

HECTOR SCHOOL OF ENGINEERING
AND MANAGEMENT
UNIVERSITY FRIDERICIANA KARLSRUHE

Module S-5/1

Insurance, Risk Analysis and Asset Liability Management

Asset Liability Management

Svetlozar T. Rachev

Literature Recommendations

- 1) Jitka Dupačová, Jan Hurt and Josef Štěpán, Stochastic Modeling in Economics and Finance, Kluwer Academic Publisher, 2002
- 2) Stavros A. Zenios, Lecture Notes: Mathematical modeling and its application in finance, <http://www.hermes.ucy.ac.cy/zenios/teaching/399.001/index.html>
- 3) F. Fabozzi and A. Konisbi, The Handbook of Asset/Liability Management: State-of-Art Investment Strategies, Risk Controls and Regulatory Required, Wiley.
- 4) Svetlozar Rachev and Stefan Mittnik, Stable Paretian Models in Finance, John Wiley & Sons Ltd., 2000

Contents

1	Introduction	1
2	Risk and Optimization	6
2.1	Risk Measures	6
2.2	Coherence	9
2.3	Risk-Return Optimization	10
2.4	CVaR Optimization	14
2.4.1	Uryasev's Optimization Shortcut	15
3	Portfolio Optimization and Stochastic Programming	18
3.1	1-stage Portfolio Optimization	18
3.2	Single-stage versus Multistage Optimization	20
3.3	Formulation of the Stochastic Program	22
3.4	Scenario Generation	25
3.5	Deterministic Equivalent Forms	28
3.6	The T-stage ALM problem	32
4	Modeling of the Risk Factors	35
4.1	Stable Distributions	36
4.1.1	Definition of Stable Random Variables	36

4.1.2	Parameters and Special Cases of the Stable Distribution	39
4.1.3	Properties of Stable Random Variables	41
4.1.4	Truncated α -Stable Distributions	43
4.2	Stable Modeling of Risk Factors	46
4.2.1	Modeling Financial Returns with Stable Distributions	46
4.3	Univariate and Multivariate Distributions	49
4.4	Fitting a Multivariate Distribution	51
4.5	Dependence Modeling and Copulas	53
5	ALM Implementation	62
5.1	Finding an adequate model	65
5.1.1	Exponentially Weighted Moving Average Models	71
5.1.2	VaR Backtesting	78
5.2	Solving the Optimization Problem	80
5.3	Efficient Frontiers	83
5.4	Postoptimality Analysis and Backtesting	85
5.4.1	Postoptimality Analysis	85
5.4.2	Portfolio Backtesting	87
6	Appendix - Tables for Empirical Analysis	94

Chapter 1

Introduction

Asset liability management (ALM) attempts to find the optimal investment strategy under uncertainty in both the asset and liability streams. In the past, the two sides of the balance sheet have usually been separated, but simultaneous consideration of assets and liabilities can be very advantageous when they have common risk factors. If assets are allocated such that they are highly correlated with the liabilities, it is possible to reduce the risk of the entire portfolio.

Traditionally, banks and insurance companies used accrual accounting for essentially all their assets and liabilities. They would take on liabilities, such as deposits, life insurance policies or annuities. They would invest the proceeds from these liabilities in assets such as loans, bonds or real estate. All assets and liabilities were held at book value. Doing so disguised possible risks arising from how the assets and liabilities were structured.

Two of the earlier ALM frameworks for constructing portfolios of fixed-income securities are dedication and immunization. Basic dedication assumes that the future liability payments are deterministic and finds an allocation such that bond income is sufficient to cover the liability payments in each time period. Achieving this type of cashflow matching in every period is likely to be costly, so traditional

immunization models match cashflows on average providing a cheaper, but usually riskier, portfolio. The immunized portfolio is constructed by matching the present values and interest rate sensitivities of the assets and liabilities, and it results in an allocation that hedges against a small parallel shift in the term structure of interest rates.

Consider the following simple example (see from riskglossary.com): A bank borrows USD 100 Mio at 3.00% for a year and lends the same money at 3.20% to a highly-rated borrower for 5 years. For simplicity, we assume that all interest rates are annually compounded and all interest accumulates to the maturity of the respective obligations. The net transaction appears profitable, since the bank is earning a 20 basis point spread. However, the transaction also entails considerable risk:

At the end of a year, the bank will have to find new financing for the loan, which will have 4 more years before it matures. If interest rates have risen, the bank may have to pay a higher rate of interest on the new financing than the fixed 3.20 it is earning on its loan. Suppose, for example that at the end of a year, an applicable 4-year interest rate is 6.00%. The bank is in serious trouble. It is going to be earning 3.20% on its loan and paying 6.00% on its financing.

Accrual accounting does not recognize the problem. The book value of the loan (the bank's asset) is:

$$100Mio \cdot 1.032 = 103.2Mio$$

The book value of the financing (the bank's liability) is:

$$100\text{Mio} \cdot 1.030 = 103.0\text{Mio}$$

Based upon accrual accounting, the bank earned USD 200,000 in the first year.

However, market value accounting recognizes the bank's predicament. The respective market values of the bank's asset and liability are:

$$100\text{Mio} \cdot 1.032^5 1.060^4 = 92.72\text{Mio}$$

Hence, from a market-value accounting standpoint, the bank has lost USD 10.28 Mio. So which result offers a better portrayal of the bank's situation, the accrual accounting profit or the market-value accounting loss? The bank is in trouble, and the market-value loss reflects this. Ultimately, accrual accounting will recognize a similar loss. The bank will have to secure financing for the loan at the new higher rate, so it will accrue the as-yet unrecognized loss over the 4 remaining years of the position.

The problem in this example was caused by a **mismatch between assets and liabilities**. Prior to the 1970's, such mismatches tended not to be a significant problem. Interest rates in developed countries experienced only modest fluctuations, so losses due to asset-liability mismatches were small or trivial. Many firms intentionally mismatched their balance sheets. Because yield curves were generally upward sloping, banks could earn a spread by borrowing short and lending long. But things started to change in the 1970s, which ushered in a period of volatile interest rates that continued into the early 1980s. US regulation which had capped the interest rates that banks could pay depositors, was abandoned

to stem a migration overseas of the market for USD deposits. Managers of many firms, who were accustomed to thinking in terms of accrual accounting, were slow to recognize the emerging risk. Some firms suffered staggering losses. Because the firms used accrual accounting, the result was not so much bankruptcies as crippled balance sheets. Firms gradually accrued the losses over the subsequent 5 or 10 years.

One of the victims of the changing conditions is the US mutual life insurance company the Equitable. During the early 1980s, the USD yield curve was inverted, with short-term interest rates spiking into the high teens. The Equitable sold a number of long-term guaranteed interest contracts (GICs) guaranteeing rates of around 16% for periods up to 10 years. During this period, GICs were routinely for principal of USD 100 Mio or more. Equitable invested the assets short-term to earn the high interest rates guaranteed on the contracts. Short-term interest rates soon came down. When the Equitable had to reinvest, it couldn't get nearly the interest rates it was paying on the GICs. The firm was crippled. Eventually, it had to demutualize and was acquired by the Axa Group.

We conclude that the the earlier framework of accrual accounting is inadequate for ALM because it misses the stochastic nature of interest rates and liabilities and the dynamic nature of investing. The two main tools that help to capture the dynamic and stochastic characteristics are **stochastic control** and **stochastic programming**. Stochastic control methods model uncertainty in a continuous-time setting through Itô processes, but a drawback is that only a few driving variables, or states, can be handled. Applications of stochastic control in ALM is for example the surplus optimization for pension funds and life insurance.

This lecture is organized as follows:

- Chapter two reviews major issues on risk and optimization. Properties of the

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) risk measure are covered.

- Chapter three provides information on single- and multistage optimization problems, stochastic programs and scenario generation.
- Chapter four discusses the choice of an adequate distribution for modeling of the risk factors with focus on the stable distributions that is able to model features like heavy-tails, skewness and excess kurtosis in the underlying risk factors. We further investigate how multivariate data can be modeled adequately.
- Chapter five provides an example of an empirical application of ALM techniques to a pension fund. Additional issues like portfolio backtesting, comparison of the results based on the underlying risk factors are treated as well.

Chapter 2

Risk and Optimization

The goal of risk-return optimization is to optimize a tradeoff between the risk and return. This chapter reviews a few risk measures and discusses how they can be implemented in simple single-stage portfolio optimization problems. The techniques for optimizing CVaR presented in this chapter will later be used in a multistage problem.

2.1 Risk Measures

The standard measure of risk for a portfolio of equities suggested by Markowitz is the *variance* of the return. A portfolio consists of weights $\omega = (\omega_1, \dots, \omega_n)'$, such that $\omega_i \geq 0$ and $\sum_{i=1}^n \omega_i = 1$, in n assets with corresponding risky returns $R = (r_1, \dots, r_n)'$. The risk associated with the portfolio return $r_p = \omega'R$ is given by $\sigma_p^2 = \omega'\Sigma\omega$, where Σ is the covariance matrix of R . While the variance of the investment return is the most traditional risk measure, a common criticism is that the variance penalizes both large gains and large losses. A modification to an

asymmetric risk measure that accounts only for large losses is the *semivariance*:

$$E([\omega'E(R) - \omega'R]^+)^2.$$

However, numerical optimization of semivariance is difficult. Another modification is the *downside formula* which measures the degree that the returns are distributed below some target return r^* :

$$E([r^* - \omega'R]^+)^2.$$

A second criticism of variance is that financial returns are typically heavy-tailed, and in that case, the variance does not even exist. A logical argument can then be made for using the *mean absolute deviation* (MAD) of the portfolio:

$$m_p = E|\omega'R - \omega'E(R)|,$$

or alternatively, the scale parameter of a stable distribution can be used in place of the variance. Stable distributions will be discussed in more detail in chapter 3.

Some other risk measures rely only on the tail of the distribution, in which case the modeling of the probability of extreme events becomes more important. The following, VaR and CVaR, are two such measures. Value at Risk (VaR) is a frequently used measure of risk for financial institutions and regulators. For a given confidence level $\beta \in (0, 1)$, VaR is the minimum value of the loss, or negative return, that is exceeded no more than $100(1-\beta)\%$ of the time. Its ease of understanding helps to make it a popular risk measure.

The following notations and definitions of VaR and CVaR resemble mostly those in [33]. For a given decision $x \in \mathbb{R}^n$, let the random variable $L(x) \in \mathbb{R}$

represent a loss, or negative return, for each x , and let $\Psi_L(x, \zeta)$ be the distribution function for $L(x)$:

$$\Psi_L(x, \zeta) = P(L(x) \leq \zeta).$$

For a given decision x , the *Value at Risk* at confidence level β is given by

$$\text{VaR}_\beta(x) = \inf \{ \zeta | \Psi_L(x, \zeta) \geq \beta \}.$$

Discussion of several other tail risk measures, including the Conditional Value at Risk (CVaR), can be found in [1] and [2]. While it is not widely used in finance, it has properties that make it a very logical alternative to VaR. These properties are referred to as *coherence* and will be described shortly.

Define a random variable $T_\beta(x)$ on the β -tail of the loss $L(x)$ through the distribution function:

$$\Psi_{T_\beta}(x, \zeta) = \begin{cases} 0 & \zeta < \text{VaR}_\beta(x) \\ \frac{\Psi_L(x, \zeta) - \beta}{1 - \beta} & \zeta \geq \text{VaR}_\beta(x) \end{cases} \quad (2.1)$$

For a given decision x , the *Conditional Value at Risk* at confidence level β is the mean of the tail random variable $T_\beta(x)$ with distribution function (2.1):

$$\text{CVaR}_\beta(x) = E(T_\beta(x)).$$

As is implied by its name, CVaR is closely related to the conditional expectation beyond VaR. In general, CVaR satisfies the inequalities

$$E(L(x) | L(x) \geq \text{VaR}_\beta(x)) \leq \text{CVaR}_\beta(x) \leq E(L(x) | L(x) > \text{VaR}_\beta(x)). \quad (2.2)$$

If there is no discontinuity in the distribution function of $L(x)$ at $\text{VaR}_\beta(x)$, then

equality holds in equation (2.2). For this reason, CVaR is also sometimes called the *Expected Tail Loss* (ETL). When there is a discontinuity, as illustrated in [33], $\text{CVaR}_\beta(x)$ splits the probability atom at $\text{VaR}_\beta(x)$ in a certain way. CVaR is defined in this manner because it has an equivalent representation that is easily optimized in the case of a discrete distribution such as in a scenario tree. This representation will be referred to as Uryasev's formula and is reviewed shortly.

2.2 Coherence

To help define a sensible risk measure, [3] introduces properties that are required of a *coherent* risk measure; however, VaR does not satisfy these properties in general. As is well known, VaR is not sub-additive: Examples have been constructed where the VaR of the sum of two portfolios is greater than the sum of individual VaRs. Lack of subadditivity is very undesirable because diversification is not promoted. However, for the special class of elliptical distribution, VaR is sub-additive and coherent [5].

The following properties of coherence are stated adhering to the axiomatic definition in [1]. If V is the space of real-valued random variables, a risk measure is a functional $\rho : V \rightarrow \mathbb{R}$. If the random variables $v, v' \in V$ are thought of as losses, then ρ is coherent if it is

- i. sub-additive: $\rho(v + v') \leq \rho(v) + \rho(v')$,
- ii. positive homogeneous: $\rho(\lambda v) = \lambda \rho(v)$, $\forall \lambda \geq 0$,
- iii. translation invariant: $\rho(v + c) = \rho(v) + c$, $\forall c \in \mathbb{R}$, and
- iv. monotonous: $\rho(v) \geq 0$, $\forall v \geq 0$.

Proof of the coherence of CVaR can be found in [2, 29, 33]. The coherence of the set of random variables $\{L(x)\}$ can be stated as a function of x when $L(x)$ is linear:

$$L(x) = x_1 Y_1 + \dots + x_n Y_n.$$

In this situation, Y_i might be a random variable representing an individual asset loss, and $L(x)$ is a random variable representing the total portfolio loss. Coherence of $\text{CVaR}_\beta(x)$ in this framework means

- i. $\text{CVaR}_\beta(x)$ is sublinear in x ,
- ii. $\text{CVaR}_\beta(x) = c$ when $L(x) = c \in \mathbb{R}$, and
- iii. $\text{CVaR}_\beta(x) \leq \text{CVaR}_\beta(x')$ when $L(x) \leq L(x')$.

See [33] for the proof.

Note that sub-additivity and positive homogeneity guarantees that a coherent risk measure is convex which is advantageous in portfolio optimization. A lack of convexity of VaR contributes to numerical difficulties in optimization. VaR is easy to work with when normality of distributions is assumed, but financial data is typically heavy-tailed. We will also consider optimization under uncertainty where discrete probability distributions arise from scenario trees. In addition to coherence, CVaR has a representation that is practical in minimization problems with scenarios generated from any distributional assumption.

2.3 Risk-Return Optimization

If the risky returns R are assumed to follow a multivariate normal distribution $N(\mu, \Sigma)$, the portfolio return $r_p = \omega' R$ is also normally distributed with mean $\mu_p = \omega' \mu$ and variance $\sigma_p^2 = \omega' \Sigma \omega$. The classical mean-variance optimization

problem is to minimize the risk of the portfolio for a minimum level of expected return:

$$\begin{aligned} \min_{\omega} \quad & \omega' \Sigma \omega \\ \text{s.t.} \quad & \omega' \mu = \mu_0, \\ & \sum_{i=1}^n \omega_i = 1. \end{aligned} \tag{2.3}$$

The solution to the above problem is easily solved with Lagrangian techniques and can be found in [6]. As μ_0 is varied, the set of portfolios trace out the mean-variance efficient frontier. If no short-selling is allowed, the restriction $\omega_i \geq 0$ is also included.

A drawback of optimization problem (2.3) is that it requires a large number of parameters to be estimated. If there are n risky assets, the covariance matrix consists of $n(n+1)/2$ elements. For instance, if the universe of assets consists of the S&P500, over 125,000 variances/covariances must be estimated. A solution, as found in [37], is to model each asset with a multifactor equation:

$$r_i = \mu_i + \beta_{i1}F_1 + \dots + \beta_{ik}F_k + \epsilon_i, \tag{2.4}$$

where F_j is the deviation of the random factor j from its mean and $cov(F_j, F_l) = 0$ for all $j \neq l$. Examples of typical factors include inflation, interest rates, and GDP. The asset specific risks ϵ_i have zero expectation, are uncorrelated, and are independent of the factors. The portfolio $r_p = \omega' R$ can be written as

$$r_p = \mu_p + \sum_{j=1}^k \beta_{pj} F_j + \epsilon_p,$$

where

$$\mu_p = \omega' \mu, \quad \beta_{pj} = \sum_{i=1}^n \omega_i \beta_{ij}, \quad \epsilon_p = \sum_{i=1}^n \omega_i \epsilon_i.$$

It follows the variance of the portfolio is

$$\sigma_p^2 = \sum_{j=1}^k \beta_{pj}^2 \sigma_{F_j}^2 + \sum_{i=1}^n \omega_i^2 \sigma_{\epsilon_i}^2.$$

The first term in the right-hand side of this equation is the systematic or market risk, and the second term is the unsystematic risk of the portfolio. If equal weight is given to each asset, $\omega_i = 1/n$, the unsystematic risk is bounded by c/n for some constant c , so this risk can be diversified away as n grows large. Using the factor model in the minimum variance optimization problem gives:

$$\begin{aligned} \min_{\omega} \quad & \sigma_p^2 = \sum_{j=1}^k \beta_{pj}^2 \sigma_{F_j}^2 + \sum_{i=1}^n \omega_i^2 \sigma_{\epsilon_i}^2 \\ \text{s.t.} \quad & \omega' \mu = \mu_0, \\ & \beta_{pj} = \sum_{i=1}^n \omega_i \beta_{ij} \\ & \sum_{i=1}^n \omega_i = 1. \end{aligned}$$

The factor sensitivities β_{ij} , factor variances, and specific risk variances can be estimated through linear regression in equation (2.4). This results in a significant reduction in the number of parameter estimates needed as compared to optimization problem (2.3).

Both of the above are quadratic optimization problem. As an alternative to mean-variance analysis, one can optimize the risk measures mentioned in the previous section. Also in [37], the author illustrates that a linear optimization problem can be achieved when the variance of the portfolio is replaced with its mean-absolute deviation m_p . Since R is multivariate normal, the relation holds that $m_p = \sqrt{\frac{2}{\pi}} \sigma_p$, so minimizing the mean-absolute deviation will produce the same optimal portfolio as minimizing the variance. In addition, the linear equivalent program is easily modified to penalize upside and downside deviations from

the mean with different weights.

The class of elliptical distributions offers special properties in portfolio theory that are useful in minimizing VaR or CVaR. The following gives a very brief review; a more complete introduction to elliptical distributions and their portfolio implications is found in [5]. For any elliptically distributed random vector R with finite variance for all univariate marginals, variance is equivalent to any positive homogeneous risk measure ρ . If $r_p = \omega'R$ and $\tilde{r}_p = \tilde{\omega}'R$ are two linear portfolios with corresponding variances σ_p^2 and $\tilde{\sigma}_p^2$:

$$\rho(r_p - \mathbb{E}(r_p)) \leq \rho(\tilde{r}_p - \mathbb{E}(\tilde{r}_p)) \iff \sigma_p^2 \leq \tilde{\sigma}_p^2.$$

In addition if ρ is translation invariant, the solution to the following risk-return optimization problems coincide:

$$\begin{array}{ll} \min_{\omega} & \sigma_p^2 \\ \text{s.t.} & r_p = \omega'R, \\ & \mathbb{E}(r_p) = \mu_0, \\ & \sum_{i=1}^n \omega_i = 1, \end{array} \quad \begin{array}{ll} \min_{\omega} & \rho(r_p) \\ \text{s.t.} & r_p = \omega'R, \\ & \mathbb{E}(r_p) = \mu_0, \\ & \sum_{i=1}^n \omega_i = 1, \end{array}$$

where μ_0 is the desired return. Therefore, under this distributional assumption, minimization of VaR, CVaR, or variance will produce the same optimal portfolios. This follows because CVaR is always coherent, and VaR is coherent for this class of distributions.

