \psi(u,\theta)|^r W(u)du.$$

Here $W(\cdot)$ is a weight function and the optimal estimate is obtained as follows:

$$\theta = \text{argmin}_\theta h(\theta).$$

More generally the objective function is of the form:

$$h(\theta) = \int_{-\infty}^{\infty} |\hat{\psi}(u) - \psi(u,\theta)|^r dF(u),$$

where $F(\cdot)$ is a distribution function. When $r = 2$, the objective function becomes

$$h(\theta) = \sum_{i=1}^{n} |\hat{\psi}(u_i) - \psi(u_i,\theta)|^2 \phi(u_i)$$

with $\phi(u) = \frac{dF(u)}{du}$. The optimal selection of discrete points $u_1, u_2, ..., u_p$ is discussed in Carrasco, Madan et. al, and Schmidt [21, 92]. We will denote such estimates as "*CFB*" where this abbreviation states for *characteristic function based* estimators.

DuMouchel (1971), Zolotarev (1986), and Nolan (2001) developed methods of asymptotic maximum likelihood for estimating the parameters of stable distributions [26, 63, 106]. The main limitation is that density functions of stable distributions do not have an explicit form. Approximate methods are based on approximating functionals of density functions. For a general problem of estimating the parameters $\alpha, \sigma, \beta, \mu$ we maximise the following expression:

$$l(\alpha,\sigma,\beta,\mu) = \sum_{i=1}^{n} \log\phi\left(X_i;\alpha,\sigma,\beta,\mu\right),$$

where $\phi(\cdot)$ is the density function that we have to approximate. The reason why asymptotic maximum likelihood methods are popular in estimating parameters is the fact that maximum likelihood estimators are consistent, efficient, and have

asymptotic normality. For large samples, if $\hat{\theta}_n$ is a maximum likelihood estimator of $\theta = (\alpha, \sigma, \beta, \mu)$ then

$$\hat{\theta}_n \sim N(\theta, n^{-1}\mathbf{B}),$$

where n is the sample size and $\mathbf{B}$ is the inverse of the Fisher information matrix $I(\theta) = \{I_{ij}(\theta)\}_{i,j=1,...,n}$ where

$$I_{ij}(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta_i} \log \phi_\theta(X) \cdot \frac{\partial}{\partial \theta_j} \log \phi_\theta(X) \right), \quad i, j = 1, ..., n.$$

So, the sample confidence intervals for each of the parameters are:

$$\hat{\theta}_i \pm z_{1-\gamma/2} \frac{\sigma_{\hat{\theta}_i}}{\sqrt{n}},$$

where $\sigma_{\hat{\theta}_1}$, $\sigma_{\hat{\theta}_2}$, $\sigma_{\hat{\theta}_3}$, and $\sigma_{\hat{\theta}_4}$ are the square roots of the diagonal entries of $\mathbf{B}$ [2]. Asymptotic maximum likelihood estimators developed by the aforementioned authors converge to maximum likelihood and their quality is determined by the rate of the convergence and easiness of their implementation. In this paper, we present a method for estimating α that is also based upon maximum likelihood estimation and can be efficiently implemented by enumerating in one dimension.

1.8.2 Parameters of Multivariate Stable Distributions

If $\mathbf{X}$ is a d-dimensional random vector that has a multivariate stable distribution with the location parameter μ, tail index α, and the *spectral measure* Γ then for $\alpha > 1$ $\boldsymbol{\mu} = \mathbb{E}\mathbf{X}$ and

$$\hat{\alpha} = \frac{1}{d} \sum_{j=1}^{d} \hat{\alpha}_j,$$

here $\hat{\alpha}_j$ is the estimate of α obtained from the observations of the marginal X_j of $\mathbf{X}$, $j = 1, .., d$ [50, 89]. However, we cannot get the *spectral measure* by separate analyses of univariate marginals because it determines the dependence structure between them [89]. There is a plethora of approaches to estimating the *spectral measure*. Nolan and Panorska [64], develop a method based upon a discrete approximation of the spectral measure. Pivato and Seco (2002) estimate the *spectral measure* by its representation as the sum of spherical harmonic functions [75]. McCulloch (2000) developed the method of estimating the *spectral measure* of a generalised bivariate stable distribution, based on a series of maximum likelihood (ML) estimates of the stable parameters of univariate projections of the data [57]. For a more detailed description of the *spectral measure* see Samorodnitsky [89].

1.8.3 Dispersion Matrix Q

As it was mentioned above, the matrix $\mathbf{Q}$ of sub-Gaussian distributions is a special case of the spectral measure. When we deal with the dispersion matrix with the rank d, we can estimate its diagonal elements q_{ii}, as the scale parameter of one-dimensional marginals X_i, $i = 1, .., d$ and the elements q_{ij} with $i \neq j$, $i, j = 1, ..d$ can be estimated from the bivariate random vector (X_i, X_j) [2]. This means that large dimensions are not a significant impediment in the case of sub-Gaussian distributions. However, in the case of general stable distributions, it

is almost impractical to estimate the *spectral measure* if the dimension is higher than 3 [57]. In Kring et. al [50], there is a method of finding a dispersion matrix with moment type estimators. Our method, whose description is provided in the following sections, is based on the maximum likelihood.

1.9 Parameter Estimation of Sub-Gaussian Distributions

If we want to estimate the parameters of a multivariate sub-Gaussian distribution with parameters α and $\mathbf{Q}$, we can do it in two stages:

- Estimation of the α parameter,
- Estimation of matrix $\mathbf{Q}$ using the estimate of α as an input,

where

$$\mathbf{Q} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 r_{12} & ... & ... & \sigma_1\sigma_d r_{1d} \\ \sigma_1\sigma_2 r_{12} & \sigma_2^2 & ... & ... & \sigma_2\sigma_d r_{2d} \\ \sigma_1\sigma_3 r_{13} & \sigma_2\sigma_3 r_{23} & ... & ... & \sigma_d\sigma_3 r_{23} \\ ... & ... & ... & ... & ... \\ \sigma_1\sigma_d r_{1d} & \sigma_2\sigma_d r_{2d} & ... & ... & \sigma_d^2 \end{pmatrix}.$$

The former task can be conducted by analyzing the marginal distributions because all of the marginals have the same α parameter. After estimating α, we can put its estimate $\hat{\alpha}$ into the formula of the characteristic function and use it to estimate the matrix $\mathbf{Q}$. Both estimation procedures will be conducted using the methodology of projections.

1.9.1 Estimation of the Tail Index.

If $\mathbf{X}$ has a d-variate sub-Gaussian distribution with the parameters α, $\mathbf{Q}$, and $\mu = \mathbf{0}$ then every marginal $X_{\{i\}}$, $i = 1, 2, .., d$ of $\mathbf{X}$ has a stable distribution $S_\alpha(\sigma_i, 0, 0)$ with the characteristic function in the form: $\psi(u) = \exp(-\sigma_i^\alpha |u|^\alpha)$. If we have an estimate $\hat{\sigma}_i$ of the scale parameter, then we approximately have $X/\hat{\sigma}_i \sim S_\alpha(1, 0, 0)$. This relation is exact when we have the exact value of the scale parameter. Therefore, in this subsection, we concentrate on only the estimation of the tail index of distributions $S_\alpha(1, 0, 0)$. If $\phi(x, \alpha)$ is a density function of the univariate stable distribution then:

$$I(\alpha) = \int_{-\infty}^{\infty} J^2(x, \alpha)\phi(x, \alpha)dx, \quad J(x, \alpha) = \frac{\partial L(x, \alpha)}{\partial \alpha} = \frac{\left(\frac{\partial \phi(x,\alpha)}{\partial \alpha}\right)}{\phi(x, \alpha)}, \tag{1.21}$$

$$\hat{\alpha}_{ML} = \left\{ \alpha : \sum_{j=1}^{n} J(X_j, \alpha) = 0 \right\}, \tag{1.22}$$

where $I(\alpha)$ is the Fisher information and $X_1, X_2, ..., X_n$ is the vector of observations. This shift to the sum follows from the expression of the maximum likelihood function

$$L(\mathbf{X}, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \ln \phi(X_j|\alpha),$$

which is an additive function.

1.9.2 The Core of the Methodology

Using the methodology of projections, we can approximate the function $J(X, \alpha)$ which enables us to obtain the estimates whose precision converges to that of the ML-estimators, and to calculate the Fisher information [41, 2]. We will express the approximation of the function $J(X, \alpha)$ in terms of

$$\{1, \exp(it_1X), \exp(it_2X), ..., \exp(it_kX)\},$$

i.e., its approximation will be in the form $J_k(X, \alpha)$ where

$$J_k(X, \alpha) = \sum_{j=0}^{k} a_j \exp(it_jX) = \sum_{j=0}^{k} a_j \cos(t_jX) + i \sum_{j=0}^{k} a_j \sin(t_jX), \qquad (1.23)$$

where $t_1, ..t_k$ are different constants that can be chosen arbitrarily. We used the constants $t_1, ..t_k$ whose absolute values are smaller than 1 and $a_1, a_2, ..., a_k$ are the unknown values that we need to estimate.

Remark. Note that we refer to the almost-sure convergence of $J_k(X, \alpha)$ to $J(X, \alpha)$ as $k \to \infty$ provided that $X \sim S_\alpha(1, 0, 0)$ [41, 95, 46].

The idea of the method described below was proposed by Kagan [41]. However, it was applied to other types of distributions where power projections were used instead of trigonometric projections. We project onto the space with a scalar product defined as follows:

If $X \sim S_\alpha(1, 0, 0)$ and t_m, t_n are constants, then a product between $\exp(it_mX)$ and $\exp(it_nX)$ is defined as

$$\langle \exp(it_mX), \exp(it_nX) \rangle = \mathbb{E} \exp(it_mX) \cdot \exp(it_nX) = \mathbb{E} \exp(iX(t_m + t_n)) =$$

$$= \int_{-\infty}^{\infty} \phi(x, \alpha) \exp(ix(t_m + t_n))dx = \exp\left(-|t_m + t_n|^\alpha\right).$$

For any projection $J_k(x, \alpha)$, it holds:

$$(J_k(x, \alpha) - J(x, \alpha)) \perp \exp(it_jx), j = 1, 2, .., k, \qquad (1.24)$$

in other words:

$$\langle (J_k(X, \alpha) - J(X, \alpha)), \exp(it_jX) \rangle = 0, j = 1, 2, .., k \qquad (1.25)$$

or

$$E((J_k(X, \alpha) - J(X, \alpha)) \cdot \exp(it_jX)) = 0, j = 1, 2, .., k. \qquad (1.26)$$

Hence, we have:

$$\int_{-\infty}^{\infty}(J_k(x,\alpha)-J(x,\alpha))\phi(x,\alpha)\exp(it_jx)dx=0$$

or

$$\int_{-\infty}^{\infty}J_k(x,\alpha)\phi(x,\alpha)\exp(it_jx)dx=\int_{-\infty}^{\infty}J(x,\alpha)\phi(x,\alpha)\exp(it_jx)dx.$$

Let us calculate each integral in the above equality separately:

$$\int_{-\infty}^{\infty}J_k(x,\alpha)\phi(x,\alpha)\exp(it_jx)dx=\int_{-\infty}^{\infty}\sum_{v=0}^{k}a_v\exp(it_vx)\phi(x,\alpha)\exp(it_jx)dx=$$

$$=\sum_{v=0}^{k}a_v\int_{-\infty}^{\infty}\exp(it_vx)\phi(x,\alpha)\exp(it_jx)dx=\sum_{v=0}^{k}a_v\int_{-\infty}^{\infty}\phi(x,\alpha)\exp(ix(t_j+t_v))dx=$$

$$=\sum_{v=0}^{k}a_v\exp\left(-|t_v+t_j|^{\alpha}\right),j=1,2,..,k.$$

We can reverse the order of the sum and the integral because the number of the items in the sum is finite. The second integral will be calculated as follows:

$$\int_{-\infty}^{\infty}J(x,\alpha)\phi(x,\alpha)\exp(it_jx)dx=\int_{-\infty}^{\infty}\frac{\left(\frac{\partial\phi(x,\alpha)}{\partial\alpha}\right)}{\phi(x,\alpha)}\phi(x,\alpha)\exp(it_jx)dx$$

$$=\int_{-\infty}^{\infty}\frac{\partial\phi(x,\alpha)}{\partial\alpha}\exp(it_jx)dx=\frac{\partial}{\partial\alpha}\int_{-\infty}^{\infty}\phi(x,\alpha)\exp(it_jx)=\frac{\partial}{\partial\alpha}\exp\left(-|t_j|\right)^{\alpha}.$$

The integral and derivative can be interchanged because of the Leibnitz rule and the fact that $\phi(x,\alpha)\geq 0$. Hence, we get the following system of linear equations:

$$\sum_{v=0}^{k}a_v\exp\left(-|t_v+t_j|\right)^{\alpha}=\frac{\partial}{\partial\alpha}\exp\left(-|t_j|\right)^{\alpha},j=1,2,..,k,$$

which can be written in the form:

$$\frac{\partial}{\partial\alpha}\exp\left(-|t_j|\right)^{\alpha}=-\exp\left(-|t_j|\right)^{\alpha}\cdot|t_j|^{\alpha}\cdot\ln|t_j|,$$

$$\sum_{v=0}^{k}a_v\exp\left(-|t_v+t_j|^{\alpha}\right)=-\exp\left(-|t_j|\right)^{\alpha}\cdot|t_j|^{\alpha}\cdot\ln|t_j|,j=1,2,..,k.$$

In the matrix form, the system looks as follows:

$$\begin{pmatrix}1 & e^{-|t_1|^{\alpha}} & ... & e^{-|t_k|^{\alpha}}\\ e^{-|t_1|^{\alpha}} & e^{-|t_1+t_1|^{\alpha}} & ... & e^{-|t_1+t_k|^{\alpha}}\\ ... & ... & & ...\\ ... & ... & & ...\\ ... & ... & & ...\\ e^{-|t_k|^{\alpha}} & e^{-|t_k+t_1|^{\alpha}} & ... & e^{-|t_k+t_k|^{\alpha}}\end{pmatrix}\cdot\begin{pmatrix}a_0\\ a_1\\ a_2\\ ..\\ ..\\ a_k\end{pmatrix}=\begin{pmatrix}0\\ -|t_1|^{\alpha}\ln|t_1|e^{-|t_1|^{\alpha}}\\ -|t_2|^{\alpha}\ln|t_2|e^{-|t_2|^{\alpha}}\\ ..\\ ..\\ -|t_k|^{\alpha}\ln|t_k|e^{-|t_k|^{\alpha}}\end{pmatrix}.$$

We take into account the fact that $\alpha \in (1, 2]$ in financial applications, and substitute its values into the system of equations where $a_1, a_2, .., a_k$ are unknown. We solve this system by manipulating the values of α and choosing such a value of the tail index for which $\left|\sum_{j=1}^{n} J_k(X_j, \alpha)\right|$ is minimal. Note that in this system of equations only $a_1, a_2, .., a_k$ are unknown.

If we denote the components of the previous equation as follows

$$\mathbf{A}(\alpha) = \begin{pmatrix} 1 & e^{-|t_1|^\alpha} & ... & e^{-|t_k|^\alpha} \\ e^{-|t_1|^\alpha} & e^{-|t_1+t_1|^\alpha} & ... & e^{-|t_1+t_k|^\alpha} \\ ... & ... & & ... \\ ... & ... & & ... \\ ... & ... & & ... \\ e^{-|t_k|^\alpha} & e^{-|t_k+t_1|^\alpha} & ... & e^{-|t_k+t_k|^\alpha} \end{pmatrix}, \quad \mathbf{b}(\alpha) = \begin{pmatrix} 0 \\ -|t_1|^\alpha \ln|t_1| e^{-|t_1|^\alpha} \\ -|t_2|^\alpha \ln|t_2| e^{-|t_2|^\alpha} \\ .. \\ .. \\ -|t_k|^\alpha \ln|t_k| e^{-|t_k|^\alpha} \end{pmatrix},$$

$\mathbf{a}(\alpha) = (\mathbf{A}(\alpha))^{-1} \cdot \mathbf{b}(\alpha)$, and

$$\overline{F(\mathbf{X})} = \{1, \frac{1}{n}\sum_{m=1}^{n} \exp(it_1 X_m), .., \frac{1}{n}\sum_{m=1}^{n} \exp(it_k X_m)\},$$

then

$$\sum_{j=1}^{n} J_k(X_j, \alpha) = \sum_{j=1}^{k} a_j \sum_{m=1}^{n} \exp(it_j X_m) = n \cdot \overline{F(\mathbf{X})} \cdot \mathbf{a}(\alpha). \tag{1.27}$$

Summary of the Methodology of Estimating α.

Note that $\overline{F(\mathbf{X})}$ can be calculated externally and used as an input into the methodology of projections. This means that the MLP procedure does not depend on the sample size. Let us denote with Ω the finite set of all considered values of α. Then we get a two-phase algorithm:

- 1) Calculation of the $\overline{F(\mathbf{X})}$,
- 2) Choosing such a value of α from Ω which minimises $|\overline{F(\mathbf{X})} \cdot \mathbf{a}(\alpha)|$.

The resulting estimator is

$$\alpha_{\text{MLP}} = \arg\min_{\alpha \in \Omega} \left|\overline{F(\mathbf{X})} \cdot \mathbf{a}(\alpha)\right|. \tag{1.28}$$

In Table 1.1, we compare the maximum likelihood projections method for $k = 15$ with the CFB estimates. The empirical mean and variance of the estimates of α are obtained from simulating 100 samples with stable distributions, and every sample contains 5000 elements. From those 100 estimates we calculate the mean and variance. In Table 1.1, $\bar{\alpha}$ and std(α) denote the mean and the standard deviation of the estimates respectively. In the case of sub-Gaussian distributions, we estimate the tail index from all of the marginals and average these values out to obtain the estimate of the tail index of the multivariate sub-Gaussian distribution [2].

1.9.3 The Estimation of the Dispersion Matrix

Without loss of generality, we can assume that matrix $\mathbf{Q}$ is of the form

$$\mathbf{Q} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

with unknown r because we can always get this form by scaling and translating of the initial random vector. Lemma 5.1 in [2] enables us to concentrate only on pairs of marginal distributions of the sub-Gaussian distribution to be able to estimate the whole matrix $\mathbf{Q}$. Let us denote the density function of the two-dimensional random vector by $\phi(x, y; r)$. Suppose that α is known because its estimate can be obtained from the univariate marginal distributions.
Let us note that if (X, Y) has a centred sub-Gaussian distribution with r, then by definition [2]

$$\mathbb{E}\left(i \cdot u_1(X+Y)\right) = \psi(u_1, u_1; r) = \exp\left\{-(2(1+r))^{\alpha/2} u_1^{\alpha}\right\}. \tag{1.29}$$

We will use the method of projections to obtain the MLP estimate of r as follows:

$$I(r) = \mathbb{E}\left[\left(\frac{\partial \phi(X, Y; r)}{\partial r}\right)^2\right], \tag{1.30}$$

$$J(x, y, r) = \frac{\partial \phi(x, y, r)}{\partial r}. \tag{1.31}$$

The ML estimate of r will be obtained by solving the equation

$$\hat{r}_{ML} = \left\{ r : \sum_{j=1}^{n} J(X_j, Y_j, r) = 0 \right\}, \tag{1.32}$$

where $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are the observations. A technique analogous to the one that we used for estimating function J for α can be used for estimating r [41]. In this case, we will obtain to another system of equations but the core of the approach is much the same. Let us project the random function $J(X, Y, r)$ onto the space $\{1, \mathrm{e}^{it_1(X+Y)}, ..., \mathrm{e}^{it_k(X+Y)}\}$, $k \in \mathbb{N}$. We assume that $t_0 = 0$. Let us define the scalar product in this space in a way similar to the case of estimating α:

$$\left\langle \mathrm{e}^{it_i(X+Y)}, \mathrm{e}^{it_j(X+Y)} \right\rangle = \mathbb{E}\left[\mathrm{e}^{it_i(X+Y)} \cdot \mathrm{e}^{it_j(X+Y)}\right] =$$
$$= \mathbb{E}\left[\mathrm{e}^{i(t_i+t_j)(X+Y)}\right] = \psi(t_i + t_j, t_i + t_j)$$

$$\psi(t_i + t_j, t_i + t_j; r) = \exp\left\{(2(1+r))^{\alpha/2} |t_i + t_j|^{\alpha}\right\}. \tag{1.33}$$

Let us approximate $J(X, Y, r)$ by

$$J_k(X, Y, r) = \sum_{j=0}^{k} a_j \cdot \mathrm{e}^{it_j(X+Y)}. \tag{1.34}$$

Because we project $J(X, Y, r)$ to the space $\{1, \mathrm{e}^{it_1(X+Y)}, ..., \mathrm{e}^{it_k(X+Y)}\}$, $k \in \mathbb{N}$, we have

$$\left(J(X, Y, r) - J_k(X, Y, r)\right) \perp \mathrm{e}^{it_j(X+Y)}, \quad j = 1, 2, ..., k. \tag{1.35}$$

After some algebra, we get from (1.35) the following system of equations:

$$\sum_{m=0}^{k} a_m \exp\left(-|t_j + t_m|^{\alpha}(2(1+r))^{\alpha/2}\right) = \tag{1.36}$$
$$-\exp\left\{-(2(1+r))^{\alpha/2} t_j^{\alpha}\right\} \alpha t_j^{\alpha}(2(1+r))^{\alpha/2-1},$$

where $j = 0, 1, ..., k$ and r is the unknown.

α	Type of estimator	$\bar{\alpha}$	$(\text{std}(\alpha))^2$	$\bar{\alpha} \pm 2\text{std}(\alpha)$
1.1	α_{MLP}	1.1052	0.017	[1.070,1.139]
	α_{CFB}	1.0997	0.020	[1.059,1.140]
1.2	α_{MLP}	1.2054	0.018	[1.167,1.243]
	α_{CFB}	1.2005	0.021	[1.059.1.140]
1.3	α_{MLP}	1.3000	0.018	[1.262,1.337]
	α_{CFB}	1.3000	0.022	[1.254,1.345]
1.4	α_{MLP}	1.4074	0.022	[1.363,1.451]
	α_{CFB}	1.3997	0.023	[1.350,1.440]
1.5	α_{MLP}	1.5010	0.021	[1.458,1.543]
	α_{CFB}	1.5002	0.024	[1.450,1.540]
1.7	α_{MLP}	1.6988	0.020	[1.638,1.759]
	α_{CFB}	1.7002	0.021	[1.637,1.762]
1.8	α_{MLP}	1.8004	0.018	[1.764,1.836]
	α_{CFB}	1.8006	0.020	[1.755,1.845]
1.9	α_{MLP}	1.8988	0.015	[1.868,1.929]
	α_{CFB}	1.9000	0.019	[1.860,1.931]

Table 1.1: Mean and Std of the MLP for $k = 15$ and the CFB Estimates.

1.9.4 The Estimation of the Tail Index of $GS_{\alpha}(1, 0, 0)$

Taking into account the shape of the characteristic function of from (1.13) and the logic of the derivations of the formula for stable distributions, we get analogous systems of equations for $GS_{\alpha}(1, 0, 0)$.

$$\sum_{v=0}^{k}\left(\frac{1}{1+|t_v + t_j|^{\alpha}} + \frac{1}{1+|t_v - t_j|^{\alpha}}\right) = -\frac{|t_j|^{\alpha}\ln|t_j|}{(1+|t_j|)^2}, \text{ where } j = 0, 1, 2, ..., k. \tag{1.37}$$

By denoting

$$\mathbf{A}(\alpha) = \left\{\frac{1}{1+|t_v + t_j|^{\alpha}} + \frac{1}{1+|t_v - t_j|^{\alpha}} \quad v, j \in \{0, 1, .., k\}\right\},$$

$$\mathbf{b}(\alpha) = \left\{-\frac{|t_j|^{\alpha}\ln|t_j|}{(1+|t_j|)^2} \quad j \in \{0, 1, .., k\}\right\},$$

and

$$\mathbf{a}(\alpha) = \mathbf{A}(\alpha)^{-1}\mathbf{b}(\alpha)$$

then

$$\alpha_{\text{MLP}} = \arg\min_{\alpha\in\Omega}\left|\overline{F(\mathbf{X})} \cdot \mathbf{a}(\alpha)\right|. \tag{1.38}$$

1.10 Parameter Estimation of $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$

Let the bivariate random vector (X, Y) have $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$. Then if $\alpha_1 = \alpha_2$ and $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$ is implementable, then we can use the method of projections for estimating r. Here we assume that the parameters $\alpha_1, \alpha_2, \beta_1$, and β_2 are estimated from univariate marginal distributions.

1.10.1 Special Case of $\mathcal{S}([\alpha, \beta_1], [\alpha, \beta_2], r)$

Let us assume that $(X, Y) \sim \mathcal{S}([\alpha, \beta_1], [\alpha, \beta_2], r)$. Then taking (1.19), we get

$$\begin{aligned}&\psi(t,t) = \exp\left(-|t|^\alpha \left((1+r)^\alpha K_1 - (1-r^\alpha)K_2\right)\right) = \\ &= \exp\left(-|t|^\alpha \omega(t, \alpha, \beta)\right)\end{aligned}$$

where

$$\begin{aligned}\omega(t, \alpha, \beta, r) = \quad & (1+r)^\alpha - (1-r^\alpha) + \\ & + i \cdot \operatorname{sign}(t) \tan(\pi\alpha/2) \left(\beta_2(1-r^\alpha) - \beta_1(1+r)^\alpha\right)\end{aligned}$$

Since in the method of projection, we consider only positive t_js, we get:

$$\begin{aligned}\omega(t, \alpha, \beta, r) = \omega(\alpha, \beta, r) = \quad & (1+r)^\alpha - (1-r^\alpha) + \\ & + i \tan(\pi\alpha/2) \left(\beta_2(1-r^\alpha) - \beta_1(1+r)^\alpha\right)\end{aligned}$$

So now we can define the scalar product:

$$\begin{aligned}&\left\langle \mathrm{e}^{it_i(X+Y)}, \mathrm{e}^{it_j(X+Y)} \right\rangle = \mathbb{E}\left[\mathrm{e}^{it_i(X+Y)} \cdot \mathrm{e}^{it_j(X+Y)}\right] = \\ &= \mathbb{E}\left[\mathrm{e}^{i(t_i+t_j)(X+Y)}\right] = \psi(t_i + t_j, t_i + t_j) = \\ &= \exp\left(-|t_i + t_j|^\alpha \omega(\alpha, \beta, r)\right).\end{aligned}$$

and

$$J_k(X, Y, r) = \sum_{j=0}^{k} a_j \cdot \mathrm{e}^{it_j(X+Y)}.$$

So the estimate of r will be obtained as

$$r = \operatorname{argmin}\left\{ r : \sum_{j=1}^{n} J_k(X_j, Y_j, r) = 0 \right\}$$

where $((X_1, Y_1), ..., (X_n, Y_n))$ are the observations.

1.10.2 General Case of $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$

When we have the data $X_1 = (X_{1,1}, X_{1,2}, .., X_{1,n})$ and $X_2 = (X_{2,1}, X_{2,2}, .., X_{2,n})$ and want to fit it by $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$, we will do it in the following steps.

1. The calculation of the parameters from $(\alpha, \sigma, \beta, \mu)$ for X_1 and X_2 and consider the linear transformations $(\hat{Y}_1, \hat{Y}_2)$ where $\hat{Y}_i = (X_i - \hat{\mu}_i)/\hat{\sigma}_i$, $i = 1, 2$, and $Y_i \sim S_{\alpha_i}(1, \beta_i, 0)$, $i = 1, 2$.

2. For the data $(\hat{Y}_1, \hat{Y}_2)$, we calculate the empirical distribution function $\psi(t_1, t_2)$.

3. We sample N-sampled realization from $\mathcal{S}(\alpha_1, \alpha_2, \beta_1, \beta_2, r)$ for $r = 0, 0.1, .., 1$ and for each sample, we calculate the empirical characteristic function $\psi(t_1, t_2; r)$.

4. The estimate of r is obtained

$$\hat{r} = \arg\min_r \int\int |\psi(t_1, t_2) - \psi(t_1, t_2; r)| dt_1 dt_2.$$

In Table 1.2, we demonstrate the estimation of r for $\mathcal{S}(1.5, 1.9, 0.2, 0.3, r_i)$ for $r_i = 0, 0.1, .., 1$.

1.11 The Fit of Financial Data

In this section, we conduct the fit of electricity and gas prices by means of the AR(1) and the Ornstein-Uhlenbeck process with different innovations. We compare stable assumption with the following alternatives:

1. The normal innovation

2. The normal inverse Gaussian (NIG) innovation

3. The hyperbolic innovation

4. The Student-t innovation

In the case of multivariate stable distributions, we test two models:

1. the sub-Gaussian,

2. the model $\mathcal{S}([\alpha_1, \beta_1], [\alpha_2, \beta_2], r)$.

r	$\bar{r}$	std(r)
0.0	0.01	0.080
0.1	0.18	0.120
0.2	0.25	0.150
0.3	0.32	0.175
0.4	0.41	0.180
0.5	0.49	0.207
0.6	0.57	0.204
0.7	0.69	0.210
0.8	0.79	0.180
0.9	0.86	0.075
1.0	0.99	0.004

Table 1.2: The estimation of r for $(Y_1, Y_2) \sim \mathcal{S}([1.5, 0.2], [1.9.0.3], r)$

We use daily data from European markets which are summarised in Figure 1.1 where the x-axis is the time from January 2014 to January 2016 . Joint models of power and gas prices can be used for the valuation of spark-spread options and gas-fired power plants. Ignoring possible dependencies between residuals, such as correlations, might lead to biased results.
We quantify the quality of the model by the following parameters:

1. The Kolmogorov Smirnov test applied on the residuals,
2. The range of the simulated model values compared to that of the real data,
3. The frequency of spikes the simulated model values compared to that of the real data,
4. The linear dependence of the simulated values compared to the correlation between electricity and gas prices.

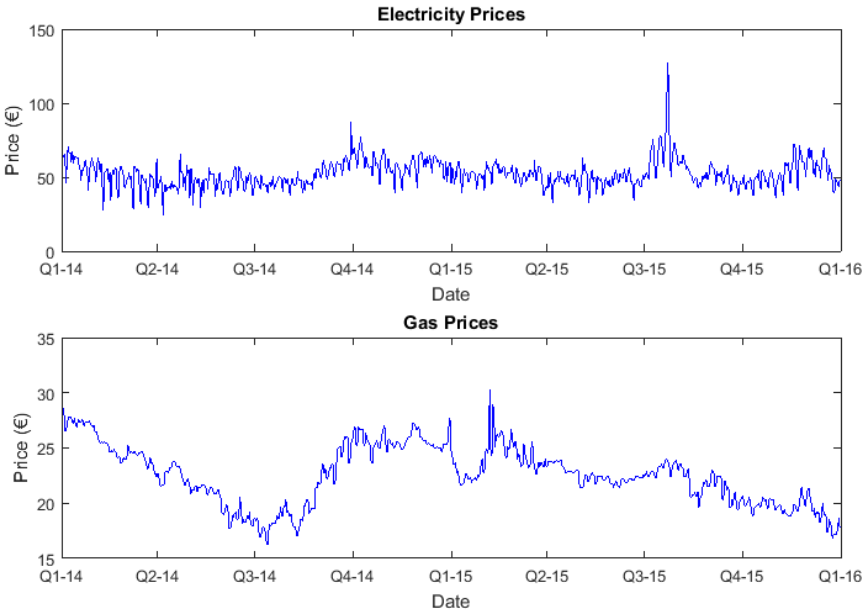


Figure 1.1: Electricity and gas prices

1.11.1 The Model of Prices

In this section we measure the fit of prices for univariate and bivariate price processes. We fit univariate data to AR(1) model with the different innovations listed in Tables 1.3, 1.4, and 1.5. For bivariate data we apply two models:

1. Two AR(1) models of gas and electricity prices, where residuals have either a bivariate normal, or a sub-Gaussian, or $\mathcal{S}$ distribution.

