# Risk Management with Extreme Value Theory

Claudia Klüppelberg
Center of Mathematical Sciences
Munich University of Technology
D-80290 Munich
Germany
email: cklu@mathematik.tu-muenchen.de
http://www.ma.tum.de/stat/

March 27, 2002

## 1    Introduction

In today's financial world, *Value-at-Risk* has become the benchmark risk measure. Following the *Basle Accord on Market Risk* (1988,1995,1996) every bank in more than 100 countries around the world has to calculate its risk exposure for every individual trading desk. The standard method prescribes: estimate the $p$-quantile of the profit/loss distribution for the next 10 days and $p = 1\%$ (or $p = 5\%$) based on observations of at least 1 year (220 trading days). Standard model is the normal model. Finally, multiply the estimated quantile by 3. This number is negative and its modulus is called *Value-at-Risk (VaR)*. The factor 3 is supposed to account for certain observed effects, also due to the model risk; it is based on backtesting procedures and can be increased by the regulatory authorities, if the backtesting proves the factor 3 to be insufficient. The importance of VaR is undebated since regulators accept this model as a basis for setting capital requirements for market risk exposure. A textbook treatment of VaR is given in Joriot [50]. Interesting articles on risk management are collected in Embrechts [32].

There were always discussions about the classical risk measure, which has traditionally been the variance, and alternatives have been suggested. They are typically based on the notion of downside risk concepts such as *lower partial moments*. The lower partial moment of order $n$ is defined as

$$\mathrm{LPM}_n(x) = \int_{-\infty}^{x} (x - r)^n dF(r), \quad x \in \mathbb{R},$$

where $F$ is the distribution function of the portfolio return. Examples can be found in Fishburn [39] or Harlow [47] including the shortfall probability ($n = 0$), which is nothing else but the VaR. An axiomatic approach to risk measures can be found in Artzner et al. [1]; cf. Embrechts [31]. For some discussion see also Rootzén and Klüppelberg [77].

Standard model in the Basle account is the normal distribution which has the property that it is sum stable, i.e. for a dynamic model we obtain

$$\mathrm{VaR}(10 \text{ days}) = \sqrt{10} \, \mathrm{VaR}(1 \text{ day}),$$

and for a multivariate model; i.e. a portfolio with weights $w_i$ for asset $i$ and correlation $\rho_{ij}$ between assets $i$ and $j$, $i, j = 1, \ldots, q$,

$$\mathrm{VaR}(\text{portfolio}) = \sqrt{\sum_{i,j=1}^{q} \rho_{ij} w_i w_j \mathrm{VaR}_i \mathrm{VaR}_j} \,.$$

However, the obvious disadvantage of the normal model is that it is wrong and can dangerously underestimate the risk. This is even visible in Figures 1.1 and 1.2.

This is the starting point of the present paper. Taking also extreme fluctuations of financial data into account we want to answer the following questions:

– How does one estimate VaR from financial time series under realistic model assumptions?

– What is the consequence of VaR as a risk measure based on a low quantile for portfolio optimization?

Statistical estimation of risk and portfolio optimization are two important issues in risk management, influenced by the choice of risk measure. Pricing of derivatives and hedging of portfolios are other important issues and the VaR has found its way also to the hedging problem. In incomplete markets, which is the setup for all "realistic" pricing models, the traditional "hedge without risk" (perfect hedge) has been replaced by a "hedge with small remaining risk" (so-called quantile-hedging); see Föllmer and Leukert [40] and Cvitanic and Karatzas [24]. This is, however, not a topic for this paper.

We first turn to the risk estimation problem.

In the simplest case, it is assumed that the only source of risk is the price of the portfolio itself, i.e. the risk is modelled in terms of price changes, which are independent and identically distributed (iid), the underlying planning horizon is $\Delta t = 1$ (1 day), and we estimate just the quantile (without multiplying by 3).
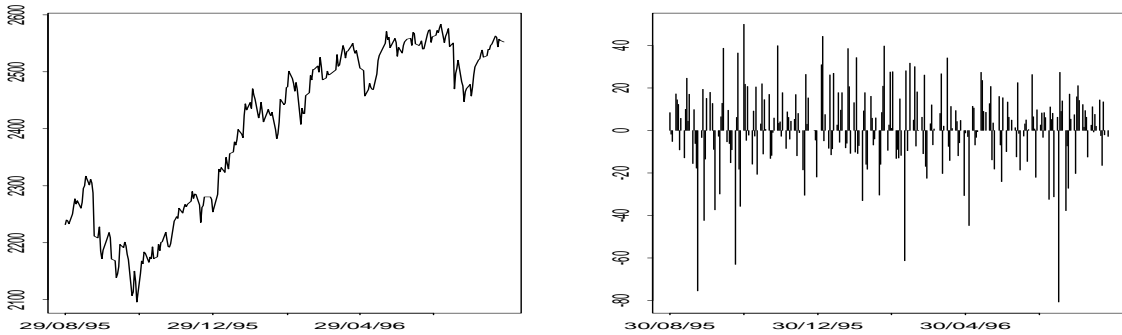


**Figure 1.1.** *DAX closing prices during 29/8/95–26/8/96 (250 data points in total). The corresponding differences, which are the daily price changes (returns), are plotted in the right-hand graph. It is obvious that the returns are not symmetric and that there are more and much more pronounced peaks (in particular negative ones) than one would expect from Gaussian data.*

Generally speaking, estimation of a small quantile is not an easy task, as one wants to make inference about the extremal behaviour of a portfolio, i.e. in an area of the sample where there is only a very small amount of data. Furthermore (and this is important to note), extrapolation
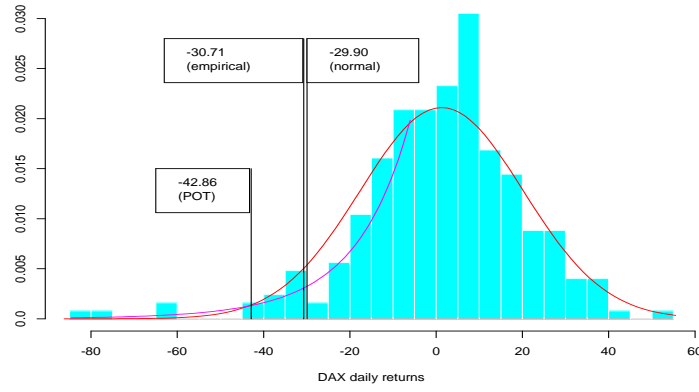
**Figure 1.2.** *Histogram of the daily price changes of the DAX closing prices with fitted normal distribution. Also fitted is a GPD distribution to the left hand tail. The corresponding quantiles are estimated by the normal quantile, the GPD quantile and the empirical quantile.*

even beyond the range of the data might be wanted, i.e. statements about an area where there are no observations at all.

Under the acronym *let the tails speak for themselves*, statistical methods have been developed which are based only on that part of the sample which carries the information about the extremal behaviour, i.e. only the smallest or largest sample values. This method is not solely based on the data but includes a probabilistic argument concerning the behaviour of the extreme sample values. This leads to a class of semiparametric distributions which can be regarded as plausible.

As a basic reference to modelling and quantifying of extreme events we refer to Embrechts, Klüppelberg and Mikosch [33], henceforth abbreviated by EKM. The DAX data example, which we analyse in Section 2 can be found in greater detail in Emmer, Klüppelberg and Trüstedt [36].

Unfortunately, most financial time series are not independent, but exhibit some very delicate temporal dependence structure, which is often captured by Markovian volatility models. Consequently, over the last decades a variety of stochastic models have been suggested as appropriate models for financial products.

In a continuous time setting the dynamics of a price or an interest rate process is often modelled as a diffusion process given by a stochastic differential equation (SDE)

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad t > 0, \quad X_0 = x, \tag{1.1}$$

where $W$ is standard Brownian motion, $\mu \in \mathbb{R}$ is the *drift term* and $\sigma > 0$ is the *diffusion coefficient or volatility*. Two standard models in finance are of the above form:

(i) The Black-Scholes model: $(X_t)_{t \geq 0}$ models the price process of an asset, here $\mu(x) = \mu x$ for $\mu \in \mathbb{R}$ and the volatility $\sigma(x) = \sigma x$ for $\sigma > 0$. The resulting model for the price process is geometric Brownian motion.

(ii) The Vasicek model: the process $(X_t)_{t \geq 0}$ models an interest rate, the drift term $\mu$ is linear and the volatility $\sigma > 0$ is some constant.

Both models can be considered in the framework of Gaussian models, however, as indicated already, financial data exhibit in general fluctuations which cannot be modelled by Gaussian

3

processes or simple transformations as in the two standard models above. In principle there are two different remedies for the problem.

A first concept sticks to Brownian motion as the driving dynamic of the process, but introduces a path-dependent, time-dependent or even stochastic volatility into the model. These models are commonly referred to as volatility models, and include diffusions given by the SDE (1.1). We investigate their extremal behaviour in Section 3.

The second concept replaces the Gaussian driving process in the Black-Scholes or Vasicek model (or any other traditional model) by a process with heavy-tailed marginals as for instance a Lévy process with non-normal noise. We consider this approach in Section 5 in the context of portfolio optimization.

A discrete time counterpart to (1.1) is the following model.

$$X_n = \mu(X_{n-1}) + \sigma(X_{n-1})\,\varepsilon_n\,, \quad n \in \mathbb{N}, \tag{1.2}$$

where $\mu$ is the conditional mean, $\sigma$ the conditional volatility and $(\varepsilon_n)_{n\in\mathbb{N}}$ are iid rvs with mean 0 and variance 1. Examples, also Markovian models of higher order, include for instance ARCH and GARCH models, which have been successfully applied in econometrics.

There is one stylized fact in financial data which models of the form (1.2) can capture in contrast to linear diffusion models of the form (1.1). This is the property of persistence in volatility. For many financial time series with high sampling frequency large changes tend to be followed by large changes, settling down after some time to a more normal behaviour. This observation has lead to models of the form

$$X_n = \sigma_n\,\varepsilon_n\,, \quad n \in \mathbb{N}, \tag{1.3}$$

where the innovations $\varepsilon_n$ are iid rvs with mean zero, and the volatility $\sigma_n$ describes the change of (conditional) variance.

The autoregressive conditionally heteroscedastic (ARCH) models are one of the specifications of (1.3). In this case the conditional variance $\sigma_n^2$ is a linear function of the squared past observations. ARCH($p$) models introduced by Engle [37] are defined by

$$\sigma_n^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{n-j}^2\,, \quad \alpha_0 > 0\,, \ \alpha_1,\ldots,\alpha_{p-1} \geq 0\,, \ \alpha_p > 0\,, \quad n \in \mathbb{N}, \tag{1.4}$$

where $p$ is the order of the ARCH process.

There are two natural extensions of this model. Bollerslev [12] proposed the so-called generalized ARCH (GARCH) processes. The conditional variance $\sigma_n^2$ is now a linear function of past values of the process $X_{n-j}^2$, $j = 1,\ldots,p$, and past values of the volatility $\sigma_{n-j}^2$, $j = 1,\ldots,q$. An interesting review article is Bollerslev, Chou and Kroner [13], a nice collection of some of the most influential articles on ARCH models is Engle [38].

The class of autoregressive (AR) models with ARCH errors introduced by Weiss [89] are another extension; these models are also called SETAR-ARCH models (self-exciting autoregressive). They are defined by

$$X_n = f(X_{n-1}, ..., X_{n-k}) + \sigma_n\,\varepsilon_n, \quad n \in \mathbb{N}, \tag{1.5}$$

where $f$ is again a linear function in its arguments and $\sigma_n$ is given by (1.4). This model combines the advantages of an AR model, which targets more on the conditional mean of $X_n$ (given the

past), and of an ARCH model, which concentrates on the conditional variance of $X_n$ (given the past).

The class of models defined by (1.5) embodies various non-linear models. In this paper we focus on the AR(1) process with ARCH(1) errors, i.e.

$$X_n = \alpha X_{n-1} + \sqrt{\beta + \lambda X_{n-1}^2}\varepsilon_n\,, \quad n \in \mathbb{N},$$

where $\alpha \in \mathbb{R}$, $\beta$, $\lambda > 0$, $(\varepsilon_n)_{n\in\mathbb{N}}$ are iid symmetric rvs with variance 1 and $X_0$ is independent of $(\varepsilon_n)_{n\in\mathbb{N}}$. This Markovian model is analytically tractable and may serve as a prototype for the larger class of models (1.5). Note also that this model is of the form (1.2).

Two early monographs on extreme value theory for stochastic processes are Leadbetter, Lindgren and Rootzén [62], henceforth abbreviated as LLR, and Berman [9]. They contain all basic results on this topic, and it is this source from which all specific results are derived.

The only models of the form (1.2), whose extremal behaviour has been analysed in detail are the ARCH(1) (by de Haan, Resnick, Rootzén and de Vries [46]; see also EKM [33], Section 8.4), the GARCH(1,1) (by Mikosch and Starica [69]), and the AR(1) model with ARCH(1)-errors (by Borkovec and Klüppelberg [18] and Borkovec [15, 16]). The interesting feature of all these models is that they are able to model heavy-tailedness as well as volatility clustering on a high level.

In Section 5 we turn to the second question posed at the beginning. We consider a portfolio optimization problem based on the VaR as a risk measure. Traditional portfolio selection as introduced by Markowitz [65] and Sharpe [82] has been based on the variance as risk measure. In contrast to the variance, the VaR captures the extreme risk. Consequently, it is to be expected that it reacts sensitive to large fluctuations in the data. This is what we investigate here.

We concentrate on the Capital-at-Risk (CaR) as a replacement of the variance in portfolio selection problems. We think of the CaR as the capital reserve in equity to set aside for future risk. The CaR of a portfolio is commonly defined as the difference between the mean of the profit-loss distribution and the VaR. We define the CaR as the difference between the mean wealth of the market (given by the riskless investment) and the VaR of our present portfolio; i.e. we consider the excess loss over the riskless investment.

We aim at closed form solutions and an economic interpretation of our results. This is why we start in a Gaussian world, represented by a Black-Scholes market, where the mean-CaR selection procedure leads to rather explicit solutions for the optimal portfolio. As a first difference to the mean-variance optimization, this approach indeed supports the commonly recommended market strategy that one should always invest in stocks for long-term investment.

As prototypes of models to allow for larger fluctuations than pure Gaussian models, we study Lévy processes, which still have independent and stationary increments, but these increments are no longer normally distributed. Such models have been used as more realistic models for price processes by Barndorff-Nielsen and Shephard [8], Eberlein and his group (see [27] and references therein) and Madan and Seneta [64]; they are meanwhile well understood. The class of normal mixture models supports the observation that volatility changes in time. This is in particular modeled by the normal inverse Gaussian model and the variance gamma model, which have also been recognised and applied in the financial industry. However, as soon as we move away from the Gaussian world, the optimization problem becomes analytically untractable and numerical solutions are called for. We present solutions for the normal mixture models mentioned above.

The data analyses, simulations and figures presented have been created with the software S-Plus. Most routines for extreme value analysis are contained in the software EVIS written by Alex McNeil and can be downloaded from *http://www.math.ethz.ch/finance/*.

# 2  Starting-kit for extreme value analysis

Let $X, X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables (rvs), representing financial losses, with distribution function (df) $F$ (we write $X \stackrel{d}{=} F$).

The classical central limit theorem states that for iid rvs such that $EX = \mu$ and $\mathrm{var}X = \sigma^2 < \infty$ the partial sums $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, satisfy

$$\lim_{n \to \infty} P\left((S_n - n\mu)/\sqrt{n\sigma^2} \le x\right) = N(x), \quad x \in \mathbb{R},$$

where $N$ is the standard normal df. This result, which holds in a much wider context than just iid data, supports the normal law for data which can be interpreted as sum or mean of many small effects, whose variance contributions are asymptotically neglible.

Consequently, the normal model is certainly questionable, whenever extreme risk has to be quantified. Empirical investigations of financial data show quite clearly that the large values, in particular the large negative values, are much more pronounced than could be explained by a normal model.

In the following we present the basic notions and ideas of extreme value theory for iid data. All this and much more can be found in EKM [33]; for more details on the DAX example we refer to Emmer, Klüppelberg and Trüstedt [36].

## 2.1  Sample maxima

The simplest extreme object of a sample is the sample maximum. Define

$$M_1 = X_1, \quad M_n = \max(X_1, \ldots, X_n), \quad n > 1.$$

Then

$$P(M_n \le x) = F^n(x), \quad x \in \mathbb{R},$$

and $M_n \uparrow x_F$ as $n \to \infty$ almost surely, where $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} \le \infty$ is the *right endpoint of $F$*.

In most cases $M_n$ can be normalized such that it converges to a limit rv, which together with the normalizing constants determines the asymptotic behaviour of the sample maxima. The following is the analogue of the CLT for maxima.

**Theorem 2.1.** [Fisher-Tippett theorem]
*Suppose we can find sequences of real numbers $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$\lim_{n \to \infty} P((M_n - b_n)/a_n \le x) = \lim_{n \to \infty} F^n(a_n x + b_n \le x) = H(x), \quad x \in \mathbb{R}, \tag{2.1}$$

*for some non-degenerate df $Q$ (we write $F \in \mathrm{MDA}(Q)$). Then $Q$ is one of the following three extreme value dfs:*

- *Frechét* $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0, \end{cases}$ *for* $\alpha > 0.$

- *Gumbel* $\Lambda(x) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}.$

- *Weibull* $\Psi_\alpha(x) = \begin{cases} \exp\left(-(-x)^\alpha\right), & x \leq 0, \\ 1, & x > 0, \end{cases}$ *for* $\alpha > 0.$

*The limit distribution $Q$ is unique up to affine transformations; we say it is of the type of $Q$.*

All commonly encountered continuous df are in $\mathrm{MDA}(Q)$ for some extreme value df $Q$; see EKM [33], pp. 153-157. Here are three examples.

**Example 2.2.** (a) Exponential distribution: $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0$, $\lambda > 0$, is in $\mathrm{MDA}(\Lambda)$ with $c_n = 1/\lambda$, $d_n = \ln n/\lambda$.

(b) Pareto distribution: $F(x) = 1 - (\frac{\kappa}{\kappa+x})^\alpha$, $x \geq 0$, $\kappa\alpha > 0$, is in $\mathrm{MDA}(\Phi_\alpha)$ with $c_n = (n/\kappa)^{1/\alpha}$, $d_n = 0$.

(c) Uniform distribution: $F(x) = x$, $x \in (0,1)$, is in $\mathrm{MDA}(\Psi_1)$ with $c_n = 1/n$, $d_n = x_F = 1$.

Taking logarithms and invoking a Taylor expansion in (2.1) we obtain

$$F \in \mathrm{MDA}(H) \quad \Longleftrightarrow \quad \lim_{n\to\infty} n\overline{F}(c_n x + d_n) = -\ln H(x) =: \tau(x), \quad x \in \mathbb{R}. \tag{2.2}$$

This MDA condition is a version of Poisson's limit theorem. It can be embedded in the more general theory of point processes as follows.

For iid rvs $X, X_1, \ldots, X_n$ and threshold $u_n$ we have

$$\mathrm{card}\{i : X_i > u_n, i = 1, \ldots, n\} \stackrel{d}{=} \mathrm{Bin}(n, P(X > u_n)).$$

Define for $n \in \mathbb{N}$

$$N_n(B) = \sum_{i=1}^{n} \varepsilon_{i/n}(B)\, I\{X_i > u_n\}, \quad B \in \mathcal{B}(0,1],$$

where $\mathcal{B}(0,1]$ denotes the Borel $\sigma$-algebra on $(0,1]$ and $\varepsilon$ the Dirac measure; i.e. $\varepsilon_{i/n}(B) = 1$ if $i/n \in B$ and 0 else. Then $N_n$ is the *time normalized point process of exceedances*.

The above equivalence (2.2) extends to the following result.

**Proposition 2.3.** *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of iid rvs with common df $F$. Let $(u_n)_{n\in\mathbb{N}}$ be threshold values tending to $x_F$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} n\,P(X > u_n) = \tau \in (0, \infty) \quad \Longleftrightarrow \quad N_n \stackrel{d}{\to} N \quad \text{Poisson process}(\tau), \quad n \to \infty.$$

From this follows the asymptotic behaviour of all upper order statistics, for instance,

$$P(M_n \leq u_n) = P(N_n((0,1]) = 0) \to P(N((0,1]) = 0) = e^{-\tau}, \quad n \to \infty.$$

## 2.2 Generalized Extreme Value Distribution (GEV)

For statistical purposes all three extreme value distributions are summarized.

**Definition 2.4.** [Jenkinson-von Mises representation]

$$H_{\xi,\beta}(x) = \left\{ \begin{array}{ll} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & \text{if } \xi \neq 0, \\ \exp\left\{-e^{-x}\right\} & \text{if } \xi = 0, \end{array} \right.$$

*where $1 + \xi x > 0$ and $\xi$ is the* shape *parameter.*

The GEV represents all three extremal types:

- $\xi > 0$ Fréchet $\quad Q_\xi((x-1)/\xi) = \Phi_{1/\xi}(x),$

- $\xi = 0$ Gumbel $\quad Q_0(x) = \Lambda(x),$

- $\xi < 0$ Weibull $\quad Q_\xi(-(x+1)/\xi) = \Psi_{-1/\xi}(x).$

Additionally, we introduce location and scale parameters $\mu \in \mathbb{R}$ and $\psi > 0$ and define $Q_{\xi;\mu,\psi}(x) = Q_\xi(x-\mu)/\psi$. Note that $Q_{\xi;\mu,\psi}$ is of the type of $Q_\xi$.

This representation is useful for any statistical method which can be based on iid maxima. These are then modelled by the GEV and the parameters are fitted leading to tail and quantile estimates; see EKM [33], Section 6.3. The method has its limitations, in particular, if the dependence structure cannot be embedded in an iid maxima model. Moreover, as for instance the method of annual maxima, it can also be a waste of data material, since it may only use annual maxima, but ignore all other large values of the sample. An excellent remedy, also for non iid data originates in the hydrology literature and has been developed and very successfully applied by Richard Smith and his collaborators for the last decades; see Smith [84] and references therein.

## 2.3 The POT–method ("Peaks Over Threshold")

We explain the POT-method and show it at work for the DAX closing prices of Figure 1.1. A superficial glimpse at the data shows already some of the so-called stylized facts of financial data. There are more peaks than can be explained by a normal model and, in particular, the negative peaks are more pronounced than the positive ones. On the other hand, the data are simple in their dependence structure; an analysis of the autocorrelations of the data, their absolute values and their squares gave no indication of dependence. Consequently, we assume that the data are iid. We want to remark, however, that many financial data are not iid, but exhibit a very delicate dependence structure; see Sections 3 and 4.

We proceed with a simple exploratory data analysis, which should stand at the beginning of every risk analysis. In a *QQ-plot*, empirical quantiles are plotted against the theoretical quantiles of a given distribution. If the chosen model is correct, nearly all data points will (if the sample size is large enough) lie on the 45-degree line. If the chosen distribution is correct up to its scale and location parameter, the plotted points will still be on a straight line, however with different slope and intersect. Linear regression gives rough estimates for the scale and location parameter, and these are often used as starting values for more sophisticated estimation methods. Figure 2.5
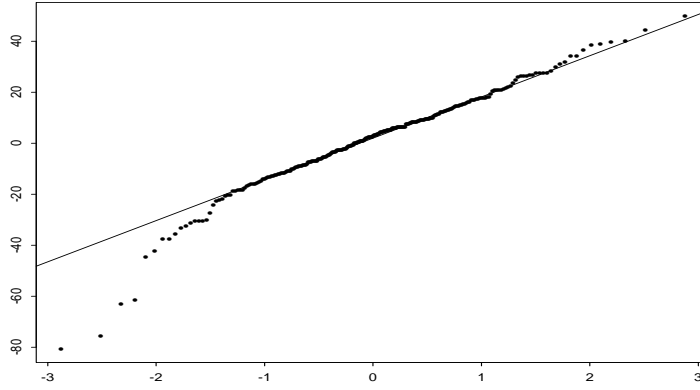
**Figure 2.5.** *Normal QQ-plot of the daily price changes of the DAX closing prices. The fit is in particular in the region of interest at the left end very bad.*

shows a normal QQ-plot of the data. The left end of the plot shows clearly that the left tail of the underlying distribution is much fatter than the left tail of a normal distribution.

Taking the modulus of the negative values of the given sample enables us to apply extreme value theory as introduced above to the left tail of the distribution of daily price changes. This is a sample of size $n = 108$ and will be the basis for the estimation of VaR.

One of the main ingredients of the POT-method is the following result.

**Theorem 2.6.** [Balkema and de Haan [3]/Pickands [71]]

$$F \in \mathrm{MDA}\left(H_\xi\right) \quad \Longleftrightarrow \quad \lim_{u \nearrow x_F} \frac{\overline{F}\left(u + x\beta(u)\right)}{\overline{F}(u)} = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{if } \xi \neq 0, \\ e^{-x} & \text{if } \xi = 0, \end{cases}$$

*where* $1 + \xi x > 0$, *for some (positive measurable) function* $\beta(u)$.

**Interpretation.** For a rv $X$ with df $F \in \mathrm{MDA}\left(H_\xi\right)$ we have

$$\lim_{u \uparrow x_F} P\left(\frac{X - u}{\beta(u)} > x \,\Big|\, X > u\right) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{if } \xi \neq 0, \\ e^{-x} & \text{if } \xi = 0, \end{cases}$$

i.e. given $X$ exceeds $u$, the scaled excess converges in distribution.

**Definition 2.7.** [Excess distribution function, mean excess function (MEF)]
*Let* $X \stackrel{d}{=} F$ *be a rv with* $x_F \leq \infty$. *For fixed* $u < x_F$ *we call*

$$F_u(x) = P\left(X - u \leq x \mid X > u\right), \quad x + u \leq x_F,$$

*the* excess df *of* $X$ *or* $F$ *over the threshold* $u$. *The function*

$$e(u) = E\left[X - u \mid X > u\right] = \int_u^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)} dt, \quad u < x_F, \tag{2.3}$$

*is called* mean excess function *of* $X$ *or* $F$.

It is easy to calculate the mean excess function of an exponential distribution, which is a constant, equal to its parameter. The mean excess function of a distribution with a tail lighter than the tail of an exponential distribution tends to zero as $u$ tends to infinity; for a distribution with tail heavier than exponential, the mean excess function tends to infinity; see Figure 6.2.4 of EKM [33].

Now let $X_1, \ldots, X_n$ denote the sample variables. As usual, $z^+ = \max(z, 0)$ denotes the positive part of $z$ and $\mathrm{card}A$ is the cardinality of the set $A$. The empirical function

$$e_n(u) = \frac{1}{\mathrm{card}\{i : X_i > u, i = 1, \ldots, n\}} \sum_{i=1}^{n} (X_i - u)^+, \quad u \geq 0,$$

estimates the *mean excess function* $e(u)$.

The right-hand side of Figure 2.8 shows the empirical mean excess function of the DAX data corresponding to the left tail. At first, the function is decreasing, but further to the right, it has an upward trend. This shows that in a neighbourhood of zero, the data might possibly be modelled by a normal distribution, but this is certainly not the case in the left tail; there, the distribution turns out to have a tail that is clearly heavier than an exponential tail.
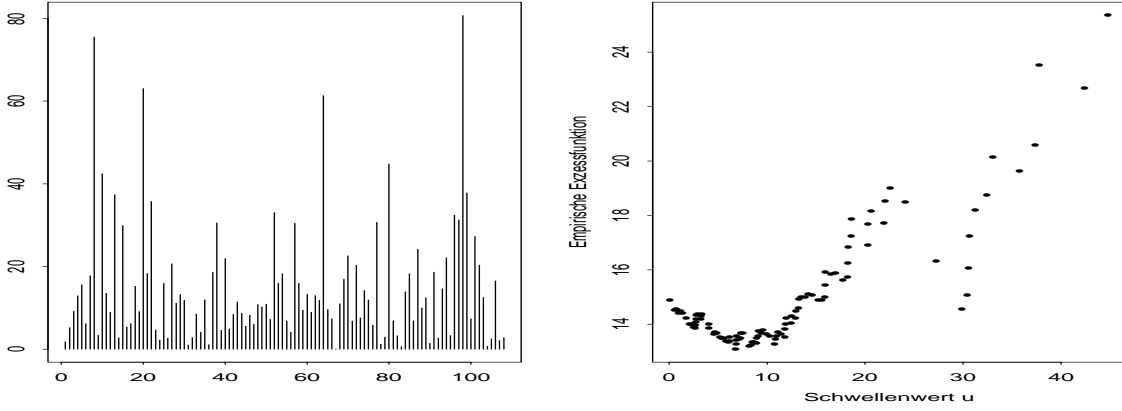


**Figure 2.8.** *The absolute negative price changes (left-hand side) and the corresponding empirical mean excess function (right-hand side) of the DAX values.*

Theorem 2.6 motivates the following definition.