The stable assumption makes portfolio optimization more difficult since variance is infinite and cannot be used as a risk measure. One natural solution is to use the scale parameter σ_p^α of the portfolio return. The scale parameter is just a generalization of the standard deviation of a normal distribution. Chapter 3 defines stable random vectors and the special case of a sub-Gaussian distribution,

which is also in the class of elliptical distributions. If Q is the dispersion matrix of the sub-Gaussian distribution, it can be shown that the CVaR and VaR of the portfolio return are both strictly increasing functions of the dispersion parameter of the portfolio return $\omega'Q\omega$. Therefore, for a sub-Gaussian random vector R , minimization of VaR or CVaR can both be achieved by the portfolio optimization problem:

$$\begin{aligned} \min_{\omega} \quad & \omega'Q\omega \\ \text{s.t.} \quad & \omega'\mu = \mu_0, \\ & \sum_{i=1}^n \omega_i = 1. \end{aligned}$$

Details of stable portfolio theory are found in [30], and a comparison of allocations under the normal and stable assumptions is found in [27].

2.4 CVaR Optimization

One would like to be able to perform risk-return analysis for a portfolio by minimizing VaR or CVaR subject to a constraint on the return for any distributional assumption. In general, VaR is difficult to optimize and is usually not used in this setting. Typically, one can model the returns with any distribution and then generate a discrete distribution of scenarios, but in this case, VaR is non-smooth and non-convex in the portfolio positions with multiple local extrema [36]. CVaR, on the other hand, has a representation that is easy to optimize both as a constraint and as an objective for a set of scenarios. Additionally, if CVaR is constrained to be small, VaR must necessarily be small. Conversely, minimization of VaR may produce very different solutions than minimization of CVaR: VaR minimization may stretch the tail of the distribution beyond VaR resulting in a poor CVaR value.

2.4.1 Uryasev's Optimization Shortcut

As defined earlier, for the decision $x \in \mathbb{R}^n$, $L(x)$ is the random variable representing the loss, or negative return, with associated $\text{VaR}_\beta(x)$ and $\text{CVaR}_\beta(x)$. To begin, define the function

$$\Gamma_\beta(x, \zeta) = \zeta + \frac{1}{1-\beta} \mathbb{E}([L(x) - \zeta]^+), \quad (2.5)$$

then CVaR is expressed as a minimization through the following optimization shortcut: $\Gamma_\beta(x, \cdot)$ is finite and continuous with

$$\text{CVaR}_\beta(x) = \min_{\zeta \in \mathbb{R}} \Gamma_\beta(x, \zeta), \quad (2.6)$$

and, in addition,

$$\text{VaR}_\beta(x) = \text{lower endpoint of } \operatorname{argmin}_\zeta \Gamma_\beta(x, \zeta).$$

Equation (2.6) will be referred to as *Uryasev's formula*. As a corollary, it can be shown that if $L(x)$ is convex in x , then $\text{CVaR}_\beta(x)$ is convex in x and $\Gamma_\beta(x, \zeta)$ is jointly convex in (x, ζ) . In addition, if a constraint set X is convex, one obtains a convex minimization problem in (x, ζ) : Minimizing $\text{CVaR}_\beta(x)$ with respect to $x \in X$ is equivalent to minimizing $\Gamma_\beta(x, \zeta)$ with respect to $(x, \zeta) \in X \times \mathbb{R}$, i.e.

$$\min_{x \in X} \text{CVaR}_\beta(x) = \min_{(x, \zeta) \in X \times \mathbb{R}} \Gamma_\beta(x, \zeta) \quad (2.7)$$

The proofs of the above results are found in [33].

Similar to mean-variance efficient frontiers, [18] illustrates risk-reward analysis using CVaR as a risk measure. If $R(x)$ is a concave reward function and X is

convex, then

$$\min_{x \in X} \text{CVaR}_\beta(x) \quad \text{subject to} \quad R(x) \geq \lambda, \quad (2.8)$$

$$\min_{x \in X} \text{CVaR}_\beta(x) - \lambda R(x), \quad \text{and} \quad (2.9)$$

$$\min_{x \in X} -R(x) \quad \text{subject to} \quad \text{CVaR}_\beta(x) \leq \lambda, \quad (2.10)$$

produce the same efficient frontiers as λ is varied. As is already shown, the optimal solution to (2.8) can be found by a joint convex optimization problem. Similarly, problems (2.9) and (2.10) produce the same optimal solution as

$$\min_{(x, \zeta) \in X \times \mathbb{R}} \Gamma_\beta(x, \zeta) - \lambda R(x),$$

and

$$\min_{x \in X} -R(x) \quad \text{subject to} \quad \Gamma_\beta(x, \zeta) \leq \lambda,$$

respectively.

An extension of these optimization procedures to risk shaping with CVaR is found in [33]. For confidence level β_i with corresponding loss tolerance λ_i , for $i = 1, \dots, I$,

$$\min_{x \in X} -R(x) \quad \text{subject to} \quad \text{CVaR}_{\beta_i}(x) \leq \lambda_i, \quad \text{for} \quad i = 1, \dots, I,$$

has the same optimal solution as

$$\min_{(x, \zeta_1, \dots, \zeta_I) \in X \times \mathbb{R} \times \dots \times \mathbb{R}} -R(x) \quad \text{subject to} \quad \Gamma_{\beta_i}(x, \zeta_i) \leq \lambda_i, \quad \text{for} \quad i = 1, \dots, I.$$

When $L(x)$ has a discrete distribution arising from, for example, a scenario

tree or sampling, equation (2.5) becomes

$$\tilde{\Gamma}_\beta(x, \zeta) = \zeta + \frac{1}{1-\beta} \sum_{i=1}^S p_i [L_i(x) - \zeta]^+, \quad (2.11)$$

where $L(x)$ takes the value $L_i(x)$ with probability p_i for $i = 1, \dots, S$. Additionally if $L(x)$ is linear, then $\tilde{\Gamma}_\beta$ is convex and piecewise linear. By introducing auxiliary variables, a CVaR optimization problem can be solved by linear programming as illustrated in the next section.

Chapter 3

Portfolio Optimization and Stochastic Programming

In this chapter we will continue the problem of portfolio optimization and introduce the stochastic programming as a solution technique.

3.1 1-stage Portfolio Optimization

We will first consider the problem of a 1-stage portfolio optimization. Hereby, we will apply Uryasev's formula to risk-return analysis with CVaR and obtains a linear programming problem.

Define

$$X = \left\{ \omega \in \mathbb{R}^n \left| \sum_{j=1}^n \omega_j = 1, \omega_j \geq 0, j = 1, \dots, n \right. \right\}, \quad (3.1)$$

where $x \in X$ represents the portfolio weights in n assets. The random return on these assets at the end of a time period is represented by $R = (r_1, \dots, r_n)'$, and the

negative return of the portfolio is given by

$$L(x) = -x'R.$$

If the mean of R is given by the vector μ , the risk-return problem is

$$\min_{x \in X} \text{CVaR}_\beta(x) \quad \text{s.t.} \quad x'\mu \geq \mu_0,$$

where μ_0 is the required portfolio return, and by varying μ_0 , the efficient frontier is obtained. This optimization problem fits into the form of equation (2.8). If the uncertainty in the return is given through the set of scenarios $\{R^1, \dots, R^S\}$ where each $R^s \in \mathbb{R}^n$ occurs with probability p^s , Uryasev's optimization shortcut produces the equivalent problem

$$\begin{aligned} \min \quad & \zeta + \frac{1}{1-\beta} \sum_{s=1}^S p^s [-x'R^s - \zeta]^+ \\ \text{s.t.} \quad & x'\mu \geq \mu_0, \\ & x \in X, \zeta \in \mathbb{R}, \end{aligned}$$

and by introducing auxiliary variables y^s , $s = 1, \dots, S$, a linear program results:

$$\begin{aligned} \min \quad & \zeta + \frac{1}{1-\beta} \sum_{s=1}^S p^s y^s \\ \text{s.t.} \quad & x'\mu \geq \mu_0, \\ & x'R^s + \zeta + y^s \geq 0, \quad s = 1, \dots, S, \\ & y^s \geq 0, \quad s = 1, \dots, S, \\ & x \in X, \zeta \in \mathbb{R}. \end{aligned}$$

This program is used to compare hedging strategies for international asset allocation in [36]. In addition, the CVaR portfolio is compared with a portfolio minimizing the mean-absolute deviation. The empirical results indicate that CVaR

and MAD produce similar risk-return frontiers in a static setting. However, in dynamic backtesting where the models are repeatedly applied over a time horizon, CVaR produces higher returns and lower volatility than MAD.

3.2 Single-stage versus Multistage Optimization

Extending the single period risk-return problem to a multi-period setting is difficult and some modifications are necessary. In a multi-period setting, one usually deals with a wealth process instead of returns so that problems will be convex and sometimes linear. The general form of a stochastic program with recourse allows **any** portfolio allocation to be made in each stage, and one typically optimizes a function of the wealth process, not the return process, over the quantities of assets held, not the portfolio weights. Instead of risk-return analysis, one can perform risk-reward analysis where the risk, for instance, is a function of the wealth process and the reward is the expected terminal wealth. This is the type of problem that is constructed in the next chapter.

Decision rules such as fixed-mixed are useful because they reduce the decision space, but they also limit the dynamic nature of the optimization problem. For instance, one multi-period extension of mean-variance analysis is found in [20]:

$$\max \quad \lambda E(w_T) - (1 - \lambda) \text{var}(w_T).$$

Here, w_T is the terminal wealth, and the max is taken over all fixed-mixed decision rules. In a fixed-mixed rule, the portfolio is reallocated in each time period to keep a certain percentage of wealth in each asset. As λ is varied between zero and one, a type of efficient frontier is obtained. While the number of decision variables are greatly reduced, the problem becomes non-convex, and a global search algorithm

is necessary.

The coherence of a risk measure in a multi-period setting is also defined in terms of a wealth process $w = (w_1, \dots, w_T)$ where w_1 is a known deterministic wealth. It is shown in [14] that a weighted average of CVaR over the time horizon is coherent: If $\text{CVaR}_\beta(-w_t)$ is the CVaR associated with the negative wealth $-w_t$, then a coherent risk measure is given by

$$\rho(w) = \rho(w_1, \dots, w_T) = \sum_{t=2}^T \mu_t \text{CVaR}_\beta(-w_t), \quad (3.2)$$

where the weights are nonnegative and sum to one. Here, coherence means that ρ is

- i. convex: $\rho(\lambda w + (1 - \lambda)\tilde{w}) \leq \lambda\rho(w) + (1 - \lambda)\rho(\tilde{w})$, $\forall \lambda \in [0, 1]$,
- ii. positive homogeneous: $\rho(\lambda w) = \lambda\rho(w)$, $\forall \lambda \geq 0$,
- iii. translation invariant: $\rho(w_1 + c, \dots, w_T + c) = \rho(w) - c$, $\forall c \in \mathbb{R}$, and
- iv. monotonous: if $w_t \leq \tilde{w}_t$ a.s. for $t = 1, \dots, T$, then $\rho(w) \geq \rho(\tilde{w})$.

When implementing the risk measure in (3.2), one can apply Uryasev's optimization shortcut in a similar manner as the previous sections: Uryasev's formula can be applied to each $\text{CVaR}_\beta(-w_t)$ where the loss L is taken to be the negative wealth $-w_t$, and the wealth in each stage is a function of some decision variables. Of course, there will also be constraints such as the balance of wealth between stages. This is illustrated in detail in the next chapter for the surplus wealth in an ALM problem.

3.3 Formulation of the Stochastic Program

Stochastic programming offers a framework that can incorporate many of the characteristics of an ALM problem. We will first discuss a general setup for stochastic programs with recourse. In the next chapter we will then apply this framework to an ALM problem for a pension fund.

In a 2-stage recourse problem, a recourse decision is made after a realization of uncertainty. The first stage has a vector of initial decisions $x_1 \in \mathbb{R}^{n_1}$ made at $t = 1$ when there is a known distribution of future uncertainty. The second stage decisions $x_2 \in \mathbb{R}^{n_2}$ adapt at $t = 2$ after the first stage uncertainty ξ_1 is realized. The second stage decisions usually also consider the distribution of future uncertainty ξ_2 realized after $t = 2$. For instance, consider an asset allocation problem: The first stage decision is the initial portfolio allocation, the uncertainty is the asset returns, and the recourse decision is the portfolio adjustments. This 2-stage recourse problem finds the optimal initial and rebalanced allocations for the given distribution of future stock movements.

This setup is described mathematically by first considering how the optimal recourse decision is determined. For a given first stage decision vector x_1 and a given realization of the first stage uncertainty ξ_1 , the best recourse decision is found through the following second stage problem:

$$\begin{aligned} \min_{x_2} \quad & q_2(x_1, x_2, \xi_1) + \mathbb{E}_{\xi_2} (Q_2(x_1, x_2, \xi_1, \xi_2) | \xi_1) \\ \text{s.t.} \quad & B_2(\xi_1)x_1 + A_2(\xi_1)x_2 = b_2(\xi_1), \\ & l_2(\xi_1) \leq x_2 \leq u_2(\xi_1) \end{aligned} \tag{3.3}$$

where

- $q_2(x_1, x_2, \xi_1)$ is a cost of decision x_2 for the given realization of the first stage uncertainty ξ_1 and the given first stage decision x_1 ,

- $Q_2(x_1, x_2, \xi_1, \xi_2)$ is the cost of decision x_2 for given realizations of uncertainties ξ_1 and ξ_2 and the given first stage decision x_1 ,
- $B_2(\xi_1)$ is the *technology matrix* that converts a first stage decision into resources in the second stage, and
- $A_2(\xi_1)$ is the *recourse matrix*.

It is possible to remove the cost function Q_2 by including the second term of the objective in the cost function q_2 . The problem is said to have *fixed recourse* when A_2 is independent of ξ_1 . The subscripts indicate at which t a value is known except in the case of ξ_t . For instance, the realizations of B_2 , A_2 , and b_2 are all known at $t = 2$, which is the beginning of the second stage, but ξ_2 is not realized until after $t = 2$.

The full 2-stage recourse problem incorporates the second stage problem as follows: With the optimal value of the second stage problem (3.3) denoted by $Q_1(x_1, \xi_1)$, the 2-stage problem minimizes the sum of a first stage cost $q_1(x_1)$ and the expected value of the second stage cost $\text{E}Q_1(x_1, \xi_1)$:

$$\begin{aligned}
 \min_{x_1} \quad & q_1(x_1) + \text{E}Q_1(x_1, \xi_1) \\
 \text{s.t.} \quad & A_1 x_1 = b_1, \\
 & l_1 \leq x_1 \leq u_1.
 \end{aligned} \tag{3.4}$$

The first set of constraints in the above problem are referred to as the first stage constraints. A good introduction to the various properties of 2-stage recourse problems, such as feasibility, is found in [4].

An obvious criticism of the 2-stage model is that it only allows one recourse decision to be made, not a sequence of decisions over the time horizon. A multi-stage recourse program can provide a more realistic model, but it is more complex

and can often be very difficult to solve numerically. As in the 2-stage problem, the initial vector of decisions x_1 is made before the first realization of uncertainty ξ_1 , and a second stage decision x_2 is then made based on x_1 and ξ_1 . In the T -stage problem, this process continues for the uncertainties ξ_t , $t = 1, \dots, T - 1$, and the decisions vectors x_t , $t = 1, \dots, T$. There is usually one additional realization of uncertainty ξ_T following the final decision x_T .

The T -stage recourse program can be defined recursively as an extension of the 2-stage program. Let the uncertainty up to and including stage t , for $t = 1, \dots, T$, be denoted by $\xi^t = \{\xi_j, j = 1, \dots, t\}$, where each ξ_j is the uncertainty realized in stage j . Similarly, let the decisions up to and including stage t be denoted by $x^t = \{x_j, j = 1, \dots, t\}$, where each x_j is the decision made for stage j . The first stage problem is essentially the same as problem (3.4):

$$\begin{aligned} \min_{x_1} \quad & q_1(x_1) + E_{\xi_1} Q_1(x^1, \xi^1) \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & l_1 \leq x_1 \leq u_1, \end{aligned} \tag{3.5}$$

with Q_t , for $t = 1, \dots, T - 1$, given by the minimization problems

$$\begin{aligned} Q_t(x^t, \xi^t) = \min_{x_{t+1}} \quad & q_{t+1}(x^{t+1}, \xi^t) + E_{\xi_{t+1}} (Q_{t+1}(x^{t+1}, \xi^{t+1}) | \xi^t) \\ \text{s.t.} \quad & B_{t+1}(\xi^t) x_t + A_{t+1}(\xi^t) x_{t+1} = b_{t+1}(\xi^t), \\ & l_{t+1}(\xi^t) \leq x_{t+1} \leq u_{t+1}(\xi^t), \end{aligned} \tag{3.6}$$

and $Q_T(x^T, \xi^T)$ is a known function, not the solution to another minimization problem. It is possible to set $Q_T = 0$ by including the second term of the objective in q_T . The above problem (3.5-3.6) is a form of the multistage recourse problem that is relevant to the ALM problem that will be presented soon. Other forms, such as that found in [13], allow the first constraint of (3.6) to depend on all

decisions up to time t :

$$\sum_{\tau=1}^t B_{t+1,\tau}(\xi^t)x_\tau + A_{t+1}(\xi^t)x_{t+1} = b_{t+1}(\xi^t), \quad (3.7)$$

but this type of constraint is not necessary.

3.4 Scenario Generation

To numerically solve the recourse problem (3.5-3.6), the distribution of (ξ_1, \dots, ξ_T) is approximated by a set of scenarios usually organized in the form of a scenario tree. Figure (3.1) contains an example of a small scenario tree similar to the one that will be used in the 2-stage ALM problem discussed later. A first stage optimal allocation is found in the node at $t = 1$, and optimal recourse allocations are found in every node at $t = 2$. In the 2-stage problem, there is no additional allocation decision made at the nodes at $t = 3$. The tree shown in the figure is called *balanced* because each node at $t = 2$ is connected to two nodes at $t = 3$.

To describe the scenario tree, assume the nodes of the scenario tree are numbered starting with the value of one at $t = 1$, and let I_t be the number of nodes up to and including those at t . Define the sets of indices $\mathcal{I}_t = \{I_{t-1} + 1, \dots, I_t\}$, for $t = 2, \dots, T + 1$, with $I_1 = 1$. A scenario s , which is a path through the scenario tree, is then represented by the set of indices (i_2, \dots, i_{T+1}) where $i_t \in \mathcal{I}_t$. Two useful functions defined on the node indices are the predecessor, $pred(\cdot)$, and the descendant, $dec(\cdot)$: $pred(i_t)$ returns the node in \mathcal{I}_{t-1} connected to i_t , and $dec(i_t)$ returns a subset of nodes in \mathcal{I}_{t+1} connected to i_t . At t , the probability of being at node $i_t \in \mathcal{I}_t$ is denoted by $p(i_t)$ so that $\sum_{i_t \in \mathcal{I}_t} p(i_t) = 1$. Sometimes it is more useful to use the transition probabilities $p(i_t, i_{t+1})$, for $i_{t+1} \in dec(i_t)$ where $\sum_{i_{t+1} \in dec(i_t)} p(i_t, i_{t+1}) = 1$.