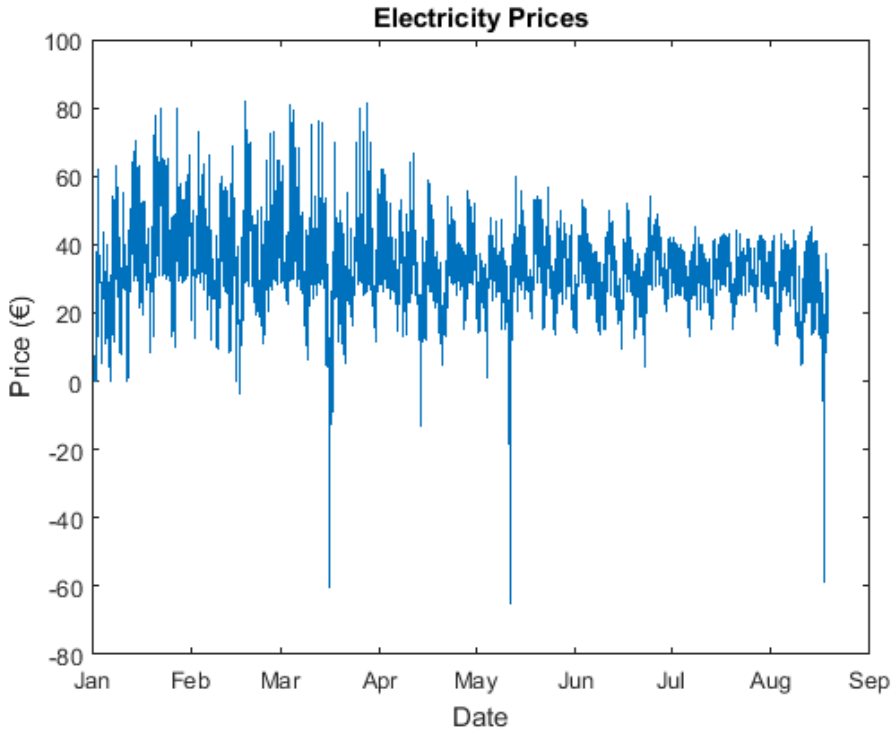


Figure 1.2: Hourly electricity prices and seasonality for 2015 January-August (5500 Hours)

2. Two OU models of gas and electricity prices, with the same assumptions on the residuals – with one exception, concerning the $\mathcal{S}$: the corresponding skewness parameters (β_{Gas} and $\beta_{\text{Electricity}}$) are zero because otherwise the stable process will not be Lévy.

1.11.2 Univariate Models

In this subsection, we consider only AR(1) models with different innovations. We model daily electricity and gas prices and additionally, we model hourly electricity prices with the same model and random factors. In Table 1.3, Table 1.4, and Table 1.5, we summarise the results of modelling prices. The null hypothesis means that the residuals have the predefined distribution. In the following list, we explain the columns of these tables:

1. **Acceptance** means whether the Kolmogorov-Smirnov tests accepted the null hypothesis (0 meaning that it is accepted) or rejected (1 meaning that it is rejected).

2. **p-value KS** means the value of the p-value generated by the Kolmogorov-Smirnov test.

3. **ks2stat** is the test statistic for the two-sample Kolmogorov-Smirnov test.

4. $\pm\mathbf{n}\boldsymbol{\sigma}$ states for the percentage of observations $\mathbf{X}$ which are outside the interval

$$[\bar{\mathbf{X}} - n \cdot \sigma, \quad \bar{\mathbf{X}} + n \cdot \sigma],$$

where $\sigma = \sqrt{\mathbf{Var}(\mathbf{X})}$ when the variance is finite, and σ is the empirical scale parameter in the case of non-Gaussian stable distributions. In Tables 1.3, 1.4, and 1.5 the symbol σ denotes the dispersion of the data.

For the simulated samples from the Gaussian, stable, and other distributions, $\pm n\sigma$ denotes the average percentage of observations outside the aforementioned interval. We note that for stable distributions, there are two parameters which cause dispersion: the scale parameter σ and the tail index α [89]. This means that in general we have to consider $\pm n(\alpha)\sigma$ where $n(\alpha)$ is the number which depends on α and the smaller the value of the tail index, the larger this value will be. In Figure 1.1, the value of the parameter α from data for gas is 1.923, and the value of α for electricity is 1.863. The value of α for data in Figure 1.2 is 1.8345. All of the estimates are obtained by using the MLP methodology. In all of the aforementioned cases the value of the tail index is close to 2, therefore we restrict our intervals to $\pm 2\sigma$, $\pm 3\sigma$, and $\pm 4\sigma$. Otherwise, we should have concerned ourselves with larger intervals.

The number of samples is 5000. The number of observations from which we calculated the statistics in Tables 1.3 and 1.4 is 720. As for Table 1.5, this number is 5500.

5. **Distance** means the standard deviation between two lines whose description is thus. In the upper panel of Figure 1.3, there are two types of lines: the black line is the season-free original time series of prices sorted in ascending order; the green lines are the simulated scenarios of the calibrated Ornstein-Uhlenbeck processes sorted in ascending order. The number of scenarios (or simulated samples) is 5000. In the lower panel of Figure 1.3, the red line is the average of the green line in the upper panel. The blue line in the lower panel is the same as the black line in the upper panel. So the **distance** in Tables 1.3, 1.4, and 1.5 states the empirical standard deviation between the blue and green lines in the lower panel of Figure 1.3.

6. **Real Data**. In Tables 1.3, 1.4, and 1.5, there is a special row at the bottom of each table which highlights corresponding characteristics for the real data. If some of the characteristics are not appropriate for the real data, we fill the corresponding cell with "–".

Innovation	Acceptance	p-value KS	ks2stat	$\pm 2\sigma$	$\pm 3\sigma$	$\pm 4\sigma$	Distance
Normal	1	0.1433	0.0380	0.0455	0.00269	0.00000	2.4985
Stable	0	0.8676	0.0222	0.0474	0.01343	0.00673	4.4987
Trunc Stable	0	0.8676	0.0222	0.0474	0.01343	0.00673	1.7842
Hyperbolic	0	0.9932	0.0158	0.0529	0.01041	0.00213	1.8545
NIG	0	0.9264	0.0202	0.0517	0.00921	0.00171	1.9845
Student-t	0	0.5155	0.0235	0.0510	0.01075	0.00283	1.8412
Real Data	-	-	-	0.0535	0.01373	0.00137	-

Table 1.3: AR(1) model of electricity prices for different innovations.

1.11.3 Main Results of the Univariate Models

1. **Daily electricity prices**
In terms of the Kolmogorov-Smirnov test, the winner is the hyperbolic mod-

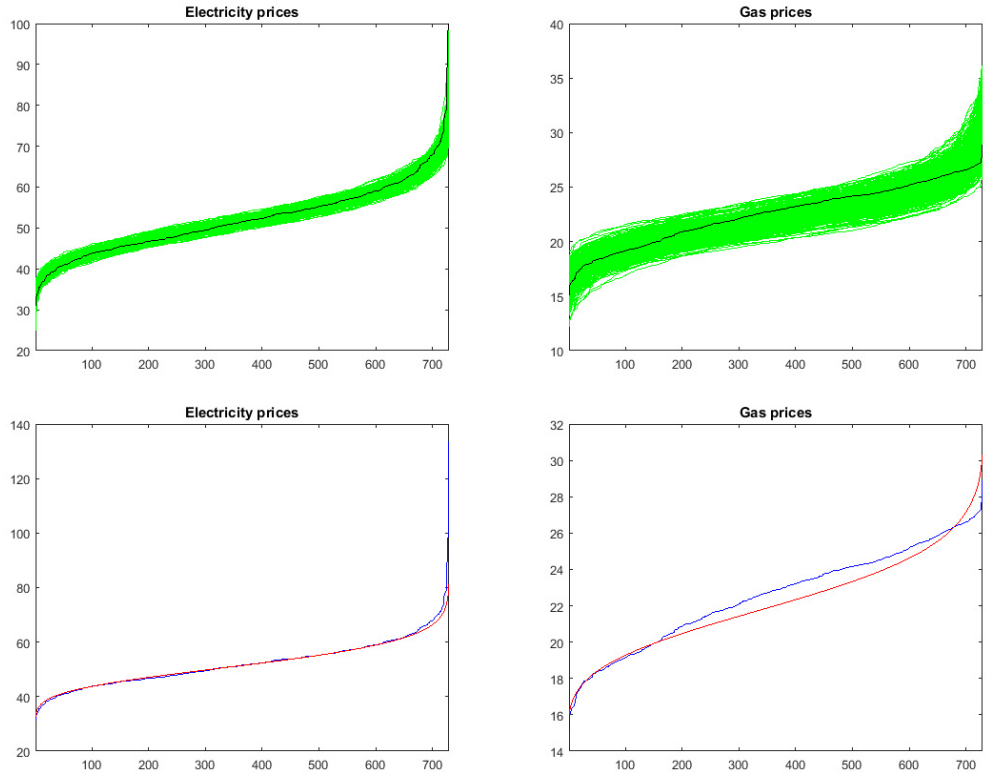


Figure 1.3: The range of the original and simulated prices

el. In terms of $\pm i\sigma$, it also performs better than its alternatives. However, in terms of the range, the stable truncated distribution performs better.

2. **Daily gas prices**
   In terms of the Kolmogorov-Smirnov test, the winner is the stable and truncated distribution; however, they do not perform as well as the alternatives in terms of $\pm i\sigma$. In terms of range, the winner is the truncated stable distribution. The Student-t model is also accepted by the Kolmogorov-Smirnov test and Student-t performs the best in terms of $\pm i\sigma$.

Innovation	Acceptance	p-value KS	ks2stat	$\pm 2\sigma$	$\pm 3\sigma$	$\pm 4\sigma$	Distance
Normal	1	0.0000	0.1380	0.0455	0.00268	0.00000	0.3421
Stable	0	0.6420	0.0275	0.0636	0.03751	0.02594	1.4125
Trunc Stable	0	0.6420	0.0275	0.0636	0.03751	0.02594	0.2415
Hyperbolic	1	0.0068	0.0062	0.0535	0.02213	0.01038	0.2954
NIG	1	0.0000	0.0932	0.0519	0.02077	0.00938	0.3125
Student-t	0	0.1716	0.0125	0.0520	0.02546	0.01516	0.2745
Real Data	-	-	-	0.0590	0.01375	0.00686	-

Table 1.4: AR(1) model of gas prices for different innovations.

Innovation	Acceptance	p-value KS	ks2stat	$\pm 2\sigma$	$\pm 3\sigma$	$\pm 4\sigma$	Distance
Normal	1	0.0000	0.1475	0.0454	0.00268	0.00006	3.1425
Stable	0	0.4119	0.0125	0.0318	0.01641	0.01043	7.2541
Trunc Stable	0	0.4119	0.0125	0.0318	0.01641	0.01043	2.1425
Hyperbolic	1	0.0010	0.0274	0.0332	0.01083	0.00401	3.7452
NIG	1	0.0000	0.1646	0.0294	0.01630	0.01020	4.2365
Student-t	0	0.1002	0.0232	0.0272	0.01066	0.00540	5.2458
Real Data	-	-	-	0.0278	0.01054	0.00527	-

Table 1.5: AR(1) model of hourly electricity prices for different innovations.

3. **Hourly electricity prices**
   In terms of the Kolmogorov-Smirnov test, the winners are the stable and truncated stable models. The truncated stable distribution performs the best in terms of the range; however the Student-t model performs the best in terms of $\pm i\sigma$.

1.12 Bivariate Models

In this section, we conduct the fit of the bivariate residuals of daily electricity and gas prices by the bivariate Ornstein-Uhlenbeck and AR(1) models. For the sake of brevity, we denote the Ornstein-Uhlenbeck process with OU. In Figure 1.4, there is the graph of the bivariate distribution function of the residuals of the electricity and gas. The axis with a higher range corresponds to the electricity prices.

1.12.1 Bivariate OU Models

For simulating a normal Ornstein-Uhlenbeck process

$$dY_t = a(\mu - Y_t)dt + \sigma d\varepsilon_t,$$

where ε_t is stable (or normal), we use Gillespie's (1996) formula [32]

$$Y_{t+\Delta t} = Y_t \exp(-a\Delta t) + \mu \cdot (1 - \exp(-a\Delta t)) + \sigma\sqrt{\frac{1 - \exp(-2a\Delta t)}{2a}}\varepsilon_t,$$
$$\varepsilon_t \sim N(0,1)$$

By using the expression for the scale index of the stable Ornstein-Uhlenbeck process [38], we can infer the analogous formula for exact simulation of this process:

$$Y_{t+\Delta t} = Y_t \exp(-a\Delta t) + \mu \cdot (1 - \exp(-a\Delta t)) + \sigma\sqrt[\alpha]{\frac{1 - \exp(-\alpha a\Delta t)}{\alpha a}}\varepsilon_t,$$
$$\varepsilon_t \sim S_\alpha(1,0,0).$$

We consider the following OU models:

1. **Gaussian OU model (Classical)**
   We model all of the univariate prices (electricity and gas) by means of the Ornstein-Uhlenbeck model

$$dX_t^E = \kappa_E\left(\gamma_E - X_t^E\right)dt + \sigma_E dW_t^E,$$
$$dX_t^G = \kappa_G\left(\gamma_G - X_t^G\right)dt + \sigma_G\rho dW_t^E + \sigma_G\sqrt{1-\rho^2}dW_t^G.$$

2. **Sub-Gaussian OU model**
   We apply the stable Ornstein-Uhlenbeck with two stable processes which

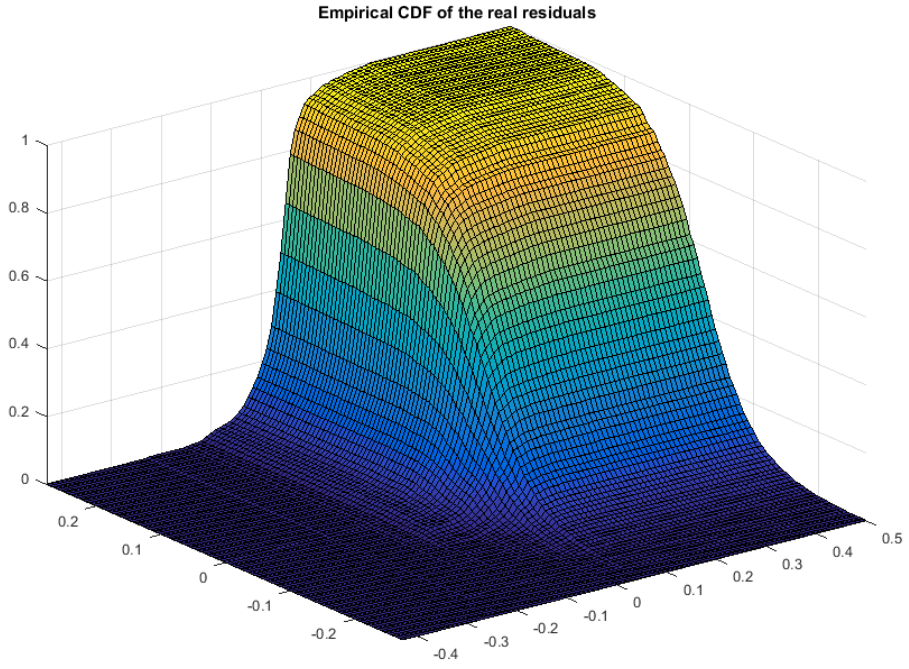


Figure 1.4: The empirical distribution function of the joint residuals of electricity and gas

we denote $\mathcal{W}_t^E$:

$$dX_t^E = \kappa_E \left(\gamma_E - X_t^E\right) dt + \sigma_E d\mathcal{W}_t^E,$$
$$dX_t^G = \kappa_G \left(\gamma_G - X_t^G\right) dt + \sigma_G \rho d\mathcal{W}_t^E + \sigma_G \sqrt{1-\rho^2} d\mathcal{W}_t^G,$$

where $\mathcal{W}_t^E = \sqrt{s_t^{(\alpha)}} W_t^E$ and $\mathcal{W}_t^G = \sqrt{s_t^{(\alpha)}} W_t^G$,
$s_t^{(\alpha)} \sim S_{\alpha/2}\left(\cos\left(\frac{\pi\alpha}{2}\right)^{2/\alpha}, 1, 0\right)$ and $W_t^E || W_t^G$.

3. $\mathcal{S}([\alpha_1, 0], [\alpha_2, 0], r)$ **OU model**
We apply the stable Ornstein-Uhlenbeck with two stable processes which

we denote $\mathcal{W}_t^E$:

$$dX_t^E = \kappa_E \left(\gamma_E - X_t^E\right) dt + \sigma_E d\mathcal{W}_t^E(\alpha_1),$$
$$dX_t^G = \kappa_G \left(\gamma_G - X_t^G\right) dt + \sigma_G \rho d\mathcal{W}_t^E(\alpha_2) + \sigma_G \sqrt[\alpha_2]{1-\rho^{\alpha_2}} d\mathcal{W}_t^G(\alpha_2),$$

where
$\mathcal{W}_1^E(\alpha_1) \sim \mathbf{stable}(\alpha_1, 0, V, W)$,
$\mathcal{W}_1^E(\alpha_2) \sim \mathbf{stable}(\alpha_2, 0, V, W)$,
$\mathcal{W}_1^G(\alpha_2) \sim \mathbf{stable}(\alpha_2, 0, V', W')$,
$\mathcal{W}_t^G(\alpha) =_d t^\alpha \mathcal{W}_1^G(\alpha), \quad \mathcal{W}_t^E(\alpha) =_d t^\alpha \mathcal{W}_1^E(\alpha)$
$V, V' \sim \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \quad W, W' \sim \exp(1)$,
$V, W, V',$ and W' are all independent.

Innovation	Acceptance	p-value KS	k2test	r
Normal	1	0.00	0.1993	0.5110
Sub-Gaussian	1	0.00	0.2241	0.4712
$\mathcal{S}$	0	0.18	0.0591	0.5014
Real data	-	-	-	0.4907

Table 1.6: Joint OU model of power and gas prices for different innovations.

Innovation	Acceptance	p-value KS	k2test	r
Normal	1	0.0	0.1993	0.5110
Sub-Gaussian	1	0.0	0.2241	0.4712
$\mathcal{S}$	0	0.2	0.0572	0.4967
Real data	-	-	-	0.4907

Table 1.7: Joint AR(1) model of power and gas prices for different innovations.

1.12.2 Bivariate AR(1) Models

We consider the following AR(1) models:

1. **Gaussian AR(1) model**
   We take the series of the residuals and calculate the mean and variance for both time series, and then calculate their correlation coefficient.

2. **Sub-Gaussian AR(1) model**
   We calculate the tail index for both of the residuals and take the average of the two αs. Then we calculate the scale and location parameters for two residuals assuming that their alpha equals to that average value. We use normalised residuals to estimate r by means of the method of projections. It can be easily shown that if (X_1, X_2) is a sub-Gaussian random vector with $r = 0$, then if $\rho \in [-1, 1]$, then $(X_1, \rho X_1 + \sqrt{1-\rho^2} X_2)$ has a sub-Gaussian distribution with $r = \rho$.

3. **$\mathcal{S}$ AR(1) model**
   From the univariate data we estimate all of the parameters of the stable

distribution and then normalise our residuals to estimate r. Here, we are not restricted by the assumption that the skewness indices are zero and that both residuals have the same tail index.

The results of modelling joint residuals of OU electricity and gas models are summarised in Table 1.6. The column r in Table 1.6 represents the average correlation between the simulated values of season-free prices.

1.12.3 Main Results Concerning the Bivariate Models

In terms of the range and the frequency of spikes, the results are identical to those from the univariate distributions, except for the sub-Gaussian distributions and $\mathcal{S}$ for the OU model. In terms of the bivariate Kolmogorov-Smirnov test, only the hypothesis that the residuals have the $\mathcal{S}$ distribution is not rejected. In terms of the correlation coefficient, the $\mathcal{S}$ distribution also performs the best: i.e the difference between the average correlation of simulated 1000 samples (with the number of elements the same as in the original data series) and the original season-free data is the smallest for this model.

Conclusion to Chapter 1

In this chapter, we introduced an approach for estimating the parameters of stable and sub-Gaussian distributions by means of the method of projections. We presented a sub-family of generalised stable distributions where the tail index may differ across the marginals, and introduced its main properties and methods for the estimation of its parameters. Finally we presented how different models fit electricity and gas price data. The best univariate models for fitting our data turned out to be truncated stable and Student-t distributions. In the case of daily electricity prices, the hyperbolic model performs the best in terms of KS statistics. However, for larger data samples, only stable truncated and Student-t distributions seem to be an appropriate choice for fitting the data. For bivariate data, the $\mathcal{S}$ model turns out to be the best choice in terms of correlation and KS statistics. We will use the results regarding the distribution $\mathcal{S}$ in Chapter 4 for the joint modelling the forward and spot prices. These joint models will be used for the valuation of gas storage units.

2. Stochastic Dominance

In practice, we often face decision-making problems which yield random payoffs. In order to make informed, preferable choices, we need to be able to compare random variables. If the utility function of the decision maker is known, then the random variable which maximises the expected utility is preferred. However, it is often difficult to find the agent's utility function. Stochastic dominance enables us to compare random variables and the underlying theory is linked with the theory of utility functions. This notion applies to pairs of random variables. The main goal of this chapter is to explore stochastic dominance in the realm of stable distributions.

The Chapter's Structure

In this chapter we first give the definition of first- and second-order stochastic dominance, and introduce the notation. Then, in Section 2.2, we introduce *Symmetric Pareto Distributions* which belong to the domain of attraction of stable distributions and use them for the proofs of the theorems, regarding the parametric regions where the stochastic dominance for stable distributions holds. In Section 2.3, we handle the parametric regions for stable distributions where stochastic dominance holds. In Section 2.4, we deal with stochastic dominance for portfolios of stable distributions. In Section 2.5, we briefly discuss the Wasserstein metric and then, we move to the conclusion for this chapter.

2.1 The Definition of Stochastic Dominance

Definition 12 (First-order stochastic dominance [10, 11])**.** *We say that the random variable X first-order stochastically dominates (FSD) the random variable Y, written $X \geq_{(1)} Y$, if*

$$F_X(z) \leq F_Y(z) \text{ for all } z, \tag{2.1}$$

where F_X and F_Y are distribution functions of X and Y, respectively.

To define higher-order stochastic dominance, we first introduce the following notation.

Notation

If a random variable X is specified by a parameter θ, which might be a vector, then we denote it by $X^{(\theta)}$. We denote its characteristics as follows:

The characteristic function:	$\psi_\theta(x)$,
The density function:	$\phi_\theta(x)$,
The distribution function:	$F_\theta(x) = F_\theta^{(1)}(x)$ and $\bar{F}_\theta^{(1)}(x) = 1 - F_\theta^{(1)}(x)$,
The m-order distribution function:	$F_\theta^{(m)}(x) = \int_{-\infty}^{x} F_\theta^{(m-1)}(y)dy$,

provided that the integral exists.

When we specify the random variable X instead of the parameter θ, we replace the subscript θ with X, i.e. $F_X^{(m)}(x) = \int_{-\infty}^{x} F_X^{(m-1)}(y)dy$, etc.

Definition 13. *We say that the random variable X second-order stochastically dominates (SSD) the random variable Y, written $X \geq_{(2)} Y$, if*

$$\int_{-\infty}^{x} F_X(z)dz \leq \int_{-\infty}^{x} F_Y(z)dz \text{ for all } x, \tag{2.2}$$

By first-order stochastic dominance it is meant that the variable is 'stochastically larger' than the variable which it first-order dominates. By second-order stochastic dominance it is meant that the variable is 'less volatile' than the variable which it second-order dominates [104].

Remark In some literature this relation (stochastic dominance inequalities) is called the Slepian inequality [90].

If there is no stochastic dominance of the first or second order between the random variables X and Y, we denote this $X \neq_{(1)} Y$ or $X \neq_{(2)} Y$, respectively.

Definition 14. *We say that for a parametric family $X^{(\theta)}$, $\theta \in \Theta$ stochastic dominance of m-order exists ($m \in \mathbb{N}$) if the m-order distribution function $F_\theta^{(m)}(\cdot)$ exists for all $\theta \in \Theta$.*

Lemma 1. *First-order stochastic dominance entails the second-order stochastic dominance, if the second-order stochastic dominance exists. The converse is not true.*

Proof. See [104]. □

Theorem 12 (FSD-Theorem [104])**.** *X is unanimously preferred to Y by all agents with a monotone increasing utility function if and only if $X \geq_{(1)} Y$.*

Proof. See [104]. □

Theorem 13 (SSD-Theorem [104])**.** *X is unanimously preferred to Y by all agents with a monotone increasing and concave utility function if and only if $X \geq_{(2)} Y$.*

Proof. See [104]. □

Theorem 14 (Ruszczynski and Dencheva)**.** *If X is a random variable and the X's $F_X^{(2)}(x)$ exists, then*

$$F_X^{(2)}(x) = \mathbb{E}(x - X)^+ \quad \forall x \in \mathbb{R},$$

where $z^+ = \max(0, z)$.

Proof. See [25]. □

Crossings

Theorem 15 (Single crossing and SSD)[104]**.** *Suppose that X and Y are non-negative random variables whose supports are compact and the same and whose distribution functions have a single crossing, with $\bar{F}_X(x) \geq \bar{F}_Y(x)$ for low and $\bar{F}_X(x) \leq \bar{F}_Y(x)$ for high x. Then*

$$X \geq_{(2)} Y \iff \mathbb{E}[X] \geq \mathbb{E}[Y].$$

Proof. See [104]. □

Theorem 16 (Crossing and SSD [68])**.** *Assume that two CDFs F_X and F_Y have a finite number $k(k \geq 1)$ of crossing points and let $t_{01}, t_{02}, ..., t_{0k}$ be their corresponding values of t where clearly $t_{01} > \min\{\min\{t : F_X(t) > 0\}, \min\{t : F_Y(t) > 0\}\}$. Then*

$$X \geq_{(2)} Y$$

if and only if

1. $\int_{-\infty}^{t_{0i}} [F_Y(z) - F_X(z)]dz \geq 0$ *for all* $i = 1, 2, ..., k$,
2. $\int_{-\infty}^{\infty} [F_Y(z) - F_X(z)]dz \geq 0$,
3. $\int_{-\infty}^{\infty} [F_Y(z) - F_X(z)]dz > 0$ *or for at least one* $t_{0h} \in \{t_{01}, t_{02}, ..., t_{0k}\}$ $\int_{-\infty}^{t_{0i}} [F_Y(z) - F_X(z)]dz > 0$

Proof. See [68]. □

2.2 Auxiliary Symmetric Pareto Distributions

Let us explore the first and second-order stochastic dominance of Pareto distributions.

Theorem 17. *For Pareto distributions $X^{\{\alpha\}}$ with the distribution function*

$$F_\alpha(x) = 1 - \frac{1}{x^\alpha}, \quad \alpha > 1, x \geq 1$$

holds

$$\text{if } \alpha_1 > \alpha_2 \text{ then } X^{\{\alpha_1\}} \leq_{(k)} X^{\{\alpha_2\}} \tag{2.3}$$

where $k = 1, 2, .., \lfloor \min(\alpha_1, \alpha_2) \rfloor$.

Proof. For all $x \geq 1$, we have

$$F_{\alpha_1}(x) - F_{\alpha_2}(x) = \left(1 - \frac{1}{x^{\alpha_1}}\right) - \left(1 - \frac{1}{x^{\alpha_2}}\right) = \frac{1}{x^{\alpha_2}} - \frac{1}{x^{\alpha_1}} \leq 0,$$

which by the continuity of $F_{\alpha_1}(\cdot)$ entails stochastic dominance of the first order hence the all of the orders that exist. □

Definition 15 (Symmetric Pareto)**.** *Let us define a symmetric Pareto distributions characterised by the parameter $\alpha > 0$ with the following characteristics*

1. *The density function:*

$$\phi_\alpha(x) = \frac{\alpha}{2}\frac{1}{(1+|x|)^{\alpha+1}}$$

2. *The distribution function:*

$$F_\alpha(x) = \begin{cases} 1 - \frac{1}{2(1+x)^\alpha} & , x \geq 0 \\ \\ \frac{1}{2(1-x)^\alpha} & , x < 0. \end{cases}$$

We denote a random variable with such a distribution or density function by $X^{(\alpha)}$.

In this distribution α is the parameter which has to be positive.