**Definition 2.9.** [Generalized Pareto distribution (GPD)]

$$G_{\xi,\beta}(x) = \left\{ \begin{array}{ll} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-x/\beta} & \text{if } \xi = 0, \end{array} \right.$$

*for $1 + \xi x > 0$. $\xi \in \mathbb{R}$ is the* shape parameter *and $\beta > 0$ is the* scale parameter.

The GPD represents three different limit excess dfs:

- $\xi > 0$ Pareto with support $x \geq 0$ ,

- $\xi = 0$ exponential with support $x \geq 0$ ,

- $\xi < 0$ Pareto with support $0 \leq x \leq -\beta/\xi$.

These results are applied to model data above a high threshold $u$ as follows:

(1) the point process of exceedances by a Poissonprocess($\tau$),
(2) the conditional excesses by a GPD($\alpha, \beta$),
(3) the stochastic quantities of (1) and (2) are independent.

## 2.4 Estimate tails and quantiles by the POT-method

Assume that $(X_n)_{n \in \mathbb{N}}$ are iid and $X_n \overset{d}{=} X \overset{d}{=} F$. For a high threshold $u$ define

$$N_u = \text{card}\{i : X_i > u, i = 1, \ldots, n\}.$$

Define $\overline{F}(u) = 1 - F(u) = P(X_i > u)$, then

$$\overline{F}_u(y) = P(X - u > y \mid X > u) = \frac{\overline{F}(u + y)}{\overline{F}(u)}, \quad y \geq 0,$$

equivalently,

$$\overline{F}(u + y) = \overline{F}(u)\overline{F}_u(y), \quad y \geq 0. \tag{2.4}$$

Estimate $\overline{F}(u)$ and $\overline{F}_u(y)$ by the POT-method:

$$\widehat{\overline{F}(u)} = \frac{1}{n} \sum_{i=1}^{n} I(X_i > u) = \frac{N_u}{n}.$$

Approximate

$$\overline{F}_u(y) \approx \left(1 + \xi \frac{y}{\beta}\right)^{-1/\xi}, \quad y \in \mathbb{R},$$

and estimate $\xi$ and $\beta$ by $\widehat{\xi}$ and $\widehat{\beta}$ (see below). This results in the following tail and quantile estimates:

- **Tail estimate**

$$\widehat{\overline{F}(u + y)} = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{y}{\widehat{\beta}}\right)^{-1/\widehat{\xi}}, \quad y \geq 0. \tag{2.5}$$

- **Quantile estimate**

$$\widehat{x}_p = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left(\left(\frac{n}{N_u}(1 - p)\right)^{-\widehat{\xi}} - 1\right), \quad p \in (0, 1). \tag{2.6}$$

A standard method to estimate the parameters $\xi$ and $\beta$ is maximum likelihood (ML) estimation. It is based on numerically maximising the likehood function for the given data, which are the excesses over a threshold $u$. However, one should bear in mind that the estimation procedure often relies on a very small data set as only the excesses will enter the estimation procedure. For this reason one cannot always rely on the asymptotic optimality properties of the ML-estimators and should therefore possibly use other estimation methods for comparison. For

example, the classic *Hill estimator* could be used as an alternative approach. For a derivation and representation of the Hill estimator as well as a comparison to other tail estimators, see e.g. EKM [33], Chapter 6.

As already mentioned, the ML estimation is based on excess data, hence making it necessary to choose a threshold parameter $u$. A useful tool here is the plot of the empirical mean excess function in Figure 2.8. Recall that for heavy-tailed distributions the mean excess function in (2.3) tends to infinity. Furthermore it can be shown that for the generalised Pareto distribution, the mean excess function is a linear function (increasing if and only if the parameter $\xi$ is positive). Hence, a possible choice of $u$ is given by the value, above which the empirical mean excess function is approximately linear. Figure 2.8 indicates a reasonable choice of $u = 10$, with corresponding $N_u = 56$. This indicates that the generalised Pareto distribution is not only a good model for the extreme negative daily price changes, but already for about half of them. The ML-estimators are then found to be

$$\widehat{\xi} = 0.186\,, \qquad \widehat{\beta} = 11.120\,,$$

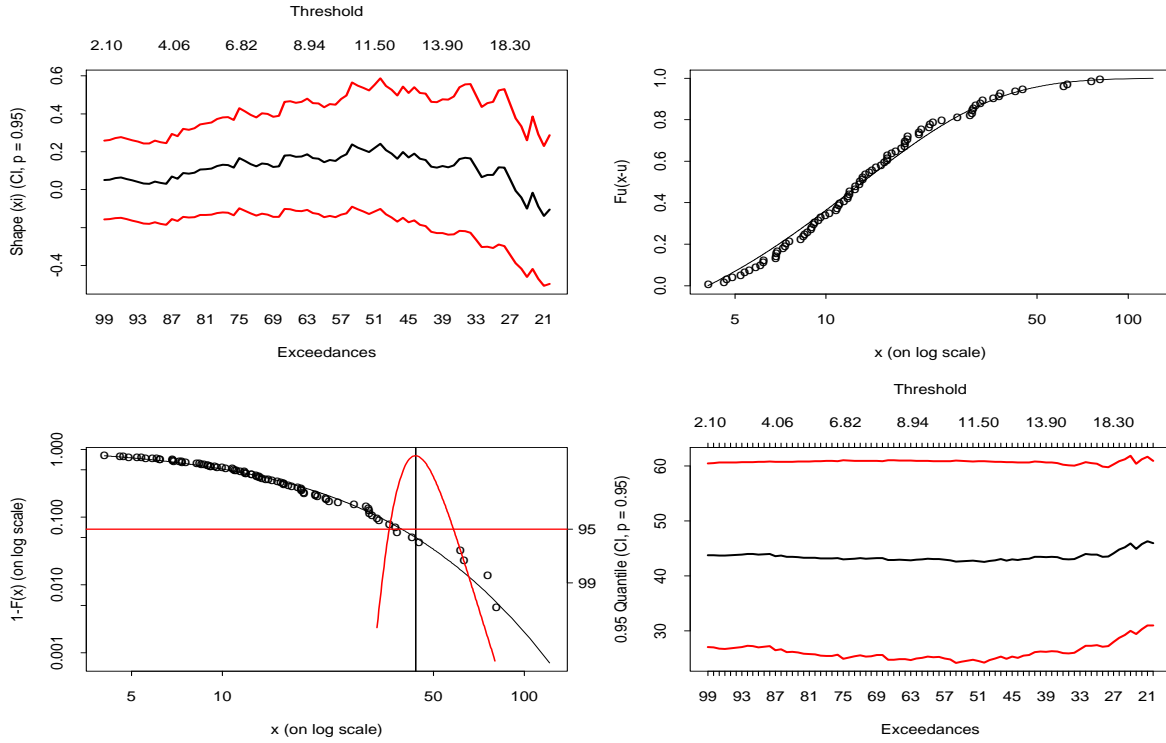which enable us to estimate the lower 5%–quantile of the daily price changes.



**Figure 2.10.** *Extreme value analysis of the data. The upper left-hand plot shows the estimated shape parameter $\widehat{\xi}$ with pointwise confidence intervals based on the normal asymptotics of the estimator, depending on different threshold values $u$. The upper right-hand plot shows the fit of the conditional df, and the lower left-hand plot the tail-fit of the DAX daily price changes: the plotted data points are the 56 largest absolute price changes; the solid curves show the estimated df and tail based on these data. In the lower left-hand plot, the vertical line marks the estimated 95%–quantile, the curve above is the corresponding profile likelihood. The lower right-hand plot shows the estimated 95%-quantile with pointwise confidence bounds depending on the threshold values.*

This estimator leads for the DAX data of Figure 1.1 to the following table.

|  | empirical | normal | GPD |
|---|---|---|---|
| VaR(1day, $p = 0.05$) | -30.654 | -29.823 | -42.856 |

As is obvious from this Table and Figure 2.10 estimation of the quantile by means of extreme value theory results in a much larger risk estimate as for the empirical and normal method.

The estimates fit the given data quite nicely, even in the far end tail. This confirms that the assumption of an underlying heavy tailed distribution is well in line with the data. In this context, the corresponding estimate of the lower 5%–quantile of the VaR seems far more plausible than those obtained under the assumption of a normal distribution.

Confidence intervals for the estimated quantile can easily be obtained from the plotted profile likelihood. The 95%–confidence interval can be read off the horizontal line. It is the interval $[34.37, 60.51]$, i.e. with probability 0.95 the 95%–quantile will lie in the interval $[34.37, 60.51]$. Not surprisingly, the confidence interval is rather wide, in particular to the right, where very few data are to be found. For the definition and mathematical properties of the profile likelihood we refer to Barndorff-Nielsen and Cox [6].

# 3    Continuous-time diffusion models

In this section, which is based on Borkovec and Klüppelberg [17], we study the extremal behaviour of diffusion processes defined by the SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad t > 0, \quad X_0 = x, \qquad (3.1)$$

where $W$ is standard Brownian motion, $\mu \in \mathbb{R}$ is the *drift term* and $\sigma > 0$ is the *diffusion coefficient or volatility*.

The stationary distributions of the processes under investigation are often well-known and one might expect that they influence the extremal behaviour of the process in some way. This is however not the case: for any pre-determined stationary distribution the process can exhibit quite different behaviour in its extremes.

Extremal behaviour of a stochastic process $(X_t)_{t \geq 0}$ is as a first step manifested in the asymptotic behaviour of the running maxima

$$M_t = \max_{0 \leq s \leq t} X_s, \quad t > 0.$$

The asymptotic distribution of $M_t$ for $t \to \infty$ has been studied by various authors, see Davis [25] for detailed references.

It is remarkable that under quite natural conditions running maxima and minima of $(X_t)_{t \geq 0}$ given by (3.1) are asymptotically independent and have the same behaviour as the extremes of iid rvs. We restrict ourselves to the investigation of maxima, the mathematical treatment for minima being similar.

The diffusion $(X_t)_{t \geq 0}$ given by the SDE (3.1) has state space $(l, r) \subseteq \mathbb{R}$. We only consider the case when the boundaries $l$ and $r$ are inaccessible and $(X_t)_{t \geq 0}$ is recurrent. We require furthermore that $\sigma^2(x) > 0$ for all $x \in (l, r)$ and that there exists some $\varepsilon > 0$ such that

$\int_{x-\varepsilon}^{x+\varepsilon} (1 + |\mu(t)|)/\sigma^2(t)dt < \infty$. These two conditions guarantee in particular that the SDE (3.1) has a weak solution which is unique in probability; see Karatzas and Shreve [52], Chapter 5.5.C.

Associated with the diffusion is the scale function $s$ and the speed measure $m$. The *scale function* is defined as

$$s(x) = \int_z^x \exp\left(-2\int_z^y \frac{\mu(t)}{\sigma^2(t)} dt\right) dy, \quad x \in (l, r), \tag{3.2}$$

where $z$ is any interior point of $(l, r)$ whose choice, by the convergence to types theorem, does not affect the extremal behaviour. For the *speed measure* $m$ we know that $m(I) > 0$ for every non-empty open subinterval $I$ of the interior of $(l, r)$. We only consider diffusions with finite speed measure $m$ and denote its total mass by $|m| = m((l, r))$. The speed measure of model (3.1) is absolutely continuous with Lebesgue density

$$m'(x) = \frac{2}{\sigma^2(x)s'(x)}, \quad x \in (l, r),$$

where $s'$ is the Lebesgue density of $s$. In this situation $(X_t)_{t \geq 0}$ is ergodic and its stationary distribution is absolutely continuous with Lebesgue density

$$h(x) = m'(x)/|m|, \quad x \in (l, r). \tag{3.3}$$

Notice that the connection between stationary distribution, speed measure, scale function, drift term and diffusion coefficient (given by (3.2)-(3.3)) allows us to construct diffusions with arbitrary stationary distribution (see Examples 3.6 and Theorems 3.4 and 3.5).

Throughout this section, we assume that the diffusion process $(X_t)_{t \geq 0}$ defined in (3.1) satisfies *the usual conditions*, which guarantee that $(X_t)_{t \geq 0}$ is ergodic with stationary density (3.3):

$$s(r) = -s(l) = \infty \quad \text{and} \quad |m| < \infty. \tag{3.4}$$

For proofs of the above relations and further results on diffusions we refer to the monographs Karatzas and Shreve [52], Revuz and Yor [76], Rogers and Williams [78], or any other advanced textbook on stochastic processes.

The following formulation can be found in Davis [25].

**Proposition 3.1.** *Let $(X_t)_{t \geq 0}$ satisfy the usual conditions (3.4). Then for any initial value $X_0 = y \in (l, r)$ and any $u_t \uparrow r$,*

$$\lim_{t \to \infty} |P^y(M_t \leq u_t) - F^t(u_t)| = 0,$$

*where $F$ is a df, defined for any $z \in (l, r)$ by*

$$F(x) = \exp\left(-\frac{1}{|m|s(x)}\right), \quad x \in (z, r). \tag{3.5}$$

*The function $s$ and the quantity $|m|$ also depend on the choice of $z$.* □

Various proofs of this result exist and we refer to Davis [25] for further references. Davis' proof is based on a representation of such a diffusion as an Ornstein-Uhlenbeck process after a random time-change. Standard techniques for extremes of Gaussian processes apply leading to the above result. (The idea is explained in the proof of Theorem 3.8).

As already noted the scale and speed measure of a diffusion $(X_t)_{t\geq 0}$ depend on the choice of $z$ and hence, are not unique. Different scale and speed measures (and therefore different $z$) lead to different df's $F$ in Proposition 3.1. They are, however, all tail-equivalent. This follows immediately by a Taylor expansion from (3.5) and the fact that $s(x) \to \infty$ as $x \uparrow r$.

**Corollary 3.2.** *Under the conditions of Proposition 3.1 the tail of the df $F$ in (3.5) satisfies*

$$\overline{F}(x) \sim \left( |m| \int_z^x s'(y)dy \right)^{-1} \sim (|m|s(x))^{-1} , \quad x \uparrow r .$$

The extremal behaviour (in particular the behaviour of the maximum) of an iid sequence with common df $F$ is determined by the far end of the right tail $\overline{F}$. In our situation the asymptotic behaviour of the maxima $M_t$ is determined by the tail of $F$ as in (3.5): if $F \in \mathrm{MDA}(Q)$ with norming constants $a_t > 0$ and $b_t \in \mathbb{R}$, then

$$a_t^{-1} (M_t - b_t) \overset{d}{\to} Q , \quad t \to \infty . \tag{3.6}$$

The notion of regular variation is central in extreme value theory and we refer to Bingham, Goldie and Teugels [11], which we henceforth abbreviate by BGT.

**Definition 3.3.** [Regular variation]
*A positive measurable function $f$ on $(0, \infty)$ is regularly varying at $\infty$ with index $\alpha$ (we write $f \in \mathcal{R}(\alpha)$) if*

$$\lim_{x \to \infty} \frac{f(tx)}{f(x)} = t^\alpha , \quad t > 0 .$$

The following results describe the different behaviour of diffusions (3.1) with stationary density $h$ by the df $F$ which governs the extreme behaviour.

**Theorem 3.4.** *Assume that the usual conditions (3.4) hold.*

(a)    *If $\mu \equiv 0$, then $S = (-\infty, \infty)$ and $\overline{F}(x) \sim cx^{-1}$ as $x \to \infty$ for some $c > 0$.*

(b)    *Let $\mu$ and $\sigma$ be differentiable functions in some left neighbourhood of $r$ such that*

$$\lim_{x \to r} \frac{d}{dx} \frac{\sigma^2(x)}{\mu(x)} = 0 \quad and \quad \lim_{x \to r} \frac{\sigma^2(x)}{\mu(x)} \exp \left( -2 \int_z^x \frac{\mu(t)}{\sigma^2(t)} dt \right) = -\infty ,$$

*then*

$$\overline{F}(x) \sim |\mu(x)|h(x) , \quad x \uparrow r .$$

**Theorem 3.5.**    *Assume that the usual conditions (3.4) hold and $r = \infty$.*

(a)    *If $\sigma^2(x) \sim x^{1-\delta}\ell(x)/h(x)$ as $x \to \infty$ for some $\delta > 0$ and $\ell \in \mathcal{R}(0)$, then*

$$\overline{F}(x) \sim \frac{\delta}{2} x^{-\delta}\ell(x), \quad x \to \infty .$$

(b)    *If $\sigma^2(x) \sim cx^{\delta-1}e^{-\alpha x^\beta}/h(x)$ as $x \to \infty$ for some $\delta \in \mathbb{R}$ and $\alpha, \beta, c > 0$, then*

$$\overline{F}(x) \sim \frac{1}{2}c\alpha\beta x^{\delta+\beta-2} \exp(-\alpha x^\beta), \quad x \to \infty .$$

15

The following example describes the simplest way to construct a diffusion process with prescribed stationary density $h$.

**Example 3.6.** Define $dX_t = \sigma(X_t)dW_t$, $t > 0$, and $X_0 = x \in (l, r)$ and $\sigma^2(x) = \sigma^2/h(x)$ for $\sigma > 0$ and some density $h$. Then $\mu(x) = 0$, $s'(x) = 1$ and $(X_t)_{t \geq 0}$ has stationary density $h$. As a consequence of Theorem 3.4(a) this example has a very special extremal behaviour, which is – independent of $h$ – the same for all $h$.

Next we investigate an analogue of the Poisson process approximation for iid data; see Proposition 2.3. Since $(X_t)_{t \geq 0}$ has sample paths with infinite variation, we introduce a discrete skeleton in terms of a point process of so-called $\varepsilon$-upcrossings of a high threshold $u$ by $(X_t)_{t \geq 0}$. For fixed $\varepsilon > 0$ the process has an $\varepsilon$-upcrossing at $t$ if it has remained below $u$ on the interval $(t - \varepsilon, t)$ and is equal to $u$ at $t$. Under weak conditions, the point process of $\varepsilon$-upcrossings, properly scaled in time and space, converges in distribution to a homogeneous Poisson process, i.e. it behaves again like exceedances of iid rvs, coming however not from the stationary distribution of $(X_t)_{t \geq 0}$, but from the df $F$ which describes the growths of the running maxima $M_t$, $t > 0$ (see Proposition 3.1).

**Definition 3.7.** Let $(X_t)_{t \geq 0}$ be a diffusion satisfying the usual conditions (3.4). Take $\varepsilon > 0$.

(a) The process $(X_t)_{t \geq 0}$ is said to have an $\varepsilon$–upcrossing of the level $u$ at $t_0 > 0$ if

$$X_t < u \quad for \quad t \in (t_0 - \epsilon, t_0) \quad and \quad X_{t_0} = u \,.$$

(b) For $t > 0$ let $N_{\varepsilon,u}(t)$ denote the number of $\varepsilon$-upcrossings of $u$ by $(X_s)_{0 \leq s \leq t}$. Then

$$N_t^*(B) = N_{\varepsilon,u_t}(tB) = \mathrm{card}\{\varepsilon\text{-upcrossings of } u_t \text{ by } (X_s)_{0 \leq s \leq t} : \frac{s}{t} \in B\}, \quad B \in \mathcal{B}(0,1]$$

is the time–normalised point process of $\varepsilon$-upcrossings on the Borel sets $\mathcal{B}(0, 1]$.  □

Immediately from the definition $\varepsilon$–upcrossings of a continuous time process correspond to exceedances of a discrete time sequence. As we known from Proposition 2.3 the point process of exceedances of iid data converge weakly to a homogeneous Poisson process. Such results also hold for more general sequences provided the dependence structure is nice enough to prevent clustering of the extremes in the limit.

For diffusions (3.1) the dependence structure of the extremes is such that the point processes of $\varepsilon$-upcrossings converge to a homogeneous Poisson process, however, the intensity is not determined by the stationary df $H$, but by the df $F$ from Proposition 3.1. This means that the $\varepsilon$-upcrossings of $(X_t)_{t \geq 0}$ are likely to behave as the exceedances of iid rvs with df $F$. The extra condition (3.7) of the following theorem relates the scale function $s$ and speed measure $m$ of $(X_t)_{t \geq 0}$ to the corresponding quantities $s_{ou}$ and $m_{ou}$ of the standard Ornstein-Uhlenbeck process, defined by

$$s_{ou}(x) = \sqrt{2\pi} \int_0^x e^{t^2/2} dt \quad and \quad m'_{ou}(x) = 1/s'_{ou}(x), \quad x \in \mathbb{R}.$$

**Theorem 3.8.** Let $(X_t)_{t \geq 0}$ satisfy the usual conditions (3.4) and $u_t \uparrow r$ such that

$$\lim_{t \to \infty} t\overline{F}(u_t) = \lim_{t \to \infty} \frac{t}{|m|s(u_t)} = \tau \in (0, \infty) \,.$$

*Assume there exists some positive constant c such that*

$$\frac{m'_{ou}\big(s_{ou}^{-1}(s(z))\big)}{s'_{ou}\big(s_{ou}^{-1}(s(z))\big)}\frac{s'(z)}{m'(z)}\geq c\,,\quad \forall z\in(l,r)\,. \tag{3.7}$$

*Then for all starting points $y\in(l,r)$ of $(X_t)_{t\geq 0}$ and $\varepsilon>0$ the time-normalised point processes $N_t^*$ of $\varepsilon$-upcrossings of the levels $u_t$ converge in distribution to $N$ as $t\uparrow\infty$, where $N$ is a homogeneous Poisson process with intensity $\tau$ on $(0,1]$.*

**Proof.** The proof invokes a random time change argument. An application of Theorem 12.4.2 of LLL [62] shows that the theorem holds for the standard Ornstein-Uhlenbeck $(O_t)_{t\geq 0}$ process. Denote by

$$Z_t = s_{ou}(O_t)\,,\quad t\geq 0\,,\quad\text{and}\quad Y_t = s(X_t)\,,\quad t\geq 0\,,$$

the Ornstein-Uhlenbeck process and our diffusion, both in natural scale. $(Y_t)_{t\geq 0}$ can then be considered as a random time change of the process $(Z_t)_{t\geq 0}$; i.e. for all $t\geq 0$,

$$Y_t = Z_{\tau_t}\quad a.s.$$

for some stochastic process $(\tau_t)_{t\geq 0}$. The random time $\tau_t$ has a representation via the local time of the process $(Y_t)_{t\geq 0}$. This is a consequence of the Dambis-Dubins-Schwarz Theorem (Revuz and Yor [76], Theorem 1.6, p. 170), Theorem 47.1 of Rogers and Williams [78], p. 277 and Exercise 2.28 of [76], p. 230. For $z\in(l,r)$ denote by $L_t(z)$ the local time of $(Y_s)_{0\leq s\leq t}$ in $z$. Then by the occupation time formula (cf. Revuz and Yor [76], p. 209)

$$\tau_t \;=\; \int_{-\infty}^{\infty} L_t(z)\,dm_{ou}(s_{ou}^{-1}(z)) \;=\; \int_0^t \frac{m'_{ou}\big(s_{ou}^{-1}(s(X_s))\big)}{s'_{ou}\big(s_{ou}^{-1}(s(X_s))\big)}\frac{s'(X_s)}{m'(X_s)}\,ds\,,\quad t\geq 0\,.$$

Notice also that $\tau_t$ is continuous and strictly increasing in $t$; i.e. it defines a random time. Under condition (3.7) we obtain

$$\tau_t - \tau_{t-\varepsilon} \geq c\varepsilon\,,\quad t\geq 0\,.$$

Moreover, Itô and McKean [49], p. 228 proved the following ergodic theorem

$$\frac{\tau_t}{t}\;\overset{\text{a.s.}}{\to}\;\frac{1}{|m|}\,.$$

The following approximations can be made precise and implies Proposition 3.1.

$$P\left(\max_{0\leq s\leq t} X_s > u_t\right) \;=\; P\left(\max_{0\leq s\leq t} Y_s > s(u_t)\right)$$

$$=\; P\left(\max_{0\leq s\leq\tau_t} Z_s > s(u_t)\right)\;\sim\; P\left(\max_{0\leq s\leq t/|m|} Z_s > s(u_t)\right)$$

$$\sim\; P\left(Z_s > s(u_t)^{t/|m|}\right)\;\sim\;\left(\exp\left(-\frac{1}{s(u_t)}\right)\right)^{t/|m|}$$

$$=\; \exp\left(-\frac{t}{s(u_t)|m|}\right)\,,\qquad t\to\infty,\quad u_t\uparrow r\,.$$

For the point process convergence we use Theorem 4.7 of Kallenberg [51] and prove that for any $y\in(l,r)$

$$\lim_{t\to\infty} P^y(N_{\varepsilon,u_t}^X(tU) = 0) = P(N(U) = 0)\,,$$

17

where $U$ is an arbitrary union of semi-open intervals. $\qquad\square$

Theorem 3.8 describes the asymptotic behaviour of the number of $\varepsilon$-upcrossings of a suitably increasing level. In particular, on average there are $\tau$ $\varepsilon$-upcrossings of $u_t$ by $(X_s)_{0\le s\le t}$ for large $t$. Notice furthermore, that we get a Poisson process in the limit which is independent of the choice of $\varepsilon > 0$.

The next lemma provides simple sufficient conditions, only on scale function and speed measure of $(X_t)_{t\ge 0}$, for (3.7). Notice that by positivity and continuity, (3.7) holds automatically on compact intervals. It remains to check this condition for $z$ in a neighbourhood of $r$ and $l$.

**Lemma 3.9.** *Assume that for $c_1, c_2 \in (0\ \infty]$*

$$\frac{1}{\ln(|s(z)|)s(z)}\left(\frac{s''(z)}{s'(z)m'(z)} - \frac{m''(z)}{(m'(z))^2}\right) \longrightarrow \begin{array}{ll} c_1 & z\uparrow r\,, \\ c_2 & z\downarrow l\,, \end{array} \tag{3.8}$$

*or (Grigelionis [45]) that for $d_1, d_2 \in (0, \infty]$*

$$\frac{s^2(z)h(z)\ln(|s(z)|)}{s'(z)} \longrightarrow \begin{array}{ll} d_1 & z\uparrow r\,, \\ d_2 & z\downarrow l\,, \end{array} \tag{3.9}$$

*then (3.7) holds.*

In the following we investigate some examples which have been prominent in the interest rate modelling. All examples have a linear drift term

$$\mu(x) = c - dx\,, \quad x \in (l, r)\,, \quad \text{for } c \in \mathbb{R}, d > 0\,,$$

which implies that the stationary version of $(X_t)_{t\ge 0}$ has mean $c/d$, provided it exists, and is mean reverting with force $d$. For financial background we refer to Lamberton and Lapeyre [61] or Merton [66].

Furthermore, $(X_t)_{t\ge 0}$ has state space $\mathbb{R}$ or $\mathbb{R}_+$, hence $F \in \text{MDA}(\Phi_\alpha)$ for some $\alpha > 0$ or $F \in \text{MDA}(\Lambda)$. Note that (3.6) implies that

$$\frac{M_t}{a_t} \quad \stackrel{d}{\to} \quad \Phi_\alpha \quad \text{if} \quad F \in \text{MDA}(\Phi_\alpha) \tag{3.10}$$

and

$$\frac{M_t - b_t}{a_t} \quad \stackrel{d}{\to} \quad \Lambda \quad \text{and} \quad \frac{M_t}{b_t} \quad \stackrel{P}{\to} \quad 1 \quad \text{if} \quad F \in \text{MDA}(\Lambda)\,. \tag{3.11}$$

Figures 3.11, 3.13, 3.15 and 3.16 show simulated sample paths of the different models. For simulation methods of solutions of SDEs see Kloeden and Platen [56]. The solid line indicates those norming constants which describe the increase of $M_t$ for large $t$, i.e. in $\text{MDA}(\Phi_\alpha)$ we plot $a_t$ (see (3.10)) and in $\text{MDA}(\Lambda)$ we plot $b_t$ (see (3.11)).

Furthermore, all models in this section except the generalised Cox-Ingersoll-Ross model with $\gamma = 1$ satisfy condition (3.8) of Lemma 3.9, hence the Poisson approximation of the $\varepsilon$–upcrossings is also explicitly given for $u_t = a_t x + b_t$ and $\tau = -\ln Q(x)$, where $Q$ is either $\Phi_\alpha$ or $\Lambda$.