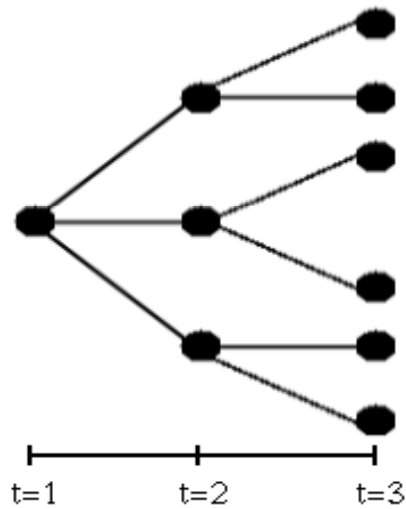


Figure 3.1: A Scenario Tree.

A topic of active research examines how to generate a good set of scenarios to represent the underlying distribution and produce good optimal decisions. The simplest approach is to just generate a very large number of scenarios by sampling from a time-series model. This is reasonable for a 1-stage problem, but recourse problems quickly become too difficult or time-consuming to solve as the number of scenarios is increased. Even with parallel implementations of solution algorithms, multistage problems must typically limit the number of scenarios. In this case, it becomes necessary to somehow generate a smaller set of “good” scenarios.

One technique in scenario selection is sequential importance sampling. The general idea behind importance sampling is to obtain scenarios that are important (in some sense) in the stochastic program. Sequential importance sampling obtains these scenarios in an iterative fashion. First, scenarios are generated for some given tree structure. The stochastic program is then solved and values for the importance sampling criterion are obtained at each node. These nodal values de-

termine where the structure of the scenario tree should be changed and/or where to resample a subtree. A more complete description of this method is in [12]. As an example, the importance sampling criterion used in [10] is the expected value of perfect information (EVPI). If the EVPI of a node is below some threshold, a new subtree emanating from that node is generated by resampling. If the EVPI is consistently below the threshold, the tree is collapsed beyond that node. If the EVPI is above the threshold for a node with no descendants, the tree is expanded beyond that node.

Discretization is an alternative to sampling from a distribution. One relatively simple technique for discretization is moment matching. For instance, to discretize the normal distribution it is possible to match the first two moments with three symmetric points. A moment matching model for a two-dimensional random vector is presented in [13] where the first and second random variables may represent the first and second stage uncertainty, respectively. To obtain the scenario values and probabilities, the first three marginal moments of both random variables are matched with the corresponding moments of the approximate distributions. In addition, the covariance between the true random variables is matched with that of the approximations. If the number of desired scenarios is large enough, and the moments are consistent, this procedure will provide a solution. However, if the moments are inconsistent, the author suggests a weighted least squares minimization problem.

As an alternative to moment matching, the discretization technique of [28] relies on the minimization of transportation metrics to approximate a continuous distribution with a discrete distribution. In this method, a desired scenario tree structure has already been determined. The goal is to minimize the difference between the optimal value of the stochastic program with the true distribution and

the optimal value of the stochastic program with the approximate distribution. This difference is termed the approximate error, and the author shows this error can be bounded through the Fortet-Mourier distance between the true and approximate probability distributions. The algorithm for the optimal discretization minimizes this bound. Through a simple 1-stage example, it is illustrated that this method performs better (in the sense of minimizing the approximation error) than moment matching.

Scenario reduction procedures can be used when a large number of scenarios are already given. An approach involving moment matching is found in [7]. A second approach involving probability metrics is found in [11] and [17]: Scenarios are recursively deleted with redistribution of the probability among the remaining scenarios by considering the Monge-Kantorovich functional.

There are many different methods to generate sample paths of the uncertain data, and not all of them initially consider a tree structure. Sample paths may come from an expert's expectation, historical observations, or any time-series model. The problem is then to convert a set of sample paths into a scenario tree. The method of clustering is described in [13]: One can group similar first stage values of the sample paths into clusters and then continue sequentially through each stage, or one can use a multi-level scheme in which the clusters consider the similarity of the entire sample path. A second method based on probability metrics which converts sample paths into a tree structure by combining scenario reduction with scenario bundling is found in [16].

3.5 Deterministic Equivalent Forms

The discrete and finite distribution of a scenario tree allows the stochastic recourse problem to be written as a deterministic program. Once a scenario tree

is constructed, each node i_t of the scenario tree determines values for $A_t(\xi^{t-1})$, $B_t(\xi^{t-1})$, $b_t(\xi^{t-1})$, $l_t(\xi^{t-1})$, $u_t(\xi^{t-1})$, and $q_t(\cdot, \xi^{t-1})$ which are denoted by A_{i_t} , B_{i_t} , b_{i_t} , l_{i_t} , u_{i_t} , and $q_{i_t}(\cdot)$. The recourse problem (3.5-3.6) can then be written as

$$\begin{aligned} \min_{x_1} \quad & q_1(x_1) + \sum_{i_2 \in \mathcal{I}_2} p(i_2) Q_{i_2}(x^1) \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & l_1 \leq x_1 \leq u_1, \end{aligned} \tag{3.8}$$

with Q_{i_t} , for $i_t \in \mathcal{I}_t$, $t = 2, \dots, T$, given by the minimization problems

$$\begin{aligned} Q_{i_t}(x^{t-1}) = \min_{x_t} \quad & q_{i_t}(x^t) + \sum_{i_{t+1} \in \text{dec}(i_t)} p(i_t, i_{t+1}) Q_{i_{t+1}}(x^t) \\ \text{s.t.} \quad & B_{i_t} x_{t-1} + A_{i_t} x_t = b_{i_t}, \\ & l_{i_t} \leq x_t \leq u_{i_t}, \end{aligned} \tag{3.9}$$

and $Q_{i_{T+1}}$ can be taken to be equal to zero.

The above (3.8-3.9) is the form of the recourse program that will be relevant when the solution method for the ALM problem is discussed later; however, there are other ways to proceed. Two other deterministic forms are now mentioned so that one can solve the ALM problem by possibly other solution algorithms. As will be shown shortly, the ALM problem will have a piecewise linear objective with linear constraints. By introducing auxiliary variables, the piecewise linear problem can be converted into a fully linear problem (with potentially a huge number of decision variables). In this case, the $q_{i_t}(\cdot)$ will take the linear form:

$$q_{i_t}(\cdot) = \langle q_{i_t}, \cdot \rangle, \tag{3.10}$$

where q_{i_t} is now a vector of appropriate dimension.

The deterministic equivalent for the linear program in arborescent form care-

fully considers the structure of the scenario tree:

$$\begin{aligned}
\min \quad & \langle q_1, x_1 \rangle + \sum_{i_2 \in \mathcal{I}_2} p(i_2) \langle q_{i_2}, x_{i_2} \rangle + \cdots + \sum_{i_T \in \mathcal{I}_T} p(i_T) \langle q_{i_T}, x_{i_T} \rangle \\
\text{subject to} \quad & \\
& A_1 x_1 = b_1, \\
& B_{i_2} x_1 + A_{i_2} x_{i_2} = b_{i_2}, \quad \forall i_2 \in \mathcal{I}_2, \\
& B_{i_3} x_{pred(i_3)} + A_{i_3} x_{i_3} = b_{i_3}, \quad \forall i_3 \in \mathcal{I}_3, \\
& \vdots \\
& B_{i_T} x_{pred(i_T)} + A_{i_T} x_{i_T} = b_{i_T}, \quad \forall i_T \in \mathcal{I}_T, \\
& l_{i_t} \leq x_{i_t} \leq u_{i_t}, \quad \forall i_t \in \mathcal{I}_t, \quad t = 1, \dots, T.
\end{aligned} \tag{3.11}$$

This arborescent form implicitly includes non-anticipatory constraints that the decision taken at t does not depend on the uncertainty that is realized in the future. Note that the decision vectors are x_{i_t} , $i_t \in \mathcal{I}_t$, $t = 1, \dots, T$, so there is one decision for each node of the scenario tree except for those at $T + 1$.

The split-variable formulation is an equivalent form that lends itself to decomposition and parallel implementation. If there are a total of S sample paths in the scenario tree, S independent subproblems are created by allowing all decisions to be scenario dependent. For the multistage case, the individual subproblem for scenario s with nodes (i_2, \dots, i_{T+1}) is

$$\min \quad \langle q_1, x_1^s \rangle + \langle q_{i_2}, x_2^s \rangle + \cdots + \langle q_{i_T}, x_T^s \rangle$$

$$\begin{aligned}
\text{s.t. } A_1 x_1^s &= b_1, \\
B_{i_2} x_1^s + A_{i_2} x_2^s &= b_{i_2}, \\
B_{i_3} x_2^s + A_{i_3} x_3^s &= b_{i_3}, \\
&\vdots \\
B_{i_T} x_{T-1}^s + A_{i_T} x_T^s &= b_{i_T},
\end{aligned} \tag{3.12}$$

plus any upper and lower bounds on x_t^s . When combining all subproblems into one problem, non-anticipatory constraints must be explicitly considered in this formulation: For any two scenarios s and s' with a common path up to and including t , $x_j^s = x_j^{s'}$, for $j = 1, \dots, t$, must be enforced. Essentially this amounts to a 0 – 1 matrix of coefficients. If p^s is the probability of scenario s , the overall split-variable representation for the multistage program is

$$\min \sum_{s=1}^S p^s (\langle q_1, x_1^s \rangle + \langle q_{i_2}, x_2^s \rangle + \dots + \langle q_{i_T}, x_T^s \rangle),$$

subject to a set of constraints (3.12) for each s , the non-anticipatory constraints, and any upper and lower bound constraints on x_t^s . As [26] states, this representation is advantageous for algorithms that temporarily ignore the non-anticipatory constraints.

Many multistage applications in finance can be posed as stochastic generalized networks. This means that each scenario subproblem of the split-variable formulation has a generalized network structure. Parallel implementation of highly efficient network algorithms can provide substantial computational advantages; however, some characteristics of a desired application, such as policy constraints, may destroy the network structure. Additionally, the arborescent form does not preserve any network structure present. Algorithms and computational studies of stochastic generalized networks is found in the work of Mulvey and Vladimirou

[22–25]. See also [8], especially for parallel implementation.

Additional resources including solutions techniques for 2-stage and multistage linear stochastic programs with recourse are in [13], [4], and [8].

3.6 The T-stage ALM problem

A specific ALM problem is now put into a form of a stochastic program with the goal of finding the allocations over a time horizon in a set of assets that optimizes a tradeoff between the risk and reward. The risk measure is a weighted average of the CVaR of the negative surplus wealth at each stage, and the reward is the expect final surplus wealth. Let the asset prices and liability price be denoted by s_t and l_t , respectively. There are n assets available at each time giving $s_t \in \mathbb{R}^n$, and there is just one liability stream giving $l_t \in \mathbb{R}$. For the T -stage problem, (s_t, l_t) are defined for $t = 1, \dots, T + 1$. The current prices known today are (s_1, l_1) , so these are not random variables; however, (s_t, l_t) is a bivariate random variable with realizations in \mathbb{R}^{n+1} known at time t for $t = 2, \dots, T + 1$. The CVaR of interest in stage t is just the CVaR of the distribution of the surplus wealth at $t + 1$. For instance, the stage 1 CVaR is determined by the distribution of surplus wealth at $t = 2$, which depends on the allocation decision at $t = 1$. For this reason, the CVaR of interest in stage t is denoted as $\text{CVaR}_\beta(-sw_{t+1})$ where sw_t is the surplus wealth at time t . The problem that will be solved can now be written as:

$$\min \quad \lambda \left(\sum_{t=1}^T \mu_t \text{CVaR}_\beta(-sw_{t+1}) \right) - (1 - \lambda)E(sw_{T+1}) \quad (3.13)$$

$$\text{s.t.} \quad \text{an initial wealth constraint,} \quad (3.14)$$

$$\text{balance of wealth constraints between time periods, and} \quad (3.15)$$

$$\text{linear transaction costs.} \quad (3.16)$$

Other constraints may include bounds on positions invested in each asset, bounds on the total transaction costs in each time period, and bounds permitting short-selling; however, these are not included in the lecture.

The above problem does not directly fit into the form (3.5-3.6), but the deterministic equivalent can be put into form (3.8-3.9) with the help of Uryasev's formula for CVaR. To begin, assume a scenario tree has already been constructed for (s_t, l_t) :

$$(s_t, l_t) = (s_{i_t}, l_{i_t}) \text{ with probability } p(i_t), \quad \forall i_t \in \mathcal{I}_t, \quad t = 1, \dots, T + 1. \quad (3.17)$$

The deterministic equivalent of the optimization problem will determine optimal asset allocations at each node of the scenario tree from $t = 1$ to $t = T$. These allocations are decision variables in the stochastic program and are denoted by a_{i_t} for $i_t \in \mathcal{I}_t$, $t = 1, \dots, T$. The distribution of sw_{t+1} depends not only on $(s_{i_{t+1}}, l_{i_{t+1}})$, $\forall i_{t+1} \in \mathcal{I}_{t+1}$, but also on the allocation decisions made at the nodes at time t . Note that this corresponds to the surplus wealth at time $t + 1$ **before** the portfolio reallocation occurs. The realization of the surplus wealth in node i_{t+1} is therefore a function of the allocation made in the node that immediately precedes i_{t+1} . With this allocation denoted by $a_{pred(i_{t+1})}$, the distribution of the surplus wealth for $t + 1 = 2, \dots, T + 1$, is

$$sw_{t+1} = \langle s_{i_{t+1}}, a_{pred(i_{t+1})} \rangle - l_{i_{t+1}} \text{ with probability } p(i_{t+1}), \quad \forall i_{t+1} \in \mathcal{I}_{t+1}.$$

For the given scenario tree, Uryasev's formula can now be applied to each CVaR:

$$\text{CVaR}_\beta(-sw_{t+1}) = \zeta_t + \frac{1}{1 - \beta} \sum_{i_{t+1} \in \mathcal{I}_{t+1}} p(i_{t+1}) [l_{i_{t+1}} - \langle s_{i_{t+1}}, a_{pred(i_{t+1})} \rangle - \zeta_t]^+, \quad (3.18)$$

where there is one auxiliary variable ζ_t introduced for each stage. To simplify things, let

$$h_{i_{t+1}}(\zeta_t, a_{pred(i_{t+1})}) = [l_{i_{t+1}} - \langle s_{i_{t+1}}, a_{pred(i_{t+1})} \rangle - \zeta_t]^+, \text{ and} \quad (3.19)$$

$$g_{i_{T+1}}(a_{pred(i_{T+1})}) = \langle s_{i_{T+1}}, a_{pred(i_{T+1})} \rangle - l_{i_{T+1}}. \quad (3.20)$$

The entire objective function is then

$$\begin{aligned} OBJ = & \lambda \sum_{t=1}^T \mu_t \zeta_t + \sum_{t=1}^T \left(\frac{\lambda \mu_t}{1 - \beta} \sum_{i_{t+1} \in \mathcal{I}_{t+1}} p(i_{t+1}) h_{i_{t+1}}(\zeta_t, a_{pred(i_{t+1})}) \right) \\ & - (1 - \lambda) \sum_{i_{T+1} \in \mathcal{I}_{T+1}} p(i_{T+1}) g_{i_{T+1}}(a_{pred(i_{T+1})}). \end{aligned} \quad (3.21)$$

Chapter 4

Modeling of the Risk Factors

An important task in ALM is the identification and adequate modeling of the underlying risk factors. The dynamic of financial risk factors is well known to often exhibit some of the following phenomena: heavy tails, skewness and high-kurtotic residuals. The recognition and description of the latter phenomena goes back to the seminal papers of Mandelbrot (1963) and Fama (1965). In this chapter we will introduce the α -stable distribution as an extension of the normal distribution. Due to its summation stability and the fact that it generalizes the Gaussian distribution, the class of α -stable distributions seems to be an ideal candidate to describe the return distribution of the considered risk factors. For further description of the stable distribution and applications of the stable distribution in financial theory see Samorodnitsky et al. (1994) or Rachev and Mitnik (1999).

4.1 Stable Distributions

4.1.1 Definition of Stable Random Variables

This section reviews some of the main features of the stable distribution as the natural extension of the Gaussian distribution. The notion of stable distribution was introduced in the 1920's by P. Lévy. A stable distribution can be defined in four equivalent ways, given in the following definitions: A random variable X follows a stable distribution, if for any positive numbers A and B there exists a positive number C and a real number D such that

$$AX_1 + BX_2 = CX + D \quad (4.1)$$

where X_1 and X_2 are independent copies of X and "=" denotes equality in distribution.

Therefore, a distribution f is stable if it is invariant under convolution, i.e., if there exist real constants $C > 0$ and D such that

$$f_{(AX_1+d_1)+(BX_2+d_2)}(s) := \int_{-\infty}^{+\infty} f(A(s-l) + d_1)f(Bl + d_2) = f(Cs + D) \quad (4.2)$$

for all real constants $A, B > 0, d_1, d_2$.

α is called the index of stability or characteristic exponent and for any stable random variable X , there is a number $\alpha \in (0, 2]$ such that the number C in 4.1.1 satisfies the following equation:

$$C^\alpha = A^\alpha + B^\alpha \quad (4.3)$$

A random variable X with index α is called α -variable. Obviously the Gaussian

distribution has an index of stability of 2.

The next definition is equivalent to 4.1.1 and considers the sum of n independent copies of a stable random variable. A random variable X has a stable distribution if for any $n \geq 2$, there is a positive real number C_n and a real number D_n such that

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} C_n X + D_n \quad (4.4)$$

where X_1, X_2, \dots, X_n are independent copies of X . Again, the number C_n and the stability index of the distribution are closely linked and we get $C_n = n^{1/\alpha}$ where the $\alpha \in (0, 2]$ is the same as in equation 4.3.

The third definition of a stable distribution is a generalisation of the central limit theorem. Stable distributions are in fact the only distributions that can be obtained as limits of normalized sums of iid random variables. A random variable X is said to be stable if it has a *domain of attraction*, i.e., if there is a sequence of random variables Y_1, Y_2, \dots and sequences of positive numbers $\{d_n\}$ and real numbers $\{c_n\}$, such that

$$\frac{Y_1 + Y_2 + \dots + Y_n}{d_n} \Rightarrow^d X. \quad (4.5)$$

The notation \Rightarrow^d denotes convergence in distribution. Definition 4.1.1 is obviously equivalent to definition 4.1.1, as one can take the Y_i s to be independent and distributed like X . As mentioned before, in the case of $\alpha = 2$, the statement is the ordinary central limit theorem. When $d_n = n^{1/\alpha}$, the Y_i s are said to belong to the *normal* domain of attraction of X .

Finally, the last equivalent way to define a stable random variable provides information about its characteristic function. A random variable X has a stable

distribution if there are parameters $0 < \alpha \leq 2, \sigma \geq 0, -1 \leq \beta \leq 1$, and μ real such that its characteristic function has the following form:

$$E(e^{iXt}) = \begin{cases} \exp(-\sigma^\alpha |t|^\alpha [1 - i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2}] + i\mu t), & \text{if } \alpha \neq 1, \\ \exp(-\sigma |t| [1 + i\beta \frac{2}{\pi} \text{sign}(t) \ln |t|] + i\mu t), & \text{if } \alpha = 1, \end{cases} \quad (4.6)$$

Definition 4.1.1 implies definition 4.1.1 what can be shown the following way:

For $\alpha \neq 1$ and X_1, X_2, \dots, X_n independent copies of the stable random variable X . Thus, we can write

$$Ee^{it(X_1+X_2+\dots+X_n)} = \exp\left(-n\sigma^\alpha |t|^\alpha \left(1 - i\beta(\arg t) \tan \frac{\pi\alpha}{2}\right) + in\mu t\right).$$

On the other hand, obviously

$$\begin{aligned} Ee^{it(c_n X + d_n)} &= e^{itd_n} Ee^{i(tc_n)X} = \\ &= e^{itd_n} \exp\left(-\sigma^\alpha |c_n t|^\alpha \left(1 - i\beta(\arg c_n t) \tan \frac{\pi\alpha}{2}\right) + i\mu c_n t\right). \end{aligned}$$

By choosing $c_n = n^{1/\alpha}$ and $d_n = \mu(n - n^{1/\alpha})$ we get the equation

$$Ee^{it(c_n X + d_n)} = Ee^{it(X_1+X_2+\dots+X_n)}$$

Since the characteristic function is uniquely defined for a random variable X we end up with the result:

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} c_n X + d_n.$$

4.1.2 Parameters and Special Cases of the Stable Distribution

A stable distribution is defined by four parameters. The dependence of a stable random variable X from its parameters we will indicate by writing:

$$X \sim S_{\alpha}(\beta, \sigma, \mu)$$

where α is the the so-called index of stability ($0 < \alpha \leq 2$). The lower the value of α the more leptocurtic is the distribution. This can be considered as a very attractive property for modeling financial asset returns. In empirical studies, the value of α for asset returns is often chosen between 1 and 2. For $\alpha > 1$, the location parameter μ is the mean of the distribution. Figure 4.1 shows the probability density function for symmetric alpha-stable random variables for different values of α .