Theorem 18. *If $\alpha_1 \neq \alpha_2$ then $X^{(\alpha_1)} \neq_{(1)} X^{(\alpha_2)}$.*

Proof. Let us consider any $x > 0$ and let us assume that $\alpha_1 < \alpha_2$. Then

$$F_{\alpha_1}(x) - F_{\alpha_2}(x) = \frac{1}{2}\left(\frac{1}{(1+x)^{\alpha_2}} - \frac{1}{(1+x)^{\alpha_1}}\right) < 0,$$
$$F_{\alpha_1}(-x) - F_{\alpha_2}(-x) = \frac{1}{2}\left(\frac{1}{(1-x)^{\alpha_1}} - \frac{1}{(1-x)^{\alpha_2}}\right) > 0,$$

which by continuity implies that $X^{(\alpha_1)}$ cannot dominate $X^{(\alpha_2)}$ and vice versa. □

Theorem 19. *$F_\alpha^{(2)}(\cdot)$ of $X^{(\alpha)}$ is as follows:*

$$F_\alpha^{(2)}(x) = x^+ + \frac{1}{2}\frac{1}{\alpha-1}\frac{1}{(|x|+1)^{\alpha-1}} \tag{2.4}$$

Proof. For $x \leq 0$, we have:

$$\begin{aligned} F_\alpha^{(2)}(x) \quad &= \int_{-\infty}^{x} F_\alpha(z)dz = \int_{-\infty}^{x} \frac{1}{2(1+|z|)^\alpha}dz = \\ &= \int_{-x}^{\infty} \frac{1}{2(1+|z|)^\alpha}dz = \int_{-x}^{\infty} \frac{1}{2(1+z)^\alpha}dz = \\ &= \frac{1}{2}\frac{(z+1)^{1-\alpha}}{1-\alpha}\Bigg|_{-x}^{\infty} = \frac{1}{2}\frac{1}{1-\alpha}\left(0-(1-x)^{1-\alpha}\right) = \\ &\frac{1}{2}\frac{1}{\alpha-1}\frac{1}{(1-x)^{\alpha-1}} \end{aligned} \tag{2.5}$$

For $x > 0$, we have:

$$\begin{aligned} F_\alpha^{(2)}(x) &= \int_{-\infty}^{x} F_\alpha(z)dz = \int_{-\infty}^{0} F_\alpha(z)dz + \int_0^x F_\alpha(z)dz = \\ &= \frac{1}{2}\frac{1}{\alpha-1} + \int_0^x \left(1 - \frac{1}{2(1+z)^\alpha}\right) dz = \\ &= \frac{1}{2}\frac{1}{\alpha-1} + x - \frac{1}{2}\int_0^x \frac{1}{(1+z)^\alpha} dz = \\ &= \frac{1}{2}\frac{1}{\alpha-1} + x - \frac{1}{2}\left(\frac{(z+1)^{1-\alpha}}{1-\alpha}\bigg|_0^x\right) = \\ &= \frac{1}{2}\frac{1}{\alpha-1} + x + \frac{1}{2}\frac{(z+1)^{1-\alpha}}{\alpha-1}\bigg|_0^x = \\ &= \frac{1}{2}\frac{1}{\alpha-1} + x + \frac{1}{2}\frac{(x+1)^{1-\alpha}}{\alpha-1} - \frac{1}{2}\frac{1}{\alpha-1} = \\ &= x + \frac{1}{2}\frac{1}{\alpha-1}\frac{1}{(x+1)^{\alpha-1}}. \end{aligned} \tag{2.6}$$

By taking (2.5) and (2.6) into account, we get for all $x \in \mathbb{R}$:

$$F_\alpha^{(2)}(x) = x^+ + \frac{1}{2}\frac{1}{\alpha-1}\frac{1}{(|x|+1)^{\alpha-1}}$$

□

Theorem 20 (Second-Order Dominance)**.** *If $\alpha_1, \alpha_2 > 1$ and $\alpha_1 < \alpha_2$ then*

$$X^{(\alpha_1)} \leq_{(2)} X^{(\alpha_2)}.$$

Proof.

$$\begin{aligned} F_{\alpha_1}^{(2)}(x) - F_{\alpha_2}^{(2)}(x) &= \left(x^+ + \frac{1}{2}\frac{1}{\alpha_1-1}\frac{1}{(|x|+1)^{\alpha_1-1}}\right) - \left(x^+ + \frac{1}{2}\frac{1}{\alpha_2-1}\frac{1}{(|x|+1)^{\alpha_2-1}}\right) \\ &= \frac{1}{2}\left(\frac{1}{\alpha_1-1}\frac{1}{(|x|+1)^{\alpha_1-1}} - \frac{1}{\alpha_2-1}\frac{1}{(|x|+1)^{\alpha_2-1}}\right) = \\ &= \frac{1}{2}\frac{1}{\alpha_1-1}\frac{1}{(|x|+1)^{\alpha_2-1}}\left((1+|x|)^{\alpha_2-\alpha_1} - \frac{\alpha_1-1}{\alpha_2-1}\right) \geq 0 \end{aligned}$$

because if $\alpha_1 < \alpha_2$, then $(\alpha_1-1)/(\alpha_2-1) < 1$ and $(1+|x|)^{\alpha_2-\alpha_1} > 1$. Hence it is clear that the inequality holds and equality takes place only if $\alpha_1 = \alpha_2$.

□

Remark: Note that for classic Pareto distributions the random variables with higher α are dominated by those with lower tail indices (provided that the corresponding characteristics exist) while in the case of symmetric Pareto distributions this is on the contrary.

Scale Parameter

Let us consider $X^{(\alpha,\sigma)} = \sigma X^{(\alpha)}$, where $\sigma > 0$.

$$\mathbb{P}(\sigma X^{(\alpha)} \leq x) = \mathbb{P}(X^{(\alpha)} \leq x/\sigma) = F_\alpha(x/\sigma).$$

After some algebra, we get

1. The density function of $X^{(\alpha,\sigma)}$ is

$$\phi_{(\alpha,\sigma)}(x) = \frac{1}{2\sigma} \frac{\alpha}{\left(1 + \frac{|x|}{\sigma}\right)^{\alpha+1}} \tag{2.7}$$

   Notation
   We denote a parametric family of distributions with the density function given in (2.7), as follows: $SP_\alpha(\sigma, 0, 0)$. "SP" states for "symmetric Pareto".

2. The second-order distribution function of $X^{(\alpha,\sigma)}$ ($\sim SP_\alpha(\sigma, 0, 0)$) is

$$F^{(2)}_{(\alpha,\sigma)}(x) = x^+ + \sigma \frac{1}{2} \frac{1}{\alpha - 1} \frac{1}{(1 + |x|/\sigma)^{\alpha-1}}$$

Lemma 2. *Let us assume that $0 < \sigma_1 < \sigma_2$ and $\alpha > 1$. Then*

$$X^{(\alpha,\sigma_1)} \geq_{(2)} X^{(\alpha,\sigma_2)}.$$

Proof.

$$F^{(2)}_{(\alpha,\sigma_1)}(x) - F^{(2)}_{(\alpha,\sigma_2)}(x) = \frac{1}{2(\alpha-1)} \left(\sigma_1 \frac{1}{(1 + |x|/\sigma_1)^{\alpha-1}} - \sigma_2 \frac{1}{(1 + |x|/\sigma_2)^{\alpha-1}} \right) =$$
$$= \frac{1}{2(\alpha-1)} \sigma_1 \frac{1}{(1 + |x|/\sigma_2)^{\alpha-1}} \left(\left(\frac{1 + |x|/\sigma_2}{1 + |x|/\sigma_1} \right)^{\alpha-1} - \frac{\sigma_2}{\sigma_1} \right) \leq 0$$

□

Stochastic Dominance in the Tail and Scale Parameters

Let us take a look, how the second-order distribution functions change when we decrease either α or σ.

$$F^{(2)}_{(\alpha_1,\sigma_1)}(x) - F^{(2)}_{(\alpha_2,\sigma_2)}(x) =$$
$$= \frac{1}{2} \left(\frac{1}{(\alpha_1 - 1)} \sigma_1 \frac{1}{(1 + |x|/\sigma_1)^{\alpha_1 - 1}} - \frac{1}{(\alpha_2 - 1)} \sigma_2 \frac{1}{(1 + |x|/\sigma_2)^{\alpha_2 - 1}} \right) =$$
$$= \frac{\sigma_1}{\alpha_1 - 1} \frac{1}{(1 + |x|/\sigma_2)^{\alpha_2 - 1}} \left(\frac{(1 + |x|/\sigma_2)^{\alpha_2 - 1}}{(1 + |x|/\sigma_1)^{\alpha_1 - 1}} - \frac{\sigma_2}{\sigma_1} \frac{\alpha_1 - 1}{\alpha_2 - 1} \right) \leq 0$$

Hence, it follows that if $\alpha_1 \geq \alpha_2$ and $\sigma_1 \leq \sigma_2$ and one of these two inequalities is strict, then $X^{(\alpha_1,\sigma_1)} \geq_{(2)} X^{(\alpha_1,\sigma_2)}$.

Skewness and Location Parameter

Let us define a random variable $X^{(\alpha,\beta)}$ by the following density function:

$$\phi_{(\alpha,\beta)}(x) = (1 + \text{sign}(x) \cdot \beta)\frac{\alpha}{2}\frac{1}{(1+|x|)^{\alpha+1}},$$

where $\beta \in [-1, 1]$. Let us calculate its distribution function:

1. For $x \leq 0$, we get:

$$F_{(\alpha,\beta)}(x) = \int_{-\infty}^{x}(1-\beta)\frac{\alpha}{2}\frac{1}{(1-z)^{\alpha+1}}dz = (1-\beta)\frac{\alpha}{2}\int_{-x}^{\infty}\frac{1}{(1+z)^{\alpha+1}}dz =$$
$$= (1-\beta)\frac{\alpha}{2}\frac{(1+z)^{-\alpha}}{-\alpha}\Bigg|_{-x}^{\infty} = (1-\beta)\frac{\alpha}{2}\frac{(1-x)^{-\alpha}}{\alpha} = \frac{1-\beta}{2(1-x)^{\alpha}}.$$

2. For $x > 0$, we get:

$$F_{(\alpha,\beta)}(x) = \int_{-\infty}^{0}\phi_{(\alpha,\beta)}(z)dz + \int_{0}^{x}f_{(\alpha,\beta)}(z)dz =$$
$$= \frac{1-\beta}{2} + \int_{0}^{x}(1+\beta)\frac{\alpha}{2}\frac{1}{(1+z)^{\alpha+1}}dz =$$
$$= \frac{1-\beta}{2} + (1+\beta)\frac{\alpha}{2}\frac{(1+z)^{-\alpha}}{-\alpha}\Bigg|_{0}^{x} =$$
$$= \frac{1-\beta}{2} + (1+\beta)\frac{1-(1+x)^{-\alpha}}{2} =$$
$$= 1 - \frac{1+\beta}{2(1+x)^{\alpha}}.$$

Hence we get:

$$F_{(\alpha,\beta)}(x) = \begin{cases} 1 - \frac{1+\beta}{2(1+x)^{\alpha}} & , x \geq 0 \\ \\ \frac{1-\beta}{2(1-x)^{\alpha}} & , x < 0. \end{cases} \tag{2.8}$$

Lemma 3. *If $\beta_1, \beta_2 \in [-1, 1]$ and $\beta_1 > \beta_2$, then for all $\alpha > 0$*

$$X^{(\alpha,\beta_1)} \geq_{(1)} X^{(\alpha,\beta_2)}.$$

Proof. We consider two cases for x:

1. Let us assume that $x > 0$. Then

$$F_{(\alpha,\beta_1)}(x) - F_{(\alpha,\beta_2)}(x) = \frac{1}{2}\frac{1+\beta_2}{(1+x)^{\alpha}} - \frac{1}{2}\frac{1+\beta_1}{(1+x)^{\alpha}} = \frac{1}{2}\frac{\beta_2-\beta_1}{(1+x)^{\alpha}} \leq 0.$$

2. Let us assume that $x \leq 0$. Then

$$F_{(\alpha,\beta_1)}(x) - F_{(\alpha,\beta_2)}(x) = \frac{1}{2}\frac{1-\beta_1}{(1-x)^{\alpha}} - \frac{1}{2}\frac{1-\beta_2}{(1-x)^{\alpha}} = \frac{1}{2}\frac{\beta_2-\beta_1}{(1-x)^{\alpha}} \leq 0.$$

□

The Expected Value of the Skewed $X^{(\alpha,\beta)}$

For $\alpha > 1$ the expected value of $X^{(\alpha,\beta)}$ is

$$\mathbb{E}X = \int_{-\infty}^{\infty} \frac{\alpha}{2}(1 + \text{sign}(x)\beta)\frac{x}{(1+|x|)^{\alpha+1}}dx = \frac{\beta}{\alpha - 1}$$

otherwise, it is infinite.

Hence, if $\alpha > 1$, then $Y^{(\alpha,\beta)} = X^{(\alpha,\beta)} - \beta/(\alpha - 1)$ has zero expected value.

Theorem 21 (The properties of $Y^{(\alpha,\beta)}$)**.** *The random variable $Y^{(\alpha,\beta)}$ with $\alpha > 1$ and $\beta \in [-1, 1]$ has*

1. *The distribution function*

$$F_{Y^{(\alpha,\beta)}}(x) = F_{(\alpha,\beta)}(x + \beta/(\alpha - 1))$$

2. *The second-order distribution function*

$$F^{(2)}_{Y^{(\alpha,\beta)}}(x) = \begin{cases} x + \frac{1+\beta}{2(\alpha-1)(1+|x+\beta/(\alpha-1)|)^{\alpha-1}} & , x \geq -\beta/(\alpha-1) \\ \\ \frac{1-\beta}{2(\alpha-1)(1+|x+\beta/(\alpha-1)|)^{\alpha-1}} & , x < -\beta/(\alpha-1). \end{cases}$$

Proof. This follows from the same approach that we used for the proof of Theorem 18. □

Let us note that for $\alpha > 1$

$$\lim_{x\to-\infty} \frac{F^{(2)}_{Y^{(\alpha,\beta_1)}}(x)}{F^{(2)}_{Y^{(\alpha,\beta_2)}}(x)} = \frac{1-\beta_1}{1-\beta_2}, \qquad \lim_{x\to\infty} \frac{F^{(2)}_{Y^{(\alpha,\beta_1)}}(x) - x}{F^{(2)}_{Y^{(\alpha,\beta_2)}}(x) - x} = \frac{1+\beta_1}{1+\beta_2} \tag{2.9}$$

which means that if $\beta_1 \neq \beta_2$, then $Y^{(\alpha,\beta_1)} \neq_{(2)} Y^{(\alpha,\beta_2)}$.

2.2.1 Simulation of the Skewed $X^{(\alpha,\beta)}$

The inverse function of $F_{\alpha,\beta}(\cdot)$ is

$$F^{-1}_{\alpha,\beta}(y) = \begin{cases} \left(\frac{1+\beta}{2(1-y)^{\alpha}}\right)^{1/\alpha} - 1 & , y \in [(1-\beta)/2, 1) \\ \\ 1 - \left(\frac{1-\beta}{2y^{\alpha}}\right)^{1/\alpha} & , y \in (0, (1-\beta)/2]. \end{cases}$$

Hence if we generate a sample $\mathbf{U} = \{U_1, U_2, ..., U_n\}$ where $n \in \mathbb{N}$, then

$$F^{-1}_{\alpha,\beta}(\mathbf{U}) = \{F^{-1}_{\alpha,\beta}(U_1), F^{-1}_{\alpha,\beta}(U_2), ..., F^{-1}_{\alpha,\beta}(U_n)\}$$

has the same distribution as $X^{(\alpha,\beta)}$ and

$$F^{-1}_{Y^{(\alpha,\beta)}}(\mathbf{U}) = \{F^{-1}_{\alpha,\beta}(U_1), F^{-1}_{\alpha,\beta}(U_2), ..., F^{-1}_{\alpha,\beta}(U_n)\} - \beta/(\alpha - 1)$$

has the same distribution as $Y^{(\alpha,\beta)}$.

2.2.2 Properties of $X^{(\alpha,\beta)}$

Lemma 4. *The random variable $X^{(\alpha,\beta)}$ has the following properties:*

1. *If $\beta = 1$, then $X^{(\alpha,1)}$ has a 'classical' Paretian distribution.*
2. *If $\beta = -1$, then $-X^{(\alpha,-1)}$ has a 'classical' Paretian distribution.*
3. *The density function of $X^{(\alpha,\beta)}$ (denoted by $\phi_{(\alpha,\beta)}(x)$) is not continuous at zero if $\beta \neq 0$.*
4. *If β is outside the interval $[-1,1]$, then neither $F_{(\alpha,\beta)}$ nor $\phi_{(\alpha,\beta)}$ can be a density or a distribution function, respectively.*

Proof. Assertions 1 and 2 follow immediately from the shape of the distribution function.
By calculating the upper and lower limits and zero, we get:

$$\lim_{x\to 0^+} \phi_{\alpha,\beta}(x) = \frac{\alpha}{2}(1+\beta),$$
$$\lim_{x\to 0^-} \phi_{\alpha,\beta}(x) = \frac{\alpha}{2}(1-\beta).$$

which equal only if $\beta = 0$.
If $\beta > 1$, then both $\phi_{\alpha,\beta}$ and $F_{\alpha,\beta}$ will be negative on $\mathbb{R}^-$. If $\beta < -1$, then both $\phi_{\alpha,\beta}$ and $F_{\alpha,\beta}$ will be negative on $\mathbb{R}^+$. □

Location Parameter

Let us note that for $\alpha > 1$,

$$\mathbb{E}X^{(\alpha)} = \mathbb{E}X^{(\alpha,0)} = 0, \tag{2.10}$$
$$\mathbb{E}X^{(\alpha,\beta)} = \frac{\beta}{\alpha-1}. \tag{2.11}$$

So by defining

$$XS^{(\alpha,1,\beta,0)} = X^{(\alpha,\beta)} - \frac{\beta}{\alpha-1} = Y^{(\alpha,\beta)}$$

we get a centred random variable with zero mean and the distribution function

$$F_{(\alpha,\beta,0)}(x) =$$
$$= F_{(\alpha,\beta)}\left(x + \frac{\beta}{(\alpha-1)}\right) = \begin{cases} 1 - \frac{1+\beta}{2\left(1+x+\frac{\beta}{(\alpha-1)}\right)^\alpha} & , x \geq -\frac{\beta}{(\alpha-1)} \\ \\ \frac{1-\beta}{2\left(1-x-\frac{\beta}{(\alpha-1)}\right)^\alpha} & , x < -\frac{\beta}{(\alpha-1)}. \end{cases} \tag{2.12}$$

So the random variable defined by

$$XS^{(\alpha,1,\beta,\mu)} = X^{(\alpha,\beta)} - \frac{\beta}{\alpha-1} + \mu = Y^{(\alpha,\beta)} + \mu$$

has mean μ and finally for $\sigma > 0$ we get the random variable

$$XS^{(\alpha,\sigma,\beta,\mu)} = \sigma\left(X^{(\alpha,\beta)} - \frac{\beta}{\alpha - 1}\right) + \mu = \sigma Y^{(\alpha,\beta)} + \mu$$

which like stable random variables is characterised by four parameters:

1. Tail index α,
2. Scale index σ,
3. Skewness parameter β,
4. Location parameter μ.

Notation

We denote the distribution of $XS^{(\alpha,\sigma,\beta,\mu)}$ by $SP_\alpha(\sigma, \beta, \mu)$.
When we assume that only one parameter is about to change, while the rest does not vary, we denote with X superscripted with this varying parameter.

Theorem 22 (The relationship between $S_\alpha(\sigma, \beta, \mu)$ and $SP_\alpha(\sigma, \beta, \mu)$)**.** *The distribution*

$$SP_\alpha\left(\sigma \sqrt[\alpha]{\frac{\sin(\pi\alpha/2)}{\pi\alpha/2}\Gamma(\alpha + 1)}, \beta, \mu\right)$$

belongs to the domain of attraction of $S_\alpha(\sigma, \beta, \mu)$.

Proof. This immediately follows from the generalised central limit theorem and (1.7). □

In Figure 2.1, we visualise two distribution functions, one of which has a stable distribution $S_\alpha(1, 0, 0)$ and the other has a distribution $SP_\alpha\left(\sqrt[\alpha]{\frac{\sin(\pi\alpha/2)}{\pi\alpha/2}\Gamma(\alpha + 1)}, 0, 0\right)$ where in both cases $\alpha = 0.691$.

2.3 Stochastic Dominance for Stable Distributions

Let $S_\alpha(\sigma, \beta, \mu)$ denote a stable distribution with the tail index equal to $\alpha \in (0, 2]$, the scale parameter $\sigma > 0$, the skewness parameter $\beta \in [-1, 1]$, and the location index $\mu \in \mathbb{R}$

Based on our empirical results where we observed that the corresponding first- and second-order distribution functions never intersected under special conditions (specified in the following hypotheses), we come up with the following two hypotheses:

Hypothesis 1. *If $X \sim S_{\alpha_1}(1,0,0)$ and $Y \sim S_{\alpha_2}(1,0,0)$, $\alpha_1, \alpha_2 > 1$, $\alpha_2 > \alpha_1$, then*

$$Y \geq_{(2)} X.$$

In other words, Y dominates X in the second order.

Hypothesis 2. *If $X \sim S_{\alpha_1}(1,1,0;0)$ and $Y \sim S_{\alpha_2}(1,1,0;0)$, $\alpha_1, \alpha_2 > 0$, $\alpha_2 > \alpha_1$, i.e. X and Y are stable random variables in zero parametrisation, (See Definition 5, Chapter 1) then*

$$X \geq_{(1)} Y.$$

Theorem 23. *Let us assume that $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\sigma > 0$, and $\mu \in \mathbb{R}$. Then for stable distributions $S_\alpha(\sigma,\beta,\mu)$ with $\beta < 1$ and $\alpha > 1$, there exist only first- and second order distribution functions of these distributions, i.e. $F^{(1)}(\cdot)$ and $F^{(2)}(\cdot)$. If $\beta < 1$ and $0 < \alpha < 1$, then only $F^{(1)}(\cdot)$ exists. If however $\beta = 1$, then for $S_\alpha(\sigma,1,\mu)$ there exists $F^{(k)}(\cdot)$ for all $k \in \mathbb{N}$.*

Proof. If $\beta > -1$, then according to (1.7), the asymptotic form of the left tail of the density is $O\left(\frac{1}{|x|^{\alpha+1}}\right)$. Hence, the asymptotic form of the distribution function $F(\cdot)$ and the second-order distribution function $F^{(2)}(\cdot)$ on $(-\infty, z)$, $z < 0$ is $O\left(\frac{1}{|x|^{\alpha}}\right)$ and $O\left(\frac{1}{|x|^{\alpha-1}}\right)$, respectively. If $\alpha > 1$, then both functions are well-defined. However, the third-order distribution function of a stable random variable has the following asymptotic form $O\left(\frac{1}{|x|^{\alpha-2}}\right)$, i.e. it will diverge whenever $\alpha < 2$. But this holds for all stable distributions. Therefore neither the third order nor any higher order of distribution function is not defined for stable distributions with $\beta < 1$. However, if $\beta = 1$, then the left tail will be lighter than any Paretian tail [66]. Therefore, for any $0 < \alpha \leq 2$, $\sigma > 0$, and $\mu \in \mathbb{R}$, $F^{(k)}(\cdot)$ of $S_\alpha(\sigma,1,\mu)$ exists for every $k \in \mathbb{N}$. □

Theorem 24. *Let X_1 and X_2 be two stable distributions which have $S_{\alpha_1}(\sigma_1,\beta_1,\mu_1)$ and $S_{\alpha_2}(\sigma_2,\beta_2,\mu_2)$, respectively with $\sigma_1, \sigma_2 > 0$. Let us also make two assumptions:*

1. *$|\beta_1| < 1$ and $|\beta_2| < 1$,*
2. *$\alpha_1 \neq \alpha_2$.*

Then $X_1 \neq_{(1)} X_2$.

Proof. The proof of this statement is analogous to the proof of Theorem 18. If none of the absolute values of β_1 and β_2 equals one, then both tails will be asymptotically Paretian. So according to (1.7)

$$\begin{aligned}
F_{X_1}(x) &\sim \frac{c_1}{|x|^{\alpha_1}} \qquad x << 0,\\
F_{X_2}(x) &\sim \frac{c_2}{|x|^{\alpha_2}} \qquad x << 0,\\
F_{X_1}(x) &\sim 1 - \frac{d_1}{|x|^{\alpha_1}} \qquad x >> 0,\\
F_{X_2}(x) &\sim 1 - \frac{d_2}{|x|^{\alpha_2}} \qquad x >> 0,
\end{aligned}$$

where $c_1, c_2, d_1,$ and d_2 are positive constants which are uniquely determined by the parameters of the corresponding stable distributions. Without loss of generality, we assume that $\alpha_1 > \alpha_2$.

$$\forall \varepsilon > 0 \quad \exists x_0 \text{ such that } \forall x < x_0 : \quad \begin{cases} \left| F_{X_1}(x) - \frac{c_1}{|x|^{\alpha_1}} \right| < \varepsilon \\ \\ \left| F_{X_2}(x) - \frac{c_2}{|x|^{\alpha_2}} \right| < \varepsilon. \end{cases} \tag{2.13}$$

$$\forall \varepsilon > 0 \quad \exists x_0 \text{ such that } \forall x > x_0 : \quad \begin{cases} \left| F_{X_1}(x) - \left(1 - \frac{d_1}{|x|^{\alpha_1}}\right) \right| < \varepsilon \\ \\ \left| F_{X_2}(x) - \left(1 - \frac{d_2}{|x|^{\alpha_2}}\right) \right| < \varepsilon. \end{cases} \tag{2.14}$$

From (2.13) it follows:

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists x_0 \text{ such that } \forall x < x_0 : \\ &\frac{c_1}{|x|^{\alpha_1}} - \frac{c_2}{|x|^{\alpha_2}} - 2\varepsilon < F_{X_1}(x) - F_{X_2}(x) < \frac{c_1}{|x|^{\alpha_1}} - \frac{c_2}{|x|^{\alpha_2}} + 2\varepsilon \end{aligned}$$

while from (2.14), we get

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists x_0 \text{ such that } \forall x > x_0 : \\ &\frac{d_2}{|x|^{\alpha_2}} - \frac{d_1}{|x|^{\alpha_1}} - 2\varepsilon < F_{X_1}(x) - F_{X_2}(x) < \frac{d_2}{|x|^{\alpha_2}} - \frac{d_1}{|x|^{\alpha_1}} + 2\varepsilon \end{aligned}$$

Hence it follows that for x_0 large enough $F_{X_1}(x) > F_{X_2}(x)$ for all $x > x_0$, and for $-y_0$ large enough $F_{X_1}(x) < F_{X_2}(x)$ for all $x < y_0$, which implies $X_1 \neq_{(1)} X_2$.

□

Theorem 25. *If $X \sim S_\alpha(\sigma_1, 0, 0)$ and $Y \sim S_\alpha(\sigma_2, 0, 0)$, $\sigma_1 < \sigma_2$, and $\alpha > 1$. Then*

$$X \geq_{(2)} Y.$$

In other words X dominates Y in the second order.

For the proof of Theorem 25, we need to prove the following two lemmas.

Lemma 5. *If X_1, X_2 are iid and $X_1, X_2 \sim SP_\alpha(\sigma_1, 0, 0)$ and Y_1, Y_2 are iid and $Y_1, Y_2 \sim SP_\alpha(\sigma_2, 0, 0)$, and $\alpha > 1$ then*

$$\begin{aligned} &\text{if } X_1 \geq_{(1)} Y_1 \rightarrow \quad X_1 + X_2 \geq_{(1)} Y_1 + Y_2, \\ &\text{if } X_1 \geq_{(2)} Y_1 \rightarrow \quad X_1 + X_2 \geq_{(2)} Y_1 + Y_2. \end{aligned}$$

Proof. Let us prove analogous statements for $X_1 + Y_2$ and $Y_1 + Y_2$. We denote the distribution functions of X_i and Y_i, $i = \{1, 2\}$ with F_X and F_Y, respectively.

If $X_1 \geq_{(1)} Y_1$, then

$$\begin{aligned} F_{X_1+Y_2}(z) - F_{X_1+X_2}(z) &= \int_{-\infty}^{\infty} F_Y(z-x)dF_X(x) - \int_{-\infty}^{\infty} F_X(z-x)dF_X(x) \\ &= \int_{-\infty}^{\infty} (F_Y(z-x) - F_X(z-x))dF_X(x) \geq 0, \end{aligned}$$

because the integrand is non-negative on $\mathbb{R}$ and this integral converges if $\alpha > 0$.

If $X_1 \geq_{(2)} Y_1$, then

$$\begin{aligned} \int_{-\infty}^{z} (F_{X_1+Y_2}(z) - F_{X_1+X_2}(z))\, dz &= \int_{-\infty}^{z} \int_{-\infty}^{\infty} (F_Y(z-x) - F_X(z-x))dF_X(x)dz \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} (F_Y(z-x) - F_X(z-x))dz \right) dF_X(x) \end{aligned}$$

because the integrand is non-negative on $\mathbb{R}$ and this integral absolutely converges if $\alpha > 1$ which explains why we can reverse the order of the integrals.

Hence, we get

$$X_1 + X_2 \geq_{(i)} X_1 + Y_2 \geq_{(i)} Y_1 + Y_2, \text{ where } i = \{1, 2\}.$$

□

Lemma 6. *If $X_1, X_2, ..., X_n$ are iid and $X_1, X_2, ..., X_n \sim SP_\alpha(\sigma_1, 0, 0)$ and $Y_1, Y_2, ..., Y_n$ are iid and $Y_1, Y_2, ..., Y_n \sim SP_\alpha(\sigma_2, 0, 0)$, and $\alpha > 1$ then*

$$\begin{aligned} &\text{if } X_1 \geq_{(1)} Y_1 \rightarrow \quad X_1 + X_2 + ... + X_n \geq_{(1)} Y_1 + Y_2 + ... + Y_n, \\ &\text{if } X_1 \geq_{(2)} Y_1 \rightarrow \quad X_1 + X_2 + ... + X_n \geq_{(2)} Y_1 + Y_2 + ... + Y_n. \end{aligned}$$

Proof. Let us make the following notation

$$\begin{aligned} \mathcal{X}_k &= X_1 + X_2 + ... + X_k \\ \mathcal{Y}_k &= Y_1 + Y_2 + ... + Y_k \end{aligned}$$

Let us assume that for some $k \in \mathbb{N}$ the statement of the theorem holds and let us prove that this also holds for $k + 1$. Let us prove first that

$$\mathcal{X}_k + X_k \geq_{(2)} \mathcal{X}_k + Y_k$$

If $X_1 \geq_{(1)} Y_1$, then

$$\begin{aligned} F_{\mathcal{X}_k+Y_{k+1}}(z) - F_{\mathcal{X}_k+X_{k+1}}(z) &= \int_{-\infty}^{\infty} F_Y(z-x)dF_{\mathcal{X}_k}(x) - \int_{-\infty}^{\infty} F_X(z-x)dF_{\mathcal{X}_k}(x) \\ &= \int_{-\infty}^{\infty} (F_Y(z-x) - F_X(z-x))dF_{\mathcal{X}_k}(x) \geq 0, \end{aligned}$$

because the integrand is non-negative on $\mathbb{R}$ and it can easily be shown that the integral converges if $\alpha > 0$.