**Example 3.10.** [The Vasicek model (Vasicek [87])]
In this model the diffusion coefficient is $\sigma(x) \equiv \sigma > 0$. The solution of the SDE (3.1) with $X_0 = x$ is given by

$$X_t = \frac{c}{d} + (x - \frac{c}{d})e^{-dt} + \sigma \int_0^t e^{-d(t-s)} dW_s\,, \quad t \ge 0\,.$$
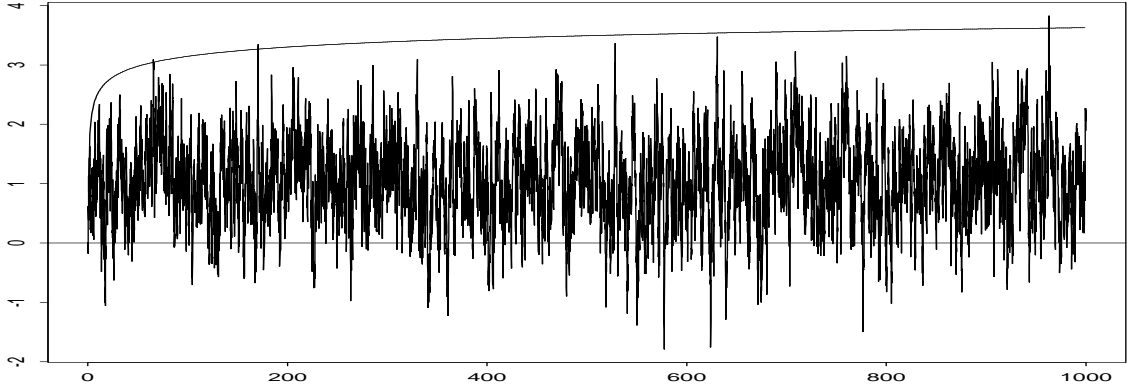
**Figure 3.11.** *Simulated sample path of the Vasicek model (with parameters $c = d = \sigma = 1$) and corresponding normalising constants $b_t$.*

$(X_t)_{t\geq 0}$ has state space $\mathbb{R}$, mean value and variance function

$$EX_t = \frac{c}{d} + (x - \frac{c}{d})e^{-dt} \to \frac{c}{d} \quad \text{and} \quad \text{var}X_t = \frac{\sigma^2}{2d}\left(1 - e^{-2dt}\right) \to \frac{\sigma^2}{2d}, \quad t \to \infty.$$

It is well-known and easy to calculate from (3.2)-(3.3) that $(X_t)_{t\geq 0}$ has a normal stationary distribution, more precisely, it is $N(\frac{c}{d}, \frac{\sigma^2}{2d})$, where $N(a, b)$ denotes the normal distribution with mean $a$ and variance $b$. The assumptions of Theorem 3.4(b) are satisfied giving

$$\overline{F}(x) \quad \sim \quad \frac{2d^2}{\sigma^2}\left(x - \frac{c}{d}\right)^2 \overline{H}(x), \quad x \to \infty,$$

where $\overline{H}(x)$ is the tail of the stationary normal distribution; hence $F$ has heavier tail than $H$. It can be shown that $F \in \text{MDA}(\Lambda)$ with norming constants

$$a_t = \frac{\sigma}{2\sqrt{d\ln t}} \quad \text{and} \quad b_t = \frac{\sigma}{\sqrt{d}}\sqrt{\ln t} + \frac{c}{d} + \frac{\sigma}{4\sqrt{d}}\frac{\ln\ln t + \ln(\sigma^2 d/2\pi)}{\sqrt{\ln t}}.$$

**Example 3.12.** [The Cox-Ingersoll-Ross model (Cox, Ingersoll and Ross [23])]
In this model $\sigma(x) = \sigma\sqrt{x}$ for $\sigma > 0$ and $2c \geq \sigma^2$. It has state space $(0, \infty)$, for $X_0 = x$ it has mean value function

$$EX_t = \frac{c}{d} + \left(x - \frac{c}{d}\right)e^{-dt} \quad \to \quad \frac{c}{d}, \quad t \to \infty$$

and variance function

$$\text{var}X_t = \frac{c\sigma^2}{2d^2}\left(1 - \left(1 + \left(x - \frac{c}{d}\right)\frac{2d}{c}\right)e^{-2dt} + \left(x - \frac{c}{d}\right)\frac{2d}{c}e^{-3dt}\right) \quad \to \quad \frac{c\sigma^2}{2d^2}, \quad t \to \infty.$$

From (3.2)-(3.3) we obtain that the stationary distribution $H$ is $\Gamma(\frac{2c}{\sigma^2}, \frac{2d}{\sigma^2})$. Theorem 3.4(b) applies giving

$$\overline{F}(x) \sim \frac{2cd}{\sigma^2}\overline{G}(x) \sim Ax\overline{H}(x), \quad x \to \infty.$$

19

where $A > 0$ and $\overline{G}(x)$ is the tail of the $\Gamma(\frac{2c}{\sigma^2} + 1, \frac{2d}{\sigma^2})$ distribution. The gamma distributions are in MDA($\Lambda$) and the norming constants for $F$ are

$$a_t = \frac{\sigma^2}{2d} \quad \text{and} \quad b_t = \frac{\sigma^2}{2d}\left(\ln t + \frac{2c}{\sigma^2}\ln\ln t + \ln\left(\frac{d}{\Gamma(2c/\sigma^2)}\right)\right).$$
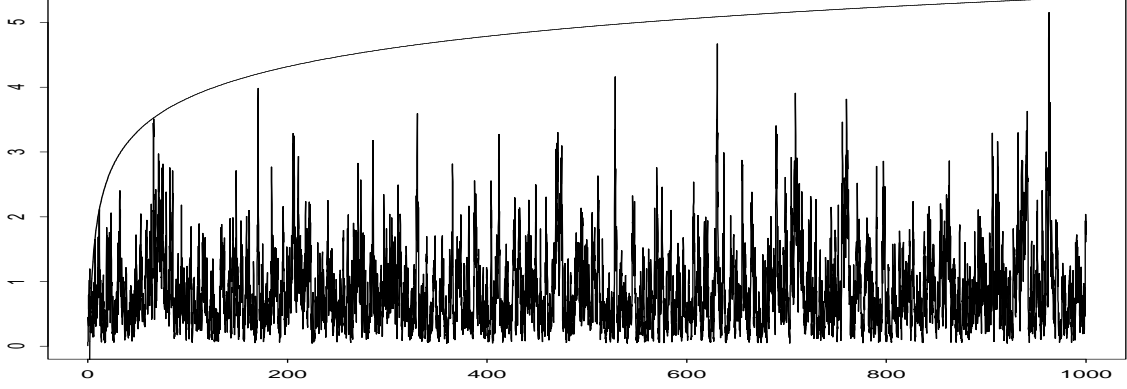


**Figure 3.13.** *Simulated sample path of the Cox-Ingersoll-Ross model with $\mu(x) = c - dx$, $c \in \mathbb{R}, d > 0$ and $\sigma(x) = \sqrt{x}$. (The chosen parameters are $c = d = \sigma = 1$). The stationary distribution is a gamma distribution. The solid line shows the corresponding norming constants $b_t$.*

**Example 3.14.** [The Generalised Cox-Ingersoll-Ross model]
In this model $\sigma(x) = \sigma x^\gamma$ for $\gamma \in [\frac{1}{2}, \infty)$. The process is ergodic with state space $(0, \infty)$.

We distinguish the following four cases:

$$
\begin{array}{lllll}
\gamma = 1/2 & : & 2c \geq \sigma^2, & d > 0 & \text{(see Example 3.12)} \\
1/2 < \gamma < 1 & : & c > 0, & d \geq 0 & \\
\gamma = 1 & : & c > 0, & d > -\sigma^2/2 & \\
\gamma > 1 & : & c > 0, d \in \mathbb{R} & \text{or} \quad c = 0, d < 0. &
\end{array}
\tag{3.12}
$$

For $\frac{1}{2} \leq \gamma \leq 1$ the mean value function of $(X_t)_{t \geq 0}$ is given by

$$
EX_t = \begin{cases}
\dfrac{c}{d} + \left(x - \dfrac{c}{d}\right)e^{-dt} & \to \dfrac{c}{d} & \text{if} \quad d > 0 \\
\dfrac{c}{d} + \left(x - \dfrac{c}{d}\right)e^{-dt} & \to \infty & \text{if} \quad d < 0 \\
x + ct & \to \infty & \text{if} \quad d = 0
\end{cases}
$$

as $t \to \infty$ where $X_0 = x$. The lack of a first moment indicates already that for certain parameter values the model can capture very large fluctuations in data, which will reflect also in the behaviour of the maxima.

- $\frac{1}{2} < \gamma < 1$

The stationary density, which can be calculated by (3.2)-(3.3), is for some norming constant $A > 0$

$$
h(x) = \frac{2}{A\sigma^2}x^{-2\gamma}\exp\left(-\frac{2}{\sigma^2}\left(\frac{c}{2\gamma - 1}x^{-(2\gamma-1)} + \frac{d}{2 - 2\gamma}x^{2-2\gamma}\right)\right), \quad x > 0.
$$

20

The assumptions of Theorem 3.4(b) are satisfied and hence

$$\overline{F}(x) \sim dxh(x) \sim Bx^{2(1-\gamma)}\overline{H}(x), \quad x \to \infty,$$

for some $B > 0$. Then $F \in \mathrm{MDA}(\Lambda)$ with norming constants

$$a_t = \frac{\sigma^2}{2d}\left(\frac{\sigma^2(1-\gamma)}{d}\ln t\right)^{\frac{2\gamma-1}{2-2\gamma}}$$

$$b_t = \left(\frac{\sigma^2(1-\gamma)}{d}\ln t\right)^{\frac{1}{2-2\gamma}}\left(1 - \frac{2\gamma-1}{(2-2\gamma)^2}\frac{\ln\left(\frac{\sigma^2(1-\gamma)}{d}\ln t\right)}{\ln t}\right) + a_t \ln\left(\frac{2d}{A\sigma^2}\right).$$

- $\gamma = 1$

In this case the solution of the SDE (3.1) with $X_0 = x$ is explicitly given by

$$X_t = e^{-(d+\frac{\sigma^2}{2})t+\sigma W_t}\left(x + c\int_0^t e^{(d+\frac{\sigma^2}{2})s-\sigma W_s}ds\right), \quad t \geq 0.$$

We obtain from (3.2)–(3.3) that the stationary density is inverse gamma:

$$h(x) = \left(\frac{\sigma^2}{2c}\right)^{-\frac{2d}{\sigma^2}-1}\left(\Gamma\left(\frac{2d}{\sigma^2}+1\right)\right)^{-1}x^{-2d/\sigma^2-2}\exp\left(-\frac{2c}{\sigma^2}x^{-1}\right), \quad x > 0.$$

Notice that $h \in \mathcal{R}(-2d/\sigma^2 - 2)$ and hence by Karamata's theorem (Theorem 1.5.11 of BGT [11]) the tail $\overline{H}$ of the stationary distribution is also regularly varying. This implies that certain moments are infinite:

$$\lim_{t\to\infty} EX_t^\delta = \begin{cases} \left(\dfrac{2c}{\sigma^2}\right)^\delta \dfrac{\Gamma\left(\frac{2d}{\sigma^2}+1-\delta\right)}{\Gamma\left(\frac{2d}{\sigma^2}+1\right)} & \text{if } \delta < \dfrac{2d}{\sigma^2}+1, \\ \infty & \text{if } \delta \geq \dfrac{2d}{\sigma^2}+1. \end{cases}$$

In particular,

$$\lim_{t\to\infty} \mathrm{var}X_t = \begin{cases} \dfrac{2c^2}{d(2d-\sigma^2)} < \infty & \text{if } \dfrac{2d}{\sigma^2} > 1, \\ \infty & \text{if } -1 < \dfrac{2d}{\sigma^2} \leq 1. \end{cases}$$

For the tail of $F$ we obtain by Theorem 3.4(b)

$$\overline{F}(x) \sim Bx^{-2d/\sigma^2-1}, \quad x \to \infty,$$

for some $B > 0$. Hence $\overline{F} \in \mathcal{B}(-1 - 2d/\sigma^2)$, equivalently, $F \in \mathrm{MDA}(\Phi_{1+2d/\sigma^2})$, with norming constants

$$a_t \sim C\,t^{1/(1+2d/\sigma^2)}, \quad t \to \infty,$$
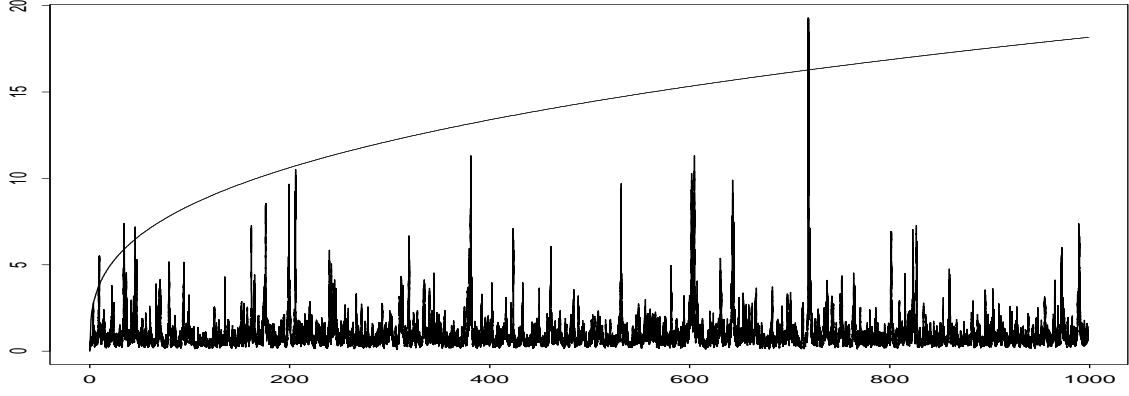
for some $C > 0$.

21

**Figure 3.15.** *Simulated sample path of the generalised Cox-Ingersoll-Ross model with $\mu(x) = c - dx$, $c \in \mathbb{R}, d > 0$ and $\sigma(x) = x^\gamma$ for $\gamma = 1$. (The chosen parameters are $c = d = \sigma = 1$). The solid line shows the corresponding norming constants $b_t$. We can calculate $\overline{F}(x) \sim C\overline{H}(x)$ as $x \to \infty$ for some $C > 0$.*

- $\gamma > 1$

Notice first that $h$ is of the same form as in the case $\frac{1}{2} < \gamma < 1$, in particular $\overline{H} \in \mathcal{R}(-2\gamma + 1)$ with $1 - 2\gamma < -1$)). We apply Theorem 3.5(a) and obtain for some $A > 0$

$$\overline{F}(x) \sim (Ax)^{-1}, \quad x \to \infty.$$

Hence $F \in \mathrm{MDA}(\Phi_1)$ with norming constants $a_t \sim t/A$. Notice that the order of increase of $a_t$ is always linear. The constant $A$, which depends on the parameters, decides about the slope.
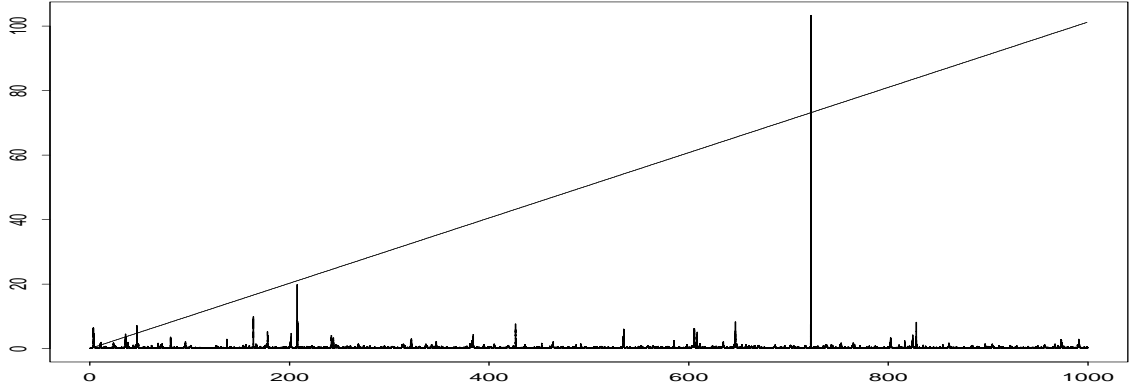


**Figure 3.16.** *Simulated sample path of the generalised Cox-Ingersoll-Ross model for $\gamma = 1.5$ (with parameters $c = d = \sigma = 1$) and the corresponding norming constants $a_t$.*
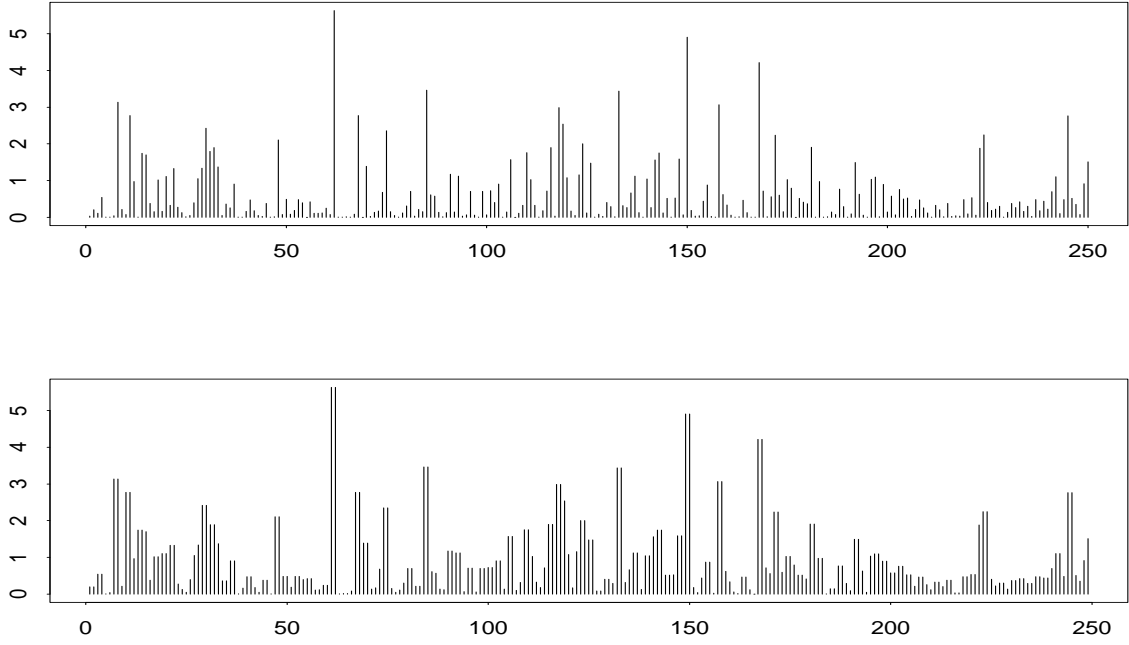
22

**Figure 4.1.** *A realisation of the sequences $(Y_n)_{n \in \mathbb{N}}$ (top) and $(X_n)_{n \in \mathbb{N}}$ (bottom) with $F$ standard exponential as discussed in Example 4.2.*

## 4    The AR(1) model with ARCH(1) errors

In this section we study the extremal behaviour of discrete time volatility models of the form

$$X_n = \mu(X_{n-1}) + \sigma(X_{n-1})\,\varepsilon_n\,, \quad n \in \mathbb{N},$$

where $\mu$ is the conditional mean, $\sigma$ the conditional volatility and $(\varepsilon_n)_{n \in \mathbb{N}}$ are iid symmetric rvs with variance 1.

As a prototype model, which can be analytically analysed we focus on the AR(1) process with ARCH(1) errors, i.e.

$$X_n = \alpha X_{n-1} + \sqrt{\beta + \lambda X_{n-1}^2}\,\varepsilon_n\,, \quad n \in \mathbb{N}, \tag{4.1}$$

where $\alpha \in \mathbb{R}$, $\beta$, $\lambda > 0$, $(\varepsilon_n)_{n \in \mathbb{N}}$ are iid symmetric rvs with variance 1 and $X_0$ is independent of $(\varepsilon_n)_{n \in \mathbb{N}}$. This section is based on Borkovec and Klüppelberg [18] and Borkovec [15]; see also [16].

Before we analyse model (4.1) we explain the influence of volatility clusters on a high level within the context of extreme value theory. We also show its consequences for risk management when estimating a high or low quantile.

We start with an introductory example, which we have found useful before.

**Example 4.2.** [EKM [33], Sections 4.4, 5.5 and 8.1]
Let $Y, Y_1, Y_2, \ldots$ be iid $Y \stackrel{d}{=} \sqrt{F}$ and define $X_n = \max(Y_n, Y_{n+1})$ for $n \in \mathbb{N}$. Then

$$P(X_n \leq x) = (P(Y_n \leq x))^2 = F(x)\,, \quad x \in \mathbb{R}.$$

23

Choose $u_n$ such that $nP(X_1 > u_n) \to \tau$ as $n \to \infty$, then $nP(Y_1 > u_n) \to \tau/2$ and

$$
\begin{aligned}
P(\max(X_1, \ldots, X_n) \le u_n) &= P(\max(Y_1, \ldots, Y_{n+1}) \le u_n) \\
&= P(\max(Y_1, \ldots, Y_n) \le u_n) F(u_n) \to e^{-\tau/2}, \quad n \to \infty.
\end{aligned}
$$

**Definition 4.3.** [Extremal index]
Let $(X_n)_{n \in \mathbb{N}}$ be strictly stationary and define as before

$$
M_1 = X_1, \quad M_n = \max(X_1, \ldots, X_n), \quad n > 1.
$$

Assume that for every $\tau > 0$ there exists a sequence $(u_n)_{n \in \mathbb{N}}$ such that

$$
\begin{aligned}
\lim_{n \to \infty} nP(X_1 > u_n) &= \tau \\
\lim_{n \to \infty} P(M_n \le u_n) &= e^{-\theta \tau}
\end{aligned}
$$

Then $\theta \in [0, 1]$ is called the extremal index of $(X_n)_{n \in \mathbb{N}}$.

The extremal index in Example 4.2 is $\theta = 1/2$. This indicates already the most intuitive interpretation of the extremal index: $1/\theta$ can be interpreted as the mean clustersize.

In the context of risk management we give an intuitive example.

**Example 4.4.** We want to calculate the VaR=VaR(10 days, $p = 0.05$) of a portfolio; i.e. for daily losses $X_i$, $i = 1, \ldots, 10$, we want to estimate $P(\max(X_1, \ldots, X_{10}) \le \text{VaR}) = 0.95$. Assume that we know

VaR(1 day, $p = 0.01$)=10 Mio and VaR(1day, $p = 0.005$)=11 Mio.

For the loss rv $X$ this means that $P(X \le 10) = 0.99$ and $P(X \le 11) = 0.995$. Denote by $Z = \max(X_1, \ldots, X_{10})$. If the $X_i$ are iid, then

$$
P(Z \le 11) = P(X \le 11)^{10} \approx 0.95,
$$

whereas for dependent $X_i$ with extremal index $\theta = 0.5$ we obtain

$$
P(Z \le 10) = P(X \le 10)^{10/2} \approx 0.95.
$$

This means that for iid data the 10-day VaR(10 days, $p = 0.05$) is higher than for dependent data.

Using the block maxima method it is easy to compare VaR estimation for independent and dependent stationary financial time series. The data are divided into, say, $N$ blocks, such that the corresponding block maxima can be considered as independent. Moreover, if the sample variables are in MDA$(Q)$ for some extreme value distribution $Q$, then the block maxima, we call them $Z_1, \ldots, Z_N$, can be viewed as an iid sample of rvs with df $Q$. Consequently, we assume that $Z_1, \ldots, Z_N$ are iid GEV distributed; i.e. introducing a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\psi > 0$,

$$
P(Z \le \text{VaR}(p)) \approx \exp\left( -\left(1 + \xi \frac{\text{VaR}(p) - \mu}{\psi}\right)^{-1/\xi} \right). \tag{4.2}
$$

Defining for given $p \in (0, 1)$ the VaR($p$) by $1 - p = P(Z \leq -\mathrm{VaR}(p))$, we obtain by inversion

$$\mathrm{VaR}(p) = \mu + \frac{\psi}{\xi} \left( (-\ln(1-p))^{-\xi} - 1 \right).$$

By Definition 4.3 dependence introduces an additional factor $\theta$ in the exponent of (4.2) giving

$$\mathrm{VaR}(p) = \mu + \frac{\psi}{\xi} \left( (-\frac{1}{\theta} \ln(1-p))^{-\xi} - 1 \right).$$

In the context of risk management we expect $\xi \geq 0$ and for $\xi = 0$ we take the limit

$$\mathrm{VaR}(p) = \mu - \psi \ln(-\frac{1}{\theta} \ln(1-p)).$$

A different method is a dependent version of the POT-method; i.e. the quantile estimate (2.6). Starting again with (2.4), the estimation of the tail in (2.5) changes, when $\overline{F}(u)$ is estimated. The empirical estimator $N_u/n$ for iid data is replaced by $N_u^b/(n\widehat{\theta}_u)$, where $N_u^b$ is the number of block maxima exceeding $u$ and $\widehat{\theta}_u$ is the estimated extremal index; see EKM [33], Section 8.1 and references therein. For the quantile estimate (2.6) this means that

$$\widehat{x}_p = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left( \left( \frac{n\widehat{\theta}_u}{N_u^b}(1-p) \right)^{-\widehat{\xi}} - 1 \right), \quad p \in (0, 1).$$

## 4.1 Stationarity and tail behaviour

In this section we present an extreme value analysis of the AR(1) process with ARCH(1) errors as given by (4.1). As a prerequisite we first need to know whether we are dealing with a stationary model and what the tail of the stationary distribution looks like.

For $\lambda = 0$ the process is an AR(1) process whose stationary distribution is determined by the innovations $(\varepsilon_n)_{n \in \mathbb{N}}$ and stationarity is guaranteed for $|\alpha| < 1$. In the ARCH(1) case (the case when $\alpha = 0$) the process is geometric ergodic provided that $\beta > 0$ and $0 < \lambda < 2e^{\gamma}$, where $\gamma$ is Euler's constant. The tail of the stationary distribution is known to be Pareto-like (see e.g. Goldie [41] or EKM [33], Section 8.4). This result was obtained by considering the square ARCH(1) process leading to a stochastic recurrence equation which fits in the setting of Kesten [53, 54] and Vervaat [88]; see also Diaconis and Freedman [26] for an interesting overview and Brandt, Franken and Lisek [19]. Goldie and Maller [42] give necessary and sufficient conditions for stationarity of stochastic processes, which are solutions of stochastic recurrence equations.

For the general case we follow the standard procedure as for instance in the case $\alpha = 0$ to find the parameter region of stationarity of the process. For the tail behaviour, however, we apply a technique, which differs completely from Kesten's renewal type arguments, by invoking the Drasin-Shea Tauberian theorem. This approach has the drawback that it ensures regular variation of the stationary tail, but gives no information on the slowly varying function. However, the method does apply to processes which do not fit into the framework of Kesten [53]. Moreover, the Tauberian approach does not depend on additional assumptions which are often very hard to check (as e.g. the existence of certain moments of the stationary distribution). Combining the Tauberian method with results in Goldie [41], we finally specify the slowly varying function as a constant.

We shall need the following assumptions on the noise variables. Denote by $\varepsilon$ a generic rv with the same df $G$ as $\varepsilon_1$. Throughout this section the following *general conditions* are in force:

- $\varepsilon$ is symmetric with variance 1,
- $\varepsilon$ is absolutely continuous with respect to Lebesgue measure with density $g$, which is positive on the whole of $\mathbb{R}$ and decreasing on $\mathbb{R}_+$ . (4.3)

We summarize in Theorem 4.7 some properties of the process $(X_n)_{n \in \mathbb{N}}$. In particular, geometric ergodicity guarantees the existence and uniqueness of a stationary distribution. For an introduction to Markov chain terminology we refer to Tweedie [86] or Meyn and Tweedie [67].

The next proposition follows easily from well-known properties of moment generating functions (one can follow the proof of the case $\alpha = 0$; see e.g. Lemma 8.4.6 of EKM [33]).