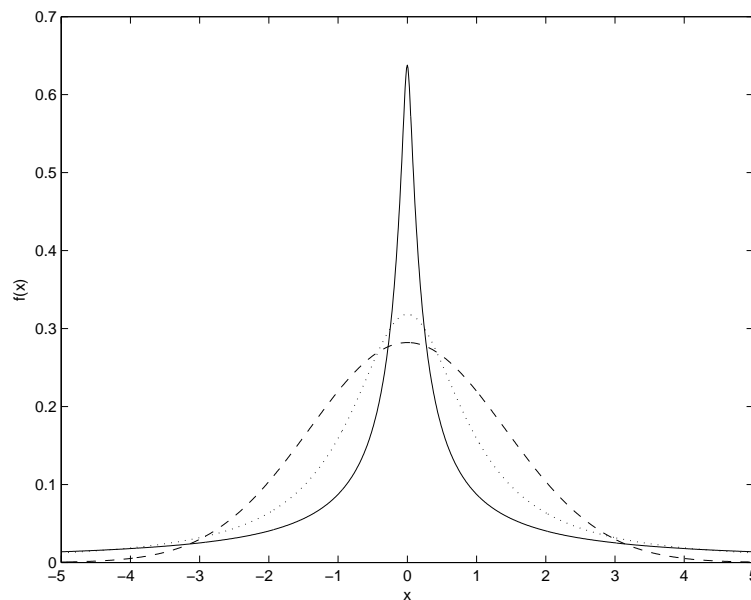


Figure 4.1: Probability density functions for standard symmetric α -stable random variables, $\alpha = 2$, $\alpha = 1$ (dotted) and $\alpha = 0.5$ (dashed).

The second parameter β is the skewness parameter ($-1 \leq \beta \leq 1$). A stable distribution with $\beta = \mu = 0$ is called a symmetric α -stable distribution ($S\alpha S$). If $\beta < 0$, the distribution is skewed to the left, if $\beta > 0$, the distribution is skewed to the right. We conclude that the stable distribution can also capture asymmetric asset returns.

σ is the scale parameter ($\sigma \geq 0$) and μ is the drift ($\mu \in \mathbf{R}$).

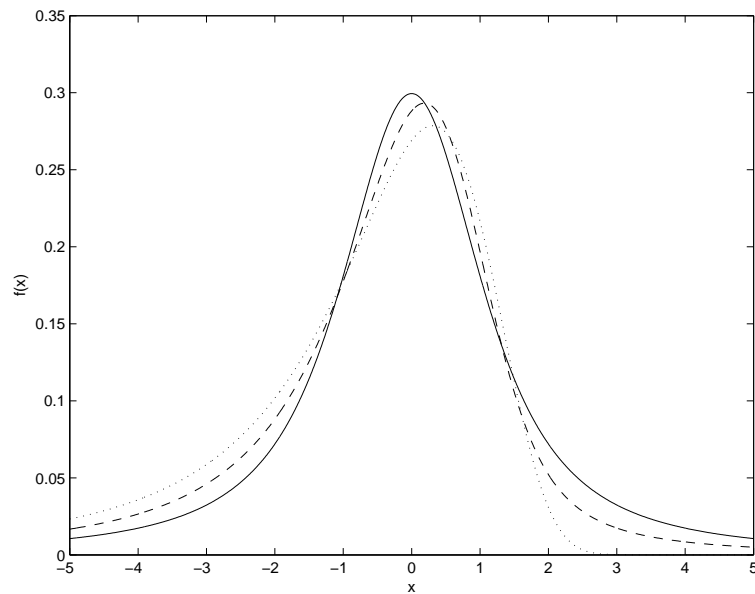


Figure 4.2: Probability density functions for stable random variables with $\alpha = 1.2$, β varying, $\beta = 0$, $\beta = -0.5$ (dashed) and $\beta = -1$ (dotted).

Figure 4.2 shows the probability density function for some skewed alpha-stable random variables with $\alpha = 1.2$.

Generally the probability density function of a stable distribution cannot be specified in explicit form. However, there are three special cases where this is possible:

- i. The *Gaussian distribution*

If the index of stability $\alpha = 2$, then the stable distribution reduces to the

Normal distribution, and it is $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$. We shall point out that the σ in definition 4.1.1 is not equal to the standard deviation. When $\alpha = 2$, the characteristic function becomes $Ee^{itX} = e^{-\sigma^2 t^2 + i\mu t}$. This is the characteristic function of a Gaussian random variable with mean μ and variance $2\sigma^2$.

ii. The *Cauchy distribution*

$S_1(\sigma, 0, \mu)$, whose density $f_1(x)$ is

$$f_1(x) = \frac{\sigma}{\pi((x - \mu)^2 + \sigma^2)} \quad (4.7)$$

If $X \sim S_1(\sigma, 0, 0)$, then for $x > 0$,

$$P(X \leq x) = 0.5 + \frac{1}{\pi} \arctan\left(\frac{x}{\sigma}\right). \quad (4.8)$$

iii. The *Lévy distribution*

$S_{1/2}(\sigma, 1, \mu)$, whose density

$$\left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x - \mu)^{3/2}} \exp\left\{-\frac{\sigma}{2(x - \mu)}\right\} \quad (4.9)$$

is concentrated on (μ, ∞)

4.1.3 Properties of Stable Random Variables

In this section we will summarize some useful properties useful of stable distributions in modeling financial data or simulation.

The first property mentioned is the so-called summation stability. Let X_1, X_2 be independent random variables with $X_i \sim (\sigma_i, \beta_i, \mu_i)$, $i = 1, 2$. Then $X_1 + X_2 \sim$

$S_\alpha(\sigma, \beta, \mu)$, with

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2. \quad (4.10)$$

for the proof we refer to Samorodnitsky et al. (1994). Thus, the sum of two alpha-stable distributed random variables with the same index α is also alpha-stable with the same index of stability α .

The second proposition concerns the parameter σ . The Gaussian distribution can be scaled by multiplication with a constant. This property extends to $0 < \alpha \leq 2$.

Let $X \sim S_\alpha(\sigma, \beta, \mu)$ and let $a \in \mathbb{R} \setminus \{0\}$. Then

$$\begin{aligned} aX &\sim S_\alpha(|a|\sigma, \arg(a)\beta, a\mu) && \text{if } \alpha \neq 1 \\ aX &\sim S_\alpha(|a|\sigma, \arg(a)\beta, a\mu - \frac{2}{\pi}a(\ln|a|\sigma\beta)) && \text{if } \alpha = 1 \end{aligned}$$

The parameter σ is therefore often called the *scale parameter*. The proof of 4.1.3 can easily be done by using the characteristic function of stable distributions

$$\begin{aligned} \ln E e^{it(aX)} &= -\sigma^\alpha |ta|^\alpha \sigma^\alpha \left(1 - i\beta \arg(ta) \tan \frac{\pi\alpha}{2}\right) + i\mu(ta) \\ &= -(\sigma|a|)^\alpha |t|^\alpha \sigma^\alpha \left(1 - i\beta \arg(a) \arg(t) \tan \frac{\pi\alpha}{2}\right) + i(\mu a)t \end{aligned}$$

.

The third proposition concerns the shift parameter μ . It was already discussed that in the case of $\alpha = 2$ the parameter μ is a shift parameter for the Gaussian distribution. The same can be inferred about μ for any admissible α . Let

$X \sim S_\alpha(\sigma, \beta, \mu)$ and let a be real constant. Then $X + a \sim S_\alpha(\sigma, \beta, \mu + a)$. This follows directly by interpreting a as a $S_\alpha(0, 0, a)$ stable random variable and applying the summation stability proposition. For $1 < \alpha \leq 2$, the shift parameter μ equals the mean.

Finally, we can also interpret the last parameter β . It can be identified as a *skewness parameter*. $X \sim S_\alpha(\sigma, \beta, \mu)$ is symmetric if and only if $\beta = 0$ and $\mu = 0$. It is symmetric about μ if and only if $\beta = 0$. We can prove this by the fact that a random variable is symmetric if and only if its characteristic function is real. By definition 4.1.1 this is the case if and only if $\beta = 0$ and $\mu = 0$. The second statement follows from property 4.1.3. In order to indicate that X is symmetric, i.e. $\beta = 0$ and $\mu = 0$, we write

$$X \sim S_\alpha S$$

4.1.4 Truncated α -Stable Distributions

Despite these advantages the stable distribution so far is only rarely used in practical implementations. A major reason for the limited use of stable distributions in applied work is that there are in general no closed-form expressions for its probability density function. Numerical approximations are nontrivial and computationally demanding. Another shortcoming in application issues is that all moments of order $\geq \alpha$ are infinite. Therefore, for some applications e.g. GARCH models with conditions on the innovations like $E(\epsilon_t) = 0$ and $V(\epsilon_t) = 1$, $t \in \mathbb{N}$ at first the stable distribution is not applicable. In the sequel, following Menn and Rachev (2004) we will give a brief introduction to a new class of probability distributions that combines the modeling flexibility of stable distributions with the existence of arbitrary moments.

A possibility to guarantee the existence of moments of order $\geq \alpha$ is to truncate

the stable distribution at certain limits and add two normally distributed tails to the distribution. Dependent on where the truncation is conducted the distribution can still be clearly more heavy-tailed than a normal distribution but may provide finite variance. This idea leads to the definition of a so-called smoothly truncated stable distribution.

Let g_θ denote the density of some α -stable distribution with parameter-vector $\Theta = (\alpha, \beta, \sigma, \mu)$ and h_i , ($i = 1, 2$) denote the densities of two normal distributions with parameters (ν_i, τ_i) , ($i = 1, 2$). Furthermore, let $a, b \in \mathbb{R}$ be two real numbers with $a \leq \mu \leq b$. The density of a smoothly truncated stable distribution (STS-distribution) is defined by:

$$f(x) = \begin{cases} h_1(x) & \text{for } x < a \\ g_\theta(x) & \text{for } a \leq x \leq b \\ h_2(x) & \text{for } x > b \end{cases}$$

In order to guarantee a well-defined continuous probability density, the following relations are imposed:

(i) Continuity:

$$h_1(a) \stackrel{!}{=} g_\theta(a) \text{ and } h_2(b) \stackrel{!}{=} g_\theta(b)$$

(ii) $P(\mathbb{R}) = 1$ and therefore

$$\begin{aligned} p_1 &:= \int_{-\infty}^a h_1(x) dx \stackrel{!}{=} \int_{-\infty}^a g_\theta(x) dx \\ p_2 &:= \int_b^{\infty} h_2(x) dx \stackrel{!}{=} \int_b^{\infty} g_\theta(x) dx \end{aligned}$$

The class of smoothly truncated stable (STS) distributions in the following will be denoted by \mathcal{S}^{trunc} , elements of \mathcal{S}^{trunc} by \cdot . Since probability distributions used for modeling white noise processes like the innovations of a time series model, are usually assumed to be standardized probability distributions with zero mean and unit variance. It remains the problem of calculation of the parameters (ν_i, τ_i) , ($i = 1, 2$) for the two normal distributions. The conditions lead to the following equations for the parameters (ν_i, τ_i) , ($i = 1, 2$):

$$\tau_1 = \frac{\varphi(\Phi^{-1}(p_1))}{g_\theta(a)} \quad \text{and} \quad \nu_1 = a - \tau_1 \Phi^{-1}(p_1) \quad (4.11)$$

$$\tau_2 = \frac{\varphi(\Phi^{-1}(p_2))}{g_\theta(b)} \quad \text{and} \quad \nu_2 = b + \tau_2 \Phi^{-1}(p_2) \quad (4.12)$$

where φ and Φ denote the density and distribution function of the standard normal distribution.

Following Menn and Rachev (2004) a useful property of α -stable distributions is the scale and translation invariance, which is transmitted to the class of STS-distributions:

$$Y := cX + d \sim S_{\tilde{\alpha}}^{[\tilde{a}, \tilde{b}]}(\tilde{\sigma}, \tilde{\beta}, \tilde{\mu}) \in \mathcal{S}^{trunc} \quad (4.13)$$

with

$$\tilde{a} = ca + d, \tilde{b} = cb + d, \tilde{\alpha} = \alpha, \tilde{\sigma} = |c|\sigma, \tilde{\beta} = \text{sign}(c)\beta, \tilde{\mu} = c\mu + d$$

The main advantage is however that the mean EX and the second moment EX^2 of a STS-distributed random variable X exists:

$$\begin{aligned}
EX &= ap_1 - \tau_1 (\Phi^{-1}(p_1)p_1 + \varphi(\Phi^{-1}(p_1))) + \\
&\quad + \int_a^b x g_\theta(x) dx + \\
&\quad + bp_2 + \tau_2 (\Phi^{-1}(p_2)p_2 + \varphi(\Phi^{-1}(p_2)))
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
EX^2 &= (\tau_1^2 + \nu_1^2)p_1 - \tau_1(a + \nu_1)\varphi(\Phi^{-1}(p_1)) + \\
&\quad + \int_a^b x^2 g_\theta(x) dx + \\
&\quad + p_2(\nu_2^2 + \tau_2^2) + \tau_2(\nu_2 + b) \cdot \varphi(\Phi^{-1}(p_2))
\end{aligned} \tag{4.15}$$

where, as above, φ denotes the density and Φ the distribution function of the standard normal distribution. p_1 and p_2 denote the cut-off-probabilities given in equation (4.1.4) and G_θ is the distribution function of the α -stable distribution with parameter-vector $\theta = (\alpha, \beta, \sigma, \mu)$.

It should be pointed out that since there exists no closed form expression for the density g_θ of a stable distribution, the mean and the variance of an STS-distribution can only be calculated with the help of numerical integration.

4.2 Stable Modeling of Risk Factors

4.2.1 Modeling Financial Returns with Stable Distributions

In this section we will give some examples on the superior fit of stable distributions to financial returns compared to the Gaussian distribution that is used in

Distribution Parameters	Stable				Gaussian	
	alpha	beta	sigma	mu	μ	σ
Unemployment Rate	1.6691	-1.0000	0.0124	0.0316	0.0337	0.0207
Working Output	1.4474	1.0000	0.0723	0.0234	-0.0074	0.1454
Gross Domestic Product	1.6325	-1.0000	0.0830	-0.0493	-0.0495	0.1986
Consumer Price Index	1.2061	0.0880	0.3130	0.0385	0.0194	0.8058
Annual Saving	1.2849	1.0000	0.0123	0.0563	0.0433	0.0283
Personal Income	2.0000	0.1427	0.0147	0.0744	0.0744	0.0210

Table 4.1: Parameters of α -stable and Gaussian fit to log-returns of several US macroeconomic time series 1960-2000.

most standard models of financial theory. Various applications of stable models in finance can be found in Rachev and Mittnik (1999). The advantages of the alpha-stable distribution for modeling financial data are manifold. Due to the summation stability the sum of stable distributed random variables with identical parameter α are again alpha-stable distributed with α .

Another advantage is the number of parameters: with four parameters the distribution provides more flexibility and is capable to explain issues of financial data like skewness, excess kurtosis or heavy tails.

Figure 4.3 and 4.4 illustrate that in most cases stable distributions provide a clearly better fit for financial variables since they can capture the kurtosis and the heavy-tailed nature of financial data. We considered the fit of the stable distribution to the yearly log-return time series of the macroeconomic variables working output per hour and GDP. Table 4.2.1 provides the goodness-of-fit measure Kolmogorov distance (KS) that measures the distance between the empirical cumulative distribution function $F_n(x)$ and the fitted CDF $F(x)$

$$KS = \max_{x \in \mathbb{R}} |F_n(x) - F(x)|. \quad (4.16)$$

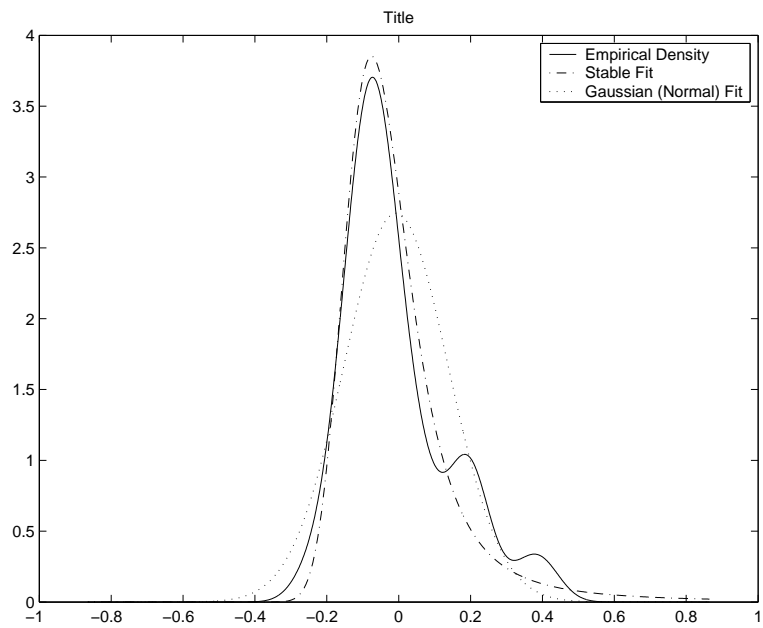


Figure 4.3: Normal and Stable fit to log return of Working Output per hour.

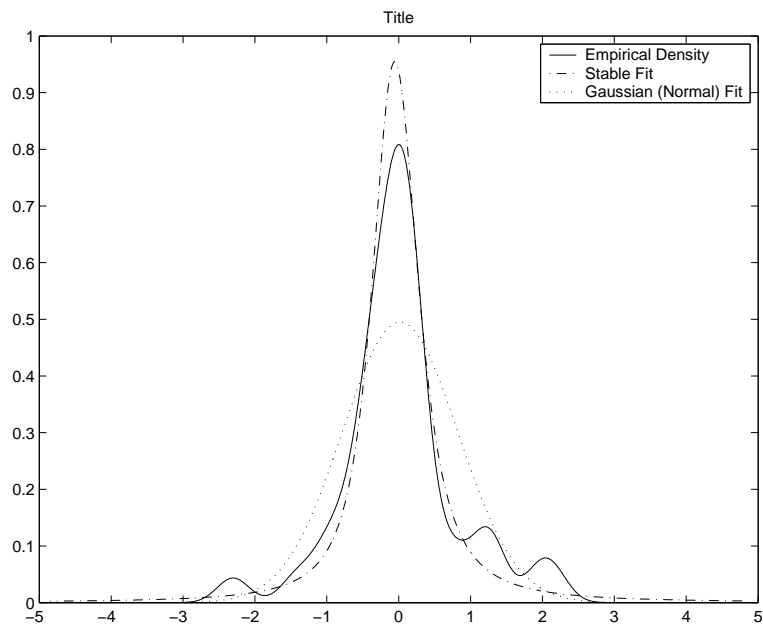


Figure 4.4: Normal and Stable fit to log return of GDP.