If $X_1 \geq_{(2)} Y_1$, then

$$\begin{aligned}
&\int_{-\infty}^{z} \left(F_{\mathcal{X}_k+Y_{k+1}}(z) - F_{\mathcal{X}_k+X_{k+1}}(z)\right) dz = \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} (F_Y(z-x) - F_X(z-x) dF_{\mathcal{X}_k}(x) \right) dz = \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} (F_Y(z-x) - F_X(z-x) dz \right) dF_{\mathcal{X}_k}(x) \geq 0,
\end{aligned}$$

because the integrand is non-negative on $\mathbb{R}$ and it can easily be shown that the integral absolutely converges if $\alpha > 1$ which explains why we can reverse the order of the integrals. In an analogous manner we get that $\mathcal{X}_k + Y_{k+1} \geq_{(2)} \mathcal{Y}_k + Y_{k+1}$. And hence

$$\mathcal{X}_k + X_{k+1} \geq_{(2)} \mathcal{X}_k + Y_{k+1} \geq_{(2)} \mathcal{Y}_k + Y_{k+1}$$

which implies

$$\mathcal{X}_{k+1} \geq_{(2)} \mathcal{Y}_{k+1},$$

so by induction, this proves the statement of the lemma. □

Proof of Theorem 25. Let $X_1^{(\sigma_1)}, X_2^{(\sigma_1)}, ..., X_n^{(\sigma_1)}$ and $Y_1^{(\sigma_2)}, Y_2^{(\sigma_2)}, ..., Y_n^{(\sigma_2)}$ be iid random variables from $SP_\alpha(\sigma_1, 0, 0)$ and $SP_\alpha(\sigma_2, 0, 0)$, respectively and let us assume that $\sigma_1 < \sigma_2$.
Let us take into account that according to Theorem 1 (Chapter 1)

$$\frac{X_1^{(\sigma_1)} + X_2^{(\sigma_1)} + ... + X_n^{(\sigma_1)}}{n^{1/\alpha}} \text{ and } \frac{Y_1^{(\sigma_2)} + Y_2^{(\sigma_2)} + ... + Y_n^{(\sigma_2)}}{n^{1/\alpha}}$$

belong to the domain of attraction of a symmetric stable distribution with the tail index equal to α. Let us also note that for every $n \in \mathbb{N}$

$$\frac{X_1^{(\sigma_1)} + X_2^{(\sigma_1)} + ... + X_n^{(\sigma_1)}}{n^{1/\alpha}} \geq_{(2)} \frac{Y_1^{(\sigma_2)} + Y_2^{(\sigma_2)} + ... + Y_n^{(\sigma_2)}}{n^{1/\alpha}}$$

because of the transitivity property and the feature that for each i

$$\frac{X_i^{(\sigma_1)}}{n^{1/\alpha}} \geq_{(2)} \frac{Y_i^{(\sigma_2)}}{n^{1/\alpha}}$$

Hence

$$\lim_{n\to\infty} \frac{X_1^{(\sigma_1)} + X_2^{(\sigma_1)} + ... + X_n^{(\sigma_1)}}{n^{1/\alpha}} \geq_{(2)} \lim_{n\to\infty} \frac{Y_1^{(\sigma_2)} + Y_2^{(\sigma_2)} + ... + Y_n^{(\sigma_2)}}{n^{1/\alpha}}$$

□

Notes on Hypothesis 1

Let $X_1^{(\alpha_1)}, X_2^{(\alpha_1)}, ..., X_n^{(\alpha_1)}$ and $Y_1^{(\alpha_2)}, Y_2^{(\alpha_2)}, ..., Y_n^{(\alpha_2)}$ be iid random variables from $SP_{\alpha_1}(1,0,0)$ and $SP_{\alpha_2}(1,0,0)$, respectively and let us assume that $\alpha_1 < \alpha_2$. Let us take it into account that

$$X_1^{(\alpha_1)} + X_2^{(\alpha_1)} + ... + X_n^{(\alpha_1)} \text{ and } Y_1^{(\alpha_2)} + Y_2^{(\alpha_2)} + ... + Y_n^{(\alpha_2)}$$

belong to the domain of attraction of a symmetric stable distributions with the tail index equal to α_1 and α_2, respectively. Let us also note that for every $n \in \mathbb{N}$

$$X_1^{(\alpha_1)} + X_2^{(\alpha_1)} + ... + X_n^{(\alpha_1)} \leq_{(2)} Y_1^{(\alpha_2)} + Y_2^{(\alpha_2)} + ... + Y_n^{(\alpha_2)}$$

because of the Lemma I and the feature that for each i

$$X_i^{(\alpha_1)} \leq_{(2)} Y_i^{(\alpha_2)}$$

However, there is no stochastic dominance between

$$\frac{X_1^{(\alpha_1)}}{n^{1/\alpha_1}} \text{ and } \frac{Y_1^{(\alpha_2)}}{n^{1/\alpha_2}}$$

because $\frac{1}{n^{1/\alpha_1}} < \frac{1}{n^{1/\alpha_2}}$. The option of calculating the distribution of the convolution of

$$X_1^{(\alpha_1)} + X_2^{(\alpha_1)} + ... + X_n^{(\alpha_1)}$$

by using the formula:

$$F_{X+Y}(z) = \int F_X(z-x)dF_Y(x)$$

is not productive because F_{X+Y} cannot be expressed in terms of elementary functions.

Auxiliary Lemmas

Let $F_\alpha^{(2)}(\cdot)$ denote the second-order distribution function of $S_\alpha(1,0,0)$.

Lemma 7. *The function $y_\alpha(x) = F_\alpha^{(2)}(x) - x^+$ is even.*

Proof. Without loss of generality, let us assume that $h > 0$ and prove that $y_\alpha(h) = y_\alpha(-h)$. So if we denote $\int_x^\infty f_\alpha(z)dz$ with $F_\alpha^*(x)$, then

$$\begin{aligned}
&y_\alpha(h) = F_\alpha^{(2)}(h) - h = \\
&= \int_{-\infty}^{-h} F_\alpha(z)dz + \int_{-h}^{0} F_\alpha(z)dz + \int_0^h F_\alpha(z)dz - h = \\
&= y_\alpha(-h) + \int_{-h}^{0} F_\alpha(z)dz + \int_0^h (1 - F_\alpha^*(z))dz - h = \\
&y_\alpha(-h) + \int_{-h}^{0} F_\alpha(z)dz - \int_0^h F_\alpha^*(z)dz = y_\alpha(-h).
\end{aligned}$$

□

Theorem 26. *The partial derivative of $F_\alpha^{(2)}(x)$ with respect to α is given as follows*

$$\frac{\partial F_\alpha^{(2)}(x)}{\partial \alpha} = -\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)t^\alpha \ln(t)\exp\left(-t^\alpha\right)dtdz$$

and Hypothesis 1 holds if and only if

$$\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)t^\alpha \ln(t)\exp\left(-t^\alpha\right)dtdz \geq 0, \quad \forall x \in \mathbb{R}^+ \text{ and } \alpha \in (1,2].$$

Proof. We express the distribution function by means of the integral of the density function which in turn is represented through the characteristic function. By applying Leibnitz integral rule, we get

$$\begin{aligned}
\frac{\partial F_\alpha^{(2)}(x)}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left(\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)\exp(-t^\alpha)\,dt\,dz\right) = \\
&= \frac{1}{\pi}\int_{-\infty}^{x}(x-z)\frac{\partial}{\partial \alpha}\left(\int_{0}^{\infty}\cos(tz)\exp(-t^\alpha)\,dt\,dz\right) = \\
&= \frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)\left(\frac{\partial}{\partial \alpha}\exp\left(-t^\alpha\right)\right)dtdz = \\
&= -\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)\left(\frac{\partial}{\partial \alpha}t^\alpha\right)\exp\left(-t^\alpha\right)dtdz = \\
&= -\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)\left(\frac{\partial}{\partial \alpha}\exp(\alpha\ln t)\right)\exp\left(-t^\alpha\right)dtdz = \\
&= -\frac{1}{\pi}\int_{-\infty}^{x}\int_{0}^{\infty}(x-z)\cos(tz)t^\alpha\ln(t)\exp\left(-t^\alpha\right)dtdz.
\end{aligned}$$

Since the derivative of $F_\alpha^{(2)}(x)$ exists for all $x \in \mathbb{R}$, $F_\alpha^{(2)}(x)$ is a decreasing function in α in the interval $(1,2]$ if and only if $\frac{\partial F_\alpha^{(2)}(x)}{\partial \alpha} \leq 0$ on $(1,2]$ for all $x \in \mathbb{R}$. □

Lemma 8. *Let us assume that $X_1 = \sigma_1 X$ and $X_2 = \sigma_2 X^*$, where X and X^* are iid, the support of X is $\mathbb{R}$, and*

$$F_X^{(2)}(x) = \int_{-\infty}^{x} F_X(z)dz$$

exists for all $x \in \mathbb{R}$, $F_X^{(2)}(x) > 0$ for all $x \in \mathbb{R}$, and $\sigma_2 > \sigma_1 > 0$. Then X_2 cannot dominate X_1 in the second order.

Proof. Let us note that by the generalised central limit theorem, the tail of $Z_1 + Z_2 + ... + Z_k$ for $Z_i, i = 1, ..., k$ being iid and $Z_1 \sim SP_\alpha(\sigma, 0, 0)$, $\sigma > 0$, $\alpha \in (0,2)$

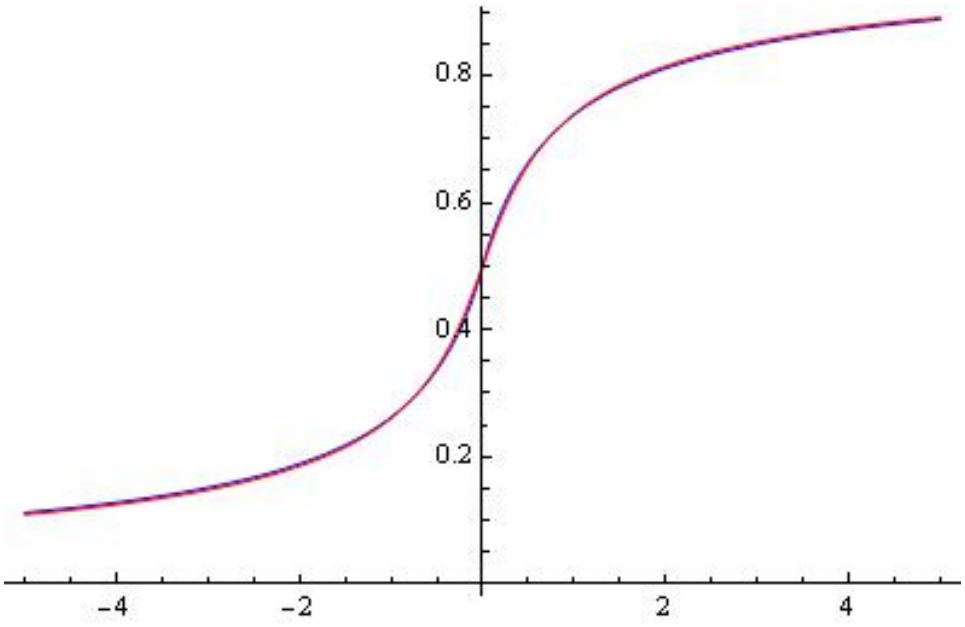


Figure 2.1: The distribution functions for $SP_{0.691}(c(0.691),0,0)$ and $S_{0.691}(1,0,0)$ where $c(\alpha)=\sqrt[\alpha]{\frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi\alpha/2}}$

.

is $O\left(\frac{1}{x^{\alpha+1}}\right)$.
Let us assume that $\sigma_2>\sigma_1>0$

$$\begin{aligned}\int_{-\infty}^{0}(F_{X_2}(z)-F_{X_1}(z))dz &= \int_{-\infty}^{0}(F_{X\sigma_2}(z)-F_{X\sigma_1}(z))dz= \\ &= \int_{-\infty}^{0}(F_X(z/\sigma_2)-F_X(z/\sigma_1))dz= \\ &= \int_{-\infty}^{0}F_X(z/\sigma_2)dz-\int_{-\infty}^{0}F_X(z/\sigma_1)dz= \\ &= \sigma_2\int_{-\infty}^{0}F_X\left(\frac{z}{\sigma_2}\right)d\left(\frac{z}{\sigma_2}\right)-\sigma_1\int_{-\infty}^{0}F_X\left(\frac{z}{\sigma_1}\right)d\left(\frac{z}{\sigma_1}\right)= \\ &= \sigma_2\int_{-\infty}^{0}F_X(y)\,dy-\sigma_1\int_{-\infty}^{0}F_X(y)\,dy= \\ &= \sigma_2F_X^{(2)}(0)-\sigma_1F_X^{(2)}(0)= \\ &= (\sigma_2-\sigma_1)F_X^{(2)}(0)>0.\end{aligned}$$

Since $F_{X_2}(0)>F_{X_1}(0)$, X_2 cannot dominate X_1 in the second order. □

Let us note that if X_1 and X_2 are random variables and $X_1\geq_{(k)}X_2$ with $k\in\mathbb{N}$, then $X_1+\mu\geq_{(k)}X_2+\mu$ where $\mu\in\mathbb{R}$ [104].

2.4 Second-Order Crossing for Stable Distributions

In our work [4], we explored the behaviour of portfolios of stable random returns in the setup of stochastic dominance. Here we present the basic inferences which we formulate in terms of the following theorem.

Theorem 27 (Portfolios of stable returns)**.** *Let us assume that* $X_1 \sim S_\alpha(\sigma_1, 0, \mu)$ *and* $X_2 \sim S_\alpha(\sigma_2, 0, \mu)$ *with* $\alpha > 1$, $\sigma_1 > 0$, $\sigma_2 > 0$, $\mu \in \mathbb{R}$ *and* $\sigma_1 < \sigma_2$. *Then*

1. $X_1 \geq_{(2)} X_2$,
2. *If*
$$a_1, a_2 \in \left[\frac{\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}, 1 \right] \quad \text{and } a_1 > a_2$$
then
$$a_1 \cdot X_1 + (1-a_1) \cdot X_2 \leq_{(2)} a_2 \cdot X_1 + (1-a_2) \cdot X_2$$
3. *If*
$$a_1, a_2 \in \left[0, \frac{\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}} \right] \quad \text{and } a_1 > a_2$$
then
$$a_1 \cdot X_1 + (1-a_1) \cdot X_2 \geq_{(2)} a_2 \cdot X_1 + (1-a_2) \cdot X_2$$

Proof. The first assertion follows from Theorem 25. Let us denote the point $\frac{\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}$ as follows:

$$a^* = \frac{\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}{1+\left(\frac{\sigma_2}{\sigma_1}\right)^{\alpha/(\alpha-1)}}$$

If $a \in [0,1]$, then according to Theorem 5 (Chapter 1):

$$a \cdot X_1 + (1-a) \cdot X_2 \sim S_\alpha\left((a^\alpha \sigma_1^\alpha + (1-a)^\alpha \sigma_2^\alpha)^{1/\alpha}, 0, \mu \right).$$

If we consider the scale parameter as a function of a, we find out that the function $\sigma(a) = (a^\alpha \sigma_1^\alpha + (1-a)^\alpha \sigma_2^\alpha)^{1/\alpha}$ has its minimum at a^*. This is because its derivative is zero at this point, and the second derivative of $\sigma(a)$ is positive for all $a \in (0,1)$, which we see from the following calculations:

$$\frac{\partial \sigma(a)}{\partial a} = \frac{a^{\alpha-1} \cdot \sigma_1^\alpha - (1-a)^{\alpha-1} \cdot \sigma_2^\alpha}{\left(a^\alpha \cdot \sigma_1^\alpha + (1-a)^\alpha \cdot \sigma_2^\alpha\right)^{1-1/\alpha}}$$

which is zero iff $a = a^*$ and it is evident that if $\alpha > 1$, $\sigma_1 > 0$, $\sigma_2 > 0$, then $a^* \in (0,1)$.

The second derivative of $\sigma(a)$ is as follows:

$$\frac{\partial^2\sigma(a)}{\partial a^2} = \frac{(\alpha-1)(\sigma_1\sigma_2)^\alpha(a\cdot(1-a))^{\alpha-2}}{(a^\alpha\sigma_1^\alpha+(1-a)^\alpha\sigma_2^\alpha)^{2-1/\alpha}},$$

which is positive if $\alpha > 1, \sigma_1 > 0, \sigma_2 > 0$, and $a \in (0,1)$. Since under the assumption of this theorem the second derivative is positive in $(0,1)$ the first derivative is negative in $[0,a^*)$ and is positive in $(a^*,1]$. At the point a^*, the function $\sigma(\cdot)$ has the minimum value, because the second derivative at this point is positive. This proves the second and the third assertion of this theorem. □

Theorem 27, enables us to choose the optimal proportion in terms of second-order stochastic dominance.

2.5 Wasserstein Metric

While stochastic dominance enables us to order random variables, the Wasserstein metric enables us to quantify distances between them which, when it is of the first order and of unit dimension, can be represented as follows [97, 7]:

$$d_{W_1^1}(\mathbb{P}_{F_1},\mathbb{P}_{F_2}) = \int_{\mathbb{R}} |F_1(z)-F_2(z)|dz,$$

where $\mathbb{P}_{F_1}$ and $\mathbb{P}_{F_2}$ are two probabilistic distributions and F_1 and F_2 are their distribution functions, respectively. A complete definition of the Wasserstein metric can be found in [97, 7, 35]. In our paper (2012 [3]), we explored the convergence properties of the first-order Wasserstein metric values for stable distributions and, in this dissertation, we use this metric for the determination of the granularity of a discrete distribution that we use for approximating a continuous function. That is to say, we increase the granularity until the value of the corresponding Wassertein metric falls below a predefined level. A similar approach is also used in [105].

Conclusion to Chapter 2

In this chapter, we derived the optimal proportion in terms of second-order stochastic dominance for portfolios of two stable random variables one of which second order dominates the other and both have the same β, μ, and α. We found parametric regions for stable distributions where we can find stochastic dominance of the second order and presented ways of simulating two stable random variables one of which dominates another one in the second order. We came up with the hypothesis that if $X_1 \sim S_{\alpha_1}(1,0,0)$ and $X_2 \sim S_{\alpha_2}(1,0,0)$ then

$$\alpha_1 > \alpha_2 \text{ implies } X_1 \geq_{(2)} X_2.$$

Apart from stochastic ordering, stochastic dominance can be used in optimization for handling probability constraints what we explore in Section 3.6.

3. Stochastic Dynamic Programming for Gas Storage Valuation

Introduction

In this chapter, we solve the problem of gas storage valuation under spot price uncertainty. We also pay attention to the fact that a gas storage unit may be connected to several markets, which implies the necessity of modelling price-spread processes, i.e. the differences between two price processes. At the end of this chapter, we also discuss forward strategies. We estimate the value of a gas storage unit by means of evaluating the state of this asset, and following Löhndorf and Wozabal [105], we define two kinds of states:

1. The environmental state which cannot be changed by our decision, i.e. spot or forward prices. It is denoted by p.

2. The resource state, which can be affected by our decision, i.e. the storage level. It is denoted by L

Our goal is to estimate the functions

$$\begin{aligned} &\text{Value function} && V_t(L,p), \quad t=1,..,T, \\ &\text{Policy function} && \text{policy}_t(L,p), \quad t=1,..,T, \end{aligned}$$

where $V_1(L,p)$ is the estimate of the gas storage unit for the initial storage level L and initial price p, which equal to the optimal value of Problem (3.4) (See Section 3.1) [79], and $\text{policy}_t(L,p)$ determines our decision on how much to inject or withdraw at time t for $L_{t-1}=L$ and $p_t=p$, and T is the considered horizon. In the following section, we provide the definition of the problem of gas storage valuation in the form of an optimisation problem. For the gas storage forward price models, the optimisation problem is defined differently, than that of the spot price models, because there is a different time scale and a gap between the time of payment and the time of delivery [105]. Additionally, the delivery might not take place because a contract might be offset.
Since we work with heavy-tailed price models, we want to reduce the influence of the necessity for discretising the price process, because any discretisation yields a light-tailed distribution. To decide whether the granularity of the discrete distribution that we use to approximate the continuous distribution is appropriate, we use the Wasserstein metric. We use this metric because it strongly relates to the distance of objective functions [71, 72]. Aside from the transition matrix, we resort to Monte Carlo methods. When we truncate the stable distribution, we also use the Wasserstein metric for choosing the points of truncation in such a way that the Wasserstein distance between the original distribution and the truncated distribution is below a predefined value δ. The first who introduced the Wasserstein metric for stochastic optimisation are W. Römisch and R. Schulz. [85, 86].

To prepare a practical tool for the estimation of a gas storage unit, we need to conduct the following steps:

1. To prepare an appropriate price model and to run tests of goodness of fit for it.

2. To create from historical observations of prices a transition matrix and to measure whether this choice is appropriate by using the Wasserstein metric. The randomness of prices in all cases is given by a lattice [105].

3. If the transition matrix is too large, we resort to post-decision value functions and Monte Carlo methods.

For the analysis of prices, we use stable, Student-t, and normal innovations. We first start with a simple SDP problem, and then consider its modifications that incorporate more features of real prices. Apart from dealing with the risk-neutral models, we also deal with risk-averse approaches. We solve these problems by means of the following methods:

- SDP with linear interpolation, which yields a lower bound of value function.

- SDP by means of linearising the problem, which yields an upper bound of the value function.

- Stochastic dual dynamic programming.

- SDP in the framework of post-decision states and value functions.

- SDP for the value function continuous in the environmental state.

The final spot-model is the spot price linked to forward prices where the forward prices are modelled by a multivariate sub-Gaussian stable distribution and the spot price also has a stable distribution with a different tail index α.
At the end of this chapter, we provide a comparison of the methods and the conclusion. We use real price data, but otherwise we use toy gas storage units, where the capacity is one million cubic meters and the rest of the parameters are described in Table 3.2. The information on the parameters of real European gas storage units can be found in http://transparency.gie.eu

The Chapter's Structure

In Section 3.1, we give the formulation of the problem in terms of Bellman equations, and discuss the problem's basic features. In section 3.2 following Powell [79], we define the concept of the post-decision states and value functions, and give a post-decision formulation of the problem of gas storage valuation. In this section, we also describe our approach, for solving the problem by means of two methods, one of which is based on Benders cuts and yields an upper bound [74], and another, which is based on linear interpolation, yields a lower bound for the problem. In section 3.3, we present our results upon solving *Simple Problem*, where the randomness of prices is determined by the transition matrix. In section 3.4, we formulate *Advanced Problem I*, whereby we incorporate the fact that some gas storage units might be connected to several markets into the problem.

In section 3.5, we address *Advanced Problem I*, by *Average Strategy*. At the end of Section 3.5, we present the comparison of the results, by means of different methods which we used to solve *Advanced Problem I*. In Section 3.6, we define *Advanced Problem II*, whose aim is to incorporate more features of prices into the model, and to link spot and forward gas prices.

In this chapter, we also discuss stochastic dominance constraints and forward strategies of gas storage valuation.

3.1 The Problem Formulation

The parameters of the problem and its notation are described in Table 3.1.
We consider the storage level in such bounds when ratchets (head effects or the dependence of the injection and withdrawal rates on the storage level) can be ignored. In practice, the equality $L = 0$ corresponds to the cushion volume which relates to the minimum pressure requirements [36].
If at time t the price is p_t and the storage level is L_t then the reward is

$$R_t(L, p, x) = \min\{(p - c^w) \cdot x, (p + c^i) \cdot x\}, \tag{3.1}$$

where

$$x \in [\max(i, (L - C)), \min(L, w)]. \tag{3.2}$$

In this section, the stochasticity of the price process is given by the transition matrix $\mathbb{P}$, which is estimated from historical data.

Here x is our decision and when it is negative, we inject and when it is positive, we withdraw.

Our objective is to maximise the composite reward from trading across all decision strategies, i.e.

$$\max \left\{ \mathbb{E} \left(\sum_{t=1}^{T} \gamma^t R_t(L_{t-1}, p_t, x_t) \right) \right\} \tag{3.3}$$

s.t

$$0 \leq L_t \leq C, \quad t = 1, ..., T,$$
$$i \leq x_t \leq w, \quad t = 1, ..., T,$$
$$L_t = L_{t-1} - x_t, \quad t = 1, ..., T,$$
$$p_t, \quad t = 1, ..., T \quad \text{is a Markovian process.}$$

where γ is the discount factor (which we assume is unit therefore, subsequently this symbol will be omitted) and x_t is our decision (which is based on the information about prices and storage levels from the beginning to time t i.e., our decision function is non-anticipatory). If the price process p_t is Markovian, then following a Bellman equation [79], we will get the value function in the following form:

C ..... The capacity of the reservoir
c^i ..... Injection cost
c^w ..... Withdrawal cost
$-i$ ..... Maximum injection
w ..... Maximum withdrawal
T ..... The length of the considered period

Table 3.1: The parameters of two reservoirs.

$$
\begin{aligned}
& V_t(L_{t-1}, p_t) = \max\{R_t(L_{t-1}, p_t, x) + \mathbb{E}(V_{t+1}(L_t, p_{t+1}))|p_t)\} \qquad (3.4)\\
& \text{s.t.}\\
& L_t = L_{t-1} - x,\\
& i \le x \le w,\\
& 0 \le L_t \le C,\\
& p_k, \quad k = 1, ..., T \quad \text{is a Markovian process.}
\end{aligned}
$$

for $t = 1, ..., T$ and $V_{T+1}(\cdot, p) \equiv 0 \ \forall p$.
In reality the price for tomorrow is not known, so the realistic version 3.4 is as follows:

$$
\begin{aligned}
& V_t(L_{t-1}, p_t) = \max\{\mathbb{E}(R_t(L_{t-1}, p_{t+1}, x) + V_{t+1}(L_t, p_{t+1}))|p_t)\} \qquad (3.5)\\
& \text{s.t.}\\
& L_t = L_{t-1} - x,\\
& i \le x \le w,\\
& 0 \le L_t \le C.
\end{aligned}
$$

We solve both problems.

3.1.1 The Variables of the Value Function

The resource variable is the storage level L which depends on our decision and the environmental variable is the price p. If the transition matrix has the size $m \times m$ then at time t there are m distinct values of prices and hence m value functions $V_t(\cdot, p_i)$, $i = 1, .., m$ [105]. We denote the space of all possible storage levels at time t as $\mathcal{L}_t$ and if this set has N elements and we want to emphasise this, then we denote it $\mathcal{L}_t(N)$. If the elements of $\mathcal{L}_t$ determine equidistant intervals, then

$$\mathcal{L}_t = \mathcal{L}_t(N) = \{0, 1/N \cdot C, ..., (N-1)/N \cdot C, C\}$$

Storage	C	c^i	c^w	i	w	T
R1	$10^6 m^3$	2 EUR	2 EUR	18000 m^3	18000 m^3	250 days
R2	$10^6 m^3$	2 EUR	2 EUR	8750 m^3	8750 m^3	250 days

Table 3.2: The values of the parameters of two reservoirs.

3.1.2 The Aim of Gas Storage Valuation

Our aim is to calculate the value of the empty storage at the beginning for every initial price level, i.e. the function $f^*(p) = V_1(0, p)$. This function naturally decreases in the initial price level, because to profit from the storage, we have to first buy gas, and then wait until its price goes up. But the higher the initial price, the longer we have to wait for an 'acceptable' price for buying (except for some special cases). If the price, however, is low, then we can buy gas immediately and then wait until the price increases.

3.1.3 Linear Interpolation Methodology

1. for every p_i and at every time t, we solve Problem (3.4) for a predefined number N of the values of L_t where $L_t \in \mathcal{L}_t(N)$ and record the values of $V_t(\cdot, p_i)$ for every value of L_t.

2. The values in the intervals $(i/N \cdot C, (i+1)/N \cdot C)$, $i = 0, .., N-1$ are calculated by linear interpolation.

The value function $V_t(L, p)$ is concave in storage level for all t and p [105, 74], and it can be easily established by the definition of a concave function that this methodology yields a lower bound of the problem, but it converges to the true value function for a given transition matrix as N goes to infinity.

We call this algorithm *Linear Interpolation*.

3.1.4 Bellman Sums

After we calculate the value and policy functions by stochastic dynamic programming, we can assign initial values of the storage level and prices and simulate the price process by the given transition matrix as is shown in Table 3.3.

Note that according to [79]

$$\lim_{M \to \infty} V_1^{\{M\}}(L, p) = V_1(L, p), \tag{3.6}$$

which means that the implementation of the optimal policy yields on average earnings estimated by the value function (the notation in (3.6) is described in Table 3.3.). We call $V_1^{(n)}(L, p)$ a *single realisation Bellman sum* (where the superscript n denotes the index of the scenario) and the variable $V_1^{\{M\}}(L, p)$ the *Bellman sum*, where M is the number of simulated scenarios (See Table 3.3).

We use Bellman sums as the validation and we implement this strategy as a Benchmark for its comparison with other methods such as post-decision SDP and ADDP which will be described and considered later.

3.1.5 A Remark on Bellman Sums

Let us note that the average Bellman sums correspond to the nested formulation of the objective function of a multi-stage problem [94]. After the value function is found, we can calculate the simulated rewards by the formula of the nested formulation of the value function

$$V_1(L_0, p_1) = \max_{x_1 \in \mathcal{X}_1} R_1(L_0, p_1, x_1) +$$
$$+\mathbb{E}\left[\sup_{x_2 \in \mathcal{X}_2(L_1)} R_2(L_1, p_2, x_2) + \mathbb{E}\left[... + \mathbb{E}\left[\sup_{x_T \in \mathcal{X}_T(L_{T-1})} R_T(L_{T-1}, p_T, x_T)\right]\right]\right]$$

When we start at time $t = 0$ and at every time $t = 1, .., T$ we make the decision by using the policy function $\text{policy}_t(L_{t-1}, p_t)$ applied on the state that we obtained by generating the randomness by using the transition matrix $\mathbb{P}$ then the average value of the resulting aggregate rewards converges to $V_1(L_0, p_1)$.

3.1.6 Markovian Property and Discretisation

In *Simple problem*, we either assume that the price process is AR(1) or is determined by a transition matrix that we calculate from the history. In Figure 3.3, there is the transition matrix of prices which we obtained from the history from PEG North market. In this simplified model the matrix is assumed to be the same for all periods and seasons. In the case of AR(1), we decide on the granularity of the discretisation by using the Wassertein metric as the measure of the distance between the corresponding empirical distribution and the discrete distribution which we use to approximate the former distribution.

Bellman Sums

0. $L_0 = 0$, $p_1 = p$
I. $n = 1$,
II. **while** $n \leq M$
  generate prices $p_2, .., p_T$,
  for $t = 1$ to T
  $x_t = \text{policy}(L_{t-1}, p_t)$,
  $L_t = L_{t-1} - x_t$,
  end for
  $V_1^{(n)}(0, p) = \sum_{t=1}^{T} \min((p_t + c^i)x_t, (p_t - c^w)x_t)$,
  $n := n + 1$,
  end while
III. The Bellman sums are calculated as follows:
  $V_1^{\{M\}}(0, p) = \frac{1}{M} \sum_{n=1}^{M} V_1^{(n)}(0, p)$.

Table 3.3: Bellman Sums.

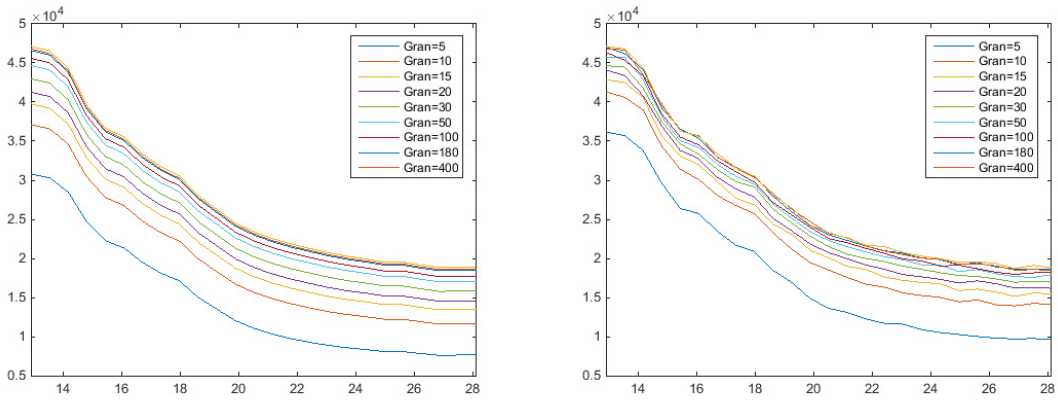


Figure 3.1: The increase in the value function $V_1(0,p)$ (The picture on the left) and the average Bellman sums $V_1^{\{10000\}}(0,p)$ (The picture on the right) for higher granularities of the resource states. The x-axis is the initial prices

3.1.7 The Convergence of Linear Interpolation

In this subsection, we consider the convergence of the method of linear interpolation as we increase the granularity of the state space, i.e. the storage level $L_t \in \mathcal{L}_t$ while other parameters remain unchanged. N is the granularity which is denoted 'Gran' in the pictures and in all cases, we get only 30 distinct values of prices. If when we have $N = 400$ the relative difference between the value function and the average Bellman sums (10000 summands) is less than 0.06%. When N is small, then the difference between the average Bellman sums and the value function is substantial. In Figure 3.1 there are graphs of the value functions (left picture) and average Bellman sums for 10000 summands (right pictures). The x-axis in Figures 3.1 and 3.2 represents the initial values of the price. The initial storage level is zero which means that we first have to buy gas in order to sell it later, therefore all of the functions are decreasing in the initial prices level. In Figure 3.2 there are graphs of the value function $V_1(0,p)$ (blue line) and the Bellman sums $V_1^{\{10000\}}(0,p)$ and we can see that the higher the granularity of the resource state, the close these functions get to each other. Before, we move to auto-regressive models of prices, we have to define post-decision states and value functions.