**Proposition 4.5.** *Let $\varepsilon$ be a rv with probability density $g$ satisfying the general conditions (4.3). Define $h_{\alpha,\lambda} : [0, \infty) \to [0, \infty]$ for $\alpha \in \mathbb{R}$ and $\lambda > 0$ by*

$$h_{\alpha,\lambda}(u) := E[|\alpha + \sqrt{\lambda}\,\varepsilon|^u], \quad u \geq 0 \,. \tag{4.4}$$

*(a) The function $h_{\alpha,\lambda}(\cdot)$ is strictly convex in $[0, T)$, where*

$$T := \inf\{u \geq 0 \,|\, E[|\sqrt{\lambda}\,\varepsilon|^u] = \infty\} \,.$$

*(b) If furthermore the parameters $\alpha$ and $\lambda$ are chosen such that*

$$h'_{\alpha,\lambda}(0) = E[\ln|\alpha + \sqrt{\lambda}\,\varepsilon|] < 0 \,, \tag{4.5}$$

*then there exists a unique solution $\kappa = \kappa(\alpha, \lambda) > 0$ to the equation $h_{\alpha,\lambda}(u) = 1$. Moreover, under $h'_{\alpha,\lambda}(0) < 0$,*

$$\kappa(\alpha, \lambda) \begin{cases} > 2 \,, & \text{if } \alpha^2 + \lambda E[\varepsilon^2] < 1 \,, \\ = 2 \,, & \text{if } \alpha^2 + \lambda E[\varepsilon^2] = 1 \,, \\ < 2 \,, & \text{if } \alpha^2 + \lambda E[\varepsilon^2] > 1 \,. \end{cases}$$

**Remark 4.6.** (a) By Jensen's inequality $\alpha^2 + \lambda\,E[\varepsilon^2] < 1$ implies $h'_{\alpha,\lambda}(0) < 0$.

(b) Proposition 4.5 holds in particular for a standard normal rv $\varepsilon$. In this case $T = \infty$.

(c) In general, it is not possible to determine explicitly which parameters $\alpha$ and $\lambda$ satisfy (4.5). If $\alpha = 0$ (i.e. in the ARCH(1)-case) and $\varepsilon \overset{d}{=} N(0, 1)$ (4.5) is satisfied if and only if $\lambda \in (0, 2e^\gamma)$, where $\gamma$ is Euler's constant (see e.g. EKM [33], Section 8.4).
For $\alpha \neq 0$, Tables 4.14-4.16 show numerical domains of $\alpha$ and $\lambda$; see Kiefersbeck [55] for more examples.

(d) Note that $\kappa$ is a function of $\alpha$ and $\lambda$. Since $\varepsilon$ is symmetric $\kappa$ does not depend on the sign of $\alpha$. For $\varepsilon \overset{d}{=} N(0, 1)$ we can show that for fixed $\lambda$ the function $\kappa$ is decreasing in $|\alpha|$. See also Table 4.14. $\qquad\square$

**Theorem 4.7.** *Consider the process $(X_n)_{n \in \mathbb{N}}$ in (4.1) with $(\varepsilon_n)_{n \in \mathbb{N}}$ satisfying the general conditions (4.3) and with parameters $\alpha$ and $\lambda$ satisfying (4.5). Then the following assertions hold:*

*(a) Let $\nu$ be the normalized Lebesgue-measure on the interval $[-M, M] \subset \mathbb{R}$; i.e. $\nu(\cdot) := \lambda(\cdot \cap [-M, M])/\lambda([-M, M])$. Then $(X_n)_{n \in \mathbb{N}}$ is an aperiodic positive $\nu$-recurrent Harris chain with regeneration set $[-M, M]$ for $M$ large enough.*

*(b) $(X_n)_{n \in \mathbb{N}}$ is geometric ergodic. In particular, $(X_n)_{n \in \mathbb{N}}$ has a unique stationary distribution and satisfies the strong mixing condition with geometric rate of convergence. The stationary distribution is continuous and symmetric.*

**Remark 4.8.** When we study the stationary distribution of $(X_n)_{n \in \mathbb{N}}$ we may w.l.o.g. assume that $\alpha \geq 0$. For a justification, consider the process $(\widetilde{X}_n)_{n \in \mathbb{N}} = ((-1)^n X_n)_{n \in \mathbb{N}}$ which solves the stochastic difference equation

$$\widetilde{X}_n = -\alpha \widetilde{X}_{n-1} + \sqrt{\beta + \lambda \widetilde{X}_{n-1}^2}\, \varepsilon_n\,, \quad n \in \mathbb{N},$$

where $(\varepsilon_n)_{n \in \mathbb{N}}$ are the same rvs as in (4.1) and $\widetilde{X}_0 = X_0$. If $\alpha < 0$, because of the symmetry of the stationary distribution, we may hence study the new process $(\widetilde{X}_n)_{n \in \mathbb{N}}$. $\qquad\square$

In order to determine the tail of the stationary distribution $F$ we need some additional technical assumptions on $g$ and $\overline{G} = 1 - G$, the density and the distribution tail of $\varepsilon$:

**$D_1$** The lower and upper Matuszewska indices of $\overline{H}$ are equal and satisfy in particular

$$
\begin{aligned}
-\infty \leq \gamma \quad &:= \quad \lim_{\nu \to \infty} \frac{\ln \limsup_{x \to \infty} \overline{H}(\nu x)/\overline{H}(x)}{\ln \nu} \\
&= \quad \lim_{\nu \to \infty} \frac{\ln \liminf_{x \to \infty} \overline{H}(\nu x)/\overline{H}(x)}{\ln \nu} \leq 0\,.
\end{aligned}
$$

**$D_2$** If $\gamma = -\infty$ then for all $\delta > 0$ there exist constants $q \in (0, 1)$ and $x_0 > 0$ such that for all $x > x_0$ and $t > x^q$

$$g\left(\frac{x \pm \alpha t}{\sqrt{\lambda t^2}}\right) \geq (1 - \delta)\, g\left(\frac{x \pm \alpha t}{\sqrt{\beta + \lambda t^2}}\right)\,. \tag{4.6}$$

If $\gamma > -\infty$ then for all $\delta > 0$ there exist constants $x_0 > 0$ and $T > 0$ such that for all $x > x_0$ and $t > T$ the inequality (4.6) holds anyway.

The definition of the lower and upper Matuszewska indices can be found e.g. in BGT [11], p. 68; for the above representation we used Theorem 2.1.5 and Corollary 2.1.6. The case $\gamma = -\infty$ corresponds to a tail which is exponentially decreasing. For $\gamma \in (-\infty, 0]$ condition $D_1$ is equivalent to the existence of constants $0 \leq c \leq C < \infty$ such that for all $\Lambda > 1$, uniformly in $\nu \in [1, \Lambda]$,

$$c(1 + o(1))\nu^\gamma \leq \frac{\overline{G}(\nu x)}{\overline{G}(x)} \leq C(1 + o(1))\nu^\gamma\,, \quad x \to \infty\,. \tag{4.7}$$

In particular, a distribution with a regularly varying tail satisfies $D_1$; the value $\gamma$ is then the tail index. Due to the equality of the Matuszewska indices and the monotonicity of $g$ we obtain easily some asymptotic properties of $\overline{G}$ and of $g$, respectively.

**Proposition 4.9.** *Suppose the general conditions (4.3) and $D_1 - D_2$ hold. Then the following holds:*

*(a) $\lim_{x \to \infty} x^m \overline{G}(x) = 0$ and $E[|\varepsilon|^m] < \infty$ for all $m < -\gamma$.*

*(b) $\lim_{x \to \infty} x^m \overline{G}(x) = \infty$ and $E[|\varepsilon|^m] = \infty$ for all $m > -\gamma$.*

*(c) $\lim_{x \to \infty} x^{m+1} g(x) = 0$ for all $m < -\gamma$.*

*(d) If $\gamma > -\infty$, there exist constants $0 < c \le C < \infty$ such that*

$$c \le \liminf_{x \to \infty} \frac{x \, g(x)}{\overline{G}(x)} \le \limsup_{x \to \infty} \frac{x \, g(x)}{\overline{G}(x)} \le C \,.$$

*Moreover, there exist constants $0 \le d \le D < \infty$ such that for all $\Lambda > 1$, uniformly in $\nu \in [1, \Lambda]$,*

$$d(1 + o(1))\nu^{\gamma - 1} \le \frac{g(\nu x)}{g(x)} \le D(1 + o(1))\nu^{\gamma - 1} \,, \quad x \to \infty \,. \tag{4.8}$$

*Furthermore, in this case (4.8) is equivalent to (4.7) or $D_1$.*

The general conditions (4.3) are fairly simple and can be checked easily, whereas $D_1$ and in particular $D_2$ seem to be quite technical and intractable. Nevertheless, numerous densities satisfy these assumptions.

**Example 4.10.** The following two families of densities satisfy the general conditions (4.3) and $D_1 - D_2$.

(a) $g_{\rho,\theta}(x) \propto \exp(-\theta^{-1}|x|^\rho)$, $x \in \mathbb{R}$, for $\rho, \theta > 0$.
Note that this family includes the Laplace (double exponential for $\rho = 1$) and the normal density with mean 0 ($\rho = 2$).

(b) $g_{a,\rho,\theta}(x) \propto (1 + x^2/\theta)^{-(\rho+1)/2}(1 + a \, \sin(2\pi \ln(1 + x^2/\theta)))$, $x \in \mathbb{R}$, for parameters $\rho > 2$, $\theta > 0$ and $a \in [0, (\rho + 1)/(\rho + 1 + 4\pi))$.
This family includes e.g. the Student-t distribution with parameter $\rho$ (set $a = 0$ and $\theta = \rho$).

The following modification of the Drasin-Shea Theorem (BGT [11], Theorem 5.2.3, p. 273) is the key to our result.

**Theorem 4.11.** *Let $k : [0, \infty) \to [0, \infty)$ be an integrable function and let $(a, b)$ be the maximal open interval (where $a < 0$) such that*

$$\widehat{k}(z) = \int_{(0,\infty)} t^{-z} k(t) \frac{dt}{t} < \infty \,, \quad \text{for } z \in (a, b) \,.$$

*If $a > -\infty$, assume $\lim_{\delta \downarrow 0} \widehat{k}(a + \delta) = \infty$, if $b < \infty$, assume $\lim_{\delta \downarrow 0} \widehat{k}(b - \delta) = \infty$. Let $H$ be a df on $\mathbb{R}_+$ with tail $\overline{H}$. If*

$$\lim_{x \to \infty} \int_{(0,\infty)} k\left(\frac{x}{t}\right) \frac{\overline{H}(t)}{\overline{H}(x)} \frac{dt}{t} = c > 0 \,,$$

*then*

$$c = \widehat{k}(\rho) \text{ for some } \rho \in (a, b) \text{ and } \overline{H}(x) \sim x^\rho l(x) \,, \quad x \to \infty \,,$$

*where $l \in \mathcal{R}(0)$.*

The following is the main theorem of this section.

**Theorem 4.12.** *Suppose $(X_n)_{n\in\mathbb{N}}$ is given by equation (4.1) with $(\varepsilon_n)_{n\in\mathbb{N}}$ satisfying the general conditions (4.3) and $D_1 - D_2$ and with parameters $\alpha$ and $\lambda$ satisfying (4.5). Let $\overline{F}(x) = P(X > x)$, $x \geq 0$, be the right tail of the stationary distribution. Then*

$$\overline{F}(x) \sim c\, x^{-\kappa}, \quad x \to \infty, \tag{4.9}$$

*where*

$$c = \frac{1}{2\kappa} \frac{E\left[\left|\alpha|X| + \sqrt{\beta + \lambda X^2}\,\varepsilon\right|^\kappa - \left|(\alpha + \sqrt{\lambda}\varepsilon)|X|\right|^\kappa\right]}{E\left[|\alpha + \sqrt{\lambda}\varepsilon|^\kappa \ln|\alpha + \sqrt{\lambda}\varepsilon|\right]}$$

*and $\kappa$ is given as the unique positive solution to*

$$E[|\alpha + \sqrt{\lambda}\varepsilon|^\kappa] = 1. \tag{4.10}$$

**Remark 4.13.** (a) Let $E[|\alpha + \sqrt{\lambda}\varepsilon|^\kappa] = h_{\alpha,\lambda}(\kappa)$ be as in Lemma 4.5. Recall that for $\varepsilon \overset{d}{=} N(0,1)$ and fixed $\lambda$, the exponent $\kappa$ is decreasing in $|\alpha|$. This means that the distribution of $X$ gets heavier tails. In particular, the AR(1) process with ARCH(1) errors has for $\alpha \neq 0$ heavier tails than the ARCH(1) process (see also Table 4.14).

(b) Theorem 4.12 together with Proposition 4.5 implies that the second moment of the stationary distribution exists if and only if $\alpha^2 + \lambda E[\varepsilon^2] < 1$. $\qquad\square$

**Idea of Proof.** Recall that $P(\varepsilon > x) = \overline{G}(x)$ with density $g$.

$$
\begin{aligned}
\overline{F}(x) &= \int_{-\infty}^\infty P\left(\alpha t + \sqrt{\beta + \lambda t^2}\,\varepsilon > x\right)\, dF(t)\\
&= \int_0^\infty \left(\overline{G}\left(\frac{x + \alpha t}{\sqrt{\beta + \lambda t^2}}\right) + \overline{G}\left(\frac{x - \alpha t}{\sqrt{\beta + \lambda t^2}}\right)\right)\, dF(t)\\
&\sim \int_0^\infty \left(g\left(\frac{x + \alpha t}{\sqrt{\beta + \lambda t^2}}\right) + g\left(\frac{x - \alpha t}{\sqrt{\beta + \lambda t^2}}\right)\right) x\overline{F}(t)\frac{dt}{t}\\
&= \int_0^\infty k\left(\frac{x}{t}\right) \overline{F}(t) \frac{dt}{t},
\end{aligned}
$$

where

$$k(x) = x\left(g\left(\frac{x + \alpha}{\sqrt{\lambda}}\right) + g\left(\frac{x - \alpha}{\sqrt{\lambda}}\right)\right), \quad x > 0,$$

then

$$\lim_{x\to\infty} \frac{1}{\overline{F}(x)} \int_0^\infty k\left(\frac{x}{t}\right) \overline{F}(t) \frac{dt}{t} = 1. \tag{4.11}$$

Define the transform

$$
\begin{aligned}
\hat{k}(z) = \int_0^\infty t^{-z} k(t)\frac{dt}{t} &= \int_0^\infty t^{-z}\left(g\left(\frac{t+\alpha}{\sqrt{\lambda}}\right) + g\left(\frac{t-\alpha}{\sqrt{\lambda}}\right)\right) dt\\
&= E[|\alpha + \sqrt{\lambda}\varepsilon|^{-z}] < \infty, \quad z \in (-\infty, 1)
\end{aligned}
$$

| $|\alpha|$ | $\lambda$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.4 | 1.8 | 2.2 | 2.6 | 3.0 | 3.4 | 3.5 |
| 0.0 | 12.89 | 6.09 | 3.82 | 2.68 | 2.00 | 1.21 | 0.77 | 0.49 | 0.29 | 0.14 | 0.03 | 0.01 |
| 0.2 | 11.00 | 5.50 | 3.54 | 2.52 | 1.89 | 1.16 | 0.74 | 0.47 | 0.28 | 0.13 | 0.02 | - |
| 0.4 | 8.14 | 4.30 | 2.88 | 2.11 | 1.61 | 1.00 | 0.64 | 0.40 | 0.23 | 0.10 | - | - |
| 0.6 | 5.45 | 3.03 | 2.12 | 1.60 | 1.24 | 0.79 | 0.50 | 0.30 | 0.15 | 0.04 | - | - |
| 0.8 | 3.02 | 1.85 | 1.37 | 1.07 | 0.85 | 0.55 | 0.33 | 0.18 | 0.06 | - | - | - |
| 1.0 | 0.96 | 0.83 | 0.70 | 0.57 | 0.47 | 0.29 | 0.15 | 0.04 | - | - | - | - |
| 1.1 | 0.12 | 0.39 | 0.40 | 0.35 | 0.29 | 0.17 | 0.07 | - | - | - | - | - |
| 1.2 | - | 0.01 | 0.12 | 0.14 | 0.12 | 0.05 | - | - | - | - | - | - |

**Table 4.14.** *Range of stationarity of the AR(1)+ARCH(1) model with parameters $\alpha$ and $\lambda$. The matrix components contain the estimated tail index $\kappa$ for standard normal noise. There is no estimate given if the estimated $\kappa$ is less than $10^{-2}$ or (4.5) is not satisfied.*

Since (4.11) holds, the conditions of the the Drasin-Shea Tauberian theorem are satisfied. Hence there exists some $\rho \in (-\infty, 1)$ such that $\widehat{k}(z) = 1$ and

$$\overline{F}(x) \sim x^{\rho}\ell(x), \quad x \to \infty.$$

But $\widehat{k}(z) = E[|\alpha + \sqrt{\lambda}\varepsilon|^{-z}] = 1$ for $\rho = -\kappa$ and hence

$$\overline{F}(x) \sim x^{-\kappa}\ell(x), \quad x \to \infty. \tag{4.12}$$

We apply now Corollary 2.4 of Goldie [41] to the process $(Y_n)_{n\in\mathbb{N}}$ given by the stochastic recurrence equation

$$Y_n = \left|\alpha Y_{n-1} + \sqrt{\beta + \lambda Y_{n-1}^2}\right|, \quad n \in \mathbb{N}, \quad \text{and} \quad Y_0 = |X_0| \quad \text{a.s.},$$

which satisfies $(Y_n) \stackrel{d}{=} (|X_n|)$. By (4.12) $EY^{\kappa-1} < \infty$ and hence the moment condition

$$E\left|(|\alpha Y + \sqrt{\beta + \lambda Y^2}\varepsilon|)^{\kappa} - (|\alpha + \sqrt{\lambda}\varepsilon|Y)^{\kappa}\right| < \infty$$

requested in Goldie [41] is satisfied. By symmetry of $X$ we conclude finally

$$\ell(x) = c = \frac{E\left[\left|\alpha|X| + \sqrt{\beta + \lambda X^2}\varepsilon\right|^{\kappa} - \left|(\alpha + \sqrt{\lambda}\varepsilon)|X|\right|^{\kappa}\right]}{2\kappa E\left[|\alpha + \sqrt{\lambda}\varepsilon|^{\kappa} \ln|\alpha + \sqrt{\lambda}\varepsilon|\right]}.$$

$\square$

## 4.2   Extreme value analysis

Theorems 4.7 and 4.12 are crucial for investigating the extremal behaviour of $(X_n)_{n\in\mathbb{N}}$. The strong mixing property implies automatically that the sequence $(X_n)_{n\in\mathbb{N}}$ satisfies the conditions $D(u_n)$ and $\Delta(u_n)$. These conditions are frequently used mixing conditions in extreme value theory, which, as we do not need them explicitly, we will not define; instead we refer to Hsing,

| $|\alpha|$ | $\lambda$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.4 | 1.8 | 2.2 | 2.6 | 2.8 |
| 0.0 | 4.00 | 2.76 | 2.00 | 1.50 | 1.14 | 0.69 | 0.41 | 0.21 | 0.07 | 0.02 |
| 0.2 | 3.93 | 2.68 | 1.92 | 1.44 | 1.10 | 0.66 | 0.39 | 0.20 | 0.06 | 0.01 |
| 0.4 | 3.70 | 2.41 | 1.70 | 1.27 | 0.97 | 0.58 | 0.32 | 0.15 | 0.03 | - |
| 0.6 | 3.14 | 1.93 | 1.36 | 1.01 | 0.77 | 0.44 | 0.23 | 0.08 | - | - |
| 0.8 | 2.10 | 1.29 | 0.92 | 0.68 | 0.51 | 0.28 | 0.11 | - | - | - |
| 1.0 | 0.78 | 0.60 | 0.45 | 0.34 | 0.24 | 0.09 | - | - | - | - |
| 1.1 | 0.19 | 0.27 | 0.22 | 0.16 | 0.10 | - | - | - | - | - |
| 1.2 | - | - | 0.02 | 0.01 | - | - | - | - | - | - |

**Table 4.15.** *Range of stationarity of the AR(1)+ARCH(1) model with parameters $\alpha$ and $\lambda$. The matrix components contain the estimated tail index $\kappa$ for student-t noise with 5 degrees of freedom. The range of stationarity has shrunk compared to the normal noise. Moreover, the corresponding tails are heavier than for normal noise (cf. Table 4.14).*

| $|\alpha|$ | $\lambda$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.4 | 1.8 | 2.2 | 2.4 |
| 0.0 | 2.43 | 1.80 | 1.35 | 1.02 | 0.78 | 0.45 | 0.23 | 0.08 | 0.01 |
| 0.2 | 2.41 | 1.76 | 1.31 | 0.99 | 0.75 | 0.43 | 0.21 | 0.06 | 0.01 |
| 0.4 | 2.31 | 1.62 | 1.18 | 0.88 | 0.66 | 0.36 | 0.16 | 0.02 | - |
| 0.6 | 2.06 | 1.35 | 0.96 | 0.70 | 0.51 | 0.26 | 0.09 | - | - |
| 0.8 | 1.50 | 0.93 | 0.64 | 0.45 | 0.32 | 0.12 | - | - | - |
| 1.0 | 0.59 | 0.41 | 0.28 | 0.18 | 0.10 | - | - | - | - |
| 1.1 | 0.13 | 0.15 | 0.10 | 0.04 | - | - | - | - | - |

**Table 4.16.** *Range of stationarity of the AR(1)+ARCH(1) model with parameters $\alpha$ and $\lambda$. The matrix components contain the estimated and tail index for student-t noise with 3 degrees of freedom. The range of stationarity has further decreased and the tails have become very heavy indeed; a third moment does not exist (cf. Tables 4.14 and 4.15).*

Hüsler and Leadbetter [48] or Perfekt [70] for precise definitions. Loosely speaking, $D(u_n)$ and $\Delta(u_n)$ give the "degree of independence" of extremes situated far apart from each other. This property together with (4.9) implies that the maximum of the process $(X_n)_{n\in\mathbb{N}}$ belongs to the domain of attraction of a Fréchet distribution $\Phi_\kappa$, where $\kappa$ is given as solution to (4.10).

In the following denote by $P^\mu$ the probability law for $(X_n)_{n\in\mathbb{N}}$ when $X_0$ starts with distribution $\mu$ and $\pi$ is the stationary distribution.

**Theorem 4.17.** [Borkovec [15]]
*Let $(X_n)_{n\in\mathbb{N}}$ be the AR(1) process with ARCH(1) errors (4.1) with noise satisfying the usual conditions and $D_1 - D_2$. Let $X_0 \stackrel{d}{=} \mu$, then*

$$\lim_{n\to\infty} P^\mu\big(n^{-1/\kappa} \max_{1\leq j\leq n} X_j \leq x\big) = \exp(-c\theta x^{-\kappa}), \quad x \geq 0,$$

*where $\kappa$ solves the equation (4.10), $c$ is the constant in the tail of the stationary distribution*

*(4.9) and*

$$\theta = \kappa \int_1^\infty P\left(\sup_{n\in\mathbb{N}} \prod_{i=1}^n (\alpha + \sqrt{\lambda}\varepsilon_i) \leq y^{-1}\right) y^{-\kappa-1} dy\,.$$

*For $x \in \mathbb{R}$ and $n \in \mathbb{N}$ let $N_n$ be the point process of exceedances of the threshold $u_n = n^{1/\kappa}x$ by $X_1, ..., X_n$. Then*

$$N_n \quad \overset{d}{\to} \quad N, \quad n \to \infty\,,$$

*where $N$ is a compound Poisson process with intensity $c\theta x^{-\kappa}$ and cluster probabilities*

$$\pi_k = \frac{\theta_k - \theta_{k+1}}{\theta}\,, \quad k \in \mathbb{N},$$

*with*

$$\theta_k = \kappa \int_1^\infty P\left(\mathrm{card}\left\{n \in \mathbb{N} : \prod_{i=1}^n (\alpha + \sqrt{\lambda}\varepsilon_i) > y^{-1}\right\} = k-1\right) y^{-\kappa-1} dy\,.$$

*In particular, $\theta_1 = \theta$.*

We want to explain the idea of the proof:

Recall first from Theorem 4.7 that $(X_n)_{n\in\mathbb{N}}$ is Harris recurrent with regeneration set $[-e^{a/2}, e^{a/2}]$ for $a$ large enough. Thus there exists a renewal point process $(T_n)_{n\geq 0}$ (e.g. the successive entrance times in $[-e^{a/2}, e^{a/2}]$), which describes the regenerative structure of $(X_n)_{n\in\mathbb{N}}$. This process $(T_n)_{n\geq 0}$ is aperiodic and has finite mean recurrence times.

Hence we can apply a coupling argument giving for any probability measure $\mu$, the stationary distribution $\pi$ and any sequence $(u_n)_{n\in\mathbb{N}}$

$$\left| P^\mu\left(\max_{1\leq k\leq n} X_k \leq u_n\right) - P^\pi\left(\max_{1\leq k\leq n} X_k \leq u_n\right)\right| \to 0\,, \quad n \to \infty\,.$$

Consequently, we suppose in the follwing that $(X_n)_{n\in\mathbb{N}}$ is stationary.

On a high level, the process $(X_n)_{n\in\mathbb{N}}$ can be linked to some random walk as follows. Define

$$S_0 = 0\,, \quad S_n = \sum_{i=1}^n \ln(\alpha + \sqrt{\lambda}\varepsilon_i)\,, \quad n \in \mathbb{N}.$$

Although it is not as natural as for pure volatility models we consider besides $(X_n)_{n\in\mathbb{N}}$ also $(X_n^2)_{n\in\mathbb{N}}$. Define the auxiliary process $(Z_n)_{n\in\mathbb{N}} := (\ln(X_n^2))_{n\in\mathbb{N}}$, which satisfies the stochastic difference equation

$$Z_n = Z_{n-1} + \ln\left((\alpha + \sqrt{\beta e^{-Z_{n-1}} + \lambda}\,\varepsilon_n)^2\right)\,, \quad n \in \mathbb{N}, \quad Z_0 = \ln(X_0^2) \quad \text{a.s.}\,.$$

Note that, since strong mixing is a property of the underlying $\sigma$-algebra of the process, $(X_n^2)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ are also strong mixing. Since $\varepsilon$ is symmetric the process $(Z_n)_{n\in\mathbb{N}}$ is independent of the sign of the parameter $\alpha$. Hence we may wlog in the following assume that $\alpha \geq 0$.

We show that $(Z_n)_{n\in\mathbb{N}}$ can be bounded by two random walks $(S_n^{l,a})_{n\in\mathbb{N}}$ and $(S_n^{u,a})_{n\in\mathbb{N}}$ from below

and above, respectively. For the construction of the two random walks $(S_n^{l,a})_{n\in\mathbb{N}}$ and $(S_n^{u,a})_{n\in\mathbb{N}}$ we define with the same notation as before

$$A_a \quad := \quad \left\{ \frac{-\alpha}{\sqrt{\beta\,e^{-a}+\lambda}-\sqrt{\beta}\,e^{-a/2}} \leq \varepsilon \leq \frac{-\alpha}{\sqrt{\beta\,e^{-a}+\lambda}+\sqrt{\beta}\,e^{-a/2}} \right\},$$

$$p(a,\varepsilon) \quad := \quad \ln\left( (\alpha + \sqrt{\beta\,e^{-a}+\lambda}\,\varepsilon)^2 \right)$$

$$q(a,\varepsilon) \quad := \quad \ln\left( 1 - \frac{2\alpha\sqrt{\beta}e^{-a/2}\varepsilon}{(\alpha+\sqrt{\beta\,e^{-a}+\lambda}\,\varepsilon)^2}\,1_{\{\varepsilon<0\}} \right),$$

$$r(a,\varepsilon) \quad := \quad \ln\left( 1 - \frac{\beta\varepsilon^2 e^{-a}}{(\alpha+\sqrt{\beta\,e^{-a}+\lambda}\,\varepsilon)^2}\,1_{\{\varepsilon<0\}} \right).$$

Note that $q(a)$ and $r(a)$ both converge to $0$ a.s. as $a\to\infty$. Define the lower and upper random walks

$$S_n^{l,a} := \sum_{j=1}^n U_j^a \quad\text{and}\quad S_n^{u,a} := \sum_{j=1}^n V_j^a, \quad n\in\mathbb{N}, \tag{4.13}$$

where for each $j=1,\dots,n$

$$U_j^a \quad := \quad -\infty\cdot 1_{A_a} + (p(a,\varepsilon_j)+r(a,\varepsilon_j))1_{A_a^c\cap\{\varepsilon_j<0\}}\ln(\alpha+\sqrt{\lambda}\varepsilon)^2)\,1_{\{\varepsilon_j\geq 0\}} \tag{4.14}$$

$$V_j^a \quad := \quad p(a,\varepsilon_j)+q(a,\varepsilon_j). \tag{4.15}$$

The following lemma summarizes some properties of the random walks defined in (4.13)-(4.15).