A short-coming of the Kolmogorov distance is that it is most sensitive around the median value and less sensitive in the tails of the distribution, where $F(x)$ is

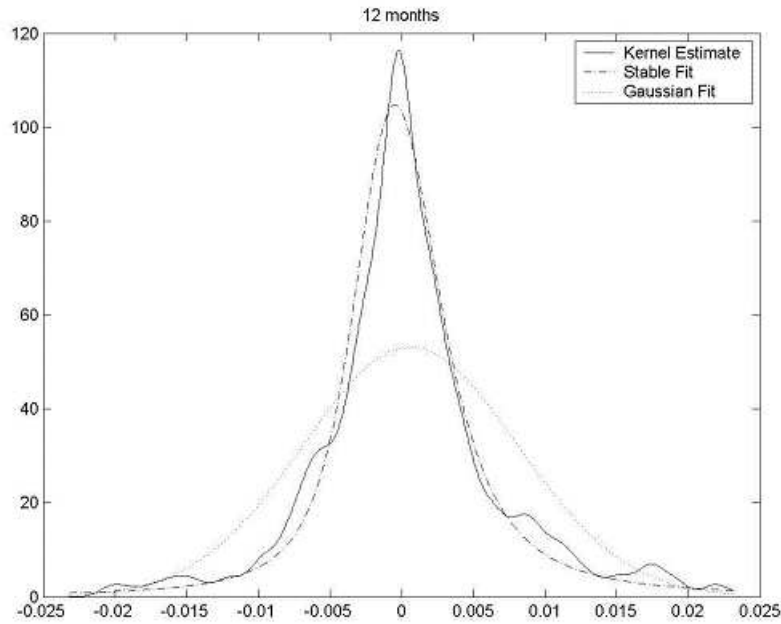


Figure 4.5: Fit of Gaussian and Stable distribution to 1 year Euribor rate

near 0 or 1. Therefore, we also considered the Anderson-Darling statistic (AD)

$$AD = \max_{x \in \mathbb{R}} \frac{|F_n(x) - F(x)|}{\sqrt{F(x)(1 - F(x))}} \quad (4.17)$$

that puts more weight to the tails of the distribution. Table 4.2.1 shows the results for the considered goodness-of-fit criteria. For most variables we find a clearly better fit of the stable distribution compared to the normal.

4.3 Univariate and Multivariate Distributions

For ALM problems, often scenarios are generated by calibrating and simulating a time-series model to multivariate data.

There are two major approaches modeling multivariate data:

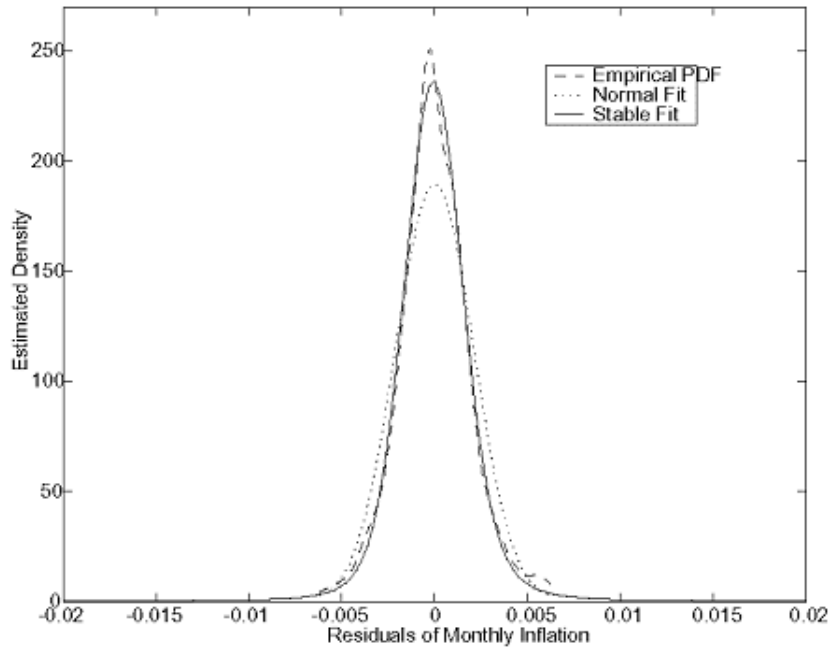


Figure 4.6: Fit of Gaussian and Stable distribution to residuals of monthly inflation

Distribution Parameters	Stable		Gaussian	
	KS	AD	KS	AD
Unemployment Rate	0.0843	0.5065	0.1333	0.4077
Working Output	0.0965	0.2869	0.1791	0.3389
Gross Domestic Product	0.0804	0.5448	0.1804	6.1008
Consumer Price Index	0.0723	0.2646	0.1833	0.5044
Annual Saving	0.0777	0.1738	0.1635	0.3521
Personal Income	0.1073	0.1864	0.0783	0.1920

Table 4.2: Goodness-of-Fit criteria Kolmogorov distance (KS) and Anderson-Darling statistic (AD) for Stable and Normal Distribution.

- Fit a multivariate distribution.
- Fit each individual time-series with a univariate distribution and use a copula to describe the dependence structure.

The second approach is more flexible in the sense that it allows any type of distribution to be fit to the individual series. For instance, one can first calibrate complex univariate models like GARCH etc. and then capture the dependence with a time-varying copula.

4.4 Fitting a Multivariate Distribution

In terms of the multivariate approach one might calibrate a vector autoregressive (VAR) model to the data. The VAR model has had much success in modeling financial and economic data. The general VAR(p) model for financial return data \tilde{R}_τ is

$$\tilde{R}_\tau = \Pi_1 \tilde{R}_{\tau-1} + \dots + \Pi_p \tilde{R}_{\tau-p} + E_\tau, \quad (4.18)$$

where the innovations process $E_\tau = (e_\tau^1, \dots, e_\tau^6)'$ is assumed to be white noise with covariance matrix Σ . It is both easy to calibrate and easy to simulate scenarios from VAR models. An introduction to modeling and estimation of VAR models can be found in [38].

To simulate the VAR model, one needs to make a distributional assumption for the innovations. After estimation of the VAR(1) model, the residuals are computed by

$$\hat{E}_\tau = \tilde{R}_\tau - \hat{\Pi}_1 \tilde{R}_{\tau-1}, \quad (4.19)$$

and the standardized residuals $\hat{\Sigma}^{-1/2} \hat{E}_\tau$ are plotted in figure (5.2). The usual assumption is that the innovations are Gaussian, in which case the standardized

residuals should be i.i.d. multivariate Normal($0, I_n$). However, based on the results on financial return data of the previous sections, it might also be promising to use a more flexible or heavy-tailed distribution like the α -stable or the truncated stable distribution.

A n -dimensional random vector Z has a multivariate stable distribution if for any $a > 0$ and $b > 0$ there exists $c > 0$ and $d \in \mathbb{R}^n$ such that

$$aZ_1 + bZ_2 \stackrel{d}{=} cZ + d,$$

where Z_1 and Z_2 are independent copies of Z and $a^\alpha + b^\alpha = c^\alpha$. The characteristic function of R is given by

$$\Phi_Z(\theta) = \begin{cases} \exp \left\{ - \int_{S_n} |\theta' s| \left(1 - i \operatorname{sign}(\theta' s) \tan \frac{\pi \alpha}{2} \right) \Gamma_Z(ds) + i \theta' \mu \right\}, & \text{if } \alpha \neq 1, \\ \exp \left\{ - \int_{S_n} |\theta' s| \left(1 + i \frac{2}{\pi} \operatorname{sign}(\theta' s) \ln |\theta' s| \right) \Gamma_Z(ds) + i \theta' \mu \right\}, & \text{if } \alpha = 1, \end{cases}$$

where θ and μ are n -dimensional vectors. The spectral measure Γ_Z is a finite measure on the sphere in \mathbb{R}^n that replaces the roles of β and σ in stable random variables. Again, α and μ are the index of stability and location parameter, respectively. A symmetric stable random vector with $\mu = 0$ is called symmetric alpha-stable ($S\alpha S$), and in this case, the stable equivalent of covariance is the covariation:

$$[\tilde{z}^1, \tilde{z}^2]_\alpha = \int_{S_2} s_1 s_2^{(\alpha-1)} \Gamma_{(\tilde{z}^1, \tilde{z}^2)}(ds),$$

where $(\tilde{z}^1, \tilde{z}^2)$ is a $S\alpha S$ vector with spectral measure $\Gamma_{(\tilde{z}^1, \tilde{z}^2)}$ and $y^{(k)} = |y|^k \operatorname{sign}(x)$. Additionally, the covariation norm is given by

$$\|\tilde{z}^i\|_\alpha = ([\tilde{z}^i, \tilde{z}^i]_\alpha)^{1/\alpha}.$$

See [30] for details on estimating the index of stability, spectral measure, and scale parameter for a general stable random vector.

4.5 Dependence Modeling and Copulas

In the elliptical world the variance-covariance approach to optimizing portfolios makes sense and VAR is a coherent measure of risk here. For this reason, the class of elliptical distributions represent an ideal environment for standard (market) risk managing approaches. However, for general multivariate distributions, correlation often gives no indication about the degree or structure of dependence here. A list of deficiencies and problems in the general case shall illustrate this point (see Embrechts et al, 1999):

- i. Correlation is simply a scalar measure of dependency; it cannot tell us everything we would like to know about the dependence structure of risks.
- ii. Possible values of correlation depend on the marginal distribution of the risks. All values between -1 and 1 are not necessarily attainable.
- iii. Perfectly positively dependent risks do not necessarily have a correlation of 1; perfectly negatively dependent risks do not necessarily have a correlation of -1.
- iv. A correlation of zero does not indicate independence of risks.
- v. Correlation is not invariant under transformations of the risks. For example, $\log(X)$ and $\log(Y)$ generally do not have the same correlation as X and Y .
- vi. Correlation is only defined when the variances of the risks are finite. It is not an appropriate dependence measure for very heavy-tailed risks where variances appear infinite.

For an illustration of point 2 and 4, consider the following example (see Embrechts et al, 1999):

Consider two rv's. X and Y that are lognormally distributed with $\mu_X = \mu_Y = 0$, $\sigma_X = 1$ and $\sigma_Y = 2$. One can show that by an arbitrary specification of the joint distribution with the given marginals, it is not possible to attain any correlation in $[-1, 1]$. In fact, there exist boundaries for a maximal and a minimal attainable correlation $[\rho_{min}, \rho_{max}]$ which in the given case is $[-0.090, 0.666]$.

Allowing σ_Y to increase, this interval becomes arbitrarily small as one can see in figure 4.7. Here, it is interesting to note that the two boundaries represent the case where the two rv's are perfectly positive dependent (the max. correlation line) or perfectly negative dependent (the min. correlation line) respectively.

Thus although the attainable interval for ρ as $\sigma_Y > 1$ converges to zero from both sides, the dependence between X and Y is by no means weak. This indicates that it is wrong to interpret small correlation as weak dependence.

A single statistical parameter like the linear correlation coefficient will not be able to capture the entire dependence structure between two rv's in the general case. At this point a general concept of describing the dependence structure within multivariate distributions is needed. Since marginal distributions are very illustrative, easy to handle and often used as basic building blocks for the design of a multivariate distribution, the idea of separating the description of the joint multivariate distribution into the marginal behaviour and the dependence structure is very attractive. One representation of the dependence structure that satisfies this concept is a copula. A copula is a function that combines the marginal distributions to form the joint multivariate distribution. A copula is the distribution function of a random vector in \mathbb{R}^n with standard uniform marginals. One can alternatively define a copula as a function and impose certain restrictions.

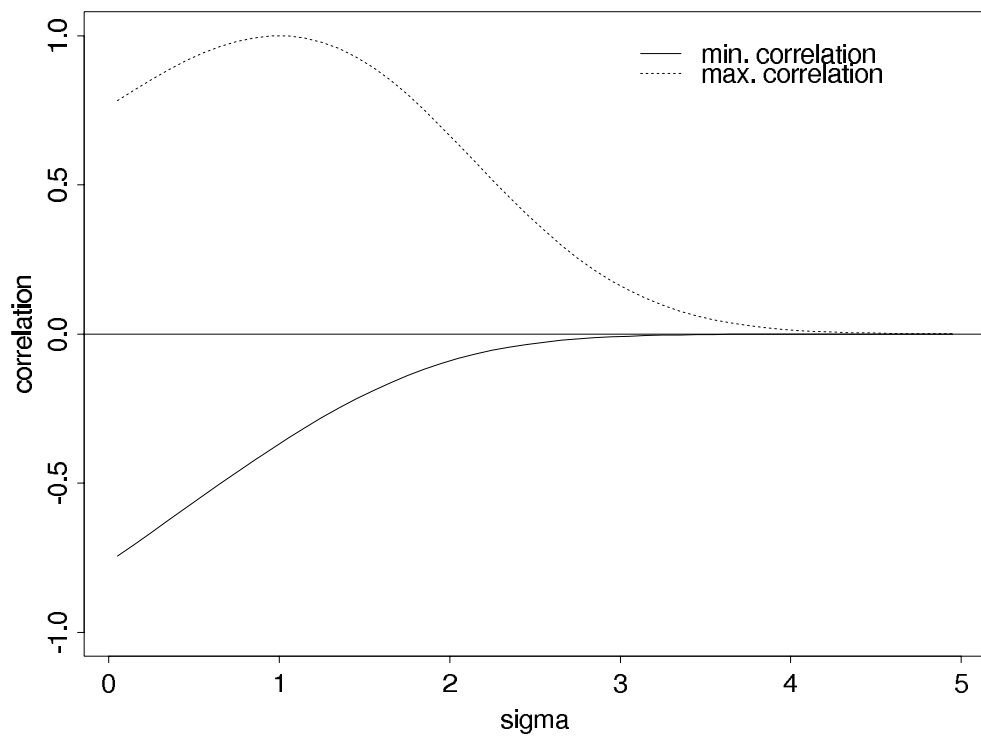


Figure 4.7: Maximum and minimum attainable correlation for $X \sim \text{Lognormal}(0, 1)$ and $Y \sim \text{Lognormal}(0, \text{sigma})$.

A copula is any real valued function $C : [0, 1]^n \rightarrow [0, 1]$, i.e. a mapping of the unit hypercube into the unit interval, which has the following three properties:

- i. $C(u_1, \dots, u_n)$ is increasing in each component of u_i .
- ii. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $i \in \{1, \dots, n\}$, $u_i \in [0, 1]$.
- iii. For all $(a_1, \dots, a_n), (b_1, \dots, b_n) \in [0, 1]^n$ with $a_i \leq b_i$:

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1+\dots+i_n} C(u_{1i_1}, \dots, u_{ni_n}) \geq 0$$

where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in \{1, \dots, n\}$.

Let $X = (X_1, \dots, X_n)'$ be a random vector of real-valued rv's whose dependence structure is completely described by the joint distribution function

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n). \quad (4.20)$$

Each rv X_i has a marginal distribution of F_i that is assumed to be continuous for simplicity. Recall that the transformation of a continuous rv X with its own distribution function F results in a rv $F(X)$ which is standardly uniformly distributed. Thus transforming equation (4.20) component-wise yields

$$\begin{aligned} F(x_1, \dots, x_n) &= P(X_1 < x_1, \dots, X_n < x_n) \\ &= P[F_1(X_1) < F_1(x_1), \dots, F_n(X_n) < F_n(x_n)] \\ &= C(F_1(x_1), \dots, F_n(x_n)), \end{aligned} \quad (4.21)$$

where the function C can be identified as a joint distribution function with standard uniform marginals — the copula of the random vector X . In equation (4.21),

it can be clearly seen, how the copula combines the marginals to the joint distribution.

Sklar's theorem provides a theoretic foundation for the copula concept:¹ [Sklar's theorem] Let F be a joint distribution function with continuous margins F_1, \dots, F_n . Then there exists a unique copula $C : [0, 1]^n \rightarrow [0, 1]$ such that for all x_1, \dots, x_n in $\mathbb{R} = [-\infty, \infty]$ (4.21) holds. Conversely, if C is a copula and F_1, \dots, F_n are distribution functions, then the function F given by (4.21) is a joint distribution function with margins F_1, \dots, F_n . For the case that the marginals F_i are not all continuous, it can be shown² that the joint distribution function can also be expressed like in equation (4.21), although C is no longer unique in this case.

Examples of copulas

- i. If the rv's X_i are independent, then the copula is just the product over the F_i

$$C^{ind}(x_1, \dots, x_n) = x_1 \cdot \dots \cdot x_n.$$

- ii. The *Gaussian* copula is

$$C_\rho^{Ga}(x, y) = \int_{-\infty}^{\Phi^{-1}(x)} \int_{-\infty}^{\Phi^{-1}(y)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left(\frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right) ds dt,$$

where $\rho \in (-1, 1)$ and $\Phi^{-1}(\alpha) = \inf\{x \mid \Phi(x) \geq \alpha\}$ is the univariate inverse standard normal distribution function. Applying C_ρ^{Ga} to two univariate standard normally distributed rv's results in a standard bivariate normal distribution with correlation coefficient ρ .

¹For further discussion see [35].

²See [35].

Note that, since the copula and the marginals can be arbitrarily combined, this (and any other) copula can be applied to *any* set of univariate rv's. The outcome will then surely not be multivariate normal, but the resulting multivariate distribution has inherited the dependence structure from the multivariate normal distribution.

iii. As a last example, the *Gumbel* or *logistic* copula

$$C_{\beta}^{Gu}(x, y) = \exp \left[- \left\{ (-\log x)^{\frac{1}{\beta}} + (-\log y)^{\frac{1}{\beta}} \right\}^{\beta} \right],$$

where $\beta \in (0, 1]$ indicates the dependence between X and Y . $\beta = 1$ gives independence and $\beta \rightarrow 0^+$ leads to perfect dependence.

According to theorem 4.5, a multivariate distribution is fully determined by its marginal distributions and a copula. Therefore, the copula contains all information about the dependence structure between the associated random variables. In the case that all marginal distributions are continuous, the copula is unique and therefore often referred to as *the dependence structure* for the given combination of multivariate and marginal distribution. If the copula is not unique because at least one of the marginal distributions is not continuous, it can still be called a *possible representation of the dependence structure*.

A very useful feature of a copula is the fact that it is invariant under increasing and continuous transformation of the marginals. If $(X_1, \dots, X_n)^t$ has copula C and T_1, \dots, T_n are increasing continuous functions, then $(T_1(X_1), \dots, T_n(X_n))^t$ also has copula C . The proof can be found for example in [15], page 6.