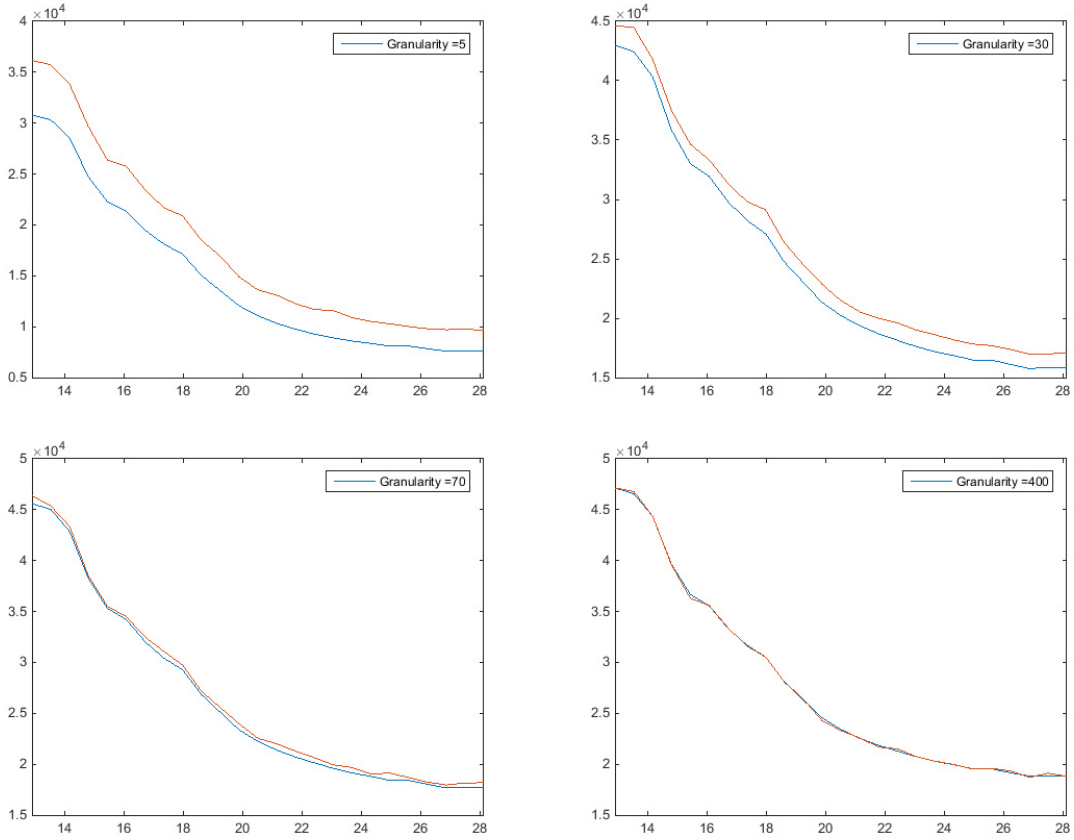


Figure 3.2: The visualisation of how the average Bellman sums $V_1^{\{10000\}}(0,p)$ (red line) approach $V_1(0,p)$ (blue line) as the granularity of the resource state increases.

Post-Decision Formulation of a Storage Problem

The concept of the post-decision state and value function is defined in Appendix. Let us denote the state of the system at time t, with $t \in \{1, .., T\}$, as follows: $Z_t = (L_t, p_t)^T$. The storage level can be affected by our decision, but we assume that we do not possess the market power. Therefore, our decisions do not influence the price. Following Powell's definition of a post-decision variable [79], we denote with $Z_t^x = (L_t^x, p_t)$ the post decision state which takes place immediately after we have made a decision, but new information has not yet been revealed. In the case of hydro storage, the post-decision state refers to the state when we made the decision at time t, but the inflow and spill at time $t+1$ (denoted f_{t+1} and s_{t+1}, respectively) has not been shown. So for the storage level in hydro reservoirs [74, 79], $L_t^x = L_t - x$, and

$$L_{t+1} = L_t - x + f_{t+1} - s_{t+1} = L_t^x + f_{t+1} - s_{t+1}.$$

However, in the case of gas storage, there is neither the inflow nor the spill, therefore

$$L_{t+1} = L_t^x.$$

Having defined the post-decision state, we can define the post-decision value function (according to [79]) – as follows:

$$V_t^x(L_{t-1}^x, p_t) = \mathbb{E}\left(V_{t+1}(L_t, p_{t+1})|(L_{t-1}^x, p_t)\right) \tag{3.7}$$

The value function can be obtained from the post-decision value function:

$$V_t(L_{t-1}, p_t) = \max_{x \in \mathcal{X}_t}\left(R_t(L_{t-1}, p_t, x) + V_{t+1}^x(L_t^x, p_{t+1})\right) \tag{3.8}$$

By substituting (3.8) into (3.7), we get

$$V_t^x(L_{t-1}^x, p_t) = \mathbb{E}\left[\max_{x \in \mathcal{X}_t}\left(R_t(L_t, p_t, x) + V_{t+1}^x(L_t^x, p_{t+1})\right)|(L_{t-1}^x, p_t)\right] \tag{3.9}$$

where p_{t+1} is the generation of price at time t, conditional on that $p_{t-1} = p$. By taking into account that:

1. In the case of gas storage $R_t(L, p, x) = \min((p + c^i)x, (p - c^w)x)$ and for the hydro $R_t(L, p, x) = p \cdot x$,
2. $L_t^x = L_{t+1}$,

we can rewrite (3.9) as follows:

$$\begin{aligned} &V_t^x(L_{t-1}, p_t) = \mathbb{E}\left[\max\left(R_t(L_{t-1}, p_t, x) + V_{t+1}^x(L_t, p_{t+1})\right)|p_t\right] \\ &L_t = L_{t-1} - x \\ &i \le x \le w, \\ &0 \le L_t \le C. \end{aligned} \tag{3.10}$$

According to Powell [79], the ability to reverse the order of the operators of maximum and expected values can provide us with a tremendous computational advantage, because instead of solving one relatively big stochastic program, we solve many small deterministic problems, and average the results out.

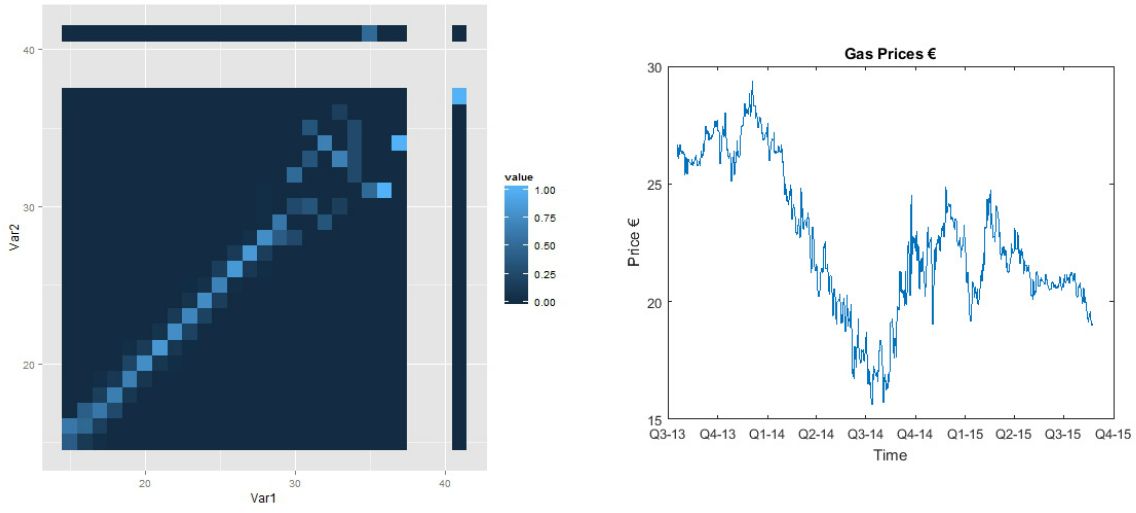


Figure 3.3: The transition matrix of PEG Nord gas prices in the shape of a heat map (left picture) and the original price history from which we transition matrix calculated (right picture).

Post-Decision Linear Interpolation Algorithm

In a similar way as that in which we defined *Linear Interpolation Algorithm*, we can define an analogous version of this algorithm in a post-decision framework.

1. for every p_t and at every time t, we solve Problem (3.10) for a predefined number N of the values of L_t where $L_t \in \mathcal{L}_t$ and record the values of $V_t(\cdot, p_i)$ for all L_t.

2. The values in the intervals $(v/N \cdot C, (v+1)/N \cdot C)$, $v = 0, .., N-1$ are calculated by linear interpolation.

It can easily be established that this methodology yields a lower bound for the problem, but it converges to the true value function for a given transition matrix, as N goes to infinity.

We call this algorithm *Post-Decision Linear Interpolation*.

Linearisation of the Gas Storage Problem

Let $\mathcal{X}_t$ denote the decision constraints at time t. Then, the problem

$$V_t^x(L_{t-1}, p_t) = \mathbb{E}\left[\max_{x, L_t \in \mathcal{X}_t} \left(\min((p_t + c^i)x, (p_t - c^w)x) + V_{t+1}^x(L_t, p_{t+1})\right) | p_t\right]$$

is solved by averaging, i.e. we approximate it with

$$V_t^x(L_{t-1}, p_t) = \frac{1}{N} \sum_{j=1}^{N} \left[\max_{x, L_t \in \mathcal{X}_t} \left(\min((p_t + c^i)x, (p_t - c^w)x) + V_{t+1}^x(L_t, p_{t+1}^{(j)})\right)\right] \tag{3.11}$$

where $p_{t+1}^{(j)}$ are prices at time $t+1$ conditional on p_t.

Spot Algorithm ADDP

At every iteration, we perform:

forw. We simulate a sample $p_1, .., p_T$,
for $t = 1$ to T
for L_{t-1} solve Problem (3.13) and record L_t
end for

backw. **for** $t = T - 1$ **downTo** 1
for $\forall p^* \in P_t$
$n(t) = \text{card}(\mathcal{V}^x_{t+1}(p^*))$

Solve
$V_t(L_{t-1}, p^*) = \max\{\theta^* + \theta_{t+1}\}$
s.t.
$L_t = L_{t-1} - x \qquad \rightarrow [a],$
$0 \leq L_t \leq C,$
$i \leq x \leq w,$
$\theta^* \leq (p^* - c^w)x,$
$\theta^* \leq (p^* + c^i)x,$
$\theta_{t+1} \leq a^{(l)}_{t+1} \cdot L_{t-1} + b^{(l)}_{t+1}, \quad l = 1, ..., n(t),$

where $(a^{(l)}_{t+1}, b^{(l)}_{t+1}) \in \mathcal{V}^x_{t+1}(p^*)$.
We add the Benders cut into the system $\mathcal{V}_t(p^*)$:
$b = V_t(L_{t-1}, p^*) - a \cdot L_{t-1}$
$\mathcal{V}_t(p^*) = \{\mathcal{V}_t(p^*) \cup (a, b)\}$
From the system of cuts $\mathcal{V}_t(p^*)$, we calculate
the post-decision system of cuts $\mathcal{V}^x_t(p^*)$ by (3.6)
end for
end for

Table 3.4: Generic ADDP algorithm for spot price strategies.

In other words, in (3.11) the right-hand side is random and the prices p_j are obtained by simulation. Every summand in (3.11) is the optimal value of the following problem

$$\begin{aligned} &\max \left(\min((p_t + c^i)x, (p_t - c^w)x) + V^x_{t+1}(L_t, p_j)\right) \\ &\text{s.t.} \\ &L_{t+1} = L_t - x, \\ &i \leq x \leq w, \\ &0 \leq L_{t+1} \leq C \end{aligned} \tag{3.12}$$

Since the constraints of all of those problems are the same, and both concavity and piecewise linearity are preserved under any linear combination, we infer that the post-decision value function for this problem has the following representation [94]

$$V^x_{t+1}(L_{t+1}, p_j) = \min_l \left(a^{t+1}_{l,j} L_{t+1} + b^{t+1}_{l,j}\right),$$

where $(a^{t+1}_{l,j}, b^{t+1}_{l,j}), l = 1, ..., m$, which are the Benders cuts [74], are assumed to be known. By introducing two slack variables θ^* and θ_j, we can transform (3.12) into a linear form:

1. $\theta^* = \min((p + c^i)x, (p - c^w)x)$,
2. $\theta_j = V^x_{t+1}(L_{t+1}, p_j) = \min_l \left(a^{t+1}_{l,j} L_{t+1} + b^{t+1}_{l,j}\right)$.

So, the final linear problem will be in the form:

$$\begin{aligned} &\max (\theta^* + \theta_j) \\ &\text{s.t.} \\ &\theta^* \leq (p_j + c^i)x, \\ &\theta^* \leq (p_j - c^w)x, \\ &\theta_j \leq a^{t+1}_{l,j} L_{t+1} + b^{t+1}_{l,j}, \quad l = 1, .., m, \\ &L_{t+1} = L_t - x, \to [a] \quad (\text{the slope of } V_t(\cdot, p_j) \text{ at } L_t) \\ &i \leq x \leq w, \\ &C_0 \leq L_{t+1} \leq C \end{aligned} \tag{3.13}$$

Our post decision value function is the average of the solutions of Problem (3.13) and this feature has to be used in the implementation of ADDP (See Table 3.4).

3.1.8 Linear Programming Algorithm

When we employ the methodology *Linear Interpolation*, we replace (3.4) with (3.13) and use (3.7) and (3.8) to obtain the value function from the post-decision value functions. Then, the resulting algorithm will produce an upper bound [105, 70] for the value function, and will be called *Linearisation Programming*.

3.1.9 ADDP for Gas Storage

An approximate dual dynamic programming algorithm (ADDP) is an iteration algorithm which is run until the criterion of the differences between an upper

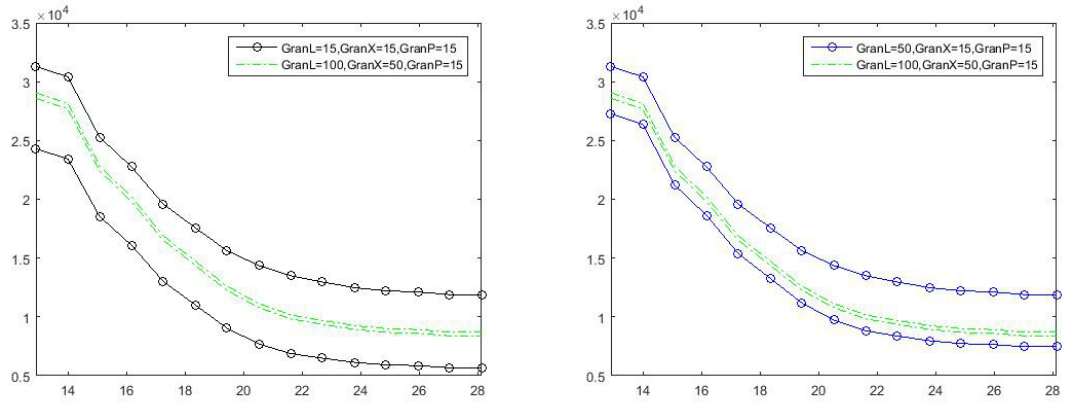


Figure 3.4: Comparison of linear interpolation and Benders cuts I.

and a lower bound is satisfied, i.e. when it is smaller than some predefined value [105, 74]. This algorithm has its roots in the SDDP algorithm developed by Pereira and Pinto [70]. In Table 3.4, $\mathcal{V}_{t+1}(p)$ and $\mathcal{V}^x_{t+1}(p)$ state for the system of Benders for the value function $V_t(\cdot, p)$ and the post-decision value function $V^x_t(\cdot, p)$, respectively. What we perform in iteration k is described in Table 3.4. For the forward path, we simulate the realizations of prices either by bootstrap and using the transition matrix, if we have the transition matrix, or by simulating the residual, if the price process is auto-regressive. We use sampling only for ADDP; in our SDP algorithms, *Alg1* and *Alg2*, we do not resort to it.

This algorithm enables us not to calculate the value function in all of the values from $\mathcal{L}_t$, but since this set is one-dimensional, we can handle the problem of gas storage valuation by means of SDP, however, in the case of forward strategies, our *resource state* consists of not only storage levels, but also of all of the open positions and for forward trading it is indispensable [105].

3.2 Simple Problem

In *Simple Problem*, we model the price uncertainty by means of two approaches:

1. Rounding all of the observed prices to whole numbers and then inferring the transition matrix by the frequencies of the transitions of one state to another one.

2. Modelling prices by AR(1) and solving it by *Alg1* and *Alg2* (i.e. using post-decision approach). Algorithms *Alg1* and *Alg2* are described in Table 3.5 and Table 3.6, respectively.

In Figure 3.3, we demonstrate the price history and the transition matrix inferred from the history which is in the form of a heat map.

3.2.1 *Simple Problem* Addressed by Transition Matrix

In Figure 3.4, there are two pairs of value functions $f(p) = V_1(0, p)$. The upper line is the line obtained by Benders cuts, and it is an upper line because Benders

cuts yield an upper bound for the problem. The lower line is the line obtained by linear interpolation; this is a lower line because linear interpolation yields a lower bound for the problem. In the legend of Figure 3.4, there are the granularities of the state, decision, and price states of the optimisation problems (through the solution of which we obtained our value functions). In Figure 3.4, the granularity of prices (GranP) is the same for both pairs, but the granularity of the storage level (GranL) and the granularity of the decision space (GranX) are different. (Note that GranX is defined only for the methodology of linear interpolation. When we use Benders cuts, the action is determined by the solution for a linear program.). The gap between two lines is thinner for higher GranL and GranX (the green lines). In our problems, the transition matrix is refers only to the environmental state. According to our results, GranP may not affect the width of gap, but it may shift the whole gap. In Figure 3.5, there are two pairs of lines where the granularity of the state and decision space is the same, but the granularity of prices is different. Note that when GranP equals five, the corresponding gap is almost above that of the value function, with GranP equal to fifteen. This is because for lower values of GranP, non-zero elements of the corresponding transition matrices are substantially larger than that of those with larger values of GranP, there are no absorption states. Hence, the Markov chain will visit the minimum value (favourable for the purchase) and the maximum value (favourable for the sale) more frequently, and thus the lower values of GranP yield upper bounds of the storage. Therefore, the solution for Simple Problem with the transition matrix is twofold:

1. As thin the solution gap as possible
2. Small changes of value function to the increases of GranP.

In the right part of Figure 3.5, there are two pairs of the value function, and we notice that the value function does not change substantially when we increase GranP from 35 to 50. The solution gap is less than one percent for GranL=200 and GranX=50. The value of the value function changes by less than one percent for GranP larger or equal to 60. When we replace the transition matrix with simulations, we calculate the expected values by average values where the number of the simulated observations is 1000.

To summarise, when solving Simple Problem, we use the following parameters:

1. GranL=200,
2. GranX=50,
3. GranP=60.

If, instead of using the transition matrix, we use *Alg1* or *Alg2*, we choose the sample sizes equal to 1000. Otherwise, the corresponding granularities remain the same.

3.2.2 Stable AR – Simple Problem

Let us consider the auto-regressive price process p_t with the parameter $a_0 > 0$ and the stable innovation $\varepsilon_t \sim S_\alpha(\sigma, 0, \mu)$ where $t \in \mathbb{N}$ and

$$p_t = a_0 \cdot p_{t-1} + \varepsilon_t.$$

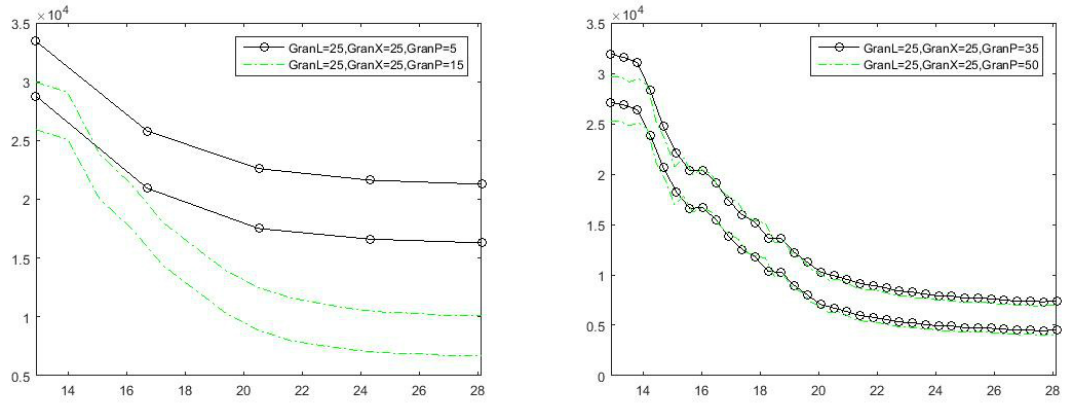


Figure 3.5: Comparison of linear interpolation and Benders cuts II.

The objective function is given as follows

$$\max \mathbb{E}\left(\sum_{t=1}^{T} \min(x_t(p_t - c^w), x_t(p_t + c^i))\right) \tag{3.14}$$
s.t.
$i \leq x_t \leq w,$
$0 \leq L_t \leq C,$
$L_t = L_{t-1} - x_t,$
$p_t = a_0 \cdot p_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim S_\alpha(\sigma, \beta, \mu),$
$t = 1, .., T, \quad L_0 = 0.$

By optimality principle [79], Problem (3.14), can be solved by using the value function for $t = 1, .., T$:

$$\begin{aligned} V_t(L, p_t) = \quad & \max \theta^* + \mathbb{E}\left(V_{t+1}(L - x_t, p_{t+1})|p_t\right) \\ & \text{s.t.} \\ & i \leq x_t \leq w, \\ & 0 \leq L_t \leq C, \\ & L_t = L_{t-1} - x_t, \\ & p_{t+1} = a_0 \cdot p_t + \varepsilon_t, \quad \varepsilon_t \sim S_\alpha(\sigma, \beta, \mu), \\ & \theta^* \leq x_t(p_t + c^i), \\ & \theta^* \leq x_t(p_t - c^w), \end{aligned} \tag{3.15}$$

and $V_{T+1}(\cdot, \cdot) \equiv 0$.

3.2.3 SDP without Resorting to the Transition Matrix

We solve Problem (3.15) by using the post decision value function which we define in the following way

$$\begin{aligned} & V_{t+1}^x(L, p) = \mathbb{E}(V_{t+1}(L, a_0 \cdot p + \varepsilon_t)|p), \\ & \varepsilon_t \sim S_\alpha(\sigma, \beta, \mu). \end{aligned}$$

If we have the representation of $V_t^x(L,p)$ in the following form

$$V_t^x(L,p) = \min_{(a,b)} \left(a \cdot L + b | (a,b) \in \mathcal{V}_t^x(p)\right),$$

then by denoting $\mathcal{V}_t(p)$ as follows

$$\mathcal{V}_t(p) = \{(a_m^{(p)}, b_m^{(p)}) \quad | \quad m = 1, ..., m^*\},$$

where m^* is the cardinality of $\mathcal{V}_t^x(\cdot, p)$, we can rewrite Problem (3.14) in a pure linear form

$$\begin{aligned} V_t(L_{t-1}, p_t) = \quad & \max \theta^* + \theta_{t+1} \\ & \text{s.t.} \\ & i \leq x_t \leq w, \\ & 0 \leq L_t \leq C, \\ & L_t = L_{t-1} - x_t, \\ & \theta^* \leq x_t(p_t + c^i), \\ & \theta^* \leq x_t(p_t - c^w), \\ & \theta_{t+1} \leq a_m^{(p_t)} L_t + b_m^{(p_t)}, \quad m = 1, ..., m^*. \end{aligned} \tag{3.16}$$

The dual solution that corresponds to constraint $L_t = L_{t-1} - x_t$ determines the slope of the function $V_t(\cdot, p)$ at point L_{t-1} [105, 74]. We denote this slope as a.

When we get an estimate of the value function $V_t(\cdot, \cdot)$, we have to use it to estimate $V_t^x(\cdot, \cdot)$. The details of this operation are described in algorithms *Alg1* and *Alg2*. Before we start describing these algorithms, we have to introduce our notation.

1. Grid$([k,l], n)$ with $l > k$ denoted the division of the interval $[k,l]$ into n equidistant parts, i.e. Grid$([k,l], n) = \{k + v \cdot (l-k)/n, \quad v = 0, .., n\}$.

2. We run a Monte Carlo simulation of the auto-regressive sample AR(1) with the length M and $p_{(M)}$ and $p^{(M)}$ are the minimum and the maximum value in the sample respectively.

In both *Alg1* and *Alg2* the value functions $V_t(L,p)$ are continuous in L and p while the post-decision value functions $V_t^x(L,p)$ are discrete in p. The values between the points in the grid of the storage level are calculated by linear interpolation in *Alg1* and Benders cuts in *Alg2*.

Auto-Regression and Transition Matrix

If the price is auto-regressive i.e. $p_t = a_0 \cdot p_{t-1} + \varepsilon_t$, with $t \in \mathbb{N}$ and Grid$([p_{(M)}, p^{(M)}], M)$ being its grid, and $\varepsilon_t \sim S_\alpha(\sigma, 0, \mu)$, then

$$p_{t+1} = a_0 \cdot p_t + \varepsilon_t \sim S_\alpha(\sigma, 0, \mu + a_0 \cdot p_t).$$

If we use a discretised version of $S_\alpha(\sigma, 0, \mu + a_0 \cdot p_t)$, we will be able to compute all of the transition probabilities $\mathbb{P}(p_{t+1}^{(j)}|p_t)$, $j = 1, .., M$, but for any finite M, the discrete stable distribution will be light-tailed, while the stable distribution is heavy-tailed. So the lower α, the higher M we need to approximate the stable distribution. We choose M as follows

$$M = \min\{M|\mathrm{d}_W(\mathbb{P}_M, \mathbb{P}) \leq \epsilon^*\}, \text{ for a predefined } \epsilon^* > 0$$

with $\mathbb{P}$ and $\mathbb{P}_M$ (distribution with M distinct values) being the empirical and discrete distributions respectively, and then apply *Linearisation Algorithm*.

3.2.4 Monte Carlo Methods

By using the law of large numbers, we get:

$$V_{t+1}^x(L_{t-1}, p_t) = \mathbb{E}\left(V_{t+1}\left(L, p_{t+1}\right)|p_t\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} V_{t+1}\left(L, p_{t+1}^j\right), \quad (3.17)$$

where p_{t+1}^j are the simulated values of prices at time $t+1$, given that at time t the price is p_t.
If we plug this expression for $V_{t+1}^x(L, p)$ in the form of the average value into the algorithms

1. *Linear Interpolation*,
2. *Linearisation*,

then we can calculate the value functions, but the former algorithm will no longer yield a lower bound and the latter algorithm will not yield an upper bound. This is because average values may substantially differ from the corresponding expected values. However, by the Law of Large Numbers this approach converges to the true value function. In Figure 3.6 these value functions are obtained by Linear Interpolation with Monte Carlo and *Alg1* where we can see that the Monte Carlo value function is jagged In the picture on the left there are two value functions one of which is obtained by Monte Carlo (the green jagged line) and another one is obtained by using transition matrix. The x-axis on the left picture is the storage level at time $t = 1$. The reservoir is R1 so at time $t = 1$, the storage level takes the value from 0 to 8750 m^3. This jaggedness affects the Bellman sums which results in their sub-optimal value what we can see in the picture on the right of Figure 3.6 where we show the value function at time 1 in the initial price. In the right part of Figure 3.6 we show the value functions obtained by the transition matrix and Monte Carlo. Post-decision yields the value function which differs from the original value function by less than a fraction of a percent. However, the Bellman sums obtained by the post-decision approach lie below the value functions.

Algorithm Alg1

1. $V_T(L,p) = (p - c^w) \cdot \min(w, L)$
2. **for** $L \in \mathcal{L}_T$ to C
   $V_T^x(L,p) = \int_{-\infty}^{\infty} V_T(L, a_0 \cdot p + y)\phi(y)dy$
   (where $\phi(\cdot)$ is the density function of ε)
   end for
3. **for** $t = T-1$ **downTo** 1
   for $p \in \text{Grid}([p_{(M)}, p^{(M)}], M)$
   for $L \in \mathcal{L}_t$
   $V_t(L,p) = \max\left\{R_t(L,p,x) + V_{t+1}^x(L-x,p)\right\}$
   end for
   end for
3a. **if** $p > p^{(M)}$
   $\text{policy}_t(L,p) = \text{policy}_t(L,p^{(m)}) = x^*$,
   $V_t(L,p) = (p - c^w) \cdot x^* + V_{t+1}^x(L - x^*, p)$
   else if $p < p_{(M)}$
   $\text{policy}_t(L,p) = \text{policy}_t(L,p_{(M)}) = x^*$,
   $V_t(L,p) = (p + c^i) \cdot x^* + V_{t+1}^x(L - x^*, p)$
   end if
3b. **for** $j = 1, .., M-1$, the values $V_t(L,p)$ for $p \in (p_j, p_{j+1})$
   and for all $L \in \mathcal{L}_t$ are linearly interpolated.
   end for
3c. **for** $p \in \text{Grid}([p_{(M)}, p^{(M)}], M)$
   for $L \in \mathcal{L}_t$
   $V_t^x(L,p) = \int_{-\infty}^{\infty} V_t(L, a_0 \cdot p + y)\phi(y)dy$
   The values of the post decision function $V_t^x(\cdot,p)$
   in points L which are outside $\mathcal{L}_t$ are approximated
   by linear interpolation.
   end for
   end for
3d. **for** $j = 1, .., M-1$, the values $V_t^x(L,p)$ **for** $p \in (p_j, p_{j+1})$
   and **for** all $L \in \mathcal{L}_t$ are linearly interpolated.
   end for
   end for

Table 3.5: Algorithm *Alg1* which provides a lower bound.