**Lemma 4.18.** *Let $a$ be large enough, $Z_0 > a$ and $N_a := \inf\{j\geq 1\,|\,Z_j\leq a\}$. Then*

(a) $Z_0 + S_k^{l,a} \leq Z_k \leq Z_0 + S_k^{u,a}$ *for all $k\leq N_a$ a.s.*

(b) $(S_n^{u,a})_{n\in\mathbb{N}}$ *and $(S_n^{l,a})_{n\in\mathbb{N}}$ are random walks with negative drift.*

(c) *Define $S_0 = 0$ and $S_k = \sum_{j=1}^k \ln((\alpha+\sqrt{\lambda}\varepsilon_j)^2)$ for $k\in\mathbb{N}$. Then*

$$S_k^{l,a} \quad \overset{P}{\to} \quad S_k \quad\text{and}\quad S_k^{u,a} \quad \overset{a.s.}{\to} \quad S_k\,, \quad a\uparrow\infty\,.$$

(d) $\sup_{k\geq 1} S_k^{l,a} \overset{d}{\to} \sup_{k\geq 1} S_k$ *and* $\sup_{k\geq 1} S_k^{u,a} \overset{a.s.}{\to} \sup_{k\geq 1} S_k$ *as $a\uparrow\infty$.*

Lemma 4.18 characterizes the behaviour of the process $(Z_n)_{n\in\mathbb{N}}$ above a high treshold $a$ and hence also the behaviour of $(X_n^2)_{n\in\mathbb{N}}$. This is the key to what follows: the process $(S_n)_{n\in\mathbb{N}}$ will completely determine the extremal behaviour of $(X_n^2)_{n\in\mathbb{N}}$.

We first need the following lemma.

**Lemma 4.19.** *Let $\gamma$ be the mixing function of $(X_n)_{n\in\mathbb{N}}$ and $(p_n)_{n\in\mathbb{N}}$ an increasing sequence such that*

$$\frac{p_n}{n}\to 0 \quad\text{and}\quad \frac{n\gamma(\sqrt{p_n})}{p_n}\to 0 \quad\text{as } n\to\infty\,. \tag{4.16}$$

*Then for $u_n = n^{2/\kappa}x$, $x > 0$,*

$$\lim_{p\to\infty}\limsup_{n\to\infty} P\left( \max_{p\leq j\leq p_n} X_j^2 > u_n \,\big|\, X_0^2 > u_n \right) = 0\,, \tag{4.17}$$

33

*and for $u_n = n^{1/\kappa} x$, $x > 0$,*

$$\lim_{p \to \infty} \limsup_{n \to \infty} P(\max_{p \leq j \leq p_n} X_j > u_n \mid X_0 > u_n) = 0 \,. \tag{4.18}$$

**Proof.** The proof of (4.17) is very technical and we refer to Borkovec [15] for details. It is, however, easy to see that (4.17) implies (4.18):

$$
\begin{aligned}
P(\max_{p \leq j \leq p_n} X_j^2 > u_n^2 \mid X_0^2 > u_n^2) &= \frac{P(\max_{p \leq j \leq p_n} X_j^2 > u_n^2, X_0^2 > u_n^2)}{P(X_0^2 > u_n^2)} \\
\geq \frac{P(\max_{p \leq j \leq p_n} X_j > u_n, X_0 > u_n)}{P(X_0 > u_n) + P(X_0 < -u_n)} &= \frac{1}{2} P(\max_{p \leq j \leq p_n} X_j > u_n \mid X_0 > u_n) \,.
\end{aligned}
$$

$\square$

**Remark 4.20.** (a) Since $(X_n)_{n \in \mathbb{N}}$ is geometric ergodic, the mixing function $\gamma$ decreases exponentially fast, hence it is not difficult to find a sequence $(p_n)_{n \in \mathbb{N}}$ to satisfy (4.16).

(b) As mentioned already, the strong mixing condition is a property of the underlying $\sigma$−field of a process. Hence $\gamma$ is also the mixing function of $(X_n^2)_{n \in \mathbb{N}}$ and $(Z_n)_{n \in \mathbb{N}}$ and we may work for all these processes with the same sequence $(p_n)_{n \in \mathbb{N}}$.

(c) In the case of a strong mixing process, conditions (4.16) are sufficient to guarantee that $(p_n)_{n \in \mathbb{N}}$ is a $\Delta(u_n)$-separating sequence. It describes somehow the interval length needed to accomplish asymptotic independence of extremal events over a high level $u_n$ in separate intervals. For a definition see Perfekt [70]. Note that $(p_n)_{n \in \mathbb{N}}$ is in the case of a strong mixing process independent of $(u_n)_{n \in \mathbb{N}}$. $\square$

The following Theorem is an extension of Theorem 3.2 of Perfekt [70], p. 543 adapted to our situation.

**Theorem 4.21.** *Suppose $(X_n)_{n \in \mathbb{N}}$ is a strongly mixing stationary Markov chain whose stationary df $F$ is symmetric with tail $\overline{F} \in \mathcal{R}(-\kappa)$ on $\mathbb{R}_+$. Suppose furthermore that*

$$\lim_{u \to \infty} P(X_1 \leq xu \mid X_0 = u) = H(x) \,, \quad x \in \mathbb{R} \,,$$

*for some df $H$. Let $(A_n)_{n \in \mathbb{N}}$ be an iid sequence with df $H$ and define $Y_n = A_n Y_{n-1}$ for $n \in \mathbb{N}$ with $Y_0$ independent of $(A_n)_{n \in \mathbb{N}}$ and $Y_0 \overset{d}{=} \mu$ given by $\mu(dx) := \kappa^{-1} x^{-1/\kappa - 1} dx$, for $x > 1$. For every $\tau > 0$ let $(u_n(\tau))_{n \in \mathbb{N}}$ be a sequence satisfying*

$$\lim_{n \to \infty} n \overline{F}(u_n(\tau)) = \tau \,.$$

*Then $(X_n)_{n \in \mathbb{N}}$ has extremal index $\theta$ given by*

$$\theta = P^\mu(\text{card}\{n \in \mathbb{N} : Y_n > 1\} = 0) \,.$$

*Moreover, for $n \in \mathbb{N}$ the time normalized point process of exceedances*

$$N_n^\tau(B) := \sum_{i=1}^{n} \varepsilon_{i/n}(\cdot) I\{X_k > u_n(\tau)\} \quad \overset{d}{\to} \quad N(B) \,, \quad B \in \mathcal{B}(0, 1] \,,$$

*where $N$ is a compound Poisson process with intensity $\theta\tau$ and jump probabilities $(\pi_k)_{k\in\mathbb{N}}$ given by*

$$\pi_k = \frac{\theta_k - \theta_{k+1}}{\theta}, \quad k \in \mathbb{N},$$

*where*

$$\theta_k = P^\mu(\mathrm{card}\{n \in \mathbb{N} : Y_n > 1\} = k - 1), \quad k \in \mathbb{N}.$$

**Proof of Theorem 4.17.** The proof is an application of Theorem 4.21. As stated already we may assume w.l.o.g. that $(X_n)_{n\in\mathbb{N}}$ is stationary. Let $x \in \mathbb{R}$ be arbitrary. Note that by (4.9)

$$\lim_{u\to\infty} P(X_1 \le ux \,|\, X_0 = u) = P(\alpha + \sqrt{\lambda}\,\varepsilon \le x), \quad x \in \mathbb{R}.$$

$(X_n)_{n\in\mathbb{N}}$ satisfies all assumptions of Theorem 4.21 and we have the extremal index

$$
\begin{aligned}
\theta &= \int_1^\infty P\left(\mathrm{card}\{n \in \mathbb{N} : (\prod_{i=1}^n (\alpha + \sqrt{\lambda}\,\varepsilon_i))Y_0 > 1\} = 0 \,|\, Y_0 = y\right) \kappa\, y^{-\kappa-1} dy \\
&= \kappa \int_1^\infty P\left(\sup_{n\ge 1}(\prod_{i=1}^n (\alpha + \sqrt{\lambda}\,\varepsilon_i)) \le y^{-1}\right) y^{-\kappa-1} dy.
\end{aligned}
$$

The cluster probabilities can be determined in the same way and hence the statement follows.
$\square$

**Remark 4.22.** (i) Notice that for the squared process the extremal index and the cluster probabilities can be described by the random walk $(S_n)_{n\in\mathbb{N}}$, namely

$$\theta_k^{(2)} = \frac{\kappa}{2} \int_0^\infty P(\mathrm{card}\{n \in \mathbb{N} \,|\, S_n > -x\} = k - 1)\, e^{-\frac{\kappa}{2}x}\, dx, \quad k \in \mathbb{N}.$$

The description of the extremal behaviour of $(X_n^2)_{n\in\mathbb{N}}$ by the random walk $(S_n)_{n\in\mathbb{N}}$ is to be expected since by Lemma 4.18 the process $(Z_n)_{n\in\mathbb{N}} = (\ln(X_n^2))_{n\in\mathbb{N}}$ behaves above a high threshold asymptotically like $(S_n)_{n\in\mathbb{N}}$. Unfortunately, this link fails for $(X_n)_{n\in\mathbb{N}}$.

(ii) Analogous to de Haan et al. [46] we may construct "estimators" for the extremal indices $\theta^{(2)}$ and $\theta_k^{(2)}$ of $(X_n^2)_{n\in\mathbb{N}}$, respectively, by

$$\widehat{\theta}^{(2)} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{\sup_{1\le j\le m} S_j^{(i)} \le -E_\kappa^{(i)}\}$$

and

$$\widehat{\theta}_k^{(2)} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{\sum_{j=1}^m \mathbf{1}\{S_j^{(i)} > -E_\kappa^{(i)}\} = k - 1\}, \quad \text{for } k \in \mathbb{N},$$

where $N$ denotes the number of independent simulated sample paths of $(S_n)_{n\in\mathbb{N}}$, $E_\kappa^{(i)}$ are i.i.d. exponential rvs with rate $\kappa$, and $m$ is chosen large enough. These estimators can be studied as in the case $\alpha = 0$ and $\varepsilon \overset{d}{=} N(0,1)$ in de Haan et al. [46]. In particular,

$$\sqrt{N}\, \frac{\theta^{(2)} - \widehat{\theta}^{(2)}}{(\theta^{(2)}(1 - \theta^{(2)}))^{1/2}} \quad \overset{d}{\to} \quad N(0,1), \quad N, m \to \infty.$$

(iii) The approach chosen in (ii) is not possible for $(X_n)_{n\in\mathbb{N}}$, because $\prod_{l=1}^{j}(\alpha + \sqrt{\lambda}\,\varepsilon_l)$ may be negative. In a similar spirit we choose as "estimators" for $\theta$ and $\theta_k$ for $(X_n)_{n\in\mathbb{N}}$

$$\widehat{\theta} = \frac{1}{N}\sum_{i=1}^{N} 1\{\sup_{1\le j\le m}\prod_{l=1}^{j}(\alpha + \sqrt{\lambda}\,\varepsilon_l) \le 1/P_\kappa^{(i)}\}$$

and

$$\widehat{\theta}_k = \frac{1}{N}\sum_{i=1}^{N} 1\{\sum_{j=1}^{m} 1\{\prod_{l=1}^{j}(\alpha + \sqrt{\lambda}\,\varepsilon_l) > 1/P_\kappa^{(i)}\} = k-1\}, \quad \text{for } k \in \mathbb{N},$$

where $N$ denotes the number of simulated paths of $(\prod_{l=1}^{n}(\alpha+\sqrt{\lambda}\,\varepsilon_l))_{n\in\mathbb{N}}$, $P_\kappa^{(i)}$ are iid Pareto rvs with shape parameter $\kappa$, i.e. with distribution function $G(x) = 1 - x^{-\kappa}$, $x \ge 1$, and $m$ is large enough. These are suggestive estimators since $\prod_{l=1}^{n}(\alpha+\sqrt{\lambda}\,\varepsilon_l) \to 0$ a.s. as $n \to \infty$ because of assumption (4.4).

(iv) Note that the extremal index $\theta$ of $(X_n)_{n\in\mathbb{N}}$ is not symmetric in $\alpha$ (see Table 4.23). This is not surprising since the clustering is for $\alpha > 0$ stronger by the autoregressive part than for $\alpha < 0$. $\qquad\square$

| $\alpha$ \ $\lambda$ | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $-1.2$ | - | 0.001 | 0.001 | 0.003 | 0.004 | 0.001 | 0.000 | - | - | - | - |
| $-1$ | 0.15 | 0.19 | 0.19 | 0.16 | 0.13 | 0.09 | 0.05 | 0.01 | - | - | - |
| $-0.8$ | 0.56 | 0.47 | 0.41 | 0.34 | 0.26 | 0.21 | 0.13 | 0.05 | 0.01 | - | - |
| $-0.6$ | 0.86 | 0.71 | 0.61 | 0.50 | 0.41 | 0.33 | 0.22 | 0.10 | 0.03 | 0.00 | - |
| $-0.4$ | 0.96 | 0.85 | 0.71 | 0.60 | 0.50 | 0.40 | 0.30 | 0.14 | 0.06 | 0.01 | - |
| $-0.2$ | 0.98 | 0.89 | 0.77 | 0.65 | 0.56 | 0.47 | 0.33 | 0.18 | 0.07 | 0.02 | 0.00 |
| $0$ | 0.98 | 0.89 | 0.78 | 0.65 | 0.55 | 0.45 | 0.33 | 0.18 | 0.08 | 0.02 | 0.00 |
| $0.2$ | 0.94 | 0.82 | 0.72 | 0.61 | 0.52 | 0.43 | 0.32 | 0.18 | 0.07 | 0.02 | 0.00 |
| $0.4$ | 0.85 | 0.72 | 0.63 | 0.53 | 0.45 | 0.37 | 0.28 | 0.13 | 0.06 | 0.01 | - |
| $0.6$ | 0.68 | 0.55 | 0.48 | 0.41 | 0.35 | 0.29 | 0.21 | 0.10 | 0.03 | 0.00 | - |
| $0.8$ | 0.39 | 0.34 | 0.32 | 0.27 | 0.22 | 0.19 | 0.12 | 0.05 | 0.01 | - | - |
| $1.0$ | 0.09 | 0.14 | 0.13 | 0.13 | 0.11 | 0.08 | 0.04 | 0.01 | - | - | - |
| $1.2$ | - | 0.000 | 0.001 | 0.003 | 0.004 | 0.001 | 0.000 | - | - | - | - |

**Table 4.23.** *"Estimated" extremal index $\theta$ of $(X_n)_{n\in\mathbb{N}}$ in the case $\varepsilon \overset{d}{=} N(0,1)$. We chose $N = m = 2\,000$. Note that the extremal index decreases as $|\alpha|$ increases and that we have no symmetry in $\alpha$.*

**Remark 4.24.** (i) Model (4.1) has a natural extension to higher order: the autoregressive model of order $q$ with ARCH($q$)-errors has been investigated in Klüppelberg and Pergamenchtchikov [58, 59]. It is also shown there that for Gaussian error variables this model is in distribution equivalent to a random coefficient model.

(ii) Such models also lead to interesting statistical theory, some can be found in econometric textbooks; see e.g. Campbell, Lo and MacKinley [21], Gouriéroux [44], Shephard [83], or Taylor [85]. In Klüppelberg et al. [57] tests for models including (4.1) are suggested. A pseudo-likelihood ratio test for the hypotheses that the model reduces to random walk or iid data is investigated and the distributional limit of the test statistic is derived. $\qquad\square$
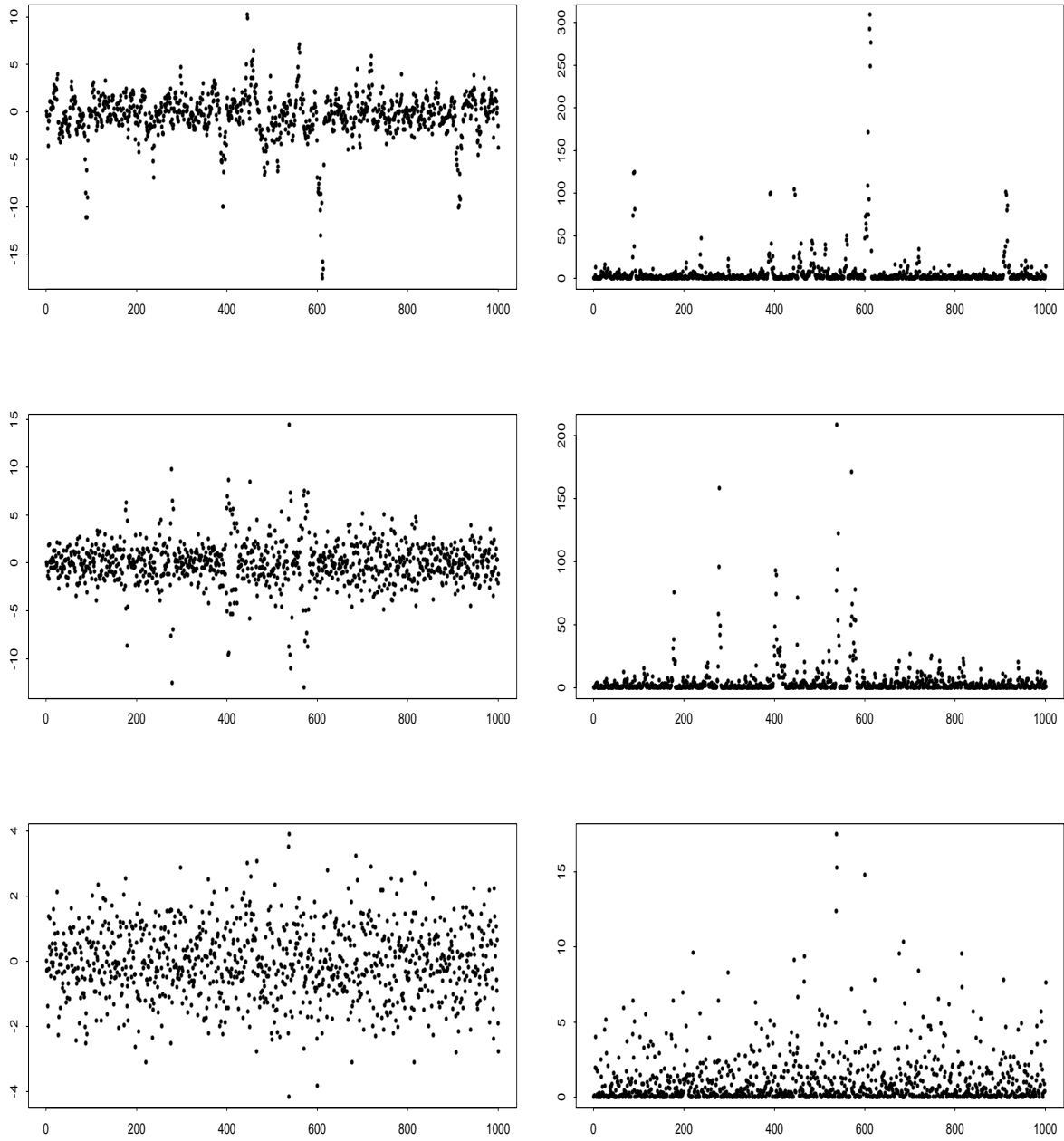
**Figure 4.25.** *Simulated sample path of $(X_n)_{n\in\mathbb{N}}$ with parameters $(\alpha,\beta,\lambda) = (0.8,1,0.2)$ (top, left), of $(X_n^2)_{n\in\mathbb{N}}$ with the same parameters (top, right), of $(X_n)_{n\in\mathbb{N}}$ with parameters $(\alpha,\beta,\lambda) = (-0.8,1,0.2)$ (middle, left), of $(X_n^2)_{n\in\mathbb{N}}$ with the same parameters (middle, right), of $(X_n)_{n\in\mathbb{N}}$ with parameters $(\alpha,\beta,\lambda) = (0,1,0.2)$ (bottom, left) and of $(X_n^2)_{n\in\mathbb{N}}$ with the same parameters (bottom,right) in the case $\varepsilon \overset{d}{=} N(0,1)$. All simulations are based on the same simulated noise sequence $(\varepsilon_n)_{n\in\mathbb{N}}$.*

37

| $\alpha$ | $\lambda$ | $\theta$ | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\pi_4$ | $\pi_5$ | $\pi_6$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.2 | 0.974 | 0.973 | 0.027 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0 | 0.6 | 0.781 | 0.799 | 0.147 | 0.036 | 0.012 | 0.005 | 0.001 |
| 0 | 1 | 0.549 | 0.607 | 0.188 | 0.107 | 0.036 | 0.034 | 0.017 |
| -0.4 | 0.2 | 0.962 | 0.962 | 0.037 | 0.001 | 0.000 | 0.000 | 0.000 |
| 0.4 | 0.2 | 0.853 | 0.867 | 0.103 | 0.026 | 0.002 | 0.002 | 0.000 |
| -0.4 | 0.6 | 0.715 | 0.747 | 0.168 | 0.048 | 0.026 | 0.006 | 0.002 |
| 0.4 | 0.6 | 0.624 | 0.676 | 0.182 | 0.066 | 0.040 | 0.019 | 0.012 |
| -0.4 | 1 | 0.497 | 0.540 | 0.210 | 0.115 | 0.075 | 0.040 | 0.004 |
| 0.4 | 1 | 0.445 | 0.533 | 0.185 | 0.080 | 0.109 | 0.032 | 0.017 |
| -0.8 | 0.2 | 0.572 | 0.626 | 0.185 | 0.111 | 0.026 | 0.033 | 0.001 |
| 0.8 | 0.2 | 0.386 | 0.470 | 0.172 | 0.148 | 0.062 | 0.068 | 0.006 |
| -0.8 | 0.6 | 0.414 | 0.520 | 0.159 | 0.134 | 0.072 | 0.043 | 0.016 |
| 0.8 | 0.6 | 0.314 | 0.443 | 0.156 | 0.110 | 0.087 | 0.073 | 0.041 |
| -0.8 | 1 | 0.273 | 0.429 | 0.137 | 0.126 | 0.106 | 0.016 | 0.012 |
| 0.8 | 1 | 0.224 | 0.346 | 0.132 | 0.114 | 0.129 | 0.045 | 0.004 |

**Table 4.26.** *"Estimated" extremal index $\theta$ and cluster probabilities $(\pi_k)_{1 \leq k \leq 6}$ of $(X_n)_{n \in \mathbb{N}}$ dependent on $\alpha$ and $\lambda$ in the case $\varepsilon \stackrel{d}{=} N(0,1)$. We chose $N = m = 2\,000$. Note that the extremal index for $\alpha > 0$ is much larger than for $\alpha < 0$.*

# 5 Optimal portfolios with bounded VaR

In this section we investigate the influence of large fluctuations and the Value-at-Risk as a risk measure, which is sensitive to such price behaviour to portfolio optimisation. It is based on Emmer, Klüppelberg and Korn [35] and Emmer and Klüppelberg [34]

Starting with the traditional Black-Scholes model, where stock prices follow a geometric Brownian motion we first study the difference between the classical risk measure, i.e. the variance, and the VaR.

Since the variance of Brownian motion increases linearly, the use of the variance as a risk measure of an investment leads to a decreasing proportion of risky assets in a portfolio, when the planning horizon increases. This is not true for the Capital-at-Risk which - as a function of the planning horizon – increases first, but decreases, when the planning horizon becomes larger. We show for the CaR that, as seems to be common wisdom in asset management, long term stock investment leads to an almost sure gain over locally riskless bond investments. In the long run stock indices are growing faster than riskless rates, despite the repeated occurrence of stock market declines. The VaR therefore supports the portfolio manager's advice that the more distant the planning horizon, the greater should be one's wealth in risky assets. Interestingly, the VaR as risk measure supports the empirical observation above and hence resolves the contradiction between theory and empirical facts.

Then we study the optimal portfolio problem for more realistic price processes, i.e. Lévy processes which model also large fluctuations. Here, as is to be expected, the VaR reacts to exactly those and consequently, the CaR. We investigate, in particular, the normal inverse Gaussian and variance gamma Lévy processes.

## 5.1 The Black-Scholes model

In this section, we consider a standard Black-Scholes type market consisting of one *riskless bond* and several *risky stocks*. Their respective prices $(P_0(t))_{t \geq 0}$ and $(P_i(t))_{t \geq 0}$ for $i = 1, \ldots, d$ evolve according to the equations

$$P_0(t) = e^{rt} \quad \text{and} \quad P_i(t) = p_i \exp\left((b_i - \frac{1}{2}\sum_{j=1}^d \sigma_{ij}^2)t + \sum_{j=1}^d \sigma_{ij}W_j(t)\right), \quad t \geq 0\,.$$

Here $W(t) = (W_1(t), \ldots, W_d(t))'$ is a standard $d$-dimensional Brownian motion, $r \in \mathbb{R}$ is the *riskless interest rate*, $b = (b_1, \ldots, b_d)'$ the vector of *stock-appreciation rates* and $\sigma = (\sigma_{ij})_{1 \leq i,j \leq d}$ is the matrix of *stock-volatilities*. For simplicity, we assume that $\sigma$ is invertible and that $b_i \geq r$ for $i = 1, \ldots, d$. Since the assets are traded on the same market, they show some correlation structure which we model by a linear combination of the same Brownian motions $W_1, \ldots, W_d$ for each traded asset. Throughout this paper we denote by $\mathbb{R}^d$ the $d$-dimensional Euclidean space. Its elements are column vectors and for $x \in \mathbb{R}^d$ we denote by $x'$ the transposed vector; analogously, for a matrix $\beta$ we denote by $\beta'$ its transposed matrix. We further denote by $|x| = (\sum_{i=1}^d x_i^2)^{1/2}$ the Euclidean norm of $x \in \mathbb{R}^d$.

We need the SDE corresponding to the price processes above.

$$
\begin{array}{rclcrcll}
dP_0(t) & = & P_0(t)r\,dt\,, & & P_0(0) & = & 1\,, & \\
dP_i(t) & = & P_i(t)\left(b_i dt + \sum_{j=1}^d \sigma_{ij}dW_j(t)\right)\,, & & P_i(0) & = & p_i\,, & i = 1, \ldots, d\,.
\end{array}
\tag{5.1}
$$

Let $\pi(t) = (\pi_1(t), \ldots, \pi_d(t))' \in \mathbb{R}^d$ be an admissible portfolio process, i.e. $\pi_i(t)$ is the fraction of the wealth $X^\pi(t)$, which is invested in asset $i$ (see Korn [60], Section 2.1 for relevant definitions). Denoting by $(X^\pi(t))_{t \geq 0}$ the *wealth process*, it follows the dynamic

$$dX^\pi(t) = X^\pi(t)\left\{((1 - \pi(t)'\underline{1})r + \pi(t)'b)dt + \pi(t)'\sigma dW(t)\right\}\,, \quad X^\pi(0) = x\,,$$

where $x \in \mathbb{R}$ denotes the *initial capital* of the investor and $\underline{1} = (1, \ldots, 1)'$ denotes the vector (of appropriate dimension) having unit components. The fraction of the investment in the bond is $\pi_0(t) = 1 - \pi(t)'\underline{1}$. Throughout the paper, we restrict ourselves to constant portfolios $\pi(t) = \pi = (\pi_1, \ldots, \pi_d)$ for all $t \in [0, T]$. This means that the fractions in the different stocks and the bond remain constant on $[0, T]$. The advantage of this is two-fold: first we obtain, at least in a Gaussian setting, explicit results; and, furthermore, the economic interpretation of the mathematical results is comparably easy. It is also important to point out that following a constant portfolio process does not mean that there is no trading. As the stock prices evolve randomly one has to trade at every time instant to keep the fractions of wealth invested in the different securities constant. Thus, following a constant portfolio process still means one must follow a dynamic trading strategy.