One application of lemma 4.5 would be that the transition from the representation of a random variable to its logarithmic representation does not change the copula. Note that the linear correlation coefficient does not have this property,

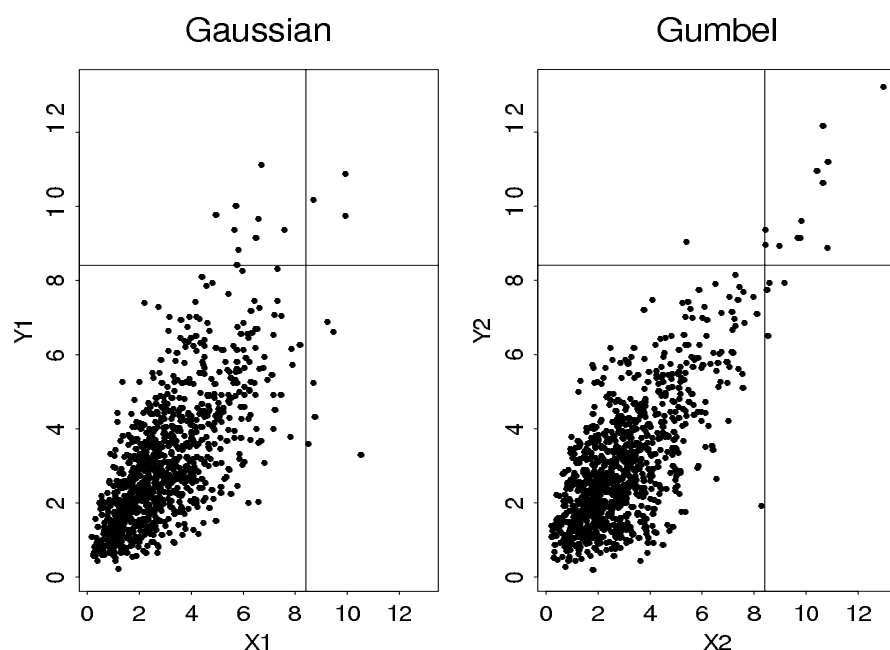


Figure 4.8: 1000 draws from two distributions that were constructed using Gamma(3,1) marginals and two different copulas, both having a linear correlation of $\rho = 0.7$.

since it is only invariant under *linear* transformation.

From the concept of a copula, it is immediately clear that the easiest way to construct a multivariate distribution using a copula is to assume some marginal distributions and apply the copula. Below there are some examples for illustration purposes.

A practical problem, however, will be set up the other way round: The multivariate distribution has to be estimated by fitting the copula to data. A discussion of this topic is beyond the scope of this lecture.

⁽³⁾ Let X and Y be two rv's that are both identical gamma (3,1) distributed. Now we apply two different copulas and compare the characteristics by simulating 1000 bivariate draws from both models. First, we use a Gaussian copula

³The example and the graph were taken from [15], page 2 and 22f.

with parameter $\rho_{Ga} = 0.7$. The second distribution is then derived by applying a Gumbel copula whose parameter β is adjusted in a way that the linear correlation coefficient for the resulting bivariate distribution is also $\rho_{Gu} = 0.7$.

Figure 4.8 shows the scatter plot of the 1000 draws for both distributions. The 99% quantile $q_{0.99}$ for the marginal Gamma distribution has been added as an indicator line for extreme values.

Note that despite the fact that both distributions have the same linear correlation coefficient, the dependence between X and Y is obviously quite different in both models. Using the Gumbel copula, extreme events have a tendency to occur together, as one can observe by comparing the number of draws where x and y exceed $q_{0.99}$ simultaneously. Those are 12 for the Gumbel and 3 for the Gaussian case.

Additionally, the probability of Y exceeding $q_{0.99}$ given that X has exceeded $q_{0.99}$ can be roughly estimated from the figure:

$$\hat{P}_{Ga}(X > q_{0.99} | Y > q_{0.99}) = \frac{3}{9} = 0.\bar{3}$$

$$\hat{P}_{Gu}(X > q_{0.99} | Y > q_{0.99}) = \frac{12}{16} = 0.75$$

This is another indicator for the increased probability for the joint occurrence of extreme events.

In the previous section we considered a bivariate distribution to show that marginal distributions and correlation are insufficient information to fully specify the joint distribution. This example was constructed in the following way, using a copula:

Let X and Y be two rv's with standard normal distributions. Obviously the outcome for the bivariate distribution when applying an arbitrary copula is not bivariate normal in general. This is only the case when choosing the Gaussian

copula $C = C_\rho^{Ga}$.

Thus, the following copula has been constructed:

$$\begin{aligned}f(x) &= \mathbf{1}_{\{(\gamma, 1-\gamma)\}}(x) + \frac{2\gamma - 1}{2\gamma} \mathbf{1}_{\{(\gamma, 1-\gamma)^c\}}(x) \\g(y) &= -\mathbf{1}_{\{(\gamma, 1-\gamma)\}}(y) - \frac{2\gamma - 1}{2\gamma} \mathbf{1}_{\{(\gamma, 1-\gamma)^c\}}(y)\end{aligned}$$

with $\gamma \in [\frac{1}{4}, \frac{1}{2}]$. For $\gamma < \frac{1}{2}$, the joint density disappears on the square $[\gamma, 1 - \gamma]^2$ such that the joint distribution is surely not bivariate normal. However, the linear correlation coefficient between X and Y exists. From symmetry considerations ($C(u, v) = c(1 - u, v), 0 \leq u, v \leq 1$) it can be deduced that $\rho_{X,Y} = 0$, irrespective of γ . Therefore, uncountably many bivariate distributions with standard normal marginals and zero correlation exist that are not bivariate normal.

Chapter 5

ALM Implementation

In the following chapter an empirical example of ALM is provided. Hereby, the T-stage ALM problem of section 3.6 will be applied to data that is representative of a defined-benefit pension fund. A liability index l_τ provided by Ryan Labs is used as a proxy for the liabilities. This is a generic index that does not correspond to the liabilities of a specific corporate defined-benefit plan, but this index helps to illustrate the current predicament of pension funds in [34]. The same reference also provides the typical asset classes invested in by pension funds: cash, bonds, equity, real estate, international stocks, international bonds, mortgage, GIC's and annuities, and private equity. Table (5.1) contains the benchmarks used for the asset classes including bonds, equities and international equities.

Given the historical data for the liability index l_τ and asset indexes s_τ^i , $i = 1, \dots, 5$ a multivariate scenario tree can be constructed. Recall that this is achieved

	Asset Class	Benchmark
s^1	Cash	Ryan Labs Cash Index
s^2	Bonds	Lehman U.S. Aggregate Bond Index
s^3	Equities	S&P500
s^4	International Equities	Morgan Stanley EAFE Index
s^5	Mortgages	Lehman Mortgage Index

Table 5.1: Benchmarks for the pension fund asset classes.

by fitting a multivariate time-series model to the return vector:

$$R_\tau = \begin{bmatrix} r_\tau^1 \\ r_\tau^2 \\ \vdots \\ r_\tau^6 \end{bmatrix} = \begin{bmatrix} l_\tau/l_{\tau-1} - 1 \\ s_\tau^1/s_{\tau-1}^1 - 1 \\ \vdots \\ s_\tau^5/s_{\tau-1}^5 - 1 \end{bmatrix}. \quad (5.1)$$

Once a time-series model is found, it is simple to generate sample paths for the returns and then convert the returns back to index values. Note that in our example τ is interpreted as time, and in the previous chapter, t is interpreted as the stage in a stochastic program. It is possible that they will coincide; however, there will usually be many smaller time periods between stages. In our application, a time-series model is fit to monthly data, but a stage covers a 6-month period.

Figure (5.1) contains the plots of the monthly returns for the components of R_τ . There are 237 data points corresponding to the returns for the months of April 1985 to December 2004. An obvious characteristic of the data is the volatility clustering, especially noticeable in the equity index. This indicates that a time-series model with time-varying volatilities is appropriate.

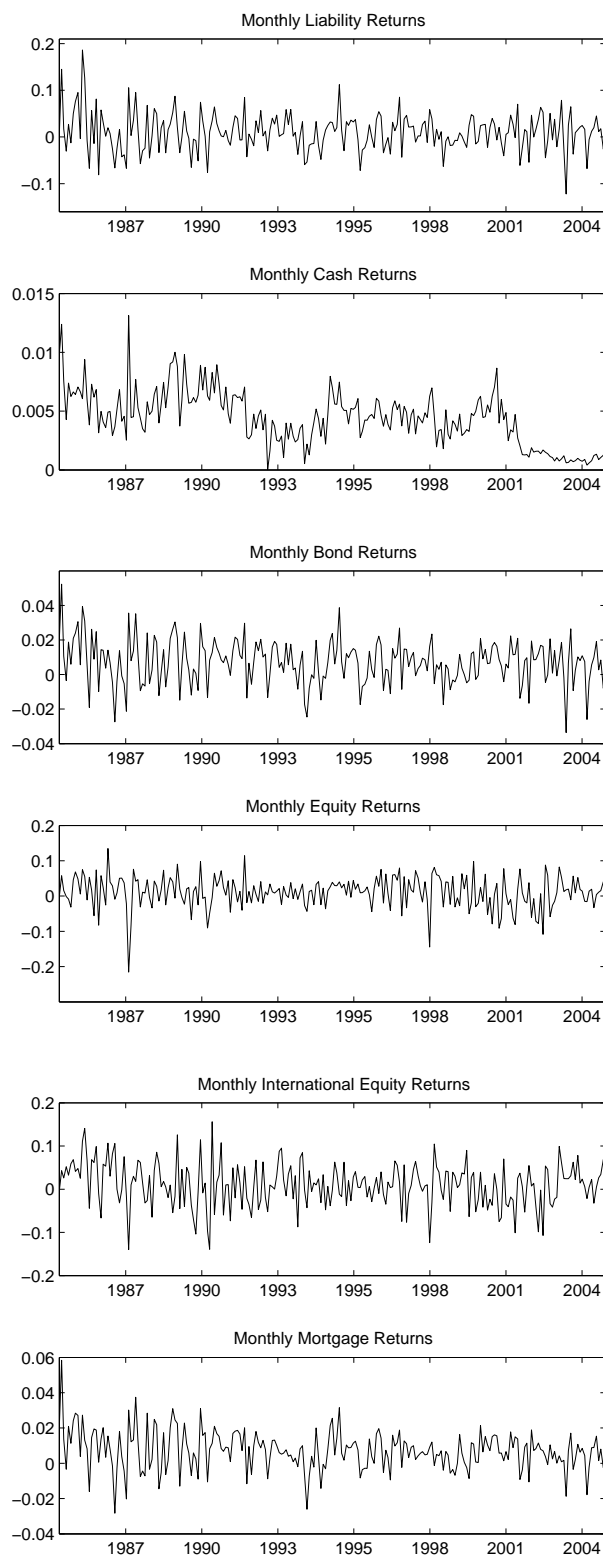


Figure 5.1: Monthly returns R_t for April 1985 to December 2004.

5.1 Finding an adequate model

As a first step in fitting a model to the data, the major trends of the individual series are removed by an exponentially weighted moving average (EWMA) process for the mean. The means of the univariate return series are assumed to follow:

$$m_{\tau}^i = \lambda_m m_{\tau-1}^i + (1 - \lambda_m) r_{\tau-1}^i, \quad \text{for } i = 1, \dots, 6, \quad (5.2)$$

where λ_m is a fixed parameter. By writing $m_{\tau} = (m_{\tau}^1, \dots, m_{\tau}^6)'$, the new return series of interest is

$$\tilde{R}_{\tau} = R_{\tau} - m_{\tau}, \quad (5.3)$$

and as the next step, a vector autoregressive (VAR) model is calibrated to \tilde{R}_{τ} . For the data at hand, the AIC indicates that the VAR of order 1 is optimal.

More generally, one may fit a multivariate autoregressive moving average (ARMA) model, however, multivariate financial data typically indicates only an autoregressive component, so it is reasonable to restrict the model to VAR. Extensions of the VAR model that additionally includes economic regime changes and long term equilibria in an ALM context may be used as well.

To find the optimal value of λ_m , a course grid was created, and for each element in the grid, the AICs of low order VAR models were compared. VAR(1) always resulted in the lowest AIC for any value of λ_m in the grid. A fine grid for λ_m was then constructed, and the AICs of the corresponding VAR(1) models were compared. This procedure gave an optimal value of $\lambda_m = 0.952$.

After estimation of the VAR(1) model, the residuals are computed by

$$\hat{E}_{\tau} = \tilde{R}_{\tau} - \hat{\Pi}_1 \tilde{R}_{\tau-1}, \quad (5.4)$$

and the standardized residuals $\hat{\Sigma}^{-1/2}\hat{E}_\tau$ are plotted in figure (5.2). The usual assumption is that the innovations are Gaussian, in which case the standardized residuals should be i.i.d. $\text{Normal}(0, I_6)$. This is clearly not the case because there is still a significant amount of volatility clustering and extreme events. The standardized residuals are aggregated into one series, and the corresponding qq-plot versus the standard normal is found in figure (5.3).

To get an idea of the variability and dependence structure of the innovations for the VAR(1) model, the estimated volatilities $\hat{\sigma}^i$ from the univariate series $\hat{e}^i = \{\hat{e}_\tau^i, \tau = 1, \dots, 237\}$, where each \hat{e}_τ^i is a component of \hat{E}_τ , are

$\hat{\sigma}^1$	$\hat{\sigma}^2$	$\hat{\sigma}^3$	$\hat{\sigma}^4$	$\hat{\sigma}^5$	$\hat{\sigma}^6$
0.0404	0.0015	0.0124	0.0450	0.0494	0.0105

and the estimated correlation of E_τ is

$$\text{Corr}_E = \begin{bmatrix} 1.0000 & 0.5176 & 0.9343 & 0.1734 & 0.0652 & 0.8134 \\ 0.5176 & 1.0000 & 0.6261 & 0.0303 & -0.0332 & 0.6168 \\ 0.9343 & 0.6261 & 1.0000 & 0.1792 & 0.0780 & 0.9350 \\ 0.1734 & 0.0303 & 0.1792 & 1.0000 & 0.5933 & 0.2026 \\ 0.0652 & -0.0332 & 0.0780 & 0.5933 & 1.0000 & 0.0874 \\ 0.8134 & 0.6168 & 0.9350 & 0.2026 & 0.0874 & 1.0000 \end{bmatrix}.$$

The first noticeable point is that the volatilities corresponding to the equity returns are the largest, the volatility corresponding to the bond returns is smaller, and the volatility corresponding to the cash returns is very small. Also, the volatility corresponding to the liability returns is almost as large as that of the equities, meaning that the liabilities of pension funds are actually quite risky. The second noticeable point is that the liability returns and bond returns are highly correlated as one would expect. This means that when the optimization program is solved

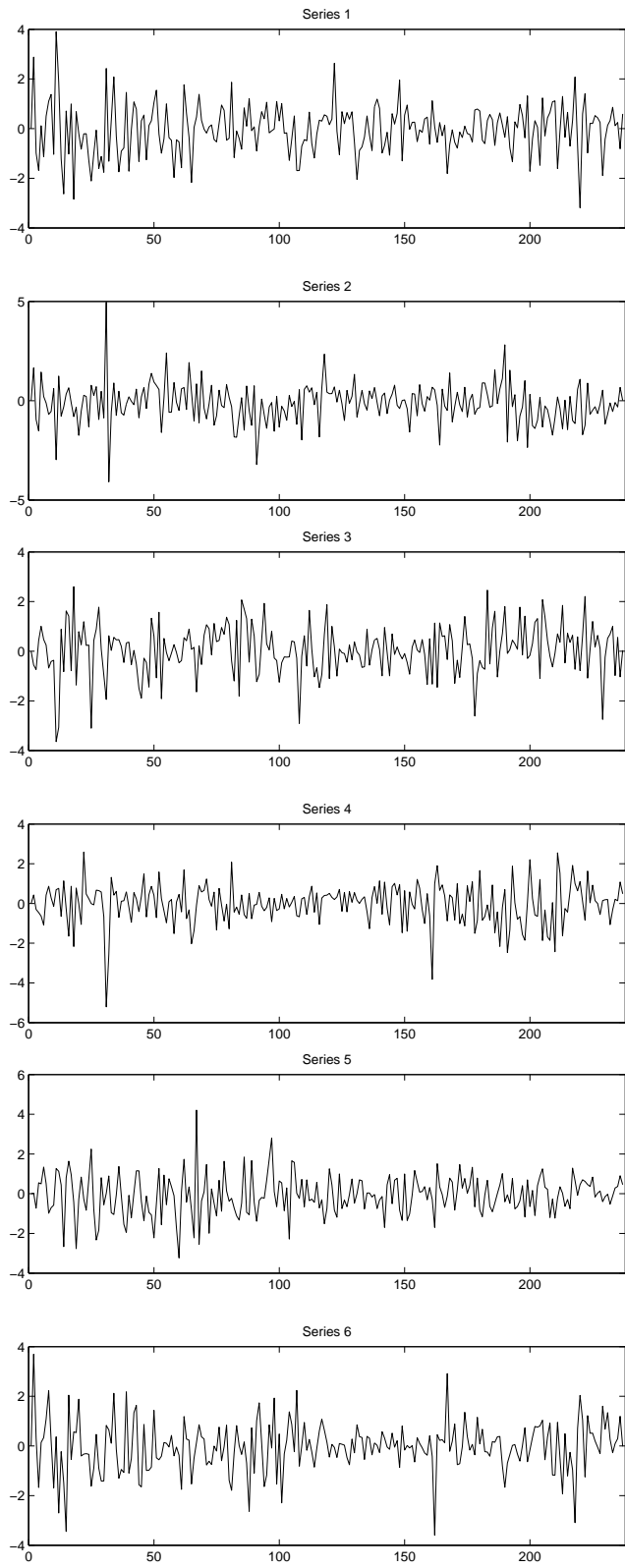


Figure 5.2: Standardized residuals $\hat{\Sigma}^{-1/2}\hat{E}_\tau$ of the VAR(1) model for R_τ .

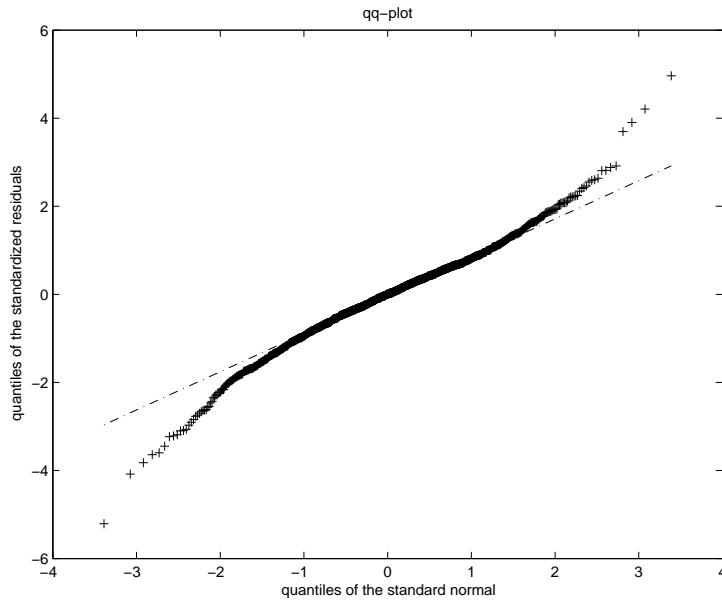


Figure 5.3: QQ-plot of the standard normal versus the standardized residuals $\hat{\Sigma}^{-1/2} \hat{E}_\tau$.

	\hat{e}^1	\hat{e}^2	\hat{e}^3	\hat{e}^4	\hat{e}^5	\hat{e}^6
$\hat{\alpha}^i$	1.8569	1.7411	1.9900	1.8727	1.9702	1.8096
$\hat{\sigma}_\alpha^i$	0.0263	0.0008	0.0087	0.0285	0.0343	0.0067

Table 5.2: Univariate ML estimates of the tail index and scale parameters for each residual series \hat{e}^i .

for the minimum risk portfolio, one could expect a large allocation in the bonds to offset the risk in the liabilities.