Algorithm Alg2

I. 1. (The same step as in *Alg1*)

II. 2. (The same step as in *Alg1*)

III. **for** $t = T - 1$ **downTo** 1

$\mathcal{V}_t(\cdot, p) = \emptyset$

for $p \in \text{Grid}([p_{(M)}, p^{(M)}], M)$

for $L \in \mathcal{L}_t$

$V_t(L, p) = \max\left\{R_t(L, p, x) + V^x_{t+1}(L - x, p)\right\}$

$x^* = \arg\max\left\{R_t(L, p, x) + V^x_{t+1}(L - x, p)\right\}$

and add the cut $(a, p^* - a \cdot L)$ to $\mathcal{V}_t(\cdot, p)$

end for

end for

$V_t(L, p) = \min\left(a \cdot L + b \quad | \quad (a, b) \in \mathcal{V}_t(\cdot, p)\right)$

IV. 3a. (The same step as in *Alg1*)

V. **for** $p \in \text{Grid}([p_{(M)}, p^{(M)}], M)$

$\mathcal{V}^x_t(\cdot, p) = \emptyset$

for $L \in \mathcal{L}_t$

$v^x_t(L, p) = \int_{-\infty}^{\infty} V_t(L, a_0 \cdot p + y)\phi(y)dy$

Find δ such that

$\frac{v^x_t(L,p) - v^x_t(L-\delta,p)}{\delta} = \frac{v^x_t(L+\delta,p) - v^x_t(L,p)}{\delta} = a,$

and add the cut $(a, v^x_t(L, p) - a \cdot L)$ to $\mathcal{V}^x_t(\cdot, p)$

end for

end for

$V^x_t(L, p) = \min\left(a \cdot L + b \quad | \quad (a, b) \in \mathcal{V}^x_t(\cdot, p)\right)$

end for

Table 3.6: Algorithm *Alg2* which provides an upper bound.

3.2.5 Stochastic Dominance Constraints

A simplified version of Problem (3.14) where $c^i = c^w = 0$, reads as follows:

$$\begin{aligned} &\max \mathbb{E}\left(\sum_{t=1}^{T} x_t p_t\right) \\ &\text{s.t.} \\ &i \leq x_t \leq w, \quad 0 \leq L_t \leq C, \quad L_t = L_{t-1} - x_t \\ &p_t = a_0 \cdot p_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim S_\alpha(\sigma, \beta, \mu), \\ &t = 1., .., T, \quad L_0 = 0. \end{aligned} \tag{3.18}$$

If in Problem (3.14) we do not know the price which we get for the bidden volume then by optimality principle [79], this can be solved by using the value function for $t = 1, .., T$ as follows:

$$\begin{aligned} &V_t(L_{t-1}, p_t) = \max\left\{\mathbb{E}(p_{t+1}x_t + V_{t+1}(L_t, p_{t+1}))|p_t)\right\} \\ &\text{s.t.} \\ &i \leq x_t \leq w, \quad 0 \leq L_t \leq C, \quad L_t = L_{t-1} - x_t \\ &p_{t+1} = a_0 \cdot p_t + \varepsilon, \quad \varepsilon \sim S_\alpha(\sigma, \beta, \mu), \end{aligned} \tag{3.19}$$

Let us assume that the noise is symmetric and centralised, i.e. $p_{t+1} = a_0 p_t + \varepsilon$ and $\varepsilon \sim S_\alpha(\sigma, 0, 0)$, $\alpha \in (1, 2], \sigma > 0$ then (3.19) will be in the form

$$\begin{aligned} &V_t(L_{t-1}, p_t) = \max\left\{a_0 p_t x_t + \mathbb{E}\left(V_{t+1}(L_t, p_{t+1})|p_t\right)\right\} \\ &\text{s.t.} \\ &i \leq x_t \leq w, \quad 0 \leq L_t \leq C, \quad L_t = L_{t-1} - x_t \end{aligned} \tag{3.20}$$

Theorem 25 (Chapter 2) enables us to include the constraint $p_{t+1}x \geq_{(2)} \varepsilon_Y$ into constraints, where $\varepsilon_Y \sim S_\alpha(\sigma_Y, 0, \mu_Y)$. Using Theorem 5 (Chapter 1) and Theorem 25, we infer that the constraint $p_{t+1}x_t \geq_{(2)} \varepsilon_Y$ implies: $a_0 p_t x_t \geq \mu_Y$ and $\sigma \cdot x_t \leq \sigma_Y$, which enables us to rewrite Problem (3.20) with second-order stochastic dominance constraint in the following form:

$$\begin{aligned} &V_t(L_{t-1}, p_t) = \max\left\{a_0 p_t x_t + \mathbb{E}\left(V_{t+1}(L_t, p_{t+1})|p_t\right)\right\} \\ &\text{s.t.} \\ &i \leq x_t \leq w, \quad 0 \leq L_t \leq C, \quad L_t = L_{t-1} - x_t \\ &a_0 p_t x_t \geq \mu_Y, \\ &\max\left(-\frac{\sigma_Y}{\sigma}, i\right) \leq x_t \leq \min\left(\frac{\sigma_Y}{\sigma}, w\right), \end{aligned} \tag{3.21}$$

which, when it is feasible, is a problem with purely linear constrains and a piecewise linear objective, i.e. for the solution of this problem we can use e.g. *Alg1*, *Alg2*, or *spot ADDP algorithm* (Table 3.4).

A Note on the Complexity of Gas Storage Valuation problems

The complexity of some problems can be such that, in practical applications, we have to solve them by means of several methods, and aim to get a similar result from each. This is because when we utilise only one method, it is

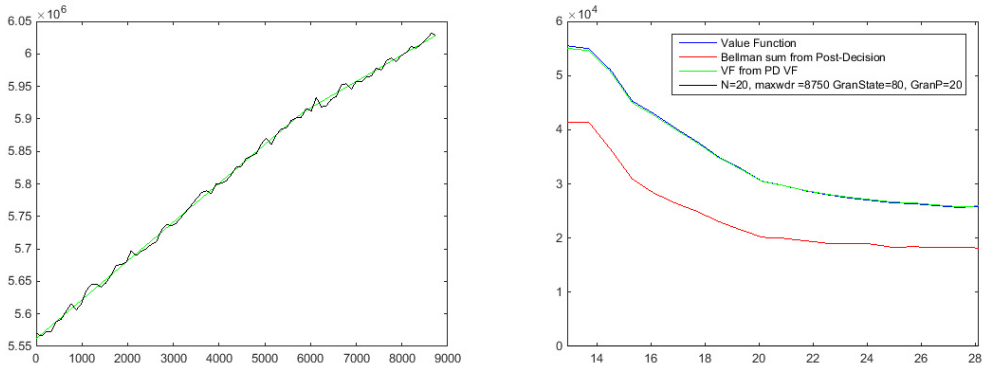


Figure 3.6: The value functions at time t obtained by *Linear Interpolation* (Green Line) and *Linear Interpolation* by Monte Carlo (Black Line). The x-axis in the picture on the left is the storage level. The x-axis on the picture on the right is the initial gas price.

possible to make a significant mistake, which would not be obvious. Realistic versions of the gas storage valuation problem belong to this class. That is why we solve it by means of two different methods. When we solve *Simple Problem*, we can use linear interpolation methodology; however, in *Advanced Problem I* and *Advanced Problem II*, we transform this methodology to obtaining Benders cuts through linear interpolation. This is due to the fact that otherwise, we face the curse of dimensionality. As for linear interpolation applied on *Simple Problem*, we can use GPUarrays within Matlab. As for Benders cuts, GPU-calculations are a more challenging task, because the Matlab built-in function *linprog* within *Optimization Toolbox* is not defined for GPUarray inputs (See http://www.mathworks.com/help/optim/ug/linprog.html).
Other factors which increase the complexity of the problem are the spot and forward prices' dependence on one another, and the fact that some gas storage units are connected to several markets. We address these features within *Advanced Problem I* and *Advanced Problem II*. We also obtain the solution by means of two methods, which at worst differ by one percent. Instead of presenting two numbers, we present their average, for the sake of brevity.

3.2.6 Gas Storage Valuation via Model Predictive Control

Another approach which we can use for gas storage valuation is *Model Predictive Control* (MPC) which is a different methodology than SDP, but is also used to address multi-stage problems. Within MPC, we make a prediction and get a vector of decisions based on this prediction. The length of this prediction depends on the prediction horizon that we select. The length of our prediction horizon equals the time that remains to the end; i.e. at the first stage, it is $T = 250$. So at the first stage, we get a deterministic problem, and if q_t, $t = 1, 2, ..., T$ is a single realisation of prices within our prediction, then the problem will be in the

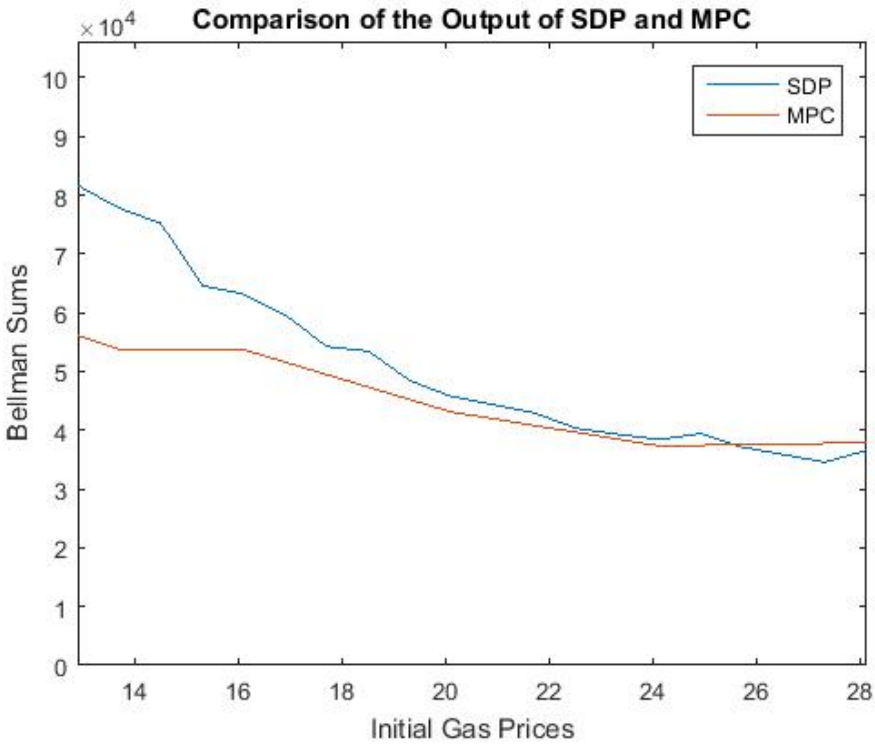


Figure 3.7: The Bellman sums for R1 with zero injection and withdrawal costs versus the average output of MPC for 50 scenarios.

following shape:

$$\max \sum_{t=1}^{T} q_t x_t$$
$$\text{s.t.}$$
$$L_t = L_{t-1} - x_t, \quad 0 \leq L_t \leq C, \quad t = 1, ..., T$$
$$i \leq x_t \leq w, \quad t = 1, ..., T.$$

Taking into account that for $t = 1, 2, ..., T$ $L_t = L_0 - \sum_{k=1}^{t} x_t$, we get a purely linear problem, which we solve by means of the simplex methodology. When we get the series of decisions $x_1, x_2, ..., x_T$, we take only the first decision x_1. Following on from this, when we get the price at time $t = 2$, we revise our predictions of prices based on this new information. On the basis of a new prediction, we create a problem analogous to that at time $t = 1$; however this time, the prediction horizon is $T - 1$. Again we select only the first decision from the $(T - 1)$-long series of decisions, and we continue so until $t = T$, inclusively. Taking only the first decision into account, and revising predictions enables us to get feedback [98, 5]. This approach gives us similar results to SDP, which are summarised in Figure 3.7. We solved the optimisation of R1 with zero injection and withdrawal costs. We compare Bellman sums of SDP and the average composite reward of 50 scenarios yielded by MPC. The source of randomness is the transition matrix. At some regions in Figure 3.7 the line yielded by MPC lies above the line of SDP. We explain this by taking the average of just 50 scenarios.

3.3 Advanced Problem

In the literature on a gas storage unit valuation, most attention is paid to trading in the American market Henry Hub, because this is the most sophisticated market. However, European markets differ from the North-American market in many ways. One such way is that, in Europe, some gas storage units are connected to several cross-border markets, and it is technically possible to inject gas from one market and, at the same time, to withdraw it from another market. The examples are as follows: 7Fields, Nafta, or Pozagas [107, 108, 109]. This enables traders to choose the highest price when they sell and the lowest prices when they buy. In this situation, utilising only one dimension of the environmental variable leads to sub-optimal solutions. In such an environment, in order to find the value of storage, we have to find the prices' inter-dependencies, and then to solve a problem with additional price variables. Let us assume that we have J spot markets. Then by taking into account the technical parameters of a gas storage, its state, and the current market prices, we can formulate our problem with the goal being to profit from trading in the expected value (or in another mathematical operator) as much as possible. Analogous situations exist for hydro storage power plants, but European markets are already coupled. This means that in most cases the prices on them are the same, although in the instances of congestion or decoupling it is possible to implement an analogous approach as in the case of a gas storage unit. In both *Advanced Problem I* and *Advanced Problem II*, we resort to post-decision approach for stochastic dynamic programming.

3.3.1 The Problem Formulation

We assume that no single injection would get outside the bounds and their sum as well; i.e. if the storage is connected to J hubs and x_j is our operation of j-s market, and w and i are maximum injection and withdrawal respectively, then

$$\begin{aligned} & i \leq \sum_{j=1}^{J} x^{(j)} \leq w, \\ & i \leq x^{(1)} \leq w, \\ & ..., \\ & i \leq x^{(J)} \leq w. \end{aligned}$$

Hence, under the assumption that the withdrawal and injection costs are zero, we get the following Bellman equation at time t

$$\begin{aligned} & V_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left(\mathbf{p}_t^T \mathbf{x} + \mathbb{E}\left(V_{t+1}(L - 1^T \mathbf{x}, \mathbf{p}_{t+1}) | \mathbf{p}_t \right) \right) \\ & \text{and if the terminal time is } T \text{ then} \\ & V_{T+1}(\cdot, \cdot) \equiv 0. \end{aligned} \tag{3.22}$$

where L is the storage level, $\mathbf{p}_t = (p_t^{(1)}, .., p_t^{(J)})^T$ are prices at J markets at time t, $\mathbf{p}_{t+1} = (p_{t+1}^{(1)}, .., p_{t+1}^{(J)})^T$ are prices at J markets at time $t+1$, and $\mathbf{x} = (x^{(1)}, .., x^{(J)})^T$

are our actions and $\mathcal{X}_t$ is the following set of constraints

$$\begin{aligned} i \leq x_j \leq w, \quad & j = 1, .., J \\ i \leq 1^T \mathbf{x} \leq w, \quad & \\ C_0 \leq L - 1^T \mathbf{x}, \quad & C_0 \text{ is the cushion volume,} \\ L - 1^T \mathbf{x} \leq C_1 \quad & C_1 \text{ is the capacity.} \end{aligned}$$

Market Data

We use the price data from the following three markets to address *Advanced Problem I*:

1. NCG which is NetConnect Germany
   www.net-connect-germany.de.
2. TTF which is Title Transfer Facility
   www.gasunietransportservices.nl/en/about-gts/gastransport/ttf
3. PEGN which is Gas exchange point - North.
   http://selectra.info/Difference-entre-PEG-Nord-et-PEG-Sud.html

The Solution to Problem (3.22)

Problem (3.22) can be solved by linearising and applying Benders cuts. However the usage of the transition matrix is limited by the number of markets. Therefore, we resort to using Monte Carlo simulations to calculate the post-decision value function

$$V_{t+1}^x(L, \mathbf{p}_t) = \mathbb{E}\left(V_{t+1}(L, \mathbf{p}_{t+1}) | \mathbf{p}_t\right).$$

If we approximate the post-decision value function $V_{t+1}^x(L, \mathbf{p}_t)$ with the following average, i.e.

$$V_{t+1}^x(L, \mathbf{p}_t) \approx V_{t+1}^N(L, \mathbf{p}_t) = \frac{1}{N} \sum_{n=1}^{N} V_{t+1}\left(L, \mathbf{p}_{t+1}^{(n)}\right)$$

where $\mathbf{p}_{t+1}^{(n)}, n = 1, .., N$ are simulated samples of the underlying random process at time t conditional on that the prices at time t are $\mathbf{p}_t$, then the resulting value function no longer has to be the upper bound of the true value function. This is because the average slope might deviate significantly from its true value, which in turn affects the intercept. This, obviously, can result the value function being under the expected value of the Bellman sums. In addition to this, this deviation at some time $t \in \{1, 2, ..., T\}$ might be projected to the time preceding t within the backward path of the SDP. When $J \leq 2$, we use a linear interpolation Benchmark. If we calculate the value function at several points, and approximate its value at other points by linear interpolation, then we will obviously obtain a lower bound of the value function. However for $J > 2$ the calculation of the value function by linear interpolation is hindered by the curse of dimensionality. So when $J > 2$, we solve Problem (3.22) by means of linearisation and Benders cuts and, in parallel, use several suboptimal solutions as benchmarks. These are described in the following section.

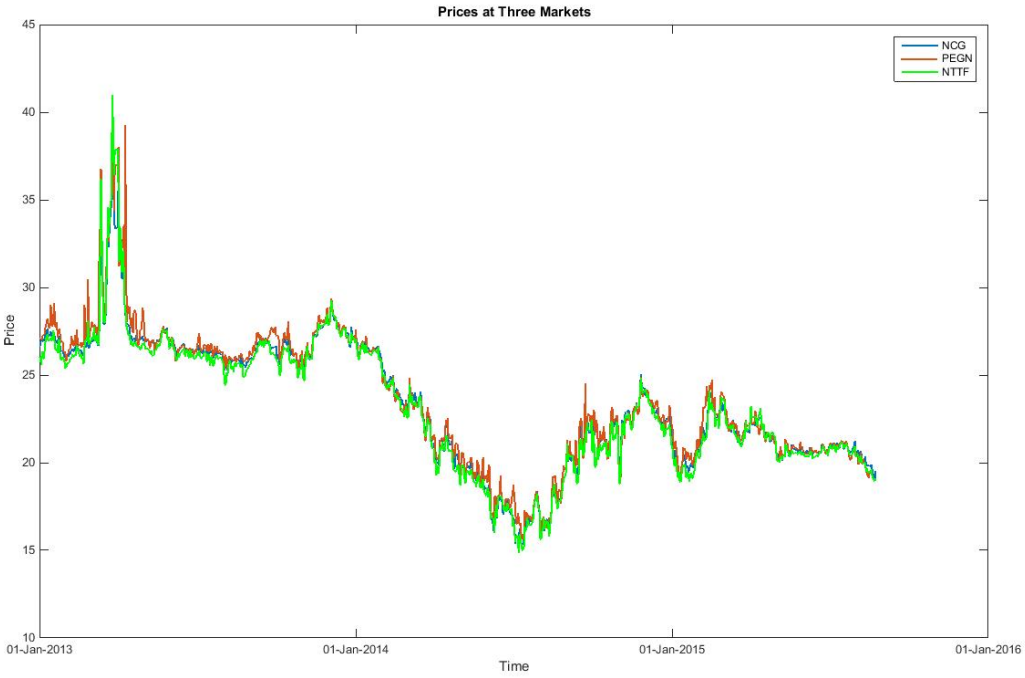


Figure 3.8: Spot prices at three markets.

3.3.2 Solutions in One Dimension

The incorporation of the connection of a gas storage unit to several markets into SDP results in additional dimensions of the environmental variable, therefore one of approximations that we consider are separate analyses of markets. By the analysis of historical spot prices for every market separately, we can infer their transition matrices and calculate the value and policy function:

$$\begin{aligned} V_t^{(j)}(L,p), \qquad & j=1,..,J, \\ \text{policy}_t^{(j)}(L,p), \qquad & j=1,..,J. \end{aligned} \tag{3.23}$$

After getting these functions, we simulate prices of the products and round their values by the closest values from the price grid. The rounded value is used as an input for the policy function. We present the following two strategies:

3.3.3 Heuristic 1.

For j-th policy function and at iteration $k \in \mathbb{N}$, we perform:

- Our initial storage level is $L_0 = 0$ and the price is $(p_1^{(1)}, ..., p_1^{(J)})^T$.
- At time $t = 1$, we take J price instances for each market $\mathbf{p}_1 = (p_1^{(1)}, .., p_1^{(J)})^T$ and choose $p_1^{(j)}$ as the input into the policy function $\text{policy}_1^{(j)}(0, \cdot)$.
- At time $t = 1$, our decision is $x_1 = \text{policy}_t^{(j)}(L_0, p_1^{(j)})$, and if $x_1 \geq 0$ then we withdraw in other words sell, so we get $\text{Reward}_1 = (\max(\mathbf{p}_1) - c^w) \cdot x_1$ and if $x_1 < 0$ then we inject, i.e. buy, so we choose the lowest price and we pay $\text{Reward}_1 = (\min(\mathbf{p}_1) + c^i) \cdot x_1$.

- $L_1 = L_0 - x_1$, $t = t + 1$.
- while $t \leq T$ do
  1. We simulate prices, $\mathbf{p}_t = (p_t^{(1)}, .., p_t^{(J)})^T$,
  2. $x_t = \text{policy}_t^{(j)} \left(L_{t-1}, p_t^{(j)} \right)$,
  3. $$\text{Reward}_t = \begin{cases} (\max(\mathbf{p}_t) - c^w) \cdot x_t, & x_t \geq 0, \\ (\min(\mathbf{p}_t) + c^i) \cdot x_t, & x_t < 0, \end{cases}$$
  4. $t := t + 1$,

  End while
- The composite profit for iteration k and the j-s policy function is $\mathbf{Reward}_{(j,k)} = \sum_{t=1}^{T} \text{Reward}_t$.

If the number of all iterations is K then we approximate the value function as follows

$$V_1^{(j)}(L_0, p) = \frac{1}{K} \sum_{k=1}^{K} \mathbf{Reward}_{(j,k)}$$

and hence the final approximation of $V_1(L_0, \mathbf{p})$ is

$$V_1(L_0, \mathbf{p}_1) \approx V_1^*(L_0, \mathbf{p}_1) = \max_{j=1,..,J} V_1^{(j)}(L_0, p^{(j)})$$

and

$$V_1(L_0, \mathbf{p}_1) \geq V_1^*(L_0, \mathbf{p}_1).$$

Let us note that the function $V_1^*(\cdot, \mathbf{p}_1)$ does not have to be concave in the storage level because the concavity of functions (concave functions have convex hypographs) is not preserved under maximizing, i.e. choosing a function out of a list with the maximum value (the hypograph of the maximum of concave functions is given as the union of the hypographs of the corresponding function which does not have to be a convex set [84]). Hence if we denote the function whose hypograph is given as the convex hull of the hypograph of the function $V_1^*(\cdot, \mathbf{p}_1)$ as $V_1^{**}(\cdot, \mathbf{p}_1)$, then

$$V_1(L_0, \mathbf{p}_1) \geq V_1^{**}(L_0, \mathbf{p}_1).$$

3.3.4 Important Remarks

For low price granularities in Problem (3.22), we usually get even lower values of value functions than those provided by *Heuristic 1.* Therefore, the results provided by these sub-optimal strategies can be used as guides to determine whether we should increase the granularity however practical examples the requirements of the granularity may be so strong, that the solving of problem (3.22) intractable.

3.3.5 Note on Dimensionality of the Price Process

Let us note that auto-regression is a simplification of the problem because the real dimensionality the price process is higher than one. For example, each spot price can be modelled by a multi-factor process where one of factors is a multi-factor forward price process. If we couple this feature with the fact that the number of the spot prices is larger than one then we get to the curse of dimensionality in the environmental state space. In order to handle this problem, we have to resort to techniques of dimensionality reduction, such as principal component analysis, approximations, and iterations. In our approach towards dimensionality reduction in environmental state space, we assume that spread processes are low-dimensional and mean-reverting, while the price process on one of markets can be multi-dimensional. In other words, one price is modelled by a sophisticated process while the rest of the prices are modelled as the sum of this price and the corresponding spread process. This approach is based on our empirical observation that all the spreads of prices (the prices in Figure 3.8) at NCG, PEGN, and TTF are mean reverting by means of the *Augmented Dickey-Fuller Test*. In addition to this, the mean reversion assumption is justified from an economic point of view: if these prices had not been mean reverting, then there would have been arbitrage opportunities. We make a simplifying assumption that the spread processes follow an auto-regressive process:

$$V_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left(\mathbf{p}_t^T \mathbf{x} + \mathbb{E}\left(V_{t+1}(L - 1^T \mathbf{x}, \mathbf{p}_{t+1}) | \mathbf{p}_t \right) \right) \tag{3.24}$$
$$\text{and if the terminal time is } T \text{ then}$$
$$V_{T+1}(\cdot, \cdot) \equiv 0.$$

where $\mathbf{p}_t = \left(p_t^{(1)}, p_t^{(1)} + \text{spread}_t^{(1)}, ..., p_t^{(1)} + \text{spread}_t^{(J-1)} \right)^T$ and $\text{spread}_t^{(i)}$, $i = 1, .., J$ is a low-dimensional spread process.

3.3.6 Realistic Features of Markets

Let us note that in reality at day ahead markets the day ahead price is not known, i.e. if we use the last spot price p_t and the current storage level L as inputs into the policy function and get the action $x = \text{policy}(L, p_t)$, then the income for time t is $p_{t+1} \cdot x$. This feature destroys the arbitrage and we denote the value function which incorporates it, as $\mathbb{V}(\cdot, \cdot)$, i.e.

$$V_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left(\mathbb{E}\left(\mathbf{x}^T \mathbf{p}_t + V_{t+1}(L - 1^T \mathbf{x}, \mathbf{p}_{t+1}) | \mathbf{p}_t \right) \right) \tag{3.25}$$
$$\text{vs}$$
$$\mathbb{V}_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left(\mathbb{E}\left(\mathbf{x}^T \mathbf{p}_{t+1} + V_{t+1}(L - 1^T \mathbf{x}, \mathbf{p}_{t+1}) | \mathbf{p}_t \right) \right) \tag{3.26}$$

Equation (3.25) corresponds to our initial conditions while Equation (3.26) corresponds to conditions when the state of real markets when the day ahead prices are not known, however it can be easily shown that these problems are equivalent if the price model has the martingale property. In the price model that we use, we can express Problem (3.26) by an equivalent problem in the form of Problem (3.25).

3.4 Average Strategy

We resort to the average only when the mean reversion rates of the spread processes are extremely high. If the current price at one market is significantly higher than the prices at another market, then it does not have to imply the arbitrage, because the prices at the following step may be different. So, if we cannot expect inertia in prices, then averaging enables us at least to earn more than one would on the worst market. This can be achieved as follows: when our policy function yielded some action x, then implementing x/J on J markets enables us to achieve the average price. In our example we consider a gas storage valuation problem where our asset is connected to three markets. We model each price by means of AR(1) process, whose residual are modelled by $\mathcal{S}$-distribution. Let us assume that the residuals at all of the markets have a common driver ε, and let us consider a triple of bivariate distributions in Table 3.5 with $\varepsilon||\varepsilon^{(1)}||\varepsilon^{(2)}||\varepsilon^{(3)}$, and the residual of the corresponding auto-regressive processes are the second component of the bivariate triples.

3.4.1 Notation for Avanced Problems I and II

Let us denote the parameter r for the residuals of NCG as r_1, the r parameter of PEGN as r_2, and the r-parameter of TTF, as r_3.