Standard Itô integration and the fact that $Ee^{sW(1)} = e^{s^2/2}, s \in \mathbb{R}$, yield the following explicit formulae for the wealth process for all $t \in [0, T]$.

$$
\begin{array}{rcll}
X^\pi(t) & = & x \exp\left((\pi'(b - r\underline{1}) + r - |\pi'\sigma|^2/2)t + \pi'\sigma W(t)\right)\,, & \tag{5.2} \\
E[X^\pi(t)] & = & x \exp\left((\pi'(b - r\underline{1}) + r)t\right)\,, & \tag{5.3} \\
\mathrm{var}(X^\pi(t)) & = & x^2 \exp\left(2(\pi'(b - r\underline{1}) + r)t\right)\left(\exp(|\pi'\sigma|^2 t) - 1\right)\,. & \tag{5.4}
\end{array}
$$

**Definition 5.1.** [Capital-at-Risk]

*Let $x$ be the initial capital and $T$ a given planning horizon. Let $z_\alpha$ be the $\alpha$-quantile of the standard normal distribution. For some portfolio $\pi \in \mathbb{R}^d$ and the corresponding terminal wealth $X^\pi(T)$, the VaR of $X^\pi(T)$ is given by*

$$
\begin{aligned}
\mathrm{VaR}(x, \pi, T) &= \inf\{z \in \mathbb{R} : P(X^\pi(T) \le z) \ge \alpha\} \\
&= x \exp\left( (\pi'(b - r\underline{1}) + r - |\pi'\sigma|^2/2)T + z_\alpha |\pi'\sigma| \sqrt{T} \right).
\end{aligned}
$$

*Then we define*

$$
\begin{aligned}
\mathrm{CaR}(x, \pi, T) &= x \exp(rT) - \mathrm{VaR}(x, \pi, T) \\
&= x \exp(rT) \\
&\quad \times \left( 1 - \exp((\pi'(b - r\underline{1}) - |\pi'\sigma|^2/2)T + z_\alpha |\pi'\sigma| \sqrt{T}) \right)
\end{aligned} \tag{5.5}
$$

*the* Capital-at-Risk *of the portfolio $\pi$ (with initial capital $x$ and planning horizon $T$).*  □

To avoid (non-relevant) subcases in some of the following results we always assume $\alpha < 0.5$ which leads to $z_\alpha < 0$.

**Remark 5.2.** (i) Our definition of the Capital-at-Risk limits the possibility of excess losses over the riskless investment.

(ii) We typically want to have a positive CaR (although it can be negative in our definition as the examples below will show) as the upper bound for the "likely losses" (in the sense that $(1 - \alpha) \times 100\%$ of occurring "losses" are smaller than $\mathrm{CaR}(x, \pi, T)$) compared to the pure bond investment. Further, we concentrate on the actual amount of losses appearing at the planning horizon $T$. This is in line with the mean-variance selection procedure enabling us to directly compare the results of the two approaches; see below.  □

In the following it will be convenient to introduce the function $f(\pi)$ for the exponent in (5.5), that is

$$
f(\pi) := z_\alpha |\pi'\sigma| \sqrt{T} - |\pi'\sigma|^2 T/2 + \pi'(b - r\underline{1})T, \quad \pi \in \mathbb{R}^d. \tag{5.6}
$$

By the obvious fact that $f(\pi) \to -\infty$ as $|\pi'\sigma| \to \infty$ we have the natural upper bound $\sup_{\pi \in \mathbb{R}^d} \mathrm{CaR}(x, \pi, T) = x \exp(rT)$; i.e., the use of extremely risky strategies (in the sense of a high norm $|\pi'\sigma|$) can lead to a CaR which is close to the total capital. The computation of the minimal CaR is done in the following proposition.

**Proposition 5.3.** *Let $\theta = |\sigma^{-1}(b - r1)|$.*

*(a) If $b_i = r$ for all $i = 1, \ldots, d$, then $f(\pi)$ attains its maximum for $\pi^* = 0$ leading to a minimum Capital-at-Risk of $\mathrm{CaR}(x, \pi^*, T) = 0$.*

*(b) If $b_i \ne r$ for some $i \in \{1, \ldots, d\}$ and $\theta \sqrt{T} < |z_\alpha|$, then again the minimal CaR equals zero and is only attained for the pure bond strategy $\pi^* = 0$.*

*(c) If $b_i \ne r$ for some $i \in \{1, \ldots, d\}$ and $\theta \sqrt{T} \ge |z_\alpha|$, then the minimal CaR is attained for*

$$
\pi^* = \left( \theta - \frac{|z_\alpha|}{\sqrt{T}} \right) \frac{(\sigma\sigma)^{-1}(b - r\underline{1})}{|\sigma^{-1}(b - r\underline{1})|} \tag{5.7}
$$

*with*

$$
\mathrm{CaR}(x, \pi^*, T) = x \exp(rT) \left( 1 - \exp\left( \frac{1}{2}(\sqrt{T}\theta - |z_\alpha|)^2 \right) \right) < 0. \tag{5.8}
$$

**Proof.** (a) follows directly from the explicit form of $f(\pi)$ under the assumption of $b_i = r$ for all $i = 1, \ldots, d$ and the fact that $\sigma$ is invertible.

(b),(c) Consider the problem of maximizing $f(\pi)$ over all $\pi$ which satisfy

$$|\pi' \sigma| = \varepsilon \tag{5.9}$$

for a fixed positive $\varepsilon$. Over the (boundary of the) ellipsoid defined by (5.9) $f(\pi)$ equals

$$f(\pi) = z_\alpha \varepsilon \sqrt{T} - \varepsilon^2 T/2 + \pi'(b - r\underline{1})T \ .$$

Thus, the problem is reduced to maximizing a linear function (in $\pi$) over the boundary of an ellipsoid. Such a problem has the explicit solution

$$\pi_\varepsilon^* = \varepsilon \frac{(\sigma \sigma')^{-1}(b - r\underline{1})}{|\sigma^{-1}(b - r\underline{1})|} \tag{5.10}$$

with

$$f(\pi_\varepsilon^*) = -\varepsilon^2 T/2 + \varepsilon \left( \theta T - |z_\alpha| \sqrt{T} \right) . \tag{5.11}$$

As every $\pi \in \mathbb{R}^d$ satisfies relation (5.9) with a suitable value of $\varepsilon$ (due to the fact that $\sigma$ is regular), we obtain the minimum CaR strategy $\pi^*$ by maximizing $f(\pi_\varepsilon^*)$ over all non-negative $\varepsilon$. Due to the form of $f(\pi_\varepsilon^*)$ the optimal $\varepsilon$ is positive if and only if the multiplier of $\varepsilon$ in representation (5.11) is positive. Thus, in the situation of Proposition 5.3(b) the assertion holds. In the situation of Proposition 5.3(c) the optimal $\varepsilon$ is given as

$$\varepsilon = \theta - \frac{|z_\alpha|}{\sqrt{T}}.$$

Inserting this into equations (5.10) and (5.11) yields the assertions (5.7) and (5.8) (with the help of equations (5.5) and (5.6)). $\qquad\square$

**Remark 5.4.** (i) Part (a) of Proposition 5.3 states that in a risk-neutral market the CaR of every strategy containing stock investment is bigger than the CaR of the pure bond strategy.

(ii) Part (c) states the (at first sight surprising) fact that the existence of at least one stock with a mean rate of return different from the riskless rate implies the existence of a stock and bond strategy with a negative CaR as soon as the planning horizon $T$ is large. Thus, even if the CaR would be the only criterion to judge an investment strategy the pure bond investment would not be optimal if the planning horizon is far away. On one hand this fact is in line with empirical results on stock and bond markets. On the other hand this shows a remarkable difference between the behaviour of the CaR and the variance as risk measures. Independent of the planning horizon and the market coefficients, pure bond investment would always be optimal with respect to the variance of the corresponding wealth process. $\qquad\square$

We now turn to a Markowitz mean-variance type optimization problem where we replace the variance constraint by a constraint on the CaR of the terminal wealth. More precisely, we solve the following problem:

$$\max_{\pi \in \mathbb{R}^d} E[X^\pi(T)] \quad \text{subject to} \quad \text{CaR}(x, \pi, T) \leq C \,, \tag{5.12}$$

where $C$ is a given constant of which we assume that it satisfies $C \leq x \exp(rT)$.

Due to the explicit representations (5.4), (5.5) and a variant of the decomposition method as applied in the proof of Proposition 5.3 we can solve problem (5.12) explicitly.

**Proposition 5.5.** *Let $\theta = |\sigma^{-1}(b - r\underline{1})|$ and assume that $b_i \neq r$ for at least one $i \in \{1, \ldots, d\}$. Assume furthermore that $C$ satisfies*

$$0 \qquad \leq C \leq x \exp(rT) \qquad \text{if } \theta\sqrt{T} < |z_\alpha|, \qquad (5.13)$$

$$x \exp(rT) \left(1 - \exp\left(\frac{1}{2}(\sqrt{T}\theta - |z_\alpha|)^2\right)\right) \leq C \leq x \exp(rT) \qquad \text{if } \theta\sqrt{T} \geq |z_\alpha|. \qquad (5.14)$$

*Then problem (5.12) has solution*

$$\pi^* = \varepsilon^* \frac{(\sigma\sigma')^{-1}(b - r\underline{1})}{|\sigma^{-1}(b - r\underline{1})|}$$

*with*

$$\varepsilon^* = (\theta + z_\alpha/\sqrt{T}) + \sqrt{(\theta + z_\alpha/\sqrt{T})^2 - 2c/T}\,,$$

*where $c = \ln\left(1 - \frac{C}{x}\exp(-rT)\right)$. The corresponding maximal expected terminal wealth under the CaR constraint equals*

$$E[X^{\pi^*}(T)] = x \exp\left(\left(r + \varepsilon^*|\sigma^{-1}(b - r\underline{1})|\right)T\right)\,. \qquad (5.15)$$

**Proof.** The requirements (5.13) and (5.14) on $C$ ensure that the CaR constraint in problem (5.12) cannot be ignored: in both cases $C$ lies between the minimum and the maximum value that CaR can attain (see also Proposition 5.3). Every admissible $\pi$ for problem (5.12) with $|\pi'\sigma| = \varepsilon$ satisfies the relation

$$(b - r\underline{1})'\pi T \geq c + \frac{1}{2}\varepsilon^2 T - z_\alpha\varepsilon\sqrt{T} \qquad (5.16)$$

which is in this case equivalent to the CaR constraint in (5.12). But again, on the set given by $|\pi'\sigma| = \varepsilon$ the linear function $(b - r\underline{1})'\pi T$ is maximized by

$$\pi_\varepsilon = \varepsilon \frac{(\sigma\sigma')^{-1}(b - r\underline{1})}{|\sigma^{-1}(b - r\underline{1})|}\,. \qquad (5.17)$$

Hence, if there is an admissible $\pi$ for problem (5.12) with $|\pi'\sigma| = \varepsilon$ then $\pi_\varepsilon$ must also be admissible. Further, due to the explicit form (5.3) of the expected terminal wealth, $\pi_\varepsilon$ also maximizes the expected terminal wealth over the ellipsoid. Consequently, to obtain $\pi$ for problem (5.12) it suffices to consider all vectors of the form $\pi_\varepsilon$ for all positive $\varepsilon$ such that requirement (5.16) is satisfied. Inserting (5.17) into the left-hand side of inequality (5.16) results in

$$(b - r\underline{1})'\pi_\varepsilon T = \varepsilon|\sigma^{-1}(b - r\underline{1})|T\,, \qquad (5.18)$$

which is an increasing linear function in $\varepsilon$ equalling zero in $\varepsilon = 0$. Therefore, we obtain the solution of problem (5.12) by determining the biggest positive $\varepsilon$ such that (5.16) is still valid. But the right-hand side of (5.18) stays above the right-hand side of (5.16) until their largest positive point of intersection which is given by

$$\varepsilon^* = (\theta + z_\alpha/\sqrt{T}) + \sqrt{(\theta + z_\alpha/\sqrt{T})^2 - 2c/T}\,,$$

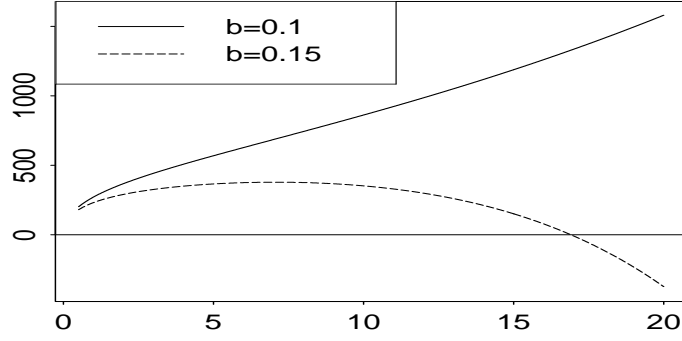The remaining assertion (5.15) can be verified by inserting $\pi^*$ into equation (5.3). $\qquad \square$

**Figure 5.7.** $\mathrm{CaR}(1\,000, 1, T)$ *of the pure stock portfolio (one risky asset only) for different appreciation rates as a function of the planning horizon $T$; $0 < T \le 20$. The volatility is $\sigma = 0.2$. The riskless rate is $r = 0.05$.*

**Remark 5.6.** Note that the optimal expected value only depends on the stocks via the norm $|\sigma^{-1}(b - r\underline{1})|$. There is no explicit dependence on the number of different stocks. We therefore interpret Proposition 5.3 as a kind of *mutual fund theorem* as there is no difference between investment in our multi-stock market and a market consisting of the bond and just one stock with appropriate market coefficients $b$ and $\sigma$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

**Example 5.8.** Figure 5.7 shows the dependence of CaR on the planning horizon $T$ illustrated by CaR($1\,000$,1,T). Note that the CaR first increases and then decreases with time, a behaviour which was already indicated by Proposition 5.3. It differs substantially from the behaviour of the variance of the pure stock strategy, which increases with $T$. Figure 5.9 illustrates the behaviour of the optimal expected terminal wealth with varying planning horizon corresponding to the pure bond strategy and the pure stock strategy as functions of the planning horizon $T$. The expected terminal wealth of the optimal portfolio even exceeds the pure stock investment. The reason for this becomes clear if we look at the corresponding portfolios. The optimal portfolio always contains a short position in the bond as long as this is tolerated by the CaR constraint. This is shown in Figure 5.10 where we have plotted the optimal portfolio together with the pure stock portfolio as function of the planning horizon. For $b = 0.15$ the optimal portfolio always contains a short position in the bond. For $b = 0.1$ and $T > 5$ the optimal portfolio (with the same CaR constraint as in Figures 5.9) again contains a long position in both bond and stock (with decreasing tendency of $\pi$ as time increases!). This is an immediate consequence of the increasing CaR of the stock price. For the smaller appreciation rate of the stock it is simply not attractive enough to take the risk of a large stock investment. Figure 5.10 shows the mean-CaR efficient frontier for the above parameters with $b = 0.1$ and fixed planning horizon $T = 5$. As expected it has a similar form as a typical mean-variance efficient frontier.

We compare now the behaviour of the optimal portfolios for the mean-CaR with solutions of a corresponding mean-variance problem. To this end we consider the following simpler optimization problem:

$$\max_{\pi \in \mathbb{R}^d} E[X^{\pi}(T)] \quad \text{subject to} \quad \mathrm{var}(X^{\pi}(T)) \le C \ . \tag{5.19}$$

**Proposition 5.11.** *If $b_i \neq r$ for at least one $i \in \{1, \dots, d\}$, then the optimal solution of the*
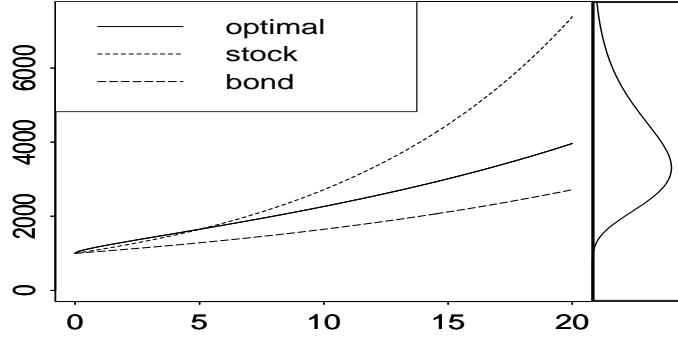
**Figure 5.9.** *Expected terminal wealth of different investment strategies depending on the planning horizon $T$, $0 \leq T \leq 20$. The parameters are $d = 1$, $r = 0.05$, $b = 0.1$, $\sigma = 0.2$, and $\alpha = 0.05$. As the upper bound $C$ of the CaR we used $\mathrm{CaR}(1\,000, 1, 5)$, the CaR of the pure stock strategy with planning horizon $T = 5$. On the right border we have plotted the density function of the wealth for the optimal portfolio.*

*mean-variance problem (5.19) is given by*

$$\widehat{\pi} = \widehat{\varepsilon} \, \frac{(\sigma\sigma')^{-1}(b - r\underline{1})}{|\sigma^{-1}(b - r\underline{1})|} \, ,$$

*where $\widehat{\varepsilon}$ is the unique positive solution of the non-linear equation*

$$rT + |\sigma^{-1}(b - r\underline{1})|\varepsilon T - \frac{1}{2} \ln\left(\frac{C}{x^2}\right) + \frac{1}{2} \ln\left(\exp(\varepsilon^2 T) - 1\right) = 0 \, .$$

*The corresponding maximal expected terminal wealth under the variance constraint equals*

$$E[X^{\widehat{\pi}}(T)] = x \exp\left((r + \widehat{\varepsilon}|\sigma^{-1}(b - r\underline{1})|)T\right) \, . \qquad\qquad \square$$

**Proof.** By using the explicit form (5.4) of the variance of the terminal wealth, we can rewrite the variance constraint in problem (5.19) as

$$(b - r\underline{1})'\pi T \leq \frac{1}{2} \ln\left(\frac{C}{x^2}\right) - \frac{1}{2} \ln\left(\exp(\varepsilon^2 T) - 1)\right) - rT =: h(\varepsilon), \qquad |\pi'\sigma| = \varepsilon \qquad (5.20)$$

for $\varepsilon > 0$. More precisely, if $\pi \in \mathbb{R}^d$ satisfies the constraints in (5.20) for one $\varepsilon > 0$ then it also satisfies the variance constraint in (5.19) and vice versa. Noting that $h(\varepsilon)$ is strictly decreasing in $\varepsilon > 0$ with

$$\lim_{\varepsilon \downarrow 0} h(\varepsilon) = \infty \quad \text{and} \quad \lim_{\varepsilon \to \infty} h(\varepsilon) = -\infty$$

we see that the left-hand side of (5.20) must be smaller than the right-hand side for small values of $\varepsilon > 0$ if we plug in $\pi_\varepsilon$ as given by equation (5.17). Recall that this was the portfolio with the highest expected terminal wealth of all portfolios $\pi$ satisfying $|\pi'\sigma| = \varepsilon$. It even maximizes $(b - r\underline{1})'\pi T$ over the set given by $|\pi'\sigma| \leq \varepsilon$. If we have equality

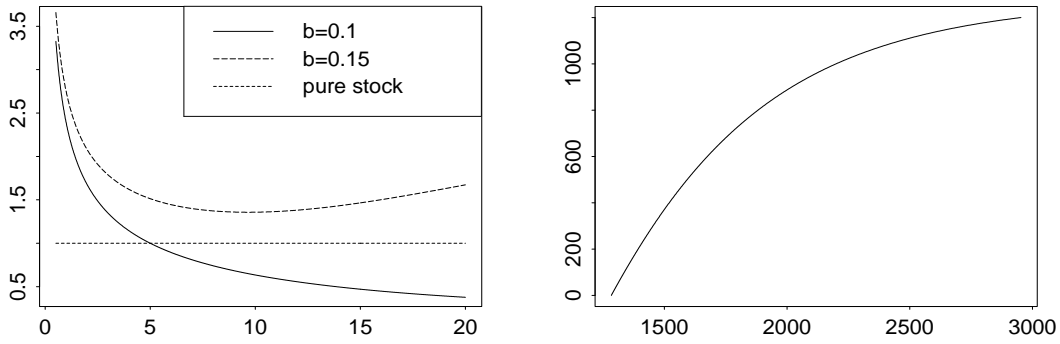$$(b - r\underline{1})'\pi_{\widehat{\varepsilon}} T = h(\widehat{\varepsilon}) \qquad\qquad (5.21)$$

**Figure 5.10.** *For different appreciation rates the leftt-hand figure shows the optimal portfolio and the pure stock portfolio. The right-hand figure shows the mean-*CaR *efficient frontier with the mean on the horizontal axis and the* CaR *on the vertical axis. The parameters are the same as in Figure 5.9.*

for the first time with increasing $\varepsilon > 0$, then this determines the optimal $\widehat{\varepsilon} > 0$. To see this, note that we have

$$E[X^\pi(T)] \leq E[X^{\pi_{\widehat{\varepsilon}}}(T)] \quad \text{for all } \pi \text{ with} \quad |\pi'\sigma| \leq \widehat{\varepsilon} \,,$$

and for all admissible $\pi$ with $\varepsilon = |\pi'\sigma| > \widehat{\varepsilon}$ we obtain

$$(b - r\underline{1})'\pi T \leq h(\varepsilon) < h(\widehat{\varepsilon}) = (b - r\underline{1})'\pi_{\widehat{\varepsilon}} T \,.$$

By solving the non-linear equation (5.21) for $\widehat{\varepsilon}$ we have thus completely determined the solution of problem (5.19) $\qquad\square$

**Example 5.12.** Figure 5.13 compares the behaviour of $\widehat{\varepsilon}$ and $\varepsilon^*$ as functions of the planning horizon $T$. We have used the same data as in Example 5.8. To make the solutions of problems (5.12) and (5.19) comparable we have chosen $C$ differently for the variance and the CaR risk measures in such a way that $\widehat{\varepsilon}$ and $\varepsilon^*$ concide for $T = 5$. Notice that $C$ for the variance problem is roughly the square of $C$ for the CaR problem taking into account that the variance measures an $L^2$-distance, whereas CaR measures an $L^1$-distance. The (of course expected) bottom line of Figure 5.13 is that with increasing time the variance constraint demands a smaller fraction of risky securities in the portfolio. This is also true for the CaR constraint for small time horizons. For larger planning horizon T ($T \geq 20$) $\varepsilon^*$ increases again due to the fact that the CaR decreases. In contrast to that, $\widehat{\varepsilon}$ decreases to 0, since the variance increases. $\qquad\square$

## 5.2   The exponential Lévy model

As in Section 5.1 we consider a standard Black-Scholes type market consisting of a *riskless bond* and several *risky stocks*, however, we assume now that their prices follow exponential Lévy processes. This is a large class of models, including besides the geometric Brownian motion also much more realistic price models. The respective prices $(P_0(t))_{t\geq 0}$ and $(P_i(t))_{t\geq 0}$ for $i = 1, \ldots, d$
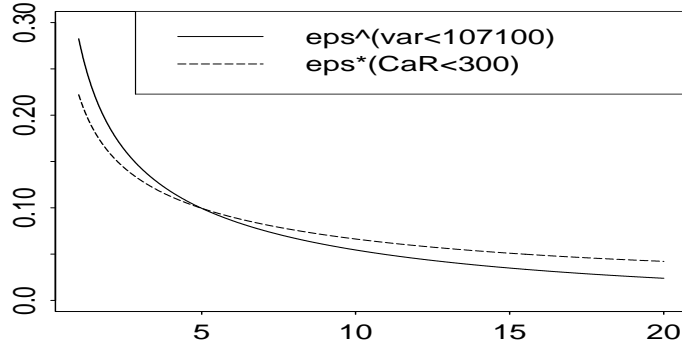
**Figure 5.13.** $\widehat{\varepsilon}$ and $\varepsilon^*$ as functions of the planning horizon; $0 < T \leq 20$. The parameters are the same as in Figure 5.9.

evolve according to the equations

$$P_0(t) = e^{rt} \quad \text{and} \quad P_i(t) = p_i \exp\left(b_i t + \sum_{j=1}^{d} \sigma_{ij} L_j(t)\right), \quad t \geq 0. \tag{5.22}$$

Here $r \in \mathbb{R}$ is the *riskless interest rate*, $b \in \mathbb{R}^d$ and $\sigma = (\sigma_{ij})_{1 \leq i,j \leq d}$ is an invertible matrix.

$(L(t))_{t \geq 0} = (L_1(t), \ldots, L_d(t))_{t \geq 0}$ is a $d$-dimensional Lévy process with independent components. Hence we assume that each $(L_i(t))_{t \geq 0}$ for $i = 1, \ldots, d$ has stationary independent increments with cadlag sample paths. We define this model analogously to the Black-Scholes model in Section 5.1, but replace the Brownian motion by a general Lévy process $L$.

Before we specify this model further we summarize some results on Lévy processes. For relevant background we refer to Bertoin [10], Protter [73] and, in particular, Sato [81]. A very interesting collection of research articles is Barndorff-Nielsen, Mikosch and Resnick [7].

Each infinitely divisible df $F$ on $\mathbb{R}^d$ generates a Lévy process $L$ by choosing $F$ as df of the $d$-dimensional vector $L(1)$. This can be seen immediately, since the characteristic function is for each $t > 0$ given by

$$E \exp(is'L(t)) = \exp(t\Psi(s)), \quad s \in \mathbb{R}^d,$$

where $\Psi$ has Lévy-Khintchine representation

$$\Psi(s) = is'a - \frac{s'\beta'\beta s}{2} + \int_{\mathbb{R}^d} (e^{is'x} - 1 - is'xI(|x| \leq 1))\nu(dx), \quad s \in \mathbb{R}^d. \tag{5.23}$$

Here $a \in \mathbb{R}^d$, $\beta'\beta$ is a non-negative definite symmetric $d \times d$-matrix, and $\nu$ is a measure on $\mathbb{R}^d$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (|x|^2 \wedge 1)\nu(dx) < \infty$, called the Lévy measure of the process $L$. The term corresponding to $xI(|x| \leq 1)$ represents a centering, without which the integral (5.23) may not converge. The characteristic triplet $(a, \beta'\beta, \nu)$ characterizes the Lévy process $L$.

According to Sato [81], Chapter 4 (see Theorem 19.2), the following holds. For each $\omega$ in the probability space, define $\Delta L(t, \omega) = (\Delta L_1(t, \omega), \ldots, \Delta L_d(t, \omega))$ with $\Delta L_j(t, \omega) = L_j(t, \omega) - L_j(t-, \omega)$ for $j = 1, \ldots, d$. For each Borel set $B \subset [0, \infty) \times \mathbb{R}^{d*}$ ($\mathbb{R}^{d*} = \mathbb{R}^d \setminus \{0\}$) set

$$M(B, \omega) = \text{card}\{t \geq 0 \ : \ (t, \Delta L(t, \omega)) \in B\}.$$

Lévy's theory says that $M$ is a Poisson random measure with intensity

$$m(dt, dx) = dt\nu(dx),$$

where $\nu$ is the Lévy measure of the process $L$. Notice that $m$ is $\sigma$-finite and $M(B, \cdot) = \infty$ a.s. when $m(B) = \infty$.

With this notation, the Lévy-Khintchine representation (5.23) corresponds to the representation

$$L(t) = at + \beta W(t) + \sum_{0 < s \le t} \Delta L(s) I(|\Delta L(s)| > 1) + \int_0^t \int_{|x| \le 1} x (M(dx, ds) - \nu(dx)ds), \ t \ge 0 \quad (5.24)$$

This means that $L(t)$ has a Brownian component $\beta W(t)$ and a pure jump part with Lévy measure $\nu$, having the interpretation that a jump of size $x$ occurs at rate $\nu(dx)$. To ensure finiteness of the integral (5.23), the small jumps are compensated by their expectation. This representation reduces in the finite variation case to

$$L(t) = \gamma t + \beta W(t) + \sum_{0 < s \le t} \Delta L(s), \quad t \ge 0, \tag{5.25}$$

where $\gamma = a - \int_{|x| \le 1} x\nu(dx)$; i.e. $L(t)$ is the independent sum of a drift term, a Brownian component and a pure jump part.

We return to model (5.22) with $L$ having characteristic triplet $(a, \beta'\beta, \nu)$, where $a$ is a $d$-dimensional vector, $\beta = \mathrm{diag}(\beta_1, \ldots, \beta_d)$ is a $d$-dimensional diagonal matrix and $\nu$ is the Lévy measure, which corresponds to the product measure of the independent components of $L$ on $\mathbb{R}^d$. This means that e.g. for $d = 2$ and a rectangle $A = (a, b] \times (c, d] \subset \mathbb{R}^2$ the Lévy measure $\nu(A) = \nu_1((a, b]) + \nu_2((c, d])$, where $\nu_i$ is the Lévy measure of $L_i$ for $i = 1, 2$. The diagonal matrix $\beta$ means that the $d$-dimensional Wiener process $W$ has independent components with different variances possible. This allows for different scaling factors in the Wiener processes and the non-Gaussian components; moreover, if some $\beta_i = 0$ the model allows for Lévy processes without Gaussian component as asset price models.