A symmetric stable distribution is fit to each of the univariate residual series of the VAR(1) model by maximum likelihood estimation. The estimates of the tail index $\hat{\alpha}^i$ and scale parameter $\hat{\sigma}_\alpha^i$ from each univariate series \hat{e}^i is given table (5.2). The estimation was restricted to symmetric distributions because of the short length of the data series. Alternatively, it is reasonable to assume that $\alpha = 1.8$ for financial data and carry out the estimation for the scale parameter alone. The empirical density of the liability return innovations is compared to

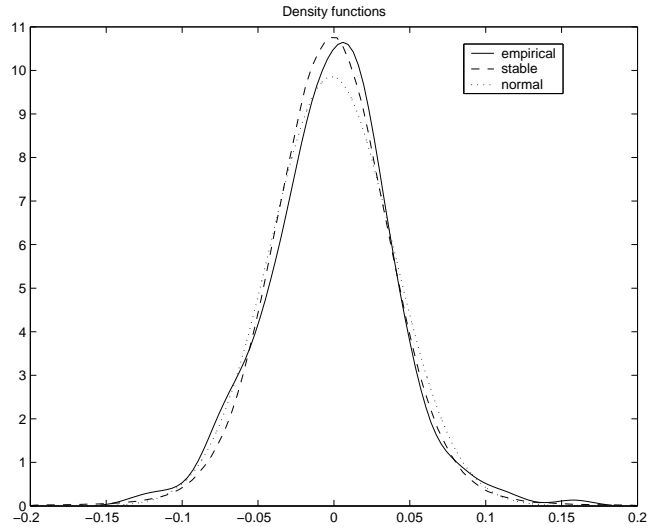


Figure 5.4: Density functions for the residuals of the liability return series.

	\hat{e}^1	\hat{e}^2	\hat{e}^3	\hat{e}^4	\hat{e}^5	\hat{e}^6
Normal	0.0505	0.0685	0.0578	0.0673	0.0422	0.0509
Stable	0.0332	0.0291	0.0572	0.0677	0.0418	0.0569

Table 5.3: Comparison of KD under the normal and stable assumptions.

both the estimated normal density and the estimated stable density in figures (5.4) and (5.5). As is seen, the stable density better matches the peak of the empirical density and has a slower decay at the tails than that of the normal density.

Two goodness-of-fit measures are employed to compare the normal fit and the stable fit of the univariate series: the Kolmogorov distance (KD) and the Anderson-Darling (AD) statistic. The KD and AD for the normal and stable estimated distributions for each of the series can be found in tables (5.3) and (5.4). The normal fit slightly outperforms the stable fit twice under the KD measure, but the stable fit is clearly superior under the AD measure.

A sub-Gaussian distribution can be fitted to the residuals E_τ of the ALM

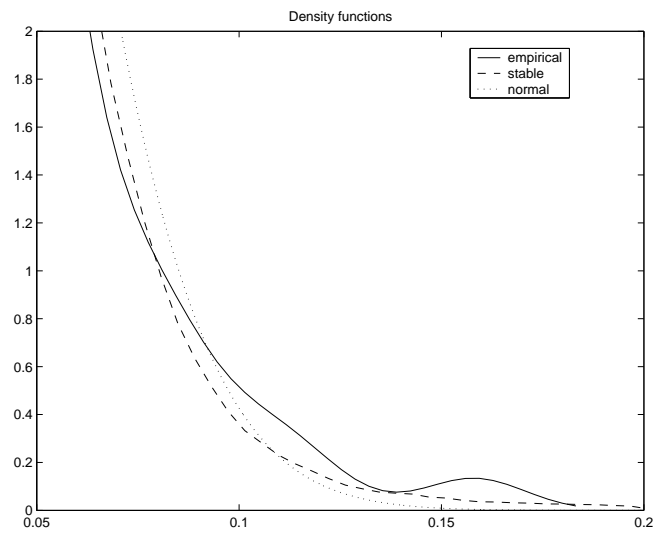


Figure 5.5: Right tail of the density functions for the residuals of the liability return series.

	\hat{e}^1	\hat{e}^2	\hat{e}^3	\hat{e}^4	\hat{e}^5	\hat{e}^6
Normal	0.6546	45.9484	0.1674	12.4620	0.1965	0.3848
Stable	0.1116	0.0947	0.1523	0.1366	0.0856	0.1151

Table 5.4: Comparison of AD under the normal and stable assumptions.

data. First, α is estimated and the univariate estimates from table (5.2), yielding $\hat{\alpha} = 1.8705$. Assuming the residuals have zero mean, a moment estimator for Q in equations is applied to $\tilde{Z}_\tau = \hat{E}_\tau$ with $p = \hat{\alpha}/3$ and $q = 1$. The resulting moment estimates of the scale parameters \hat{q}_{jj} are:

\hat{q}_{11}	\hat{q}_{22}	\hat{q}_{33}	\hat{q}_{44}	\hat{q}_{55}	\hat{q}_{66}
0.0257	0.0008	0.0080	0.0282	0.0319	0.0066

They can be compared with the ML estimates in table (5.2). The moment estimate for Q given by the above equations is not symmetric, but a symmetric estimate is given by $\hat{Q} = ((\hat{q}_{ij}^2 + \hat{q}_{ji}^2)/2)$. The standardized residuals $\hat{Q}^{-1/2}\hat{E}_\tau$ are also computed and are plotted in figure (5.6). In this case, the data points should all be temporally and serially independent realizations of a $S_{1.8705}(1, 0, 0)$ random variable. This is clearly not the case because there is a significant amount of volatility clustering. The qq-plot of the stable random variable versus the aggregated standardized residuals is found in figure (5.7). This plot appears closer to linear than the qq-plot with the standard normal in (5.3), which indicates the sub-Gaussian provides a better fit than the multivariate normal; however, neither of these capture the time-varying nature of the innovations.

5.1.1 Exponentially Weighted Moving Average

Models

To account for the volatility clustering, different types of models are implemented: The first assumes the innovations are Gaussian with a time-varying covariance matrix, and the second assumes the innovations are sub-Gaussian with a time-varying dispersion matrix.

Given a multivariate data set $\{E_\tau, \tau = 1, \dots, \tau_m\}$ with zero mean, the sample

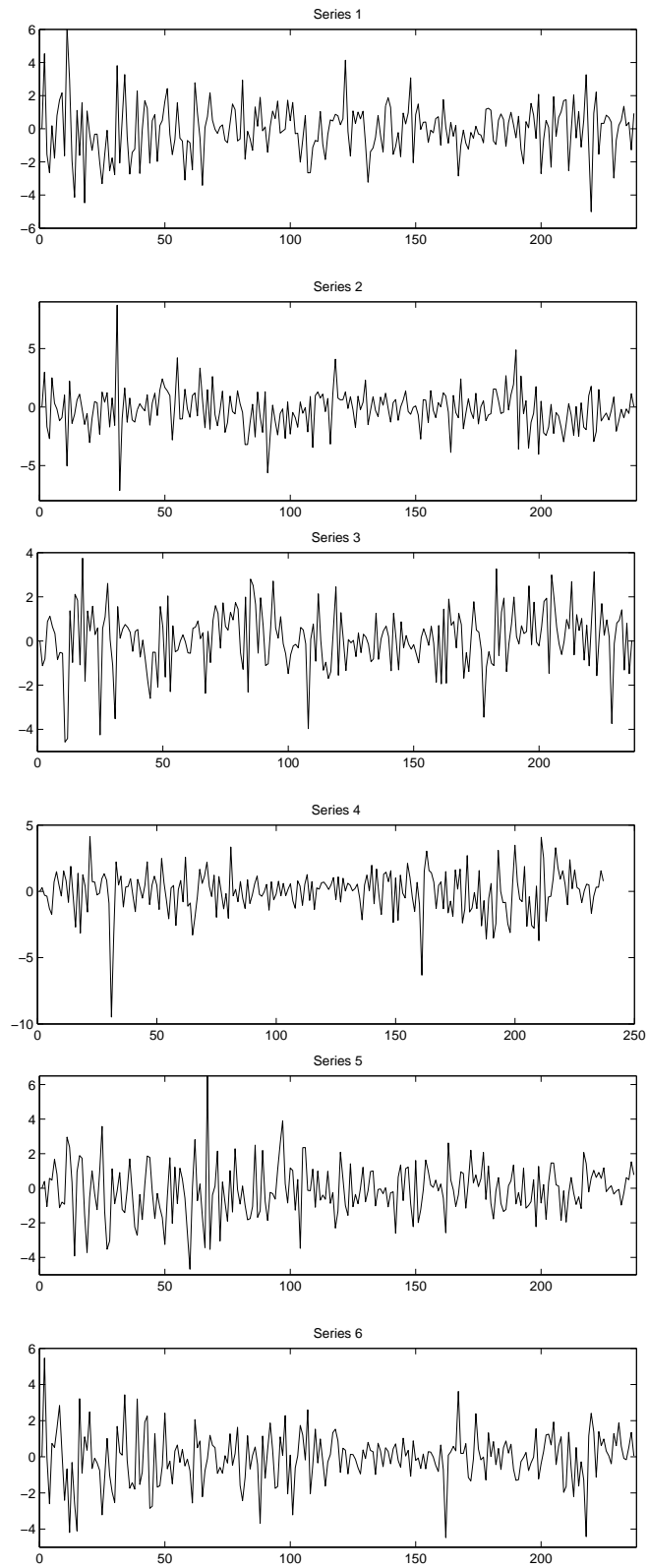


Figure 5.6: Standardized residuals $\hat{Q}^{-1/2}\hat{E}_\tau$ of the VAR(1) model for R_τ .

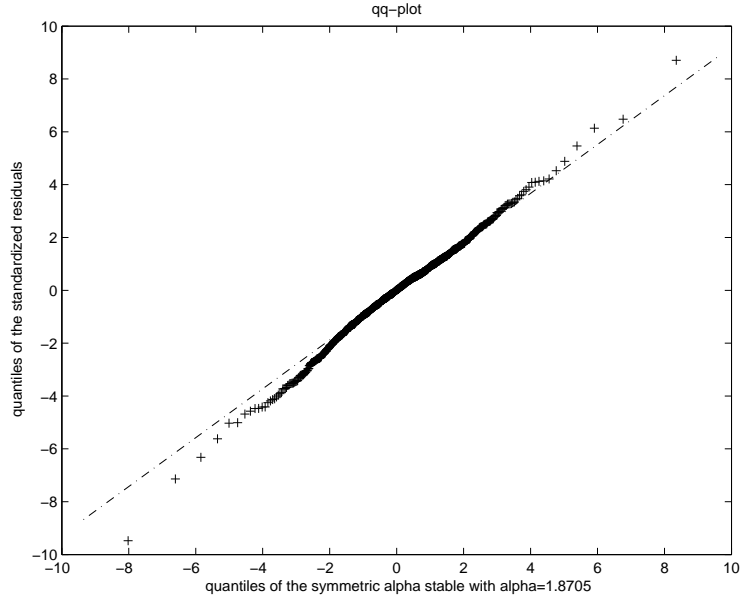


Figure 5.7: QQ-plot of the symmetric stable with $\alpha = 1.8705$ versus the standardized residuals $\hat{Q}^{-1/2} \hat{E}_\tau$.

estimate of the covariance is just

$$\hat{\Sigma} = \frac{1}{\tau_m - 1} \sum_{\tau=1}^{\tau_m} E_\tau E'_\tau. \quad (5.5)$$

Note that there is equal weight applied to each observation of the data set. To allow a time-varying volatility estimate, the covariance estimate at time τ is allowed to depend on the data before time τ , and the weights are assumed to decay exponentially from the most recent observation:

$$\hat{\Sigma}_{\tau|\tau-1} = (1 - \lambda_e) (E_{\tau-1} E'_{\tau-1} + \lambda_e E_{\tau-2} E'_{\tau-2} + \lambda_e^2 E_{\tau-3} E'_{\tau-3} + \dots),$$

where $0 < \lambda_e < 1$ and the weights are chosen so that they sum to one for an

infinite series. The estimate can also be written in the recursive form

$$\hat{\Sigma}_{\tau|\tau-1} = (1 - \lambda_e)E_{\tau-1}E'_{\tau-1} + \lambda_e\hat{\Sigma}_{\tau-1|\tau-2}, \quad (5.6)$$

which is known as the exponentially weighted moving average (EWMA) covariance model with decay factor λ_e . In practice, an initial covariance $\hat{\Sigma}_{0|-1}$ is needed to estimate λ_e , compute standardized residuals, and simulate sample paths. The approach used is to estimate $\Sigma_{0|-1}$ from the sample covariance (5.5) of the initial 10% of the data set.

RiskMetrics [21] offers an estimation technique for λ_e based on the root mean squared prediction error (RMSE) of $(e_{\tau}^i)^2$:

$$RMSE_2^i(\lambda_e) = \sqrt{\frac{1}{\tau_m} \sum_{\tau=1}^{\tau_m} \left((e_{\tau}^i)^2 - \hat{\sigma}_{\tau|\tau-1,ii}^2(\lambda_e) \right)^2}, \quad (5.7)$$

where $\hat{\sigma}_{\tau|\tau-1,ii}^2(\lambda_e)$ is a diagonal component of $\hat{\Sigma}_{\tau|\tau-1}$ in equation (5.6). Since the data series is assumed to have zero mean, $E_{\tau-1}(e_{\tau}^i)^2 = \sigma_{\tau|\tau-1,ii}^2$, so the prediction error of $(e_{\tau}^i)^2$ is the difference of terms inside the square root in equation (5.7). A single optimal estimate λ_e^* for the decay factor is computed from the RMSE of each univariate series through the formulas:

$$\lambda_e^* = \sum_{i=1}^n \phi_i \lambda_i^*, \quad (5.8)$$

where

$$\lambda_i^* = \underset{\lambda}{\operatorname{argmin}} RMSE_2^i(\lambda), \quad \theta_i = \frac{RMSE_2^i(\lambda_i^*)}{\sum_{k=1}^n RMSE_2^k(\lambda_k^*)}, \quad \phi_i = \frac{\theta_i^{-1}}{\sum_{k=1}^n \theta_k^{-1}}. \quad (5.9)$$

Using this technique, RiskMetrics recommends typical parameter values of $\lambda_e =$

.94 for daily data and $\lambda_e = .97$ for monthly data.

Stable EWMA

These same ideas are used in the sub-Gaussian case in [19] by allowing a time-varying dispersion matrix. Similar to as above, exponential weights are applied to the moment estimators, yielding the equations:

$$\hat{q}_{\tau|\tau-1,jj}^p = (1 - \lambda_e) |e_{\tau-1}^j|^p A(p) + \lambda_e \hat{q}_{\tau-1|\tau-2,jj}^p \quad (5.10)$$

$$B_{\tau|\tau-1,ij} = (1 - \lambda_e) e_{\tau-1}^i (e_{\tau-1}^j)^{(q-1)} A(q) + \lambda_e B_{\tau-1|\tau-2,ij} \quad (5.11)$$

$$\hat{q}_{\tau|\tau-1,ij}^2 = B_{\tau|\tau-1,ij} \hat{q}_{\tau|\tau-1,jj}^{2-q}, \quad i \neq j, \quad (5.12)$$

and the symmetric estimator for the time-varying dispersion matrix is given by $\hat{Q}_{\tau|\tau-1} = (\hat{q}_{\tau|\tau-1,ij}^2 + \hat{q}_{\tau|\tau-1,ji}^2) / 2$. This model is referred to as the stable exponentially weighted moving average model (SEWMA). The authors also extend the estimation technique for the decay factor by considering the prediction error of $|e_{\tau}^i|^p$. They note that $E_{\tau-1}(|e_{\tau}^i|^p) = q_{\tau|\tau-1,ii}^p / A(p)$ and suggest to minimize the following RMSE error for each univariate series:

$$RMSE_p^i(\lambda_e) = \sqrt{\frac{1}{\tau_m} \sum_{\tau=1}^{\tau_m} \left(A(p) |e_{\tau}^i|^p - \hat{q}_{\tau|\tau-1,ii}^p(\lambda_e) \right)^2}. \quad (5.13)$$

The single optimal decay factor λ_e^* is then found by replacing $RMSE_2^i$ with $RMSE_p^i$ in equations (5.8-5.9). Using the VAR(1) residuals of the ALM data, this technique is applied in both the normal and sub-Gaussian cases with $p = \alpha/3$. A grid for λ was constructed with increments of 0.001, and $RMSE_p^i(\lambda)$ was minimized over this grid. In both cases, a value of $\lambda_e = 0.95$ for equations (5.10-5.12) is found to be appropriate. The exact values of λ_e^* are found in table (5.5).

There are difficulties in implementing the SEWMA model for the ALM resid-

	α	p	λ_e^*
Normal	2	0.6667	0.9496
Stable	1.8705	0.6235	0.9494

Table 5.5: Comparison of the optimal decay factor λ_e^* under the normal and stable assumptions using the selection criterion based on $RMSE_p^i$.

uals: While the estimate $\hat{Q}_{\tau|\tau-1}$ is defined to be symmetric, there is no guarantee that it is positive definite. In the case of the ALM residuals, the eigenvalues are often negative and often very near zero. The negative eigenvalues are easily dealt with by using an incomplete Cholesky decomposition when computing the standardized residuals and generating sample paths. The eigenvalues very near zero, on the other hand, will cause the standardized residuals to explode beyond any reasonable value. The likely cause of this inadequate estimate of the dispersion matrix is the short length of the data series. For this reason, the scenarios generated according to the SEWMA model were not inputted into the ALM optimization problem.

Stable Subordination EWMA

To overcome the difficulties of the SEWMA model, a more *ad hoc* approach is taken by modeling the time-varying sub-Gaussian distribution in terms of a governing Gaussian distribution and the scale parameters of the individual univariate series. First, one needs the following result: if

$$g \sim N(0, \sigma_g^2), \quad y \sim S_\alpha(\sigma_y, 0, 0), \quad s \sim S_{\alpha/2} \left(\frac{2\sigma_y^2}{\sigma_g^2} \left(\cos \left(\frac{\pi\alpha}{4} \right) \right)^{2/\alpha}, 1, 0 \right),$$

and s and g are independent, then

$$y \stackrel{d}{=} \sqrt{s}g.$$

See [32] and the reference therein. If the governing Gaussian distribution G_τ for the multivariate data has a time-varying covariance matrix $\Sigma_{\tau|\tau-1} = (\sigma_{\tau|\tau-1,ij}^2)$ and each univariate series is modeled with an α_i -stable random variable with time-varying scale parameter $q_{\tau|\tau-1,i}$, the previous results suggest a way to model E_τ with a time-varying sub-Gaussian-like distribution:

$$E_\tau \stackrel{d}{=} \begin{bmatrix} \sqrt{s_\tau^1} g_\tau^1 \\ \vdots \\ \sqrt{s_\tau^n} g_\tau^n \end{bmatrix}, \quad (5.14)$$

$$G_\tau = \begin{bmatrix} g_\tau^1 \\ \vdots \\ g_\tau^n \end{bmatrix} \sim N(0, \Sigma_{\tau|\tau-1}), \quad (5.15)$$

$$s_\tau^i \sim S_{\alpha_i/2} \left(\frac{2q_{\tau|\tau-1,i}^2}{\sigma_{\tau|\tau-1,ii}^2} \left(\cos \left(\frac{\pi\alpha_i}{4} \right) \right)^{2/\alpha_i}, 1, 0 \right). \quad (5.16)$$

When generating a sample for E_τ , the samples of s_τ^i , $i = 1, \dots, n$, are taken from the same random seed so that the above equations will be close to the sub-Gaussian representation where the same subordinator multiplies each component of the normal random vector. In the above equations, the covariance of the governing Gaussian distribution captures the dependence between the series, and each subordinator s_τ^i is chosen to give the proper tail index and scale parameter for each of the univariate series. Recall that for the sub-Gaussian distribution, all marginals have the same tail index, so the above equations are actually an extension that allow different tail indexes, α_i , for the marginals. The scale parameters and covariance matrix are estimated from EWMA equations already seen. The time-varying

estimate for the scale parameter is given by:

$$\hat{\sigma}_{\tau|\tau-1,i}^{p_i} = (1 - \lambda_e) |e_{\tau-1}^j|^{p_i} A(p_i) + \lambda_e \hat{\sigma}_{\tau-1|\tau-2,i}^{p_i}, \quad (5.17)$$

which similar to equation (5.10), and it is reasonable to take $p_i = \alpha_i/3$ for financial series. To obtain the estimate for the covariance of the governing Gaussian, the data set of E_τ is first truncated at 5% and 95% to remove the effects of extreme events. The estimate is then obtained from the truncated series E_τ^* by:

$$\hat{\Sigma}_{\tau|\tau-1} = (1 - \lambda_e) E_{\tau-1}^* (E_{\tau-1}^*)' + \lambda_e \hat{\Sigma}_{\tau-1|\tau-2}. \quad (5.18)$$

The optimal value of λ_e is best calibrated through backtesting, or alternatively, the RiskMetric technique for $RMS E_p^i$ can be carried over. The latter approach is used for the ALM data, which gives $\lambda_e = 0.95$ again. This model will be referred to as the stable subordination exponentially weighted moving average model (SSEWMA).