3.4.2 Implementation of Average Strategy

If $\bar{z}_t^j$ with

$$j \in \{\text{NCG}, \text{PEGN}, \text{TTF}\}$$

are auto-regressive processes whose coefficient is A and its residuals are the second components of the aforementioned triples, then after some algebra, we get

$$\frac{1}{3}\left(z_{t+1}^{\text{NCG}} + z_{t+1}^{\text{PEGN}} + z_{t+1}^{\text{TTF}}\right) = \frac{A}{3}\left(z_t^{\text{NCG}} + z_t^{\text{PEGN}} + z_t^{\text{TTF}}\right) + \varepsilon_t^{(*)},$$

where

$$\begin{aligned}
&\varepsilon_t^{(*)} \sim S_\alpha(\sigma^{(*)}, 0, 0)\\
&\sigma^{(*)} = \frac{1}{3}\sqrt[\alpha]{(r\sigma)^\alpha + (q_3\sigma_1)^\alpha + (q_2\sigma_2)^\alpha + (q_3\sigma_3)^\alpha}\\
&\text{where}\\
&r = r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3\\
&q_j = 1 - p_j, \text{ where } j = \{1, 2, 3\}.
\end{aligned}$$

$\left(\varepsilon, \quad r_1\varepsilon + (1 - r_1^\alpha)^{1/\alpha}\,\varepsilon^{(1)}\right)$	$\sim$	$\mathcal{S}([\alpha, \beta], [\alpha, \beta_1], r_1)$
$\left(\varepsilon, \quad r_2\varepsilon + (1 - r_2^\alpha)^{1/\alpha}\,\varepsilon^{(2)}\right)$	$\sim$	$\mathcal{S}([\alpha, \beta], [\alpha, \beta_2], r_2)$
$\left(\varepsilon, \quad r_3\varepsilon + (1 - r_3^\alpha)^{1/\alpha}\,\varepsilon^{(3)}\right)$	$\sim$	$\mathcal{S}([\alpha, \beta], [\alpha, \beta_3], r_3)$

Table 3.7: The triple of distributions

And hence the problem of the average strategy takes the form:

$$\begin{aligned}
&\mathbb{V}_t(L, \bar{p}_t) = \max_{\mathbf{x}\in\mathcal{X}_t} \left(\mathbb{E}\left(x\bar{p}_{t+1} + \mathbb{V}_{t+1}(L-x, \bar{p}_{t+1})|\bar{p}_t\right)\right), \\
&i \le x \le w, \\
&0 \le L - x \le C, \\
&\text{where} \\
&\bar{p}_{t+1} = A\bar{p}_t + \varepsilon_t^{(*)},
\end{aligned}$$

which because $\mathbb{E}\varepsilon_t^{(*)} = 0$ is equivalent to

$$\begin{aligned}
&V_t(L, \bar{p}_t) = \max_{\mathbf{x}\in\mathcal{X}_t} \left(x\bar{p}_t A + \mathbb{E}\left(V_{t+1}(L-x, \bar{p}_{t+1})|\bar{p}_t\right)\right), \\
&i \le x \le w, \\
&0 \le L - x \le C, \\
&\text{where} \\
&\bar{p}_{t+1} = A\bar{p}_t + \varepsilon_t^{(*)},
\end{aligned}$$

Approach	ε	$\max_p V_1(0,p)$	$\min_p V_1(0,p)$	$\max_p \mathbb{V}_1(0,p)$	$\min_p \mathbb{V}_1(0,p)$
Direct	Gaus	36648	8497	35930	8259
Direct	sub-G	41863	10435	40839	10046
Direct	$\mathcal{S}$	39652	9526	39057	8725
Heuristic	Gaus	34672	8036	34586	7814
Heuristic	Sub-G	39035	9561	38450	9349
Heuristic	$\mathcal{S}$	37519	8582	36916	8352
Average	Gaus	26377	5007	26150	4126
Average	Sub-G	26525	4825	25450	3755
Average	$\mathcal{S}$	26694	4883	26709	4180
Market1	Gaus	25293	4017	24877	3975
Market1	Stable	26117	4338	25327	4117
Market2	Gaus	25817	4456	25107	4317
Market2	Stable	27225	4490	26406	3778
Market3	Gaus	26548	4335	26298	4378
Market3	Stable	25933	4783	25631	4136

Table 3.8: Comparison of Bellman sums for Advanced Problem I. (R2)

Advanced problem I

We took spot price data from NCG, PEGN, and TTF gas markets for the period from January 1 of 2013 to August 1 of 2015. This data is shown in Figure 3.8. According to the augmented Dickey-Fuller test the spreads of prices possess the mean reversion. Therefore, it is reasonable to model only one price, for example NCG, and assume that the rest of prices are given as the sum of the NCG price

plus an auto-regressive process. In Advanced problem I, we assume that NCG prices are modelled by AR(1), and the difference between NCG and PEGN and NCG and TTF are also modelled by AR(1). Hence if we denote the NCG prices as $p_t^{(1)}$, PEGN as $p_t^{(2)}$, and TTF as $p_t^{(3)}$, then

$$\begin{aligned}
p_t^{(1)} &= a_1 \cdot p_{t-1}^{(1)} + \varepsilon_t^{(1)}, \\
p_t^{(2)} &= p_t^{(1)} + a_2 \cdot (p_{t-1}^{(2)} - p_{t-1}^{(1)}) + \varepsilon_t^{(2)}, \\
p_t^{(3)} &= p_t^{(1)} + a_3 \cdot (p_{t-1}^{(3)} - p_{t-1}^{(1)}) + \varepsilon_t^{(3)}, \\
\varepsilon_t^{(j)} &\sim S_\alpha(\sigma_j, 0, 0), \quad j = 1, 2, 3.
\end{aligned}$$

The linear combinations of prices are as follows:

$$\begin{aligned}
&p_t^{(1)} x_1 + p_t^{(2)} x_2 + p_t^{(3)} x_3 = \\
&= x_1 a_1 p_{t-1}^{(1)} + x_2 \left(a_2 p_{t-1}^{(2)} + (a_1 - a_2) p_{t-1}^{(1)} \right) + x_3 \left(a_3 p_{t-1}^{(3)} + (a_1 - a_3) p_{t-1}^{(1)} \right) + \\
&(x_1 + x_2 + x_3) \varepsilon_t^{(1)} + x_2 \varepsilon_t^{(2)} + x_3 \varepsilon_t^{(3)}
\end{aligned}$$

The conditional expectation is then given as follows:

$$\begin{aligned}
&\mathbb{E}\left(p_t^{(1)} x_1 + p_t^{(2)} x_2 + p_t^{(3)} x_3 | \quad p_{t-1}^{(1)}, p_{t-1}^{(2)}, p_{t-1}^{(3)} \right) = \\
&= x_1 a_1 p_{t-1}^{(1)} + x_2 \left(a_2 p_{t-1}^{(2)} + (a_1 - a_2) p_{t-1}^{(1)} \right) + x_3 \left(a_3 p_{t-1}^{(3)} + (a_1 - a_3) p_{t-1}^{(1)} \right).
\end{aligned}$$

Hence Problem (3.26) for these settings will be

$$\begin{aligned}
&\mathbb{V}_t(L_t, p_t^{(1)}, p_t^{(2)}, p_t^{(3)}) = \\
&\max \left\{ x_1 a_1 p_t^{(1)} + x_2 \left(a_2 p_t^{(2)} + b_1 p_t^{(1)} \right) + x_3 \left(a_3 p_t^{(3)} + b_2 p_t^{(1)} \right) + \right. \\
&\left. + \mathbb{E}\left(\mathbb{V}_{t+1}(L_t, p_{t+1}^{(1)}, p_{t+1}^{(2)}, p_{t+1}^{(3)}) | p_t^{(1)}, p_t^{(2)}, p_t^{(3)} \right) \right\} \qquad (3.27) \\
&\text{s.t.} \\
&L_t = L - x_1 - x_2 - x_3, \\
&0 \le L_t \le C, \\
&i \le x_1 + x_2 + x_3 \le w, \\
&i \le x_j \le w, \quad j = 1, 2, 3, \\
&p_{t+1}^{(1)} = a_1 \cdot p_t^{(1)} + \varepsilon_t^{(1)}, \\
&p_{t+1}^{(2)} = p_{t+1}^{(1)} + a_2 \cdot (p_t^{(2)} - p_t^{(1)}) + \varepsilon_t^{(2)}, \\
&p_{t+1}^{(3)} = p_{t+1}^{(1)} + a_3 \cdot (p_t^{(3)} - p_t^{(1)}) + \varepsilon_t^{(3)}, \\
&\varepsilon_t^{(j)} \sim S_\alpha(\sigma_j, 0, 0), \quad j = 1, 2, 3, \\
&b_1 = (a_1 - a_2), \quad b_2 = (a_1 - a_3).
\end{aligned}$$

Problem (3.25) for these settings is the same except for the objective function which is

$$\begin{aligned} &V_t(L_{t-1}, p_t^{(1)}, p_t^{(2)}, p_t^{(3)}) = \\ &\max\Big\{x_1 p_t^{(1)} + x_2 p_t^{(2)} + x_3 p_t^{(3)} + \\ &+\mathbb{E}\left(V_{t+1}(L_t, p_{t+1}^{(1)}, p_{t+1}^{(2)}, p_{t+1}^{(3)}) | p_t^{(1)}, p_t^{(2)}, p_t^{(3)}\right)\Big\}. \end{aligned} \tag{3.28}$$

The value functions in Problems (3.27) and (3.28) are of the type: $V_t(L, \mathbf{p}) : \mathbb{R}^4 \to \mathbb{R}$, $t = 1, .., T$, which is practically implementable by SDP. Assume that each of the variables of the value function in 3.28 takes 50 possible values, which means that within every time period $t = 1, ..., T$, we calculate $50^4 = 6250000$ values, and for every triple of $50^3 = 125000$ prices, we calculate its system of Benders cuts.

3.4.3 The Summary of Methods

In Table 3.8, there is the summary of methods used for the valuation of a gas storage connected to three markets. We need to explain the notation in Table 3.8:

1. ε states for the innovation in AR(1) process.
2. Gaus means that the innovations are Gaussian.
3. sub-G means that the innovations are sub-Gaussian.
4. Direct means that the calculations of the value function are conducted directly by (3.25) or (3.26).
5. Heuristic states for the method described in Section 3.3.3.
6. Average states for trading the average value on J markets.
7. Market1, Market2, and Market3 refer to NCG, PEGN, and TTF respectively. Market1 means that we operate only on a single market, i.e. market NCG and implement Alg2.
8. $\min_p V_1(0, p)$ and $\min_p \mathbb{V}_1(0, p)$ are the minimum value of the value function of Problems (3.25) and (3.26), respectively in terms of all initial values of prices.
9. $\max_p V_1(0, p)$ and $\max_p \mathbb{V}_1(0, p)$ are the maximum value of the value function of Problems (3.25) and (3.26), respectively in terms of all initial values of prices, each of which has 25 distinct values, i.e. there are 25^3 distinct values of the three initial prices.

In Table 3.11, we summarise our results of the valuation of a gas storage unit which is connected to three markets with $T = 250$, within the parameters of *R1* which can pump and withdraw at the same time. The price process in *Advanced Problem I* is always AR(1). What we see in Table 3.11 is

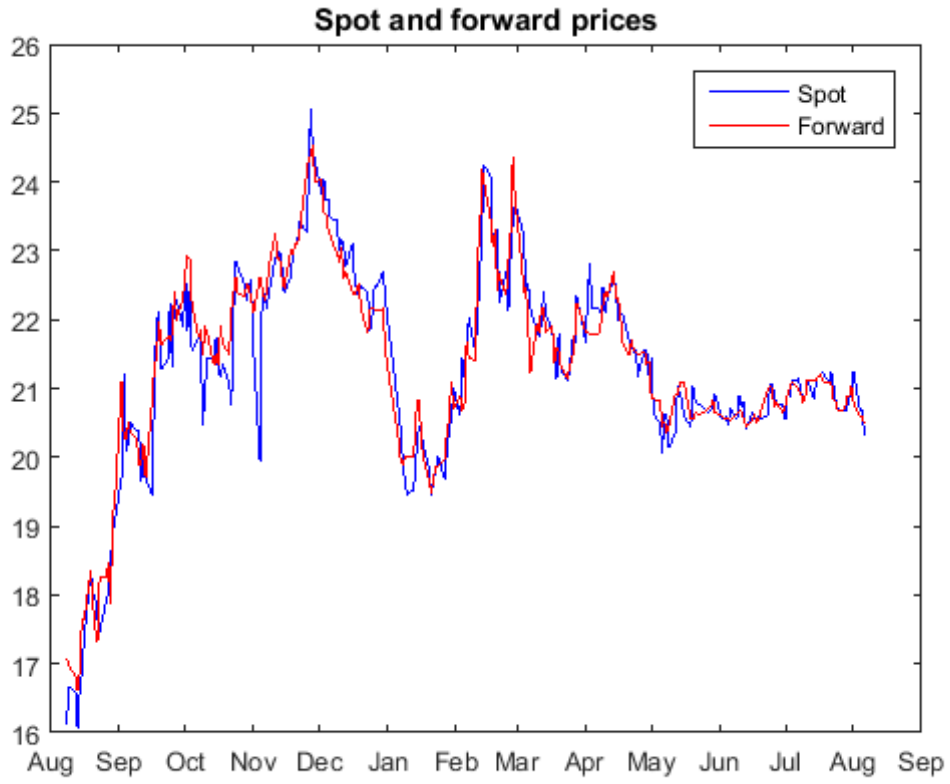


Figure 3.9: NCG spot (blue) and forward (red) prices from August 2014 to August 2015.

1. The importance of taking heavy tails into account.
2. The importance of taking advantage of the connection with other markets.
3. The heuristic yields in all cases more than 90% precision, but it takes by two orders of magnitude shorter (See Table 3.12).

Most of the methods used for the valuation of a gas storage unit are based on the data from Henry Hub, but the American markets are generally more sophisticated than European markets. On the other hand, they lack many features seen in Europe, for example, the connection of gas storage units to several markets. Since AR(1) is not the best choice for the price model, we analyse more sophisticated models in Advanced problem II.

3.5 Advanced Problem II

Figure 3.9 shows spot prices, and the price of the corresponding forward product with the nearest maturity (M1) on NCG. The correlation between the spot prices and forward M1 curve exceeds 99 percent; however the forward curve does not exhibit spikes. In addition to this, according to the augmented Dickey-Fuller test the differences between each of the spot prices and forward M1 prices on NCG, PEGN, TTF also possess the mean reverting property. Therefore each of the spot prices will be modelled as the sum of the forward curve and a spread process which will be AR(1). Let us denote the forward price as y_t, $t = 1, 2, ..., T$.

In Table 3.9, we compare 15 different Gaussian models in term of the Akaike information criterion, the Bayesian information criterion, and the Kolmogorov-Smirnov test of the residuals. The best performance in terms of AIC and BIC criteria is provided by ARMA(1,1)-GARCH(1,1). However, the next best performance is provided by AR(1)-GARCH(1,1), whereby in terms of both AIC and BIC criteria the corresponding numbers differ by less than one percent. In addition to this, this model is simpler, and its choice results in reducing the complexity of the problem therefore, we choose AR(1)-GARCH(1,1) as the model of nearest forward-price model, and hence get:

$$r_t = \ln\left(\frac{y_{t+1}}{y_t}\right),$$

$$\begin{aligned} &r_t = \mu + a \cdot r_{t-1} + \epsilon_t, \\ &\epsilon_t = \sigma_t \cdot \varepsilon_t, \quad \sigma_t^2 = \omega + \gamma \cdot \epsilon_{t-1}^2 + \delta \cdot \sigma_{t-1}^2, \\ &\varepsilon_t \sim N(0,1). \end{aligned}$$

Hence, from the returns, we can deduce the model for prices:

$$y_{t+1} = y_t \exp\left(\mu + a \cdot r_{t-1} + \sqrt{\omega + \gamma \cdot \epsilon_{t-1}^2 + \delta \cdot \sigma_{t-1}^2} \cdot \varepsilon_t\right)$$

Therefore, the models of the spot prices are follows:

$$\begin{aligned} &y_t^{(j)} = y_t + z_t^{(j)}, \\ &\text{where} \\ &z_t^{(j)} \sim \text{ stable } AR(1) \\ &\text{where } j = \{\text{NCG}, \text{PEGN}, \text{TTF}\} \end{aligned}$$

3.5.1 Average Strategy

The mean reversion rates of the spread process for NCG, PEGN, and TTF differ by less than 0.05%. Therefore we round all of them to their mean to facilitate the modelling of the average of all the spreads. Hence, after some applications of

algebra, we get:

$$
\begin{aligned}
&V_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left(\mathbb{E}\left(x(y_{t+1} + \bar{z}_{t+1}) + V_{t+1}(L - x, \mathbf{p}_{t+1}) | \left[\bar{z}_t, r_{t-1} \right] \right) \right) \\
&\text{s.t.} \\
&i \leq x \leq w, \\
&0 \leq L - x \leq C, \\
&\text{where} \\
&\mathbf{p}_t = [y_t, r_{t-1}, \epsilon_{t-1}, \sigma_{t-1}, \bar{z}_t], \\
&\mathbf{p}_{t+1} = [y_{t+1}, r_t, \epsilon_t, \sigma_t, \bar{z}_{t+1}], \\
&\sigma_t^2 = \omega + \gamma \cdot \epsilon_{t-1}^2 + \delta \cdot \sigma_{t-1}^2, \\
&\epsilon_t = \sigma_t \varepsilon_t, \\
&r_t = \mu + r_{t-1} a + \epsilon_t, \\
&\bar{z}_{t+1} = A \bar{z}_t + \sigma \varepsilon_t^{(*)}, \\
&y_{t+1} = y_t \exp\left(r_t \right), \\
&\left(\varepsilon_t, \varepsilon_t^{(*)} \right) \sim \mathcal{S}([2, 0], [\alpha, \beta], r_0).
\end{aligned}
$$

The joint residual has $\mathcal{S}([2, 0], [\alpha, \beta], r_0)$, because the innovation in the forward process is Gaussian. This problem can be reformulated as follows

$$
\begin{aligned}
&V_t(L, \mathbf{p}_t) = \max_{\mathbf{x} \in \mathcal{X}_t} \left\{ x \cdot (y_t q_t + A \bar{z}_t) + \mathbb{E}(V_{t+1}(L - x, \mathbf{p}_{t+1}) | \left[\bar{z}_t, r_{t-1} \right]) \right\} \\
&\text{s.t.} \\
&i \leq x \leq w, \\
&0 \leq L - x \leq C, \\
&\text{where} \\
&\mathbf{p}_t = [y_t, r_{t-1}, \epsilon_{t-1}, \sigma_{t-1}, \bar{z}_t], \\
&\mathbf{p}_{t+1} = [y_{t+1}, r_t, \epsilon_t, \sigma_t, \bar{z}_{t+1}], \\
&q_t = \mathbb{E}\left(\exp(r_t) \right) = \exp\left(\mu + r_{t-1} a + \frac{\sigma_t^2 \sigma_F^2}{2} \right), \\
&\sigma_t^2 = \omega + \gamma \cdot \epsilon_{t-1}^2 + \delta \cdot \sigma_{t-1}^2, \\
&\epsilon_t = \sigma_t \varepsilon_t, \\
&r_t = \mu + r_{t-1} a + \epsilon_t, \\
&\bar{z}_{t+1} = A \bar{z}_t + \sigma \varepsilon_t^{(*)}, \\
&y_{t+1} = y_t \exp\left(r_t \right), \\
&\left(\varepsilon_t, \varepsilon_t^{(*)} \right) \sim \mathcal{S}([2, 0], [\alpha, \beta], r_0).
\end{aligned}
$$

Which because of the convolution features of stable distributions requires only one additional variable to the forward model instead of three variables.

The formulation of Advanced Problem II

In our settings the linear combinations of spot prices are as follows:

$$\begin{aligned}
&x_1(y_{t+1} + z_{t+1}^{(1)}) + x_2(y_{t+1} + z_{t+1}^{(2)}) + x_3(y_{t+1} + z_{t+1}^{3)}) = \\
&= x_1(y_{t+1} + Az_t^{(1)}) + x_2(y_{t+1} + Az_t^{(2)}) + x_3(y_{t+1} + Az_t^{(3)}) + \\
&+x_1\varepsilon_t^{(1)} + x_2\varepsilon_t^{(2)} + x_3\varepsilon_t^{(3)}.
\end{aligned}$$

and hence the problem reads:

$$\begin{aligned}
&V_t(L, \mathbf{p}_t) = \\
&\max\Big\{x_1(y_{t+1} + Az_t^{(1)}) + x_2(y_{t+1} + Az_t^{(2)}) + \\
&+x_3(y_{t+1} + Az_t^{(3)}) + \mathbb{E}\left(V_{t+1}(L_t, \mathbf{p}_{t+1})|\mathbf{p}_t\right)\Big\} \\
&\text{s.t.} \\
&L_t = L - x_1 - x_2 - x_3, \\
&i \le x_j \le w, \quad j = 1, 2, 3, \\
&0 \le L_t \le C, \\
&\text{where} \\
&\mathbf{p}_t = [y_t, r_{t-1}, \epsilon_{t-1}, \sigma_{t-1}, z_t^{(1)}, z_t^{(2)}, z_t^{(3)}], \\
&\mathbf{p}_{t+1} = [y_{t+1}, r_t, \epsilon_t, \sigma_t, z_{t+1}^{(1)}, z_{t+1}^{(2)}, z_{t+1}^{(3)}], \\
&\sigma_t^2 = \omega + \gamma \cdot \epsilon_{t-1}^2 + \delta \cdot \sigma_{t-1}^2, \\
&\epsilon_t = \sigma_t\varepsilon_t, \\
&r_t = \mu + r_{t-1}a + \epsilon_t, \\
&\bar{z}_{t+1} = A\bar{z}_t + \sigma\varepsilon_t^{(*)}, \\
&y_{t+1} = y_t \exp\left(r_t\right), \\
&(\varepsilon_t, \varepsilon_t^{(*)}) \sim \mathcal{S}([2, 0], [\alpha, \beta], r_0).
\end{aligned}$$

which is the problem with 8 dimensions. We solve it by means of heuristic. The main added value of this approach consists in linking spot and forward prices which is crucial in implementing spot-forward strategies.

3.6 Usage of Stochastic Dominance

Apart from stochastic ordering, stochastic dominance can be used in optimisation for handling probability constraints. Let us denote the constraints of Problem (3.25) $\mathcal{X}_t$ for the sake of brevity. So the general practical problem of gas storage valuation reads:

$$\begin{aligned}
V_t(L_{t-1}, \mathbf{p}_t) \qquad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T\mathbf{x} + V_{t+1}\left(L_t, \mathbf{p}_{t+1}\right)|\mathbf{p}_t\right) \\
&\text{s.t.} \\
&(\mathbf{x}, L_t) \in \mathcal{X}_t
\end{aligned}$$

Although, we know our decision $\mathbf{x}$ at time t, we do not know, how much we will earn or lose from it; i.e. the variable $\mathbf{p}_{t+1}^T\mathbf{x}$ is uncertain, and this is the source

Model	AIC	BIC	KS
AR(1)-GARCH(1,1)	163.7740	216.9516	0.4915
ARMA(1,1)-GARCH(1,1)	162.3882	215.5659	0.7073
AR(1)	237.8734	291.0510	0.5707
ARIMA(1,1,0)	246.5014	299.6791	0.3734
ARIMA(1,1,1)	244.4552	297.6329	0.2889
ARIMA(1,1,0)-GARCH(1,1)	246.5014	299.6791	0.3233
MA(1)	318.7770	371.9547	0.6581
MA(1)-GARCH(1,1)	697.5914	750.7692	0.4087
MAX(1)-GARCH(1,1)	697.5012	750.741	0.3371
ARMAX(1,1)-GARCH(1,1)	237.8733	291.0510	0.4672
AR(2)-GARCH(1,1)	325.1485	378.3262	0.4292
AR(2)-GARCH(2,2)	318.8896	372.0673	0.5659
AR(2)	237.8666	291.0442	0.5503
ARMA(2,2)-GARCH(2,2)	324.7371	377.9248	0.5566
ARIMA(4,1,0)	240.4481	293.6257	0.8741

Table 3.9: Comparison of different models.

Approach	ε	$\max_p V_1(0,p)$	$\min_p V_1(0,p)$	$\max_p \mathbb{V}_1(0,p)$	$\min_p \mathbb{V}_1(0,p)$
Heuristic	Sub-G	32175	8299	31862	7674
Heuristic	$\mathcal{S}$	31689	7661	30940	6818
Heuristic	Gaus	31194	7790	30482	6932
Average	Sub-G	24648	3746	23267	3694
Average	$\mathcal{S}$	23710	3822	23345	3685
Average	Gaus	24282	3960	23380	3757
Market1	Gaus	23913	4401	23492	3714
Market1	Stable	24484	4351	24071	3633
Market2	Gaus	24399	4169	24131	3519
Market2	Stable	24563	4024	23499	3515
Market3	Gaus	23727	4284	23477	3709
Market3	Stable	24773	3944	23664	3696

Table 3.10: Comparison of Bellman sums for Advanced Problem II. (R2)

μ	a	ω	γ	δ	σ_F	a_S	σ_S	r
$-4.18 \cdot 10^{-4}$	0.09	$7.59 \cdot 10^{-6}$	0.20	0.80	0.031	0.81	0.13	0.55

Table 3.11: Parameters of AR(1)-GARCH(1,1) for spot and forward prices

of risk. We now aim to mitigate the risk of loss from open positions. When we decide on the size of injection or withdrawal from (or to) different markets, we do not know the tomorrow's prices $\mathbf{p}_{t+1}$ however, it is reasonable to add the constraint that the probability of some predefined loss will be smaller than some predefined value, i.e.:

$$\mathbb{P}\left(\mathbf{p}_{t+1}^T\mathbf{x} \geq \delta\right) \geq p_\delta, \quad p_\delta \in [0,1].$$

We assume that the prices in each market follow a forward model perturbed by a stable AR(1) process, i.e.

$$p_{t+1,j} = y_{t+1} + z_{t+1}^{(j)}$$

for $j = 1, ..., J$. In other words, y_{t+1} is a normal forward model, while $\mathbf{z}_{t+1} = \{z_{t+1}^{(1)}, ..., z_{t+1}^{(J)}\}$ is the source of the price risk.

Let us note that if $z_{t+1}^{(j)} \sim S_\alpha(\sigma_j, 0, \mu_j)$ for $j = 1, 2, ..., J$, then

$$Z_{\mathbf{x}} = \mathbf{z}_{t+1}^T\mathbf{x} = \sum_{j=1}^{J} z_{t+1}^{(j)} x_j \sim S_\alpha\left(\left(\sum_{j=1}^{J}(\sigma_j|x_j|)^\alpha\right)^{1/\alpha}, 0, \sum_{j=1}^{J} x_j\mu_j\right)$$

At every time t, we want to avoid the immediate risks associated with opening positions, so we want the scalar product $\mathbf{p}_{t+1}^T\mathbf{x}$ to be above some predefined loss δ with a predefined probability p_δ. So it is reasonable to add constraint (3.29) into the optimisation problem,i.e.:

$$\begin{aligned} V_t(L_{t-1}, \mathbf{p}_t) \quad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T\mathbf{x} + V_{t+1}\left(L_t, \mathbf{p}_{t+1}\right)|\mathbf{p}_t\right) \\ &\text{s.t.} \\ &(\mathbf{x}, L_t) \in \mathcal{X}_t \\ &\mathbb{P}\left(\mathbf{z}_{t+1}^T\mathbf{x} \geq \delta\right) \geq p_\delta, \quad \forall \delta \in \mathbb{S}, \end{aligned} \tag{3.29}$$

where $\mathbb{S}$ can be either a singleton, a countable finite set or any subset of $\mathbb{R}$ and p_δ is the number in $[0, 1]$. It is reasonable to assume that p_δ is a distribution function of some random variable Y, i.e. $y_\delta = F_Y^*(\delta)$ and let us denote $Z_{\mathbf{x}} = \mathbf{z}_{t+1}^T\mathbf{x}$. Taking this notation into account, we can rewrite constraint (3.29) as follows:

$$F_{Z_{\mathbf{x}}}^*(\delta) \geq F_Y^*(\delta), \quad \forall \delta \in \mathbb{S},$$

and hence the problem can be rewritten

$$\begin{aligned} V_t(L_{t-1}, \mathbf{p}_t) \quad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T\mathbf{x} + V_{t+1}\left(L_t, \mathbf{p}_{t+1}\right)|\mathbf{p}_t\right) \\ &\text{s.t.} \\ &(\mathbf{x}, L_t) \in \mathcal{X}_t \\ &F_{Z_{\mathbf{x}}}(\delta) \leq F_Y(\delta) \quad \forall \delta \in \mathbb{S}, \end{aligned} \tag{3.30}$$

The relaxed versions of the problem with probabilistic constraints are as follows:

$$\begin{aligned} V_t^{(k)}(L_{t-1}, \mathbf{p}_t) \quad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T\mathbf{x} + V_{t+1}^{(k)}\left(L_t, \mathbf{p}_{t+1}\right)|\mathbf{p}_t\right) \\ &\text{s.t.} \\ &(\mathbf{x}, L_t) \in \mathcal{X}_t \\ &F_Z(\delta) \leq_{(k)} F_Y(\delta) \quad \forall \delta \in \mathbb{S}, \end{aligned} \tag{3.31}$$

where $k > 1$. When $k = 2$, we get:

$$\begin{aligned} V_t^{(2)}(L_{t-1}, \mathbf{p}_t) \quad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T \mathbf{x} + V_{t+1}^{(2)}\left(L_t, \mathbf{p}_{t+1}\right) | \mathbf{p}_t\right) \\ &\text{s.t.} \\ &(\mathbf{x}, L_t) \in \mathcal{X}_t \\ &\mathbb{E}\left(\delta - Z_{\mathbf{x}}\right)_+ \leq \mathbb{E}\left(\delta - Y\right)_+ \quad \forall \delta \in \mathbb{S}, \end{aligned} \tag{3.32}$$

By discretising the random variables, we will be able to preserve both the convexity and the linearity of the problem. Therefore, one option is the discrete stable distributions studied by Klebanov and Slamová [95]. However, in the case of general continuous stable distributions, the linearity will be broken, while the convexity will be preserved.

Let us consider a special example:

1. $\mathbb{S} = \mathbb{R}$,
2. $Y \sim S_\alpha(\sigma_Y, 0, 0)$,
3. $\alpha > 1$.

Then by using Theorem 5 (Chapter 1) we get the relaxed problem (i.e. the incorporation of the second-order constraints instead of the first-order constraints) in the following form:

$$\begin{aligned} V_t^{(2)}(L_{t-1}, \mathbf{p}_t) \quad &= \max \mathbb{E}\left(\mathbf{p}_{t+1}^T \mathbf{x} + V_{t+1}^{(2)}\left(L_t, \mathbf{p}_{t+1}\right) | \mathbf{p}_t\right) \\ &\text{s.t.} \\ &(\mathbf{x}, L_t) \in \mathcal{X}_t \\ &\sigma_Y - \sqrt[\alpha]{\sum_{j=1}^{J} (\sigma_j |x_j|)^\alpha} \geq 0 \end{aligned} \tag{3.33}$$

which is a convex problem, however it is not a linear problem. If, however, $Y \sim S_{\alpha_1}(\sigma_Y, 0, 0)$ and $\alpha_1 > \alpha$, then by the empirical Hypothesis 1 (Chapter 2), this second-order stochastic dominance constraint is infeasible.

In Ruszczynski and Dencheva (2009) [94] an analogous example is shown for portfolio optimisation.

3.7 Forward Strategies

In this section, we briefly discuss forward strategies for the valuation of a gas storage unit. Intrinsic and rolling intrinsic valuation yield a risk-free value for a gas storage unit. Therefore, if we take into account that investors are usually risk-averse, then ignoring this opportunity leads to biased results.

3.7.1 Intrinsic Value

Intrinsic value of a gas storage is obtained by solving the following problem

$$\begin{aligned} &\max_{v_{ij}} \sum_{ij} v_{ij} \Delta F_{ij} \qquad (3.34) \\ &\text{s.t.} \\ &v_{ij} \geq 0 \\ &W_j = \sum_i v_{ij} \leq W_{\max} \\ &I_i = \sum_j v_{ij} \leq I_{\max} \\ &C_i = I_i - W_i, \quad 0 \leq C_i \leq C \end{aligned}$$

where ΔF_{ij} is the discounted price spread for injection in month i and withdrawal in month j; v_{ij} is the position in spread ΔF_{ij}; I_i is the total injection at month i; W_j is the total withdrawal in month j; C_i is the storage level at month i which equals total injection minus total withdrawal; C is the storage facility capacity; $I_{\max}$ is the maximum daily injection rate; $W_{\max}$ is the maximum daily withdrawal rate.

3.7.2 Rolling Intrinsic Value

In the objective (3.34) we take into account only the forward curve that was available at the beginning. In practice, this mostly implies trading summer-winder spreads [36]. This implies that we do not profit from a possible change in the forward curve; hence intrinsic valuation yields a suboptimal solution. Grey and Khandewal (2004) [34] proposed using a possible rebalancing of the forward positions when we ascertain that we can get a higher value. This is the core and the bottom line of the rolling intrinsic valuation. It starts with the intrinsic valuation and, whenever the forward curve changes, it recalculates the new intrinsic value. If the difference between the new and old profit is higher than the rebalancing cost, then we will rebalance our open positions.

Advanced Problem I	Direct	Heuristic	Average	Single Market
	16212	211	186	152
Advanced Problem II	–	Heuristic	Average	Single Market
	–	27439	23128	17234

Table 3.12: The speed of the calculations by different methods (in seconds).

3.7.3 Löhndorf's and Wozabal's ADDP algorithm

Note that Intrinsic and Rolling intrinsic valuations are risk free because the exchange guarantees the settlement of the contract. All what we do is profiting from the time spreads of forward curves. So the the rolling intrinsic valuation is a certain benchmark, and we can improve our results only at the cost of higher

risks. We can apply ADDP algorithm on a gas storage unit valuation in the framework of the risk measures. If we use the following risk measure

$$\rho_{\alpha,\gamma}(X) = \gamma \cdot \text{CVaR}_\alpha(X) + (1-\gamma)\mathbb{E}(X)$$
$$\text{and we define N-CVaR}_{\alpha,\gamma} \text{ recursively}$$
$$\text{N-CVaR}_{\alpha,\gamma}(X_1, .., X_T) = X_1 + \rho_{\alpha,\gamma}(X_2 + \rho_{\alpha,\gamma}(X_3 + ...))$$

then the application of the linear ADDP algorithm proposed by Wozabal and Löhndorf on the problem of forward gas storage valuation where the expected value is replaced with N-CVaR$_{0,1}$ yields the rolling intrinsic strategy [105]. If, however, $\gamma = 0$ then we deal with the risk-neutral strategy. Since forward trading is less risky than spot trading and might yield certain incomes, we suspect that the optimal strategy of a gas storage unit valuation is spot-forward. In the case of spot strategies, the resource state is only the storage level, while in the case of forward strategies, the resource state is not only the storage level, but also all of the open positions. This means that there are different setups, and it is a challenging task to combine them in a single model. We propose the following spot-forward strategy, which evidently yields a sub-optimal solution:

1. If C is the volume of the storage then for $p \in [0, 1]$ we assign the volume $p \cdot C$ only to spot trading and the volume $(1-p) \cdot C$ only to forward trading.