In order to derive the wealth process of a portfolio we need the corresponding SDE. By Itô's formula (see e.g. Protter [73], $P_i$, $i = 1, \ldots, d$, is the solution to the SDE

$$
\begin{aligned}
dP_i(t) &= P_i(t-)(b_i dt + d\widehat{L}_i(t)) \\
&= P_i(t-) \left( \left( b_i + \frac{1}{2} \sum_{j=1}^d (\sigma_{ij} \beta_j)^2 \right) dt + \sum_{j=1}^d \sigma_{ij} (dL_j(t) - \Delta L_j(t)) \right. \\
&\qquad\qquad \left. + \exp\left( \sum_{j=1}^d \sigma_{ij} \Delta L_j(t) \right) - 1 \right), \quad t > 0, \quad P_i(0) = p_i.
\end{aligned}
\tag{5.26}
$$

**Remark 5.14.** (i) Note the similarity but also the difference to the geometric Brownian motion model (5.1). Again the Wiener process introduces an Itô term in the drift component of the SDE. However, there is a main change in the jumps of the Lévy processes. First note that, because of the independence, jumps of the different processes $L_1, \ldots, L_d$ occur at different times (see Sato [81], Exercise E12.10 on p. 67). Then every jump of one of the original processes is replaced: a jump of size $\sum_{j=1}^d \sigma_{ij} \Delta L_j$ is replaced by a jump of size $\exp(\sum_{j=1}^d \sigma_{ij} \Delta L_j) - 1$.

(ii) Note also that $\widehat{L}_i$ is such that

$$\exp\Big(\sum_{j=1}^{d}\sigma_{ij}L_j\Big) = \mathcal{E}(\widehat{L}_i)\,,\quad i = 1,\dots,d\,,$$

where $\mathcal{E}$ denotes the *stochastic exponential* of a process. $\qquad\square$

We shall use the following lemma which relates the characteristic triplet of an exponential Lévy process and its stochastic exponential in $\mathbb{R}$.

**Lemma 5.15.** (Goll and Kallsen [43])
*If $L$ is a real-valued Lévy process with characteristic triplet $(a,\beta,\nu)$, then also $\widehat{L}$ defined by $e^L = \mathcal{E}(\widehat{L})$ is a Lévy process with characteristic triplet $(\widehat{a},\widehat{\beta},\widehat{\nu})$ given by*

$$\begin{aligned}
\widehat{a} - a &= \frac{1}{2}\beta^2 + \int \big((e^x - 1)1_{\{(|e^x-1|<1\}} - x1_{\{|x|<1\}}\big)\nu(dx)\\
\widehat{\beta} &= \beta\\
\widehat{\nu}(\Lambda) &= \nu(\{x \in \mathbb{R} : e^x - 1 \in \Lambda\})\ \text{for any Borel set}\ \Lambda \subset \mathbb{R}^*\,.
\end{aligned}$$

As in the Black-Scholes model before, we restrict ourselves to constant portfolios; i.e. $\pi(t) = \pi$, $t \in [0,T]$, for some fixed planning horizon $T$. In order to avoid negative wealth we require that $\pi \in [0,1]^d$ and $\pi'\underline{1} \le 1$. Denoting by $(X^\pi(t))_{t\ge 0}$ the *wealth process*, it follows the dynamic

$$dX^\pi(t) = X^\pi(t-)\Big(((1 - \pi'\underline{1})r + \pi'b)dt + \pi'd\widehat{L}(t)\Big)\,,\quad t > 0\,,\quad X^\pi(0) = x\,,$$

where $x \in \mathbb{R}$ denotes the *initial capital* of the investor.

Using Itô's formula, this SDE has solution

$$X^\pi(t) = x\exp((r + \pi'(b - r\underline{1}))t)\mathcal{E}(\pi'\widehat{L}(t))\,,\quad t \ge 0\,. \tag{5.27}$$

One important consequence of this represenation is the fact that a jump $\Delta L(t)$ is transformed into a jump $\Delta X^\pi(t) = \ln(1 + \pi'(e^{\sigma\Delta L(t)} - 1)) > \ln(1 - \pi'\underline{1})$ and hence we also require for the portfolio that $\pi'\underline{1} \le 1$.

From (5.26) it is clear that $(X^\pi(t))_{t\ge 0}$ cannot have a nice and simple representation as in the case of geometric Brownian motion; see (5.2). In any case, $(X^\pi(t))_{t\ge 0}$ is again an exponential Lévy process and we calculate the characteristic triplet of its logarithm.

**Lemma 5.16.** *Consider model (5.22) with Lévy process $L$ and characteristic triplet $(a,\beta,\nu)$. Define for the $d \times d$-matrix $\sigma\beta$ the vector $[\sigma\beta]^2$ with components*

$$[\sigma\beta]_i^2 = \sum_{j=1}^{d}(\sigma_{ij}\beta_j)^2\,,\quad i = 1,\dots,d\,.$$

*The process $(\ln X^\pi(t))_{t\ge 0}$ is a Lévy process with triplet $(a_X,\beta_X,\nu_X)$ given by*

$$\begin{aligned}
a_X &= r + \pi'(b - r\underline{1} + [\sigma\beta]^2/2 + \sigma a) - |\pi'\sigma\beta|^2/2\\
&\quad + \int_{\mathbb{R}^d}\big(\ln(1 + \pi'(e^{\sigma x} - \underline{1}))1_{\{|\ln(1+\pi'(e^{\sigma x}-1))|\le 1\}} - \pi'\sigma x1_{\{|x|\le 1\}}\big)\nu(dx)\,,\\
\beta_X &= |\pi'\sigma\beta|\,,\\
\nu_X(A) &= \nu(\{x \in \mathbb{R}^d : \ln(1 + \pi'(e^{\sigma x} - 1)) \in A\})\ \text{for any Borel set}\ A \subset \mathbb{R}^*\,.
\end{aligned}$$

In the finite variation case we obtain

$$\ln \mathcal{E}(\pi'\widehat{L}(t)) = \gamma_X t + \pi'\sigma\beta W(t) + \sum_{0 < s \leq t} \ln\left(1 + \sum_{i=1}^{d} \pi_i \Delta\widehat{L}_i(s)\right)$$

$$= \gamma_X t + \pi'\sigma\beta W(t) + \sum_{0 < s \leq t} \ln\left(1 + \sum_{i=1}^{d} \pi_i\left(\exp(\sum_{j=1}^{d}\sigma_{ij}\Delta L_j(s)) - 1\right)\right), \, t \geq 0,$$

where

$$\gamma_X = \pi'(\sigma\gamma + [\sigma\beta]^2/2) - |\pi'\sigma\beta|^2/2,$$

and $\gamma = a - \int_{|x| \leq 1} x\nu(dx)$ as in (5.25).

By Lemma 5.16 $\ln X^\pi(t)$ has characteristic function $E\exp(is \ln X^\pi(t)) = \exp(t\psi_X(s))$, $s \in \mathbb{R}$. If it can be analytically extended around $s = 0$ in $\mathbb{C}$, then by Theorem 25.17 in Sato [81] we obtain for all $k \in \mathbb{N}$

$$E[(X^\pi(t))^k] = \exp(t\Psi_X(-ik)), \quad t \geq 0. \tag{5.28}$$

In particular, $E\exp(s \ln X^\pi(t)) = E[(X^\pi(t))^s] < \infty$ for one and hence all $t > 0$ if and only if $\int_{|x| > 1} e^{sx} \nu_X(dx) < \infty$.

**Proposition 5.17.** *Let $L = (L_1, \ldots, L_d)$ be a Lévy process with independent components and assume that for all $j = 1, \ldots, d$ the rv $L_j(1)$ has finite moment generating function $\widehat{f}_j$ such that $\widehat{f}_j(\sigma_{ij}) = E\exp(\sigma_{ij}L_j(1)) < \infty$ for $i, j = 1, \ldots, d$. Denote*

$$\widehat{f}(\sigma) := E\exp(\sigma L(1)) = \left(E\exp\left(\sum_{j=1}^{d}\sigma_{1j}L_j(1)\right), \ldots, E\exp\left(\sum_{j=1}^{d}\sigma_{dj}L_j(1)\right)\right). \tag{5.29}$$

*Let $X^\pi(t)$ be as in equation (5.27). Then*

$$E[X^\pi(t)] = x\exp\left(t\left(r + \pi'\left(b - r\underline{1} + \ln\widehat{f}(\sigma)\right)\right)\right),$$

$$\text{var}(X^\pi(t)) = x^2\exp\left(2t\left(r + \pi'\left(b - r1 + \ln\widehat{f}(\sigma)\right)\right)\right)\left(\exp(t\pi'A\pi) - 1\right)$$

*where $A$ is a $d \times d$-matrix with components*

$$A_{ij} = E\exp\left(\sum_{l=1}^{d}(\sigma_{il} + \sigma_{jl})L_l(1)\right) - E\exp\left(\sum_{l=1}^{d}\sigma_{il}L_l(1)\right) - E\exp\left(\sum_{l=1}^{d}\sigma_{jl}L_l(1)\right), \quad 1 \leq i, j \leq d.$$

**Proof.** Formula (5.28) reduces for $k = 1$ and $k = 2$ somewhat, giving together with the expression of $\nu_X$ in terms of $\nu$ of Lemma 5.16,

$$E[X^\pi(t)] = x\exp\left(t\left(r + \pi'\left(b - r\underline{1} + \frac{1}{2}[\sigma\beta]^2 + \sigma a\right.\right.\right.$$

$$\left.\left.\left. + \int_{\mathbb{R}^d}(e^{\sigma x} - 1 - \sigma x 1_{\{|x| < 1\}})\nu(dx)\right)\right)\right),$$

$$\text{var}(X^\pi(t)) = x^2\exp\left(2t\left(r + \pi'\left(b - r\underline{1} + \frac{1}{2}[\sigma\beta]^2 + \sigma a\right.\right.\right.$$

$$\left.\left.\left. + \int_{\mathbb{R}^d}(e^{\sigma x} - 1 - \sigma x 1_{\{|x| < 1\}})\nu(dx)\right)\right)\right)$$

$$\times\left(\exp\left(t\left(|\pi'\sigma\beta|^2 + \int_{\mathbb{R}^d}(\pi'(e^{\sigma x} - \underline{1}))^2\nu(dx)\right)\right) - 1\right).$$

For $i = 1, \ldots, d$ denote by $e_i$ the $i$-th unit vector in $\mathbb{R}^d$. Then the $i$-th component of (5.29) is obtained by

$$E \exp(\sum_{l=1}^{d} \sigma_{il} L_l(1)) = \exp((\sigma a + [\sigma \beta]^2/2 + \int (e^{\sigma x} - 1 - \sigma x 1_{\{|x|<1\}}) \nu(dx))_i) \, .$$

which corresponds to the $i$-th component of $\ln(E \exp(\sigma L(1)))$. The formula for the variance is obtained analogously. $\qquad \square$

**Remark 5.18.** Note that for $l = 1, \ldots, d$ $(i = \sqrt{-1})$

$$\ln(E \exp(\sum_{j=1}^{d} \sigma_{lj} L_j(1))) = \sum_{j=1}^{d} \ln \widehat{f}_j(\sigma_{lj}) = \ln E[\mathcal{E}(\widehat{L}_l)(1)] = \sum_{j=1}^{d} \Psi(-i\sigma_{lj}) \, .$$

This implies in particular $E\mathcal{E}(\pi' \widehat{L}(t)) = (\prod_{l=1}^{d}(E[\mathcal{E}(\widehat{L}_l(t))])^{\pi_l})$, $\qquad \square$

## 5.3 Portfolio optimization

We consider now the portfolio optimization problem using the Capital-at-Risk as risk measure in the more general setting of Lévy processes. The definition of the CaR from Definition 5.1 adapted to the more general situation reads as follows.

**Definition 5.19.** [Capital-at-Risk]
*Let $x$ be the initial capital and $T$ a given planning horizon. Let $z_\alpha$ be the $\alpha$-quantile of $\mathcal{E}(\pi' \widehat{L}(T))$ for some portfolio $\pi \in \mathbb{R}^d$ and $X^\pi(T)$ the corresponding terminal wealth. Then the VaR of $X^\pi(T)$ is given by*

$$
\begin{aligned}
\mathrm{VaR}(x, \pi, T) &= \inf\{z \in \mathbb{R} : P(X^\pi(T) \le z) \ge \alpha\} \\
&= x z_\alpha \exp((\pi'(b - r\underline{1}) + r)T)
\end{aligned}
$$

*and we define*

$$
\begin{aligned}
\mathrm{CaR}(x, \pi, T) &= x \exp(rT) - \mathrm{VaR}(x, \pi, T) \\
&= x \exp(rT)(1 - z_\alpha \exp(\pi'(b - r\underline{1})T))
\end{aligned}
$$

*the* Capital-at-Risk *of the portfolio $\pi$ (with initial capital $x$ and planning horizon $T$).* $\qquad \square$

We consider now the following optimization problem.

$$\max_{\pi \in [0,1]^d, \, \pi'\underline{1} \le 1} E[X^\pi(T)] \quad \text{subject to} \quad \mathrm{CaR}(x, \pi, T) \le C \, .$$

In general, quantiles of Lévy processes cannot be calculated explicitly. Usually, the df of $X^\pi(T)$ is not known explicitly. At first sight there are various possibilities for approximations and we discuss their applicability for quantile estimation below.

For simplicity we restrict ourselves to $d = 1$, i.e. the portfolio consists of the bond and one risky asset, which is modelled by the exponential Lévy process

$$P(t) = p \exp(bt + L(t)) \quad t \ge 0 \, ,$$

where $L$ has characteristic function $Ee^{isL(t)} = e^{t\Psi(s)}$, $s \in \mathbb{R}$. We set $\pi_1 = \pi$ and $X^\pi(t)$ reduces to

$$X^\pi(t) = x\exp((r + \pi(b - r))t)\mathcal{E}(\pi\widehat{L}(t)), \quad t \geq 0, \quad X^\pi(t) = x,$$

where $(\ln\mathcal{E}(\pi\widehat{L}(t)))_{t\geq 0}$ is a Lévy process with characteristic triplet $(a_X, \beta_X, \nu_X)$ given by

$$
\begin{aligned}
a_X &= \pi(a - \frac{1}{2}(1-\pi)\beta^2) + \int\Big(\ln(1 + \pi(e^x - 1))1(|\ln(1 + \pi(e^x - 1))| \leq 1) - \pi x 1(|x| \leq 1)\Big)\nu(dx),\\
\beta_X &= \pi\beta,\\
\nu_X(A) &= \nu(\{x \in \mathbb{R} : \ln(1 + \pi(e^x - 1)) \in A\}) \text{ for any Borel set } A \subset \mathbb{R}^*.
\end{aligned}
$$

Setting $\Psi(-si) = \ln Ee^{sL(1)}$ for $s \in \mathbb{R}$ such that the moment generating function is finite, also the existing moments reduce for $t \geq 0$ to

$$
\begin{aligned}
E[X^\pi(t)] &= x\exp(t(r + \pi(b - r + \Psi(-i))))\\
\mathrm{var}(X^\pi(t)) &= x\exp(2t((r + \pi(b - r + \Psi(-i))))\big(\exp(\pi^2 t(\Psi(-2i) - 2\Psi(-i))) - 1\big)
\end{aligned}
$$

We obtain in the case of a jump part of finite variation for $t \geq 0$,

$$
E[X^\pi(t)] = x\exp((r + \pi(b - r + \frac{1}{2}\beta^2 + \gamma + \widehat{\mu}))t), \tag{5.30}
$$

$$
\begin{aligned}
\mathrm{var}(X^\pi(t)) &= x^2\exp\left(2t\big(r + \pi(b - r + \gamma + \widehat{\mu} + \frac{1}{2}\beta^2)\big)\right)\\
&\quad \times \big(\exp\big(\pi^2 t(\beta^2 + \widehat{\mu}_2 - 2\widehat{\mu})\big) - 1\big),
\end{aligned}\tag{5.31}
$$

where $\widehat{\mu} = \int(e^x - 1)\nu(dx)$, $\widehat{\mu}_2 = \int(e^{2x} - 1)\nu(dx)$, and $\gamma = a - \int_{|x|<1} x\nu(dx)$.

In the following we discuss some estimation methods for the CaR, which means that we have to estimate a small quantile of $\mathcal{E}(\pi'\widehat{L}(T))$; see Definition 5.19.

Simulation methods of Lévy processes are often based on infinite series representations; see Rosinski [79] and references therein. In principle, such methods can be applied here to simulate independent copies of $X^\pi(T)$ and estimate the quantile by its empirical counterpart. Such methods are based on the Lévy measure $\nu_X$, which we derived in Lemma 5.16. There are, however, two serious drawbacks. The first is that low and high quantiles are even in straightforward models not well estimated by their empirical counterparts; the second is that the infinite series has to be truncated, which obviously is another source of inaccuracy.

We invoke instead an idea used for instance by Bondesson [14] and Rydberg [80] for simulation purposes and made mathematically precise by Asmussen and Rosinski [2]. Before we apply their result to approximate a low quantile as the VaR above we explain first the idea. The intuition behind is that small jumps $(< \varepsilon)$ may be approximated by Brownian motion, whereas large ones $(\geq \varepsilon)$ constitute a compound Poisson process $N^\varepsilon$. This normal approximation works for various, but not for all models. In particular, it fails for the exponential variance-gamma model, which has become an important model also in practice. We formulate therefore a more general result.

For a Lévy process with representation (5.24) we write for small $\varepsilon > 0$,

$$
\begin{aligned}
L(t) &= \mu(\varepsilon)t + \beta W(t) + N^\varepsilon(t) + \int_0^t\int_{|x|<\varepsilon} x(M(ds, dx) - ds\nu(dx))\\
&\approx \mu(\varepsilon)t + \beta W(t) + N^\varepsilon(t) + \sigma(\varepsilon)V(t), \quad t \geq 0,
\end{aligned}\tag{5.32}
$$

where $V$ is some (hopefully simple) Lévy process and

$$\sigma^2(\varepsilon) = \int_{|x|<\varepsilon} x^2 \nu(dx), \tag{5.33}$$

$$\mu(\varepsilon) = a - \int_{\varepsilon \le |x| \le 1} x\nu(dx), \tag{5.34}$$

$$N^\varepsilon(t) = \sum_{s \le t} \Delta L(s) 1_{\{|\Delta L(s)| \ge \varepsilon\}}. \tag{5.35}$$

The approximation (5.32) can be made precise. It is a consequence of a functional central limit theorem, provided that for $\varepsilon \to 0$

$$\sigma(\varepsilon)^{-1} \int_0^t \int_{|x|<\varepsilon} x(M(ds,dx) - ds\nu(dx)) = \sigma(\varepsilon)^{-1}(L(t) - L_\varepsilon(t)) \overset{d}{\to} W'(t), \quad t \ge 0, \tag{5.36}$$

where

$$L_\varepsilon(t) = \mu(\varepsilon)t + \beta W(t) + N_\varepsilon(t), \quad t \ge 0. \tag{5.37}$$

We denote by $\overset{d}{\to}$ weak convergence in $D[0, \infty)$ with the supremum norm uniformly on compacta; see Pollard [72].

Since the Brownian component and the jump component of a Lévy process are independent, (5.36) justifies approximation in distribution (5.32).

We want to invoke this result to approximate quantiles of $\mathcal{E}(\pi'\widehat{L}(T))$. We do this in two steps: firstly, we approximate $\mathcal{E}(\pi'\widehat{L}(T))$, secondly, we use that convergence of dfs implies also convergence of their generalized inverses. This gives the approximation of the quantiles.

**Theorem 5.20.** [Emmer and Klüppelberg [34]]
*Let $Y$ be any Lévy process with Lévy measure $\nu$. Let $\mathcal{E}^\leftarrow(\exp(Y(\cdot)) = Z(\cdot)$ be such that $\mathcal{E}Z(\cdot) = \exp Y(\cdot)$ with characteristic triplets given in Lemma 5.15. Let furthermore, $\sigma(\cdot)$ be defined as in (5.33), and $Y_\varepsilon$ and $Z_\varepsilon$ as $L_\varepsilon$ in (5.37), respectively.*

*Let $V$ be a Lévy process. Equivalent are for $\varepsilon \to 0$*

$$\sigma(\varepsilon)^{-1}(Y(t) - Y_\varepsilon(t)) \quad \overset{d}{\to} \quad V(t), \quad t \ge 0, \tag{5.38}$$

$$(\pi\sigma(\varepsilon))^{-1}(\ln \mathcal{E}(\pi Z(t)) - \ln \mathcal{E}(\pi Z_\varepsilon(t))) \quad \overset{d}{\to} \quad V(t), \quad t \ge 0. \tag{5.39}$$

For the proof we need the following theorem.

**Theorem 5.21.** *Let $Z^\varepsilon$, $\varepsilon > 0$, be Lévy processes without Brownian component and $Y^\varepsilon = \ln \mathcal{E}(Z^\varepsilon)$ their logarithmic stochastic exponentials with characteristic triplets $(a_Z, \beta_Z, \nu_Z)$ and $(a_Y, \beta_Y, \nu_Y)$ as defined in Lemma 5.15. Let $g : \mathbb{R} \to \mathbb{R}_+$ with $g(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $V$ some Lévy process. Then equivalent are as $\varepsilon \to 0$,*

$$\frac{Z^\varepsilon(t)}{g(\varepsilon)} \quad \overset{d}{\to} \quad V(t), \quad t \ge 0,$$

$$\frac{Y^\varepsilon(t)}{g(\varepsilon)} \quad \overset{d}{\to} \quad V(t), \quad t \ge 0.$$

**Proof of (5.38) $\Leftrightarrow$ (5.39).** Setting $g(\varepsilon) := \sigma(\varepsilon)$ and $Y^\varepsilon := Y - Y_\varepsilon$ in Theorem 5.21 we obtain that (5.38) holds if and only if

$$\sigma(\varepsilon)^{-1}\mathcal{E}^{\leftarrow}(\exp(Y(t) - Y_\varepsilon(t))) \quad \overset{d}{\to} \quad W(t), \quad t \geq 0. \tag{5.40}$$

Applying Theorem 5.21 to $g(\varepsilon) := \pi\sigma(\varepsilon)$ and $Z_\varepsilon := \pi\mathcal{E}^{\leftarrow}(\exp(Y(t) - Y_\varepsilon(t)))$ leads to the equivalence of (5.40) and

$$(\pi\sigma(\varepsilon))^{-1}\ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp(Y(t) - Y_\varepsilon(t)))) \quad \overset{d}{\to} \quad W(t), \quad t \geq 0.$$

The identity

$$\ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp(Y(t) - Y_\varepsilon(t)))) = \ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp(Y(t)))) - \ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp Y_\varepsilon(t))), \quad t \geq 0, \tag{5.41}$$

which can be proven by calculating all three logarithmic exponentials by Itô's formula (see Emmer and Klüppelberg [34]), leads to the equivalence with (5.39). $\qquad\square$

In the finite variation case (5.36), i.e. (5.38) can be rewritten to

$$\sigma(\varepsilon)^{-1}\left(\sum_{0 < s \leq t}\Delta L(s)I(|\Delta L(s)| < \varepsilon) - E\left[\sum_{0 < s \leq t}\Delta L(s)I(|\Delta L(s)| < \varepsilon)\right]\right) \overset{d}{\to} V(t), \quad t \geq 0,$$

which shows immediately the connection to the classical central limit theorem.

We apply (5.39) and (5.41) to approximate $\ln\mathcal{E}(\pi\widehat{L})$ for $\pi \in (0,1]$ as follows.

$$\ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp(L(t)))) \quad \approx \quad \ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(L_\varepsilon(t)))) + \pi\sigma(\varepsilon)V(t), \quad t \geq 0,$$

and hence we obtain

$$\ln\mathcal{E}(\pi\mathcal{E}^{\leftarrow}(\exp(L(t)))) \quad \approx \quad \gamma_\pi^\varepsilon t + \pi\beta W(t) + M_\pi^\varepsilon(t) + \pi\sigma(\varepsilon)V(t), \quad t \geq 0,$$

where

$$\begin{aligned} \gamma_\pi^\varepsilon &= \pi(\mu(\varepsilon) + (1 - \pi)\beta^2/2), \\ M_\pi^\varepsilon(t) &= \sum_{s \leq t}\ln(1 + \pi(\exp(\Delta L(s)1_{\{|\Delta L(s)| > \varepsilon\}}) - 1)), \end{aligned}$$

i.e. $M_\pi^\varepsilon$ is a compound Poisson process with jump measure

$$\nu_{M_\pi^\varepsilon}(\Lambda) = \nu_L(\{x : \ln(1 + \pi(e^x - 1)) \in \Lambda\}\setminus(-\varepsilon, \varepsilon))$$

for any Borel set $\Lambda \subset \mathbb{R}^*$. $\qquad\square$

By Proposition 0.1 of Resnick [75] we obtain the corresponding approximation for the $\alpha$-quantile $z_\alpha$ of $\mathcal{E}(\pi\widehat{L}(T))$.

**Proposition 5.22.** *With the quantities as defined above we obtain*

$$z_\alpha \approx z_\alpha^\varepsilon(\pi) = \inf\{z \in \mathbb{R} : P(\gamma_\pi^\varepsilon T + \pi\beta W(T) + M_\pi^\varepsilon(T) + \pi\sigma_L(\varepsilon)V(T) \leq \ln z) \geq \alpha\},$$

*giving the following approximations*

$$\begin{aligned} \mathrm{VaR}(x, \pi, T) &\approx xz_\alpha^\varepsilon(\pi)\exp((\pi(b - r) + r)T), \\ \mathrm{CaR}(x, \pi, T) &\approx xe^{rT}\left(1 - z_\alpha^\varepsilon(\pi)e^{\pi(b-r)T}\right). \end{aligned}$$

The following corollary characterises the normal approximation.

**Corollary 5.23.** [Asmussen and Rosinski [2]]

(a) *V is standard Brownian motion if and only if*

$$\sigma(h\sigma(\varepsilon) \ \wedge \ \varepsilon) \sim \sigma(\varepsilon) \ \text{for each } h > 0 \,. \tag{5.42}$$

(b) *Condition (5.42) holds if* $\lim_{\varepsilon\downarrow 0} \sigma(\varepsilon)/\varepsilon = \infty \,.$

(c) *If the Lévy measure has no atoms in a neighbourhood of 0, then condition (5.42) is equivalent to* $\lim_{\varepsilon\downarrow 0} \sigma(\varepsilon)/\varepsilon = \infty \,.$

Provided the above condition is satisfied we have reduced the problem of estimating a quantile of a complicated Lévy process to the estimation of a quantile of the sum of the compound Poisson rv $M_\pi^\varepsilon(T)$ and the normal rv $\widetilde{W}(T) := \pi'(\beta^2 + \sigma_L^2(\varepsilon))^{1/2}W(T)$. We calculate the density of $M_\pi^\varepsilon(T) + \widetilde{W}(T)$ using the Fast Fourier Transform method, henceforth abbreviated as FFT. By independence, we have for the characteristic function of $M_\pi^\varepsilon(T) + \widetilde{W}(T)$

$$\phi(u) = \phi_{M_\pi^\varepsilon(T)}(u)\phi_{\widetilde{W}(T)}(u)\,, \quad u \in \mathbb{R}. \tag{5.43}$$

Denote by $h_{M_\pi^\varepsilon}$ the Lévy density of $M_\pi^\varepsilon$, which we assume to exist, then we obtain

$$\phi_{M_\pi^\varepsilon(T)}(u) = \exp(T\nu_{M_\pi^\varepsilon}(\mathbb{R})(\phi_Y(u) - 1))\,, \quad u \in \mathbb{R},$$

where

$$\phi_Y(u) = \frac{1}{\nu_{M_\pi^\varepsilon}(\mathbb{R})} \int e^{iux} h_{M_\pi^\varepsilon}(x)dx\,, \quad u \in \mathbb{R}. \tag{5.44}$$

Furthermore, by normality,

$$\phi_{\widetilde{W}(T)}(u) = \exp(-Tu^2\pi^2(\beta^2 + \sigma_L^2(\varepsilon)))\,, \quad u \in \mathbb{R}.$$

We approximate the integral in (5.44) by the trapezoid rule, choosing a number $n$ (a power of 2) of intervals and a step size $\Delta x$. Set $g = h_{M_\pi^\varepsilon}/\nu_{M_\pi^\varepsilon}(\mathbb{R})$. We truncate the integral $\phi_Y$ and obtain

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{iux}g(x)dx \ &\approx \ \int_{-(n/2)\Delta x}^{(n/2-1)\Delta x} e^{iux}g(x)dx \\
&\approx \ \sum_{k=-(n/2)}^{n/2-1)} e^{iuk\Delta x}g((k\Delta x)\Delta x \\
&= \ \sum_{k=0}^{n-1} e^{iu(k-n/2)\Delta x}g((k-n/2)\Delta x)\Delta x \\
&= \ \Delta x e^{-iun\Delta x/2} \sum_{k=0}^{n-1} e^{iuk\Delta x}g((k-n/2)\Delta x)\,.
\end{aligned}
$$

For $g_k := g((k-n/2)\Delta x)$, $k = 0,\ldots,n-1$, the sum is the discrete Fourier transform of the complex numbers $g_k$ and can be calculated by the FFT algorithm for $u_k = 2\pi k/(n\Delta x)$, $k = 0,\ldots,n-1$, simultaneously (see e.g. Brigham [20], Chapter 10). This results in an approximation for $\phi$ in (5.43). By the inverse FFT we obtain the density of $M_\pi^\varepsilon(T) + \widetilde{W}(T)$.