5.1.2 VaR Backtesting

The forecasting performances of the EWMA and SSEWMA models are examined by comparing the predicted VaRs with the observed returns as in [31]. From the definition of VaR, the null hypothesis to test is:

$$P(r_\tau < -\text{VaR}_\beta(\tau)) = 1 - \beta, \quad (5.19)$$

for a return series $\{r_\tau\}$. This hypothesis is tested for each ALM return series $r^i = \{r_\tau^i, \tau = 1, \dots, 237\}$, $i = 1, \dots, 6$, and for various values of β .

In this backtesting analysis, both the VAR(1)-EWMA and VAR(1)-SSEWMA

models are fit to a moving window of 100 data points. Since it is difficult to estimate the tail index of the stable distribution with such a short time-series, it is assumed that $\alpha_i = 1.8$ for each of the univariate series in the SSEWMA model. Let $\overline{\text{VaR}}_\beta(\tau)$, for $\tau = 101, \dots, 237$, be the estimate of $\text{VaR}_\beta(\tau)$ from a model calibrated to $\{r_{\tilde{\tau}}, \tilde{\tau} = \tau - 100, \dots, \tau - 1\}$. If equation (5.19) holds, then

$$\chi_\tau = \mathbf{1}(r_\tau < -\overline{\text{VaR}}_\beta(\tau)) = \begin{cases} 1 & \text{with probability } 1 - \beta, \\ 0 & \text{with probability } \beta, \end{cases} \quad (5.20)$$

where $\mathbf{1}(\cdot)$ is the indicator function, and the total number of VaR exceedings has a binomial distribution:

$$\mathcal{X} = \sum_{\tau=101}^{237} \chi_\tau \sim \text{Bin}(137, 1 - \beta). \quad (5.21)$$

The testing rule is to reject the null hypothesis at level of significance $100\delta\%$ if

$$\sum_{k=1}^{\mathcal{X}} \binom{137}{k} (1 - \beta)^k \beta^{137-k} \leq \delta/2, \quad (5.22)$$

or,

$$\sum_{k=1}^{\mathcal{X}} \binom{137}{k} (1 - \beta)^k \beta^{137-k} \geq 1 - \delta/2. \quad (5.23)$$

The number of exceedings and the corresponding p-values for each ALM return series are contained in tables (5.6-5.7). The conclusions are:

- At level of significance 99%, neither the EWMA or the SSEWMA model is rejected for any value of β .
- At level of significance 95%, the EWMA model is rejected three times for $\beta = 0.99$ and once for $\beta = 0.95$ while the SSEWMA model is never rejected.

β	Exceedings and p-values					
	r^1	r^2	r^3	r^4	r^5	r^6
0.99	4 (0.0252)	3 (0.0990)	4 (0.0252)	4 (0.0252)	3 (0.0990)	3 (0.0990)
0.95	12 (0.0405)	11 (0.0850)	9 (0.2984)	8 (0.4955)	10 (0.1657)	6 (0.9379)
0.90	16 (0.4168)	13 (0.9851)	14 (0.7920)	16 (0.4168)	14 (0.7920)	12 (0.7586)
0.80	26 (0.8639)	28 (0.7987)	26 (0.8639)	32 (0.2773)	28 (0.7987)	26 (0.8639)

Table 5.6: Number of VaR_β exceedings in 137 data points with corresponding p-values under the normal assumption.

β	Exceedings and p-values					
	r^1	r^2	r^3	r^4	r^5	r^6
0.99	1 (0.7968)	2 (0.3171)	2 (0.3171)	3 (0.0990)	2 (0.3171)	3 (0.0990)
0.95	11 (0.0850)	11 (0.0850)	9 (0.2984)	9 (0.2984)	10 (0.1657)	8 (0.4955)
0.90	16 (0.4168)	14 (0.7920)	17 (0.2808)	18 (0.1800)	18 (0.1800)	16 (0.4168)
0.80	28 (0.7987)	31 (0.3783)	27 (0.9659)	32 (0.2773)	29 (0.6417)	27 (0.9659)

Table 5.7: Number of VaR_β exceedings in 137 data points with corresponding p-values under the stable assumption ($\alpha = 1.8$).

This indicates that the normal distribution is overly optimistic in predicting the occurrence of the largest losses, and the stable distribution results in a more reliable forecast.

- For $\beta = 0.90$ and 0.80 , the EWMA model produces reasonably large p-values, which just indicates that the normal distribution could be suitable for forecasting more toward the middle of the distribution.

Overall, the SSEWMA model provides a better fit to the tails and is preferable based on the examination of the p-values.

5.2 Solving the Optimization Problem

The ALM optimization problem is now solved using scenarios generated from the time-series models of the previous section. First, efficient frontiers are developed from the 2-stage problem with scenarios based on the EWMA and SSEWMA

models, and postoptimality analysis is briefly discussed. Then, backtesting is carried out to compare the performance of the 1-stage problem versus the 2-stage recourse problem and the normal assumption versus the stable assumption. The results from varying the distributional assumption are mixed, but the 2-stage recourse problem outperforms the 1-stage problem. Before presenting these results, the parameters of the optimization problem are first specified.

For pension funds, decisions are made approximately on an annual basis, so a stage in the stochastic program should correspond to 12 months. A twelve month stage left too few data points in the backtesting, so the decision was made to shorten the stage to cover a six month period. In addition to giving more points for comparison in the backtesting, the time-series models should generate more reliable scenarios over the shorter time period.

For the 2-stage problem, a balanced scenario tree is generated with 10^4 first stage scenarios and 10^7 second stage scenarios, giving 10^3 second stage nodes connected to each first stage node. This huge number of scenarios gives fairly reliable optimal allocations, and memory limitations did not allow much larger scenario trees to be considered. The first stage scenarios were created by simulating 10^4 sample paths of the time-series model out to six months, and the second stage scenarios were created by simulating another 10^3 sample paths out an additional six months for each of first stage scenarios. Scenario reduction and bundling using the methods of probability metrics was also attempted in order to create a better set of first stage scenarios, but these methods could not handle sample paths of this number with the given hardware.

It is necessary to convert the generated sample paths of the returns back to the index values of the benchmarks. This is not a problem when using the normal distribution, but it does cause some small difficulties when using the stable

distribution. Since the returns have infinite variance under the stable assumption and are temporally dependent, the sample paths of the corresponding index values will explode. For this reason, the stable return scenarios are truncated at levels corresponding to p-values of 0.001 and 0.999 of the estimated distribution. This eliminates the explosion of the index values while still fitting the tail of the return distribution better than the normal assumption.

For the efficient frontiers and at the start date of the backtesting, it is assumed that the pension fund is fully funded: the total asset wealth and the liability obligation are both taken to be \$1,000, and because of the structure of the deterministic equivalent form of the optimization problem, any pension fund that is fully funded will have the same optimal allocations (as a percent of the asset wealth). For instance, a fund with an initial \$1,000,000 in both asset wealth and liability obligation has the same optimal allocations as one with \$1,000 in both.

Including transaction costs, the optimal allocations depend also on the initial allocation, not just the generated scenarios and initial wealth. In this case, it is assumed that the fund initially holds 40% of its wealth in bonds and 60% of its wealth in equities. A reasonable assumption for the trading costs, as a percent of wealth traded, is obtained from data on mutual funds in [9]. In our example, the median trading cost (TC) is 0.70% of fund assets per year:

$$TC \approx 0.0070 \cdot \text{Fund Assets.}$$

The turnover, defined as the ratio of annual fund sales to the fund assets, is determined to have a median of 0.70:

$$\text{Fund Sales} \approx 0.70 \cdot \text{Fund Assets.}$$

Assuming that the fund buys approximately as much as it sells, then

$$\text{Traded Wealth} \approx 2 \cdot \text{Fund Sales.}$$

Combining equations yields

$$\text{TC} \approx \frac{0.0070}{2 \cdot 0.70} \text{Traded Wealth,}$$

or trading costs are approximately 0.5% of the traded wealth. Additionally assuming that the transaction costs are the same for each of the five ALM asset classes, the values of $TCB^i = TCS^i = 0.005$, for $i = 1, \dots, 5$, are used in the optimization problem.

5.3 Efficient Frontiers

The numerical results of the efficient frontiers for the 2-stage recourse problem are now given. Recall that the risk measure for the 2-stage problem is:

$$\rho_2 = \mu_1 \text{CVaR}_\beta(-sw_2) + \mu_2 \text{CVaR}_\beta(-sw_3), \quad (5.24)$$

where sw_{t+1} is the surplus wealth at the end of stage t (and $sw_1 = 0$ since the pension fund is initially fully funded). A confidence level of $\beta = 0.95$ is used in this section to emphasize the differences between the normal and stable assumptions. For the remainder, it is taken that $\mu_1 = \mu_2 = .5$, and studies in assigning different weights to the CVaR at different stages is saved for a later time. Since the reward is the expected surplus wealth at the end of the second stage, $E(sw_3)$, the efficient frontier is obtained by varying λ in the minimization objective: $\lambda\rho_2 - (1-\lambda)E(sw_3)$. Figure (5.8) contains three different efficient frontiers:

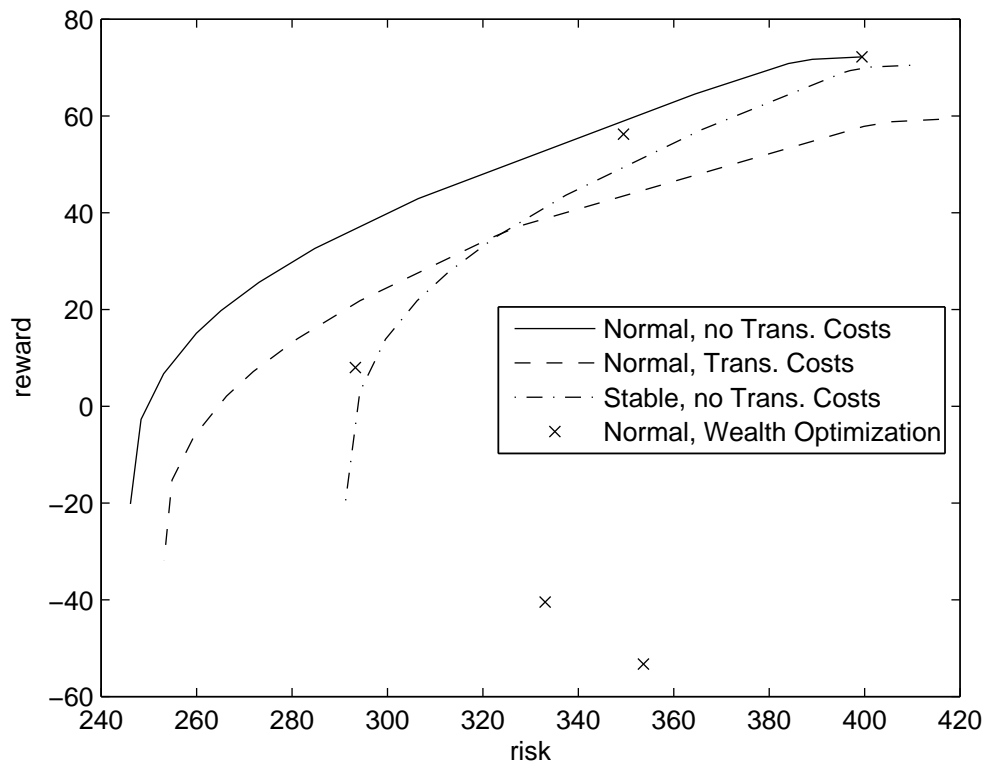


Figure 5.8: Efficient frontiers under the normal assumption and stable assumption for $\beta = 0.95$.

- Optimization without transaction costs and scenarios generated from the normal assumption (EWMA model).
- Optimization with transaction costs and the same set of scenarios generated from the normal assumption.
- Optimization without transaction costs and scenarios generated from the stable assumption (SSEWMA model) using the tail index estimates from table (5.2).

The optimal allocations, as percents of the initial wealth, can be found in the tables (6.1-6.3) in the appendix of the lecture notes.

In all three cases and for any value of λ , the optimal first stage allocations are some combination of the bond and international equity indexes. The portfolio that maximizes the expected final surplus wealth ($\lambda = 0$) invests entirely in the international equity index, and the minimum risk portfolio ($\lambda = 1$) invests entirely in the bond index. A couple immediate comments can be made about the figure. Since the stable distribution has a higher probability of extreme events, the frontier of the stable distribution lies below that of the normal distribution. The inclusion of transaction costs also moves the efficient frontier downward, and the distance it moves for various values of λ depends on the initial allocation.

A few risk-reward points obtained by replacing the surplus wealth with the wealth in the optimization problem, under the normal assumption, are also included in figure (5.8). The optimal allocations, found in table (6.4), are very different in this case: The minimum risk portfolio has a very large proportion of wealth invested in the cash index. When the corresponding ρ_2 and $E(sw_3)$ are calculated, the points for the risk-averse portfolios lie far below the efficient frontier. This illustrates the advantage of considering the liabilities and assets together in the same optimization problem. Maximizing the expected final wealth and maximizing the expected final surplus wealth result in the same values of ρ_2 and $E(sw_3)$ because of the linearity of the problem.

5.4 Postoptimality Analysis and Backtesting

5.4.1 Postoptimality Analysis

The basic postoptimality analysis examines how the optimal value of a stochastic program changes as the initial probability distribution P_1 becomes contaminated with another probability distribution P_2 . Usually problems of the following form

are considered:

$$\phi(P_1) = \min_{x_1 \in X} F(x_1, P_1) \quad (5.25)$$

where P_1 is a discrete probability distribution of scenarios, X does not depend on P_1 , x_1 are the scenario independent first stage decision variables, and F is convex in x_1 and linear in P_1 . The original probability distribution is assumed to become contaminated through

$$P_\psi = (1 - \psi)P_1 + \psi P_2, \quad \text{with} \quad 0 < \psi < 1. \quad (5.26)$$

This means that the scenarios of both distributions are aggregated into one set of scenarios where the probabilities of the scenarios in P_1 are weighted by $1 - \psi$ and the probabilities of the scenarios in P_2 are weighed by ψ . If the optimal solution is denoted by

$$x_1(P_1) = \arg \min_{x_1 \in X} F(x_1, P_1), \quad (5.27)$$

a set of bounds for the optimal value of the stochastic program under the contaminated distribution, $\phi(P_\psi)$, are given by

$$\begin{aligned} (1 - \psi)\phi(P_1) + \psi\phi(P_2) &\leq \phi(P_\psi) \\ &\leq \min \{ (1 - \psi)\phi(P_1) + \psi F(x_1(P_1), P_2), (1 - \psi)F(x_1(P_2), P_1) + \psi\phi(P_2) \}, \end{aligned} \quad (5.28)$$

where $F(x_1(P_1), P_2)$ is the value of the objective under distribution P_2 when the first stage decision is $x_1(P_1)$ (there is still an implicit minimization over the second stage variables). $F(x_1(P_2), P_1)$ is found in a similar manner.

It is not difficult to verify that the ALM problem can be written in the above form. This contamination method can be easily applied to the situation where P_1 corresponds to the set of scenarios generated from the normal assumption and P_2

corresponds to the set of scenarios generated from the stable assumption. In the case of the minimum risk portfolio ($\lambda = 1$), the optimal objective value coincides with the minimum risk value. Let

$$\rho_2^n = \phi(P_1), \quad \text{and} \quad \rho_2^s = \phi(P_2), \quad (5.29)$$

correspond to the 2-stage risk under the normal and stable distributions, respectively. Also, denote the risk under distribution P_ψ by ρ_2^ψ . As seen in tables (6.1-6.2), the optimal allocations under both the normal assumption and the stable assumption invest all the wealth in the bond index. It follows that $F(x(P_2), P_1) = \rho_2^n$ and $F(x(P_1), P_2) = \rho_2^s$, and the bounds in equation (5.28) produce

$$\begin{aligned} \rho_2^\psi &= (1 - \psi)\rho_2^n + \psi\rho_2^s, \\ &= (1 - \psi) \cdot 246.13 + \psi \cdot 291.21. \end{aligned}$$

The minimum risk in the 2-stage program is then easily calculated when scenarios under the normal assumption and stable assumption are combined. The general contamination technique can also be applied for any value of λ , but direct information about the risk can no longer be calculated.

5.4.2 Portfolio Backtesting

Finally, some backtesting results will be presented. The first round includes transaction costs, and the initial conditions for each run of the optimization problem come from the previous period considered. This provides a realistic comparison for the 1-stage problem versus the 2-stage problem, but it is difficult to calculate the realized risk using the risk measure that was optimized. In the second round, the transaction costs are removed and the initial conditions are reset every run of

the optimization problem. This allows the realized risk to be directly calculated in terms of the optimized risk measure and provides a better comparison for the distributional assumptions; however, this setup favors the 1-stage problem over the 2-stage problem because the second stage becomes irrelevant.

Dynamic Backtesting: 1-stage versus 2-stage

This section performs the dynamic backtesting of the minimum risk 1-stage and 2-stage portfolios with transaction costs. The 2-stage problem finds the optimal allocations that minimize ρ_2 , and the 1-stage problem finds that optimal allocation that minimizes

$$\rho_1 = \text{CVaR}_\beta(-sw_2). \quad (5.30)$$

For a given distributional assumption, the same sets of scenarios are used when solving the 1-stage and 2-stage problems: The 1-stage problem is just restricted to considering the 10^4 first stage scenarios.

The time-series models are fit to a moving window of 100 data points under both the normal and stable assumptions using the EWMA and SSEWMA models, respectively. Running the optimization problems with scenarios generated from the time-series models fit to the first 100 monthly data points give optimal allocations for the six month period beginning in July, 1993. It is again assumed that the pension fund is initially fully funded with 40% of wealth in the bond index and 60% of wealth in the equity index. The window is then shifted forward by 6 data points, and the optimization problems output optimal allocations for January, 1994. The asset wealths resulting from the previous allocations, and those allocations themselves, are used as the initial conditions for the new optimization problems. This setup means that the 2-stage problem is run on a rolling horizon: Since new scenarios are generated every 6 months, only the first stage allocations are actually implemented.

Since it is difficult to obtain a good estimate for the tail index of a stable distribution with only 100 data points, it is assumed that $\alpha = 1.8$ in the SSEWMA model. The backtesting, therefore, gives a comparison of the normal assumption with the stable assumption for this particular value of the tail index.

The window is shifted 21 times resulting in a final surplus wealth for July, 2004. Since this results in only 22 values of the surplus wealth for comparison, the confidence level of CVaR was reduced to $\beta = 0.80$ in ρ_1 and ρ_2 . To measure the relative performances, it is necessary to calculate the risk of the realized surplus wealths. However, it is not reasonable to directly calculate the CVaR of these values because the surplus wealth that is used as the initial condition in the optimization problems varies over the time horizon and is different for the different assumptions. It is also not possible to calculate the CVaR of the return of the surplus wealth because the surplus