2. The forward and the spot prices are linked and for the modelling of the former we use principal component analysis.

3. If we denote the optimal value of the gas storage by v and the optimal value yielded by this strategy by $v(p)$, then it is obvious that

$$\max_{0 \le p \le 1} v(p) \le v.$$

This approach enables us to implement spot and forward strategies on parallel, and to obtain a useful lower bound to the problem. However, its implementation goes beyond the scope of this work and is part of further research.

Conclusion to Chapter 3

In this chapter, we introduced the problem of a gas storage unit valuation and demonstrated our approaches to solving it by means of *Simple Problem*, *Advanced Problem I*, and *Advanced Problem II*. We mentioned the importance of being aware that some European gas storage facilities may be connected to several markets and ignoring these connections leads to sub-optimal values of gas storage units. In *Advanced Problem II*, we compared different models of gas prices in terms of Akaike and Bayesian information criteria. The usage of stable distributions enables us to reduce the dimension of the environmental state of the *Advanced Problem II*, and to introduce stochastic dominance constraints without violating of the convexity of the problem. By observing the mean reversion of the considered spreads, we can model only one forward price (by an advanced

price model) and the spot prices can be represented as the forward price plus a low-dimensional spread process, which also results in dimensionality reduction. We showed that if all of the spread processes are AR(1) and all of their residuals are stable with the same tail index, then the complexity of the average problem, i.e. Problem (3.29) will be identical to that of the *Single Market Problem*. We also introduced a Heuristic approach, which enables us to get a suboptimal solution about 77 times faster than by applying direct procedures of valuation (See Table 3.12). All of the prices have to be discretised, and we measure the quality of discretisation by means of the Wasserstein metric except for Monte Carlo methods. In our algorithms *Alg1* and *Alg2* we resort to Monte Carlo methods, however they can also be implemented in the setup of the transition matrix.

Conclusion

In this dissertation, we explored the features of heavy-tailed distributions and studied stochastic dominance. In applications, we dealt with the problems of gas storage valuation under price uncertainty, and paid attention to the theories which we used to examine this problem. These are theories such as heavy-tailed distributions, and, in particular, stable distributions; stochastic dominance in the framework of stable distributions, and stochastic dynamic programming.

We empirically came up with Hypothesis 1, which states that if X_1 and X_2 have $S_{\alpha_1}(1,0,0)$ and $S_{\alpha_2}(1,0,0)$, respectively, and $\alpha_1 > \alpha_2$ then $X_1 \geq_{(2)} X_2$. We also came empirically to Hypothesis 2, which states if X_1 and X_2 have $S_{\alpha_1}(1,1,0;0)$ and $S_{\alpha_2}(1,1,0;0)$, respectively, and $\alpha_1 > \alpha_2$ then $X_1 \leq_{(2)} X_2$, i.e. in 0-parametrisation for $\beta = 1$ the decrease in the tail index leads to a second-order stochastically dominant random variable.

We proved that the decrease in the scale parameter of symmetric stable distributions leads to a second-order stochastically dominant random variable and used this result for the proof of the theorem about the stochastic dominance of portfolios of stable returns. For portfolios of two symmetric stable returns with the same tail index, skewness, and location parameter where one return second-order dominates another return, we found out what is the optimal proportion of the dominating return in the portfolio in terms of "$\geq_{(2)}$" relation.

We introduced a two-phase method for parameter estimation, which we call the *Method of projection* and which is based on Kagan's methodology [41, 1]. This methodology basically lies in calculating the means of the harmonic functions of the observations, and then solving linear equations. The value of the parameters entailed by this methodology converges to the value entailed by the maximum likelihood methodology as k in (1.23) tends to infinity. We showed that this methodology can be used for the estimation of the tail index of stable and geometric stable distributions, the dispersion matrix of sub-Gaussian distributions, and the parameter r of special cases of $\mathcal{S}$ distributions.

In Chapter 3, we presented how to incorporate certain stochastic dominance constraint into problems of gas storage valuation without destroying the convexity of the problem, provided that the innovations are stable. In Tables 3.8, we presented the importance of taking into account both the heavy tails in prices and the connections of gas storage units to several cross-border markets. The last feature is technically implementable and if legal obstacles disappear, then based on the results in Table 3.8 and Table 3.10, we suspect that this will lead to revising the values of certain gas storage units.

By increasing the complexity of the problem of gas storage valuation, from *Simple Problem* to *Advanced Problem I*, and to *Advanced Problem II*, we demonstrated how this problem becomes intractable within the setup of stochastic dynamic programming. The increase in time consumption when we switch from *Advanced Problem I* to *Advanced Problem II* is demonstrated in Table 3.12. The incorporation of more features of real prices into models of gas storage valuation than available in *Advanced Problem II*, and methods for addressing this problem (possibly other methods than SDP) is a topic for further research alongside other

topics in the following list:

- Theoretical results regarding the relationship between distributions $\mathcal{S}$ and general operator-stable distributions.
- Finding the proof of Hypothesis 1 and Hypothesis 2 (Chapter 2).
- Stochastic dominance for multivariate stable distributions and other types of heavy-tailed distributions.
- Exploration of stochastic dominance for tempered stable distributions.
- Elaborating algorithms which are based on the decision policy to address the problems of not only gas storage valuation, but also hydro storage valuation.
- Elaborating spot-forward strategies for gas storage valuation.

APPENDIX

Post-Decision State Variables and Value Functions

In this section, we provide the post-decision approach which enables us to avoid the second course of dimensionality by using Monte Carlo simulations [79]. The main benefit of the post-decision approach is the ability to reverse the order of the operators of 'max' and '$\mathbb{E}$'. Rather than solving a large problem, this enables us to solve a large number of small problems. With regards to many practical problems, this may give us a tremendous computational advantage [79].

Let us recall Bellman's equation for state variables, which we denote as Z_t, and the reward function $R_t(Z_t, x_t)$, $t = 1, .., T$ which is expressed as follows

$$V_t(Z_t) = \max_{x \in \mathcal{X}_t} \left(R_t(Z_t, x) + \mathbb{E}[V_{t+1}(Z_{t+1})|Z_t] \right), \tag{3.35}$$

where $\mathcal{X}_t$ denotes the feasible set. Note that the operator '$\mathbb{E}$' is preceded by 'max'. This expected value is conditional on the state Z_t. In terms of the transition matrix, problem (3.35) looks as follows:

$$V_t(Z_t) = \max_{x \in \mathcal{X}_t} \left(R_t(Z_t, x) + \sum_{z'} \mathbb{P}(z'|Z_t) V_{t+1}(z') \right).$$

In problems of dynamic programming at time t for the state Z_t, we first make a decision x_t, then wait for new information ω_{t+1} and hereby we obtain Z_{t+1}. In other words:

$$Z_{t+1} = S^M(Z_t, x_t, \omega_{t+1}) = z',$$

where $S^M(\cdot, \cdot, \cdot)$ is a function which based on the current state, our decision, and new information, thus creating the state in the following timestep.

Let us note that two states, Z_t and Z_{t+1}, have an interstate. This is a post-decision state, which we denote as Z_t^x, i.e. $Z_t^x = S^M(Z_t, x)$, where $S^M(\cdot, \cdot)$ is a function which based on the current state and our decision, thus creates the post-decision state.

In the inventory problem [79, 94], this state corresponds to the situation where the inventory manager has already acquired the goods, but does not yet know what tomorrow's demand will be. Sometimes, $Z_t^x = Z_t$. This happens when our decision cannot affect the state, because the state is environmental – such as a market price. Sometimes $Z_t^x = Z_{t+1}$ and this situation occurs in the problem of gas storage (The resource state).
If we shift from state variables to post-decision variables, we also have to reconsider the value function. For post-decision variables Z_t^x there exist post-decision value functions $V_t^x(\cdot)$, and there is a unique correspondence between them [79]:

$$V_t^x(Z_{t-1}^x) = \mathbb{E}\left(V_t(Z_t)|Z_{t-1}^x.\right) \tag{3.36}$$

In the case of the and hydro storage, there is a linear dependence between Z_t and Z_t^x, which means that their σ-algebras coincide.

If $\sigma(Z_t) = \sigma(Z_t^x)$ then

$$\mathbb{E}(V_{t+1}(Z_{t+1})|Z_t^x) = \mathbb{E}(V_{t+1}(Z_{t+1})|S^M(Z_t, x)) = \mathbb{E}(V_{t+1}(Z_{t+1})|Z_t) \tag{3.37}$$

and

$$V_t(Z_t) = \max_{x\in\mathcal{X}_t}\left(R_t(Z_t, x_t) + V_{t+1}^x(Z_t^x)\right). \tag{3.38}$$

Equations (3.36) and (3.38) enable us to express the value function in terms of a post-decision value function and vice versa. However, if we plug (3.38) into (3.36), we will get

$$V_t^x(Z_{t-1}^x) = \mathbb{E}\left(\max_{x\in\mathcal{X}_t}\left(R_t(Z_t, x) + V_{t+1}^x(Z_t^x)\right)|Z_{t-1}^x\right) \tag{3.39}$$

Note that the order of $\mathbb{E}$ and max has been switched and the resulting optimization problems $\max_{x\in\mathcal{X}_t}\left(R_t(Z_t, x) + V_t^x(Z_t^x)\right)$ have become deterministic. We do not have to deal with a transition matrix and we can achieve an arbitrary precision by sampling. The main achievement of this approach is summarised in the following formulae [79]:

$$\begin{aligned} V_t^x(Z_{t-1}^x) = &\quad \mathbb{E}\left(\max_{x\in\mathcal{X}_t}\left(R_t(Z_t, x) + V_{t+1}^x(Z_t^x)\right)|Z_{t-1}^x\right) \\ V_t(Z_t) = &\quad \max_{x\in\mathcal{X}_t}\mathbb{E}\left(\left(R_t(Z_t, x) + V_{t+1}(Z_{t+1})\right)|Z_t\right). \end{aligned}$$

The policy function can thus be calculated as follows:

$$\begin{aligned} &\text{For } t = 1, .., T \\ &\text{policy}_t(Z_t) = \arg\max_{x\in\mathcal{X}_t}\left(R_t(Z_t, x_t) + V_{t+1}^x(Z_t^x)\right) \end{aligned} \tag{3.40}$$

Bibliography

[1] V. Omelchenko (2010): *Elliptical Stable Distributions.* Mathematical Methods in Economics, (M. Houda, J. Friebelova eds.), pp: 483–488.

[2] V. Omelchenko (2014): *Parameter Estimation of Sub-Gaussian Stable Distributions.* Kybernetika, Vol. 50, No. 6, pp: 1065–1076.

[3] V. Omelchenko (2012): *Behavior and Convergence of Wasserstein Metric in the Framework of Stable Distributions.* Bulletin of the Czech Econometric Society Vol 30, pp: 124–138.

[4] V. Omelchenko, V. Kaňková (2015): *Empirical Estimates in Stochastic Programs with Probability and Second-Order Stochastic Dominance Constraints.* AMUC. Vol. 84, 2, pp: 267–281.

[5] M. Arnold, G. Andersson (2011): *Model Predictive Control of Energy Storage Including Uncertain Forecasts.* Proceedings of the Power Systems Computation Conference (PSCC), Stockholm, Sweden, pp: 22–26.

[6] K. Alexeev, V. Borman, V. Varshavskiy, A. Kvanin, Yu. Lebedinskii, V. Tronin, I. Tronin, V. Troyana (2015): *Stable Distributions and Rare Events Statistics in Analysis of the Carbon Fibers Strength and Determination of the Critical Stages of Conversion.* Physics Procedia, Vol. 72, 2015, pp: 62–65.

[7] E. Barrio, E. Giné, C. Matrán (1999): *Central Limit Theorems for the Wasserstein Distance Between the Empirical and True Distributions.* Annals of Probability, 27 , 2, pp: 1009–1071.

[8] O. Barndorff-Nielsen, P. Blaesild (1981): *Hyperbolic Distributions and Ramifications: Contributions to Theory and Application.*
Chapter: Statistical Distributions in Scientific Work Vol. 79 of the series NATO Advanced Study Institutes Series, pp: 19–44.

[9] O. Barndorff-Nielsen (1997): *Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling.* Scandinavian Journal of Statistics, Vol 23, pp: 1–13.

[10] V. S. Bawa (1978): *Safety-First, Stochastic Dominance, and Optimal Portfolio Choice.* The Journal of Financial and Quantitative Analysis. Vol. 13, No. 2, pp: 255–271.

[11] V. S. Bawa (1980): *On Stochastic Dominance and Estimation Risk.* Haim Levy (ed.). Reseach in Finance II, JAI Press, New York, 1980, pp: 57–67.

[12] M. L. Bianchi, S. Rachev, F. J. Fabozzi (2015): *Tempered Stable Ornstein-Uhlenbeck Processes: A Practical View.* Communications in Statistics-Simulation and Computation. Taylor and Francis.

[13] B.M. Bibby, M. Sorensen (2003): *Hyperbolic Processes in Finance.* Handbook of Heavy Tailed Distributions in Finance, Elsevier Science. Amsterdam, pp: 212–244.

[14] R. Blattberg and N. Gonedes (1974): *A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices.* Journal of Business 47, pp: 244–280.

[15] C. Blanco, D. Soronow, P. Stefiszyn (2002): *Multi-Factor Models of the Forward Price Curve.* Lacima Publication.

[16] S. Borak, W. Härdle, R. Weron (2005): *Stable Distributions.* In: P. Cizek, W. Härdle, R. Weron (eds) Statistical tools for finance and insurance. Springer, Berlin, pp: 21–44.

[17] P. Bougerol, N. Picard (1992): *Strict Stationarity of Generalized Autoregressive Processes.* Annals of Probability, Vol. 20 (4), pp: 1714–1730.

[18] P. Bougerol N. Picard (1992): *Stationarity of GARCH Processes and Some Nonnegative Time Series.* Journal of Econometrics 52, pp: 115–127.

[19] P. Carr, H. Geman, D.B. Madan, M. Yor (2002): *The Fine Structure of Asset Returns: An Empirical Investigation.* The Journal of Business, Vol. 75(2) pp: 305–333.

[20] M. Branda, M. Kopa (2016): *DEA Models Equivalent to General N-th Order Stochastic Dominance Efficiency Tests.* Operations Research Letters, Vol. 44, pp: 285–289.

[21] M. Carrasco, J. Florens (2000): *Generalization of GMM to a Continuum Moment Condition.* Econometric Theory, Vol. 16, pp: 767–834.

[22] M. Carrasco, R. Kotchoni (2013): *Efficient Estimation Using the Characteristic Function.* CIRANO - Scientific Publications, 2013s-22.

[23] J.M. Chambers, C.L. Mallows, and B.W. Stuck (1976): *A Method for Simulating Stable Random Variables.* Journal of the American Statistical Association Vol. 71, pp: 340–344.

[24] B. N. Cheng, S. Rachev (1995): *Multivariate Stable Commodities in Financial Markets.* Mathematical Finance, Vol 5, pp: 133–159.

[25] Dencheva, D. Ruszczynski (2003): *Optimization with Stochastic Dominance Constraints.* SIAM Journal on Optimization, Vol. 14, pp: 548–566.

[26] W.H. DuMouchel (1971): *Stable Distributions in Statistical Inference. PhD Thesis.* University of Ann Arbor, Ann Arbor, MI.

[27] B. Grechu (2014): *A Simple SSD-Efficiency Test.* Optimization Letters. Vol 8, 7, pp: 2135–2143.

[28] E. Fama (1965): *Portfolio Analysis in a Stable Paretian Market.* Management Science, Vol. 11, pp: 404–419.

[29] N. Farsad, W. Guo, C.-B. Chae, and A. W. Eckford (2015): *Stable Distributions as Noise Models for Molecular Communication.* EEE Global Communication Conference (GLOBCOM).

[30] A. Feuerverger (1990): *An Efficiency Result for the Empirical Characteristic Function in Stationary Time-Series Models.* Canadian Journal of Statistics, Vol. 18, pp: 155–161.

[31] J. Gil-Pelaez (1951): *Note on the Inversion Theorem.* Biometrika, Vol. 38, pp: 481–482.

[32] D. T. Gillespie (1996): *Exact Numerical Simulation of the Ornstein-Uhlenbeck Process and its Integral.* Physical Review E. 54, S. 20842091.

[33] Grabchak, M. and Samorodnitsky, G. (2010): *Do Financial Returns Have Finite or Infinite Variance? A Paradox and an Explanation.* Quantitative Finance, Vol. 10, 8, pp: 883–893.

[34] J. Gray, P. Khandewal (2004): *Realistic Gas Storage Models II: Trading Strategies.* Commodities Now, pp: 1–5.

[35] P. Gross, R. M. Kovalcevic, G. Pflug (2014): *Energy Markets, Handbook of Risk Management.* Energy Production and Trading, Springer, New York, pp: 3–24.

[36] P. Henaff, I. Laachir, F. Russo (2013): *Gas Storage Valuation and Hedging: a Quantification of the Model Risk* (working paper). arXiv:1312.3789, 2013.

[37] B.M. Hill (1975): *A Simple General Approach to Inference about the Tail of a Stable Distribution.* Annals of Statistics Vol. 3(5), pp: 1163–1174.

[38] Y. Hu, H. Long (2009): *Least Squares Estimator for Ornstein-Uhlenbeck Processes Driven by α-Stable Motions.* Stochastic Processes and their Applications. Vol. 119, Issue 8, pp: 2465–2480.

[39] C. de Jong (2015): *Gas Storage Valuation and Optimization.* Journal of Natural Gas Science and Engineering, Vol. 24, pp: 365–378.

[40] Z.J. Jurek, J.D. Mason (1993): *Operator-Limit Distributions in Probability Theory.* Wiley. New York, NY.

[41] A. Kagan (2009): *Fisher Information Contained in a Finite-dimensional Linear Space, and a Properly Formulated Version of the Method of Moments (in Russian).* Problemy Peredachi Informacii 12, Moscow , pp: 15–29.

[42] A. Kalemanova, B. Schmid, R. Werner (2007): *The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing.* Journal of Derivatives.

[43] V. Kaňková, M. Houda (2006): *Empirical Estimates in Stochastic Programming.* Proceedings of the Prague Stochastics 2006, (M. Hušková, and M. Janžura P. Lachout, eds.), Matfyzpress, Prague, pp: 426–436.

[44] H. Kesten, F. Spitzer (1984): *Convergence in Distribution for Products of Random Matrices.* Zeitschrift Fiir Wahrscheinlichkeitstheorie und Ver'wandte Gebiete, Vol. 67, pp: 363–386.

[45] L.B. Klebanov (2003): *Heavy Tailed Distributions.* Matfyzpress, Prague.

[46] L.B. Klebanov, J.A. Melamed (1979): *One Linear Method of Estimation of Parameters.* Proceedings of 2-nd Prague Symposium on Asymptotic Statistics, Prague, pp: 259–270.

[47] L. Klebanov and I. Volchenkova (2015): *Heavy Tailed Distributions in Finance: Reality or Myth? Amateurs Viewpoint.* ArXiv: 1507.07735 Vol. 1, pp: 1–17.

[48] I.A. Koutrovelis (1980): *Regression-Type Estimation of the Parameters of Stable Laws.* Journal of the American Statistical Association, 75, pp: 918–928.

[49] M. Kopa, P. Chovanec (2008): *A Second-Order Stochastic Dominance Portfolio Efficiency Measure.* Kybernetika. Vol. 44, 2, pp: 243–258.

[50] S. Kring, S.T. Rachev, M. Höchstötter, F.Z. Fabozzi (2009): *Risk Assessment: Decisions in Banking and Finance*, Chapter 6: Estimation of α-Stable Sub-Gaussian Distributions for Asset Returns. Heidelberg: Springer/Physika, pp: 111-152.

[51] L. Landau, L. (1944): *On the Energy Loss of Fast Particles by Ionization.* J. Phys. (USSR) 8: 201.

[52] J. E. Gentle (2003): *Random Number Generation and Monte Carlo Methods.* Statistics and Computing (2nd ed.). New York: Springer, pp 196. doi:10.1007/b97336. ISBN 978-0-387-00178-4.

[53] Holtsmark, J. (1919): *Über die Verbreiterung von Spektrallinien.* Annalen der Physik (in German), Vol. 363 (7), pp: 577–630.

[54] B. Mandelbrot (1963): *The Variation of Certain Speculative Prices.* Journal of Business, Vol. 36, pp: 394–419.

[55] M. L. Mansfielda (2016): *Numerical Tools for Obtaining Power-Law Representations of Heavy-Tailed Datasets.* European Physics Journal B, 89:16

[56] J.H. McCulloch (1986): *Simple Consistent Estimators of Stable Distribution Parameters.* Communications in Statistics – Simulation and Computation, Vol. 15, pp: 1109–1136.

[57] J.H. McCulloch (2000): *Estimation of the Bivariate Stable Spectral Representation by the Projection Method* Computational Economics 16, Kluwer Academic Publishers, pp: 47–62.

[58] M.M. Meerschaft, H.P. Scheffler (2001): *Limit Distributions for Sums of Independent Random Vectors.* Wiley. New York. NY.

[59] C. Menn, S.T. Rachev (2005): *Smoothly Truncated Stable Distributions. GARCH-Models and Option Pricing.* Mathematical Methods of Operations Research, Vol. 69, issue 3, pp: 411–438

[60] S. Mittnik, M.S. Paolella (2003): *Prediction of Financial Downside-Risk with Heavy-Tailed Conditional Distributions.* Handbook of Heavy Tailed Distributions in Finance. Elsevier. Amsterdam, pp: 385–403.

[61] S. Mittnik, S. Rachev (1996): *Tail Estmation of the Stable Index Alpha. Applied Mathematics.* Letters 9(3), pp: 53–56.

[62] C. Mugele, S. Rachev, S. Trück (2005): *Stable Modeling of Different European Power Markets.* Investment Management and Financial Innovations 3.

[63] J. P. Nolan: (2001): *Maximum Likelihood Estimation and Diagnostics for Stable Distributions, in O. E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.).* Lévy Processes, Brickhauser, Boston.

[64] J. P. Nolan and A. K. Panorska (1997): *Data Analysis for Heavy-Tailed Multivariate Samples.* Communications in Statistics: Stochastic Models, pp: 687–702.

[65] J. P. Nolan (2003): *Modeling Financial Data with Stable Distributions.* Handbook of Heavy Tailed Distributions in Finance. Elsevier. Amsterdam, pp: 105–128.

[66] J. P. Nolan (2009): *Stable Distributions – Models for Heavy Tailed Data (PDF).* Retrieved 2009-02-21

[67] S. Ortobelli, I. Huber, S. Rachev, E.S. Schwarz (2003): *Portfolio Choice Theory with Non-Gaussian Distributed Return.* Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance. Elsevier. Amsterdam. pp: 547–594.

[68] E.E. Osuna (2012): *Crossing Points of Distributions and a Theorem That Relates Them to Second-Order Stochastic Dominance* Statistics and Probability Letters 82, pp: 758–764.

[69] Paraschiv F. (2013): *Price Dynamics in Electricity Markets.* In Handbook of Risk Management in Energy Production and Trading, Springer, New York, pp: 47–69.

[70] Pereira M.V.F., Pinto L.M.V.G. (1991): *Multistage Stochastic Optimization Applied to Energy Planning.* Math Program, Vol. 52, pp: 359–375.

[71] G. C. Pflug, A. Pichler (2014): *Multistage Stochastic Optimization.* Springer, pp: 46–95.

[72] G. C. Pflug (2001): *Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization.* Mathematical Programming Series B, Vol. 89, 2, pp: 251–271.

[73] G. C. Pflug, W. Römisch (2007): *Modeling, Measuring and Managing Risk*, World Scientific, Singapore.

[74] Philpott A. B., Guan Z. (2008): *On the Convergence of Stochastic Dual Dynamic Programming and Related Methods.* Operations Research Letters, 36, pp: 450–455.

[75] M. Pivato, L. Seco (2003): *Estimating the Spectral Measure of a Multivariate Stable Distribution via Spherical Harmonic Analysis.* Journal of Multivariate Analysis, Vol. 87, Issue 2, pp: 219–240.

[76] T. Post, Y. Fang, M. Kopa (2015): *Linear Tests for DARA Stochastic Dominance*, Management Science, Vol. 61, 7, pp: 1615–1629.

[77] T. Post, M. Kopa (2016): *Portfolio Choice based on Third-degree Stochastic Dominance.* Management Science, to appear.

[78] Postek K.S., Hertog D. (2016) *Multi-stage Adjustable Robust Mixed-Integer Optimization via Iterative Splitting of the Uncertainty set.* Tilburg: Operations research, CentER Discussion Paper, Series 2016-006, pp: 2014–2031.

[79] W. Powell (2007): *Approximate Dynamic Programming, Solving the Curses of Dimensionality.* Wiley.

[80] S. T. Rachev, E.S. Schwarz, I. Khindanova (2003): *Stable Modeling of Market and Credit Value at Risk.* Handbook of Heavy Tailed Distributions in Finance. Elsevier. Amsterdam, pp: 255–264.

[81] S. T. Rachev (1991): *Probability Metrics and the Stability of Stochastic Models.* Wiley, West Sussex.

[82] S. T. Rachev, S.V. Stoyanov, F.J. Fabozzi (2011): *A Probability Metrics Approach to Financial Risk Measures.* Wiley. London.

[83] S.T. Rachev, S.T., S. Trück, R. Weron (2004): *Risk Management in the Power Markets (part III): Advanced Spot Price Models and VaR.* Risknews 05/2004, pp: 67–71 (in German).

[84] R.T. Rockafellar, R. J-B. Wets (2009): *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften 317, Springer-Verlag.

[85] W. Römisch, R. Schultz (1991): *Distribution Sensitivity in Stochastic Programming.* Mathematical Programming. Vol. 50, pp: 197–226.

[86] W. Römisch, R. Schultz (1991): *Stability Analysis for Stochastic Programs.* Annals of Operation Research. Vol. 30, pp: 197–226.

[87] J. Rosinski (2007): *Tempering Stable Processes.* Stochastic Processes and Their Applications, 117(6), pp: 677–707.

[88] J. Rosinski, J.L. Sinclair (2010): *Generalized Tempered Stable Processes.* In J.K. Misiewicz, editor, Stability in Probability, Banach Center Publication, Vol. 90, pp: 153–170.

[89] G. Samorodnitsky, M.S. Taqqu (1994): *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

[90] G. Samorodnitsky, and M.S. Taqqu (1993): *Stochastic Monotonicity and Slepian-type Inequalities for Infinitely Divisible and Stable Random Vectors.* Annals of Probability, Vol. 21, pp: 143–160.

[91] N. Safarov, C. Atkinson (2016): *Natural Gas Storage Valuation and Optimization under Time-Inhomogeneous Exponential Lévy Processes.* International Journal of Computer Mathematics.

[92] P. Schmidt (1982): *An Improved Version of Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions.* Econometrica,50 ,501-524.

[93] Shapiro A. (2011) *Analysis of Stochastic Dual Dynamic Programming Method.* European Journal of Operational Research, Vol. 209, Elsevier, pp: 63–72.

[94] A. Shapiro, D. Dencheva, D. Ruszczynski (2009): *Lectures on Stochastic Programming (Modeling and Theory).* Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia.

[95] L. Slámová, L. Klebanov (2012): *Modeling Financial Returns by Discrete Stable Distributions* Proceedings of 30-th International Conference Mathematical Methods in Economics.

[96] V. Uchaikin, V. Zolotarev (1999): *Chance and Stability: Stable Distributions and their Applications.* De Gruyter.

[97] S. S. Vallander (1973): *Calculation of the Wasserstein Distance between Probability Distributions on the Line.* Theory of Probability and its Applications, pp: 784–786.

[98] S. Vichik, F. Borrelli (2014): *Solving Linear and Quadratic Programs with an Analog Circuit.* Computers and Chemical Engineering, Vol. 70, pp: 160–171.

[99] A. Weron, R Weron (2006): *Computer Simulation of Lévy α-Stable Variables and Processes.* In: Chaos – The Interplay Between Stochastic and Deterministic Behaviour, pp: 379–392.

[100] R. Weron (2013): *Forecasting Wholesale Electricity Prices: A Review of Time Series Models, Principles of Modelling, Forecasting and Decision-Makin.* FindEcon Monograph Series, WUL, Lodz.

[101] R. Weron (2006): *Modeling and Forecasting Electricity Loads and Prices.* John Wiley and Sons.

[102] R. Weron (2005): *Heavy Tails and Electricity Prices.* The Deutsche Bundesbank's 2005 Annual Fall Conference. Eltville.

[103] O. Wolfgang, A. Haugstad, B. Mo, A. Gjelsvik, I. Wangensteen, G. Doorman (2009): *Hydro Reservoir Handling in Norway before and after Deregulation.* Energy 34(10), pp: 1642–1651. doi:10.1016/j.energy.2009.07.025.

[104] E. Wolfstetter (1996): *Stochastic Dominance: Theory and Applications.* No 1996,40, SFB 373 Discussion Papers from Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

[105] D. Wozabal, N. Löhndorf (2015): *Optimal Gas Storage Valuation and Futures Trading under a High-Dimensional Price Process.* Optimization Online.

[106] V. Zolotarev (1986): *On Representation of Stable Laws by Integrals Selected Translation.* Mathematical Statistics and Probability, Vol. 6, pp: 84–88.

[107] http://www.icis.com/resources/news/2012/09/17/9596191/egs-to-offer-natural-gas-storage-product-linked-to-german-hub-prices/

[108] http://www.pozagas.sk/en/

[109] http://nafta.sk/en/about-gas-storage

List of Tables

List of Abbreviations

SDP	Stochastic Dynamic Programming
ADP	Approximate Dynamic Programming
SDDP	Stochastic Dual Dynamic Programming
ADDP	Approximate Dual Dynamic Programming
AR	Auto-regression
MA	Moving Average
ARMA	Auto-regressive Moving Average
GARCH	Generalised Auto-regressive Conditional Heteroskedasticity
OU	Ornstein-Uhlenbeck
NIG	Normal Inverse Gaussian
...	
$S_\alpha(\sigma, \beta, \mu)$	A stable distribution with parameters α, σ, β, and μ.
std	Standard deviation
$X \vert\vert Y$	Means that the random variables X and Y are independent
iid	Independent identically distributed
iff	If and only if
$X \geq_{(k)} Y$	The random variable X dominates Y in the k-th order
$X \neq_{(k)} Y$	Neither $X \geq_{(k)} Y$ nor $Y \geq_{(k)} X$
P_t	The grid of all possible prices at time t for SDP or ADDP
$\mathbb{P}$	The probability
$\mathbb{E}$	The expected value
CV@R	Conditional Value at Risk
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$\mathcal{X}_t$	A feasible set for the decision variable
$F^*(\cdot)$	If $F(\cdot)$ is a distribution function, then $F^*(x) = 1 - F(x)$, $x \in \mathbb{R}$.