**Example 5.24.** *[Exponential Brownian motion with jumps]*
Here the Lévy process is the sum of a Brownian motion with drift $(\beta W(t) + \gamma t)_{t\geq 0}$, and a compound Poisson process $(L(t))_{t\geq 0}$, with Poisson intensity $c$ and $p$ as distribution of the jump heights $(Y_i)_{i\in\mathbb{N}}$. For illustratrive purpose we restrict this example to one compound Poisson process, we could as well take several different ones, see e.g. [35]. The drift $\gamma = -\frac{1}{2}\beta^2 - \widehat{\mu}$ is chosen such that it compensates the jumps. The Lévy measure is $\nu(dx) = cp(dx)$ and hence $\widehat{\mu} = c(\widehat{g}(1) - 1)$ and $\widehat{\mu}_2 = c(\widehat{g}(2) - 1)$, where $g$ is the moment generating function of $Y_1$, which we assume to exist at the required points. By (5.30) and (5.31) we obtain for $t \geq 0$

$$
\begin{aligned}
X^\pi(t) &= x\exp\left(t\left(r + \pi(b - r) - \pi\widehat{\mu} - \frac{1}{2}\pi^2\beta^2\right) + \pi\beta W(t)\right)\prod_{i=1}^{N(t)}(1 + \pi(e^{Y_i} - 1)), \\
E[X^\pi(t)] &= x\exp(t(r + \pi(b - r))), \\
\mathrm{var}(X^\pi(t)) &= x^2\exp(2t(r + \pi(b - r)))\left(\exp(\pi^2 t(\beta^2 + c(\widehat{g}(2) - 2\widehat{g}(1) + 1))) - 1\right).
\end{aligned}
$$

Note that for $c = 0$ the model reduces to exponential Brownian motion; i.e.

$$
X^\pi(t) = x\exp\left(t\left(r + \pi(b - r) - \frac{1}{2}\pi^2\beta^2\right) + \pi\beta W(t)\right).
$$

On the other hand, if $\beta = 0$ the model reduces to exponential compound Poisson process; i.e.

$$
X^\pi(t) = x\exp\left(t\left(r + \pi(b - r) - \pi\widehat{\mu}\right)\right)\prod_{i=1}^{N(t)}(1 + \pi(e^{Y_i} - 1)).
$$

**Example 5.25.** *[Exponential normal inverse Gaussian (NIG) Lévy process]*
This normal mixture model has been suggested by Barndorff-Nielsen [5, 4]; see also Eberlein and collaborators [28, 29, 30] It has the representation

$$
L(t) = \rho t + \lambda\sigma^2(t) + \sigma(t)\varepsilon, \quad t \geq 0,
$$

where $\rho, \lambda \in \mathbb{R}$, $\varepsilon$ is a standardnormal rv and $(\sigma^2(t))_{t\geq 0}$ has inverse Gaussian increments. The process $(L(t))_{t\geq 0}$ is uniquely determined by the distribution of the increment $L(1)$ which is NIG (see Barndorff-Nielsen [5]). This means that $L(1) \overset{d}{=} N(\rho + \lambda Z, Z)$, where $N(a, b)$ denotes a normal rv with mean $a$ and variance $b$ and $Z$ is inverse Gauss distributed; more precisely, the density of $L(1)$ is given by

$$
nig(x, \alpha, \lambda, \rho, \delta) := \frac{\alpha}{\pi}\exp\left(\delta\sqrt{\alpha^2 - \lambda^2} + \lambda(x - \rho)\right)\frac{K_1(\delta\alpha g(x - \rho))}{g(x - \rho)}, \quad x \in \mathbb{R},
$$

where $\alpha \geq |\lambda| \geq 0$, $\delta > 0$, $\rho \in \mathbb{R}$, $g(x) = \sqrt{\delta^2 + x^2}$ and

$$
K_1(s) = \frac{1}{2}\int_0^\infty \exp\left(-\frac{1}{2}s(x + x^{-1})\right)dx
$$

is the modified Bessel function of the third kind. The parameter $\alpha$ is a steepness parameter, i.e. for larger $\alpha$ we get less large and small jumps and more jumps of middle height, $\delta$ is a scale parameter, $\lambda$ is a symmetry parameter and $\rho$ a location parameter. For $\rho = \lambda = 0$ (symmetry around 0) the characteristic triplet of a NIG Lévy process is given by $(0, 0, \nu)$ with

$$
\nu(dx) = \frac{\delta\alpha}{\pi}|x|^{-1}K_1(\alpha|x|)dx, \quad x \in \mathbb{R}^*.
$$

We can calculate $\widehat{\mu}$ and $\widehat{\mu}_2$ via the moment generating function of $L(1)$, which is for the NIG distribution given by

$$E \exp(sL(1)) = \exp(\delta(\alpha - \sqrt{\alpha^2 - s^2})), \quad |s| < \alpha,$$

(see e.g. Raible [74], Example 1.6) and hence because of symmetry,

$$\begin{aligned}
\widehat{\mu} &= \delta(\alpha - \sqrt{\alpha^2 - 1}), \\
\widehat{\mu}_2 &= \delta(\alpha - \sqrt{\alpha^2 - 4}).
\end{aligned}$$

Plugging these results into (5.30) and (5.31), and choosing $b_{nig} = b_{BS} - \delta(\alpha - \sqrt{\alpha^2 - 1})$ ($b_{BS}$ is the quantity $b$ from Example 5.24, such that the expectation of an asset in the NIG model is the same as for the exponential Brownian motion, we obtain for $t \geq 0$,

$$\begin{aligned}
X^\pi(t) &= x \exp(t(r + \pi(b_{BS} - r - \delta(\alpha - \sqrt{\alpha^2 - 1})))) \prod_{0 < s \leq t} \left(1 + \pi(e^{\Delta L(s)} - 1)\right), \\
E[X^\pi(t)] &= x \exp(t(r + \pi(b_{BS} - r))), \\
\text{var}(X^\pi(t)) &= x^2 \exp(2t(\pi(b_{BS} - r)) + r)) \left(\exp\left(\delta\pi^2 t(2\sqrt{\alpha^2 - 1} - \alpha - \sqrt{\alpha^2 - 4})\right) - 1\right).
\end{aligned}$$

By Corollary 5.23, for the exponential normal inverse Gaussian Lévy process the normal approximation for small jumps is allowed since $\sigma^2(\varepsilon) \sim (2\delta/\pi)\varepsilon$ as $\varepsilon \to 0$

For an estimate of the $\alpha$-quantile we invoke Proposition 5.22 and use FFT.

Figures 5.26 show sample paths a geometric NIG-Lévy process with certain parameter values.
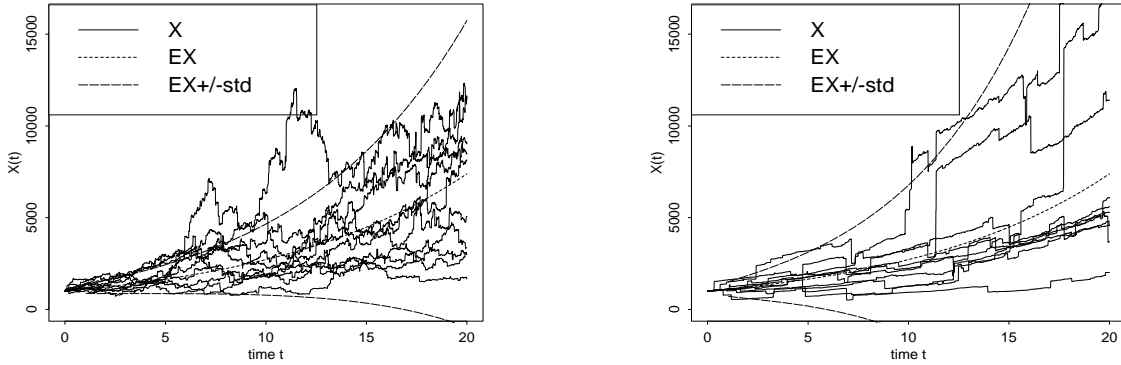


**Figure 5.26.** *Ten sample paths of the exponential NIG Lévy process with $\alpha = 8$ and $\delta = 0.32$ (left) and with $\alpha = 2$ and $\delta = 0.08$ (right), its expectation $\mathcal{E}(L(T))$ and expectation$\pm$standard deviation for $x = 1\,000$, $b = 0.1$, and $r = 0.05$.*

**Example 5.27.** *[Exponential variance gamma (VG) model]*
This normal mixture model has been suggested by Madan and Seneta [64], its non-symmetric version can be found in Madan, Carr and Chang [63]. An interesting empirical investigation has been conducted by Carr et al. [22]. The non-symmetric model is defined as follows.

$$L(t) = \rho t + \lambda\sigma^2(t) + \sigma(t)\varepsilon, \quad t \geq 0,$$

where $\rho, \lambda \in \mathbb{R}$, $\varepsilon$ is a standardnormal rv and $(\sigma^2(t))_{t \geq 0}$ has gamma increments, more precisely, $\sigma^2(s) \stackrel{d}{=} \Gamma(\alpha s, \theta)$ for parameters $\alpha > 0$ and $\theta > 0$; i.e. it has density

$$h(x) = \frac{x^{\alpha s - 1}}{\Gamma(\alpha s)\theta^{\alpha s}} e^{-x/\theta}, \quad x > 0.$$

The characteristic function of $L(1)$ is given by

$$E\exp(isL(1)) = \frac{\exp(is\rho t)}{(1 - is\theta\lambda + s^2\theta/2)^{\alpha t}} \quad s \in \mathbb{R}.$$

The Lévy process $L$ is a pure jump process with Lévy density

$$\nu(dx) = \frac{\alpha}{|x|} \exp\left(-\sqrt{\frac{2}{\theta} + \lambda^2}\,|x| + \lambda x\right) dx, \quad x \in \mathbb{R}^*.$$

We obtain as before

$$
\begin{aligned}
X^\pi(t) &= x\exp(t(r + \pi(b - r + \rho)))\prod_{s \leq t}(1 + \pi(e^{\Delta L(s)} - 1)) \\
E[X^\pi(t)] &= x(1 - \theta\lambda - \theta/2)^{-\alpha\pi t}\exp(t(r + \pi(b - r + \rho))) \\
\mathrm{var}(X^\pi(t)) &= x^2(1 - \theta\lambda - \theta/2)^{-2\alpha\pi t}\exp(2t(r + \pi(b - r + \rho))) \\
&\quad \times \left(\left(\frac{(1 - \theta\lambda - \theta/2)^2}{1 - 2\theta\lambda - 2\theta}\right)^{\alpha\pi^2 t} - 1\right).
\end{aligned}
$$

For our figures we choose $\rho = \alpha\ln(1 + \theta\lambda - \theta/2)$ and $b = b_{BS}$ such that $E[X^\pi(t)] = x\exp((r + \pi(b_{BS} - r))t)$. In order to find an approximation for the VaR we calculate $\sigma^2(\varepsilon) \sim \alpha\varepsilon^2$ as $\varepsilon \to 0$. Since its Lévy measure has no atoms in a neighbourhood of 0, by Corollary 5.23, the normal approximation for small jumps is not allowed.

However, there is another limit process to allow for approximation of the small jumps: for $\varepsilon \to 0$

$$\sigma(\varepsilon)^{-1}(L(t) - L_\varepsilon(t)) \stackrel{d}{\to} V(t), \quad t \geq 0,$$

where $V$ is a Lévy process with characteristic triplet $(0, 0, \nu_V)$ where the Lévy measure $\nu_V$ has density $\nu_V(dv) = (\alpha/v)1_{(-1/\sqrt{\alpha}, 1/\sqrt{\alpha})}(v)dv$. This means that the following approximation is valid

$$z_\alpha \approx z_\alpha^\varepsilon(\pi) = \inf\{z \in \mathbb{R} : P(\gamma_\pi^\varepsilon T + M_\pi^\varepsilon + \pi\sigma(\varepsilon)V(T) \leq \ln z) \geq \alpha\},$$

giving again approximations as in Proposition 5.22.

**Remark 5.30.** When we want to perform a portfolio optimization for the different exponential Lévy models as price processes, then certain structures can be exploited. Note e.g. that the expected wealth process is increasing in $\pi$; hence the optimal portfolio is always the largest $\pi$ such that the risk bound is satisfied. For Lévy processes additionally $\pi \leq 1$ has to be satisfied. Such $\pi$ can always easily be found by a simple numerical iteration procedure.

Next note that to make results comparable we have chosen all mean portfolio processes equal.

(a) *Mean-variance optimization:* Since NIG and VG models have so many parameters we can always choose them so that all variances are equal in the different examples. Then, of course,
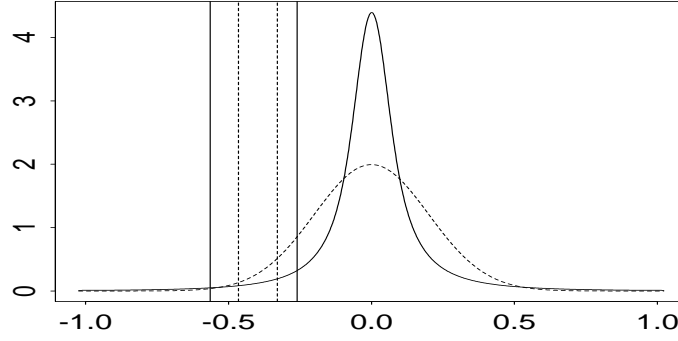
**Figure 5.28.** *Density of $L(1)$ for the NIG model with parameters $\alpha = 2$, $\delta = 0.08$, $\lambda = \rho = 0$, $x = 1\,000$, $b = 0.1$ and $r = 0.05$. The normal density with the same variance 0.04 is plotted for comparison. Moreover, the respective 1% quantiles (left vertical lines) and 5%-quantiles (richt vertical lines) are plotted. All solid lines correspond to the NIG model, all dotted ones to the normal model.*
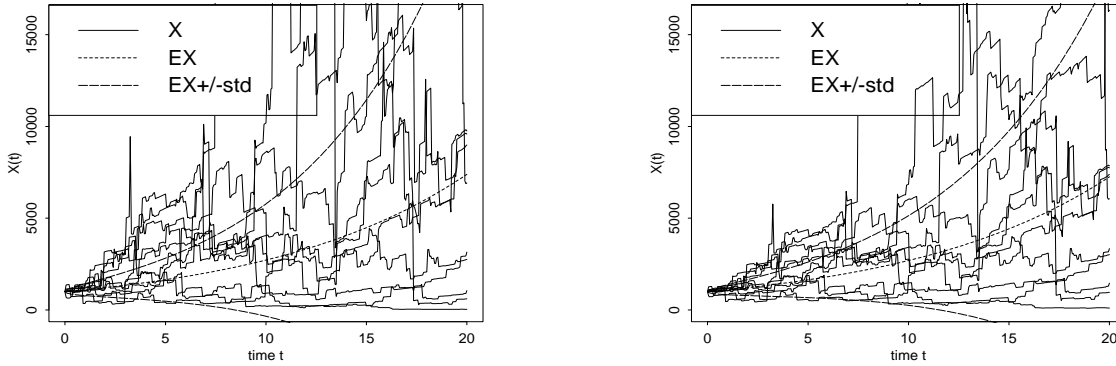


**Figure 5.29.** *Ten sample paths of the exponential VG Lévy process with $\alpha = 0.1$, $\delta = 0$, $\theta = 0.35$ and $\mu = -0.019$ (left) and with $\alpha = 0.2$, $\delta = 0$, $\theta = 0.2$ and $\mu = -0.022$ (right), its expectation $E[L(T)]$ and expectation±standard deviation for $x = 1\,000$, $b = 0.1$, and $r = 0.05$.*

the mean-variance optimization problem always leads to the same result.

(c) *Mean-CaR optimization:* Here the shape of the distribution in the left tail enters; see Figure 5.28. The heavier the tail at the corresponding $\alpha$-quantile, the higher the risk, i.e. the more cautious the investment $\pi$ into the risky stock.

For more details see Emmer and Klüppelberg [34]                                    □

# References

[1] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) Coherent measures of risk. *Math. Finance* **9**, 203-228.

[2] Asmussen, S. and Rosinski, J. (2000) Approximation of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.* **38**, 482-493.

[3] Balkema, A.A. and Haan, L. de (1974) Residual lifetime at great age. *Ann. Probab.* 2, 792-804.

[4] Barndorff-Nielsen, O.E. (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1-14.

[5] Barndorff-Nielsen, O.E. (1998) Processes of normal inverse Gaussian type. *Finance and Stochastics* **2**, 41-68.

[6] Barndorff-Nielsen, O.E. and Cox, D. (1994) *Inference and Asymptotics.* Chapman and Hall, London.

[7] Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I. (Eds.) (2001) *Lévy Processes - Theory and Applications.* Birkhäuser. Boston.

[8] Barndorff-Nielsen, O.E. and Shephard, N. (2001) Modelling by Lévy processes for financial econometrics. In: Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S. (eds) *Lévy Processes –Theory and Applications.* Birkhäuser, Boston.

[9] Berman, S.M. (1992) *Sojourns and Extremes of Stochastic Processes.* Wadsworth, Belmond, California.

[10] Bertoin, J. (1996) *Lévy Processes.* Cambridge University Press. Cambridge.

[11] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

[12] Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journ. Econometrics* **31**, 307–327.

[13] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992) ARCH modelling in finance: a review of the theory and empirical evidence. *Journ. Econometrics* **52**, 5-59.

[14] Bondesson, L. (1982) On simulation from infinitely divisible distributions. *Adv. Appl. Probab.* **14**, 855-869.

[15] Borkovec, M. (2000) Extremal behaviour of the autoregressive process with ARCH(1) errors. *Stoch. Proc. Appl.* **85**, 189-207.

[16] Borkovec, M. (2001) Asymptotic behaviour of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH(1) errors. *Bernoulli* **7**, 847-872.

[17] Borkovec, M. and Klüppelberg, C. (1998) Extremal behaviour of diffusion models in finance. *Extremes* **1**, 47-80.

[18] Borkovec, M. and Klüppelberg, C. (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220-1241.

[19] Brandt, A., Franken, P. and Lisek, B. (1990) *Stationary Stochastic Models.* Wiley, Chichester.

[20] Brigham, E.O. (1974) *The Fast Fourier Transform.* Prentice-Hall. Englewood Cliffs, N.J.

[21] Campbell, H,Y, Lo, A.W. and MacKinley, A.C. (1997) *The Econometrics of Financial Markets.* Princeton University Press. Princeton, NJ.

[22] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2001) The fine structure of asset returns: an empirical investigation. *Forthcoming in Journal of Business.*

[23] Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of the term structure of interest rates. *Econometrica* **53**, 385-408.

[24] Cvitanic, J. and Karatzas, I. (1999) On dynamic measures of risk. *Finance and Stochastics* **3**, 451-482.

[25] Davis, R.A. (1982) Maximum and minimum of one-dimensional diffusions. *Stoch. Proc. Appl.* **13**, 1-9.

[26] Diaconis, P. and Freedman, D. (1999) Iterated random functions. *SIAM Review* **41**, 45-76.

[27] Eberlein, E. (2001) Application of generalized hyperbolic Lévy motion to finance. In: Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S. (eds) *Lévy Processes –Theory and Applications.* Birkhäuser, Boston.

[28] Eberlein, E. and Keller, U. (1995) Hyperbolic distributions in finance. *Bernoulli* **1**, 281-299.

[29] Eberlein, E., Keller, U. and Prause, K. (1998) New insights into smile, mispricing and value at risk: the hyperbolic model. *Journal of Business* **71**, 371-405.

[30] Eberlein, E. and Raible, S. (1999) Term structure models driven by general Lévy processes. *Math. Finance.* **9**, 31-53.

[31] Embrechts. P. (2002) Semstat presentation.

[32] Embrechts. P. (Ed.) (2000) *Extremes and Integrated Risk Management.* UBS Warburg and Risk Books.

[33] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[34] Emmer, S. and Klüppelberg, C. (2001) Optimal portfolios when stock prices follow an exponential Lévy process. Preprint. Munich University of Technology. *www.ma.tum.de/stat/*

[35] Emmer, S., Klüppelberg, C. and Korn, R. (2000) Optimal portfolios with bounded Capital-at-Risk. *Math. Finance* **11**, 365-384.

[36] Emmer, S., Klüppelberg, C., Trüstedt, M. (1998) VaR - ein Mass für das extreme Risiko. *Solutions* **2**, 53-63. English version available at *http://www.ma.tum.de/stat/*

[37] Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1007.

[38] Engle, R.F. (Ed.) (1995) *ARCH. Selected Readings.* Oxford University Press, Oxford.

[39] Fishburn, P.C. (1977) Mean-risk analysis with risk associated with below-target returns. *American Economic Review* **67**, 116-126.

[40] Föllmer, H. and Leukert, P. (1999): Quantile hedging. *Finance and Stochastics* **3**, 251-273.

[41] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Prob.* **1**, 126–166.

[42] Goldie, C.M. and Maller, R. (2000) Stability of perpetuities. *Ann. Probab.* **28**, 1195-1218.

[43] Goll, T. and Kallsen, J. (2000) Optimal portfolios for logarithmic utility. *Stoch. Proc. Appl.* **89**, 31-48.

[44] Gouriéroux, C. (1997) *ARCH Models and Financial Applications.* Springer, New York.

[45] Grigelionis, B. (2001) On point measures of $\varepsilon$-upcrossings for stationary diffusions. Preprint. Vilnius University.

[46] Haan, L. de, Resnick, S.I., Rootzén, H. and Vries, C. G. de (1989) Extremal behaviour of solutions to a stochastic difference equation, with applications to ARCH processes. *Stoch. Proc. Appl.* **32**, 213-224.

[47] Harlow, W. (1991) Asset allocation in a downside-risk framework. *Financial Analyst Journal*, 28-40.

[48] Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988) On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78**, 97-112.

[49] Itô, K.I. and McKean, H.P. (1974) *Diffusion Processes and Their Sample Paths.* Springer, Berlin.

[50] Joriot, P. (1997) *Value at Risk: the New Benchmark for Controlling Market Risk.* Irwin. Chicago.

[51] Kallenberg, O. (1983) *Random Measures.* Akademie Verlag, Berlin.

[52] Karatzas, I. and Shreve, S.E. (1988) *Brownian Motion and Stochastic Calculus.* Springer, New York.

[53] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrixes. *Acta Math.* **131**, 207-248.

[54] Kesten, H. (1974) Renewal theory for functional of a Markov chain with general state space. *Ann. Probab.* **2**, 355-386.

[55] Kiefersbeck, N. (1999) Stationarität und Tailindex bei zeitdiskreten Volatilitätsmodellen. Diploma Thesis. Munich University of Technology. *www.ma.tum.de/stat/*

[56] Kloeden, P.E. and Platen, E. (1992) *Numerical Solutions of Stochastic Differential Equations.* Springer, Berlin.

[57] Klüppelberg, C., Maller R.A., Van De Vyver M., Wee D. (2000) Testing for reduction to random walk in autoregressive conditional heteroscedasticity models. Submitted for publication.

[58] Klüppelberg, C. and Pergamenchtchikov, S. (2001) Renewal theory for functionals of a Markov chain with compact state space. Submitted for publication. *www.ma.tum.de/stat/*

[59] Klüppelberg, C. and Pergamenchtchikov, S. (2001) The tail of the stationary distribution of a random coefficient AR($q$) process with applications to an ARCH($q$) process. Sumitted for publication. *www.ma.tum.de/stat/*

[60] Korn, R. (1997): *Optimal Portfolios.* World Scientific, Singapore.

[61] Lamberton, D. and Lapeyre, B. (1991) *Introduction au calcul stochastique appliqué à la finance.* Edition Marketing S.A., Ellilpses, SMAI, Paris. English translation by Rabeau, N. and Mantion, F. (1996) *Introduction to Stochastic Calculus Applied to Finance.* Chapman and Hall, London.

[62] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer. Berlin.

[63] Madan, D.B., Carr, P. and Chang, E. (1998) The variance gamma process and option pricing. *Forthcoming in the European Finance Review.*

[64] Madan, D.B. and Seneta, E. (1990) The variance gamma (VG) model for share market returns. *Journal of Business* **63**, 511-524.

[65] Markowitz, H. (1959): *Portfolio Selection − Efficient Diversification of Investments.* New York: Wiley.

[66] Merton, R.C. (1994) *Continuous−time Finance.* Revised Ed. Blackwell, Cambridge, MA.

[67] Meyn, S. and Tweedie, R. (1993) *Markov Chains and Stochastic Stability.* Springer, New York.

[68] Mikosch, T. (2001) Modeling financial time series. Semstat presentation.

[69] Mikosch, T. and Starica, C. (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427-1451. *www.math.ku.dk/∼mikosch*

5de matrixs aléatoires. equations aux différences

[70] Perfekt, R. (1994) Extremal behaviour of stationary Markov chains with applications. *Ann. Appl. Probab.* **4**, 529-548.

[71] Pickands, J. III (1975) Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.

[72] Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer, New York.

[73] Protter, P. (1990): *Stochastic Integrals and Differential Equations.* New York: Springer.

[74] Raible, S. (2000) *Lévy Processes in Finance: Theory, Numberics and Empirical Facts.* PhD thesis, University of Freiburg; *http://www.freidok.uni-freiburg.de/volltexte/51*

[75] Resnick, S.I. (1987) *Extreme Values, Regular Variation and Point Processes.* Springer. New York.

[76] Revuz, D. and Yor, M. (1991) *Continuous Martingales and Brownian Motion.* Springer, Berlin.

[77] Rootzén, H. und Klüppelberg, C. (1999) A single number can't hedge against economic catastrophes. *Ambio* **28**, No 6, 550-555. Royal Swedish Academy of Sciences.

[78] Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov Processes, and Martingales, Vol. 2., 2nd Ed.* Cambridge University Press, Cambridge.

[79] Rosinski, J. (2001) Series representations of Lévy processes from the perspective of point processes. In: Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S. (eds) *Lévy Processes −Theory and Applications.* Birkhäuser, Boston.

[80] Rydberg, T. (1997) The normal inverse Gaussian Lévy process: simulation and approximation. *Stoch. Models* **13**, 887-910.

[81] Sato, K. I. (1999) *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press. Cambridge.

[82] Sharpe, W. (1964) A theory of market equilibrium under condition of risk. *Journal of Finance* **19**, 1536-1575.

[83] Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (Eds.) *Likelihood, Time Series with Econometric and other Applications.* Chapman and Hall, London.

[84] Smith, R. (2001) Semstat presentation.

[85] Taylor, S. (1995) *Modelling Financial Time Series.* Wiley, Chichester.

[86] Tweedie, R.L. (1976) Criteria for classifying general Markov chains. *Adv. Appl. Prob.* **8**, 737–771.

[87] Vasicek, O.A. (1977) An equilibrium characterization of the term structure. *J. Fin. Econ.* **5**, 177-188.

[88] Vervaat, W. (1979) On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Prob.* **11**, 750–783.

[89] Weiss, A.A. (1984) ARMA models with ARCH errors. *J. Time Ser. Analysis* **3**, 129–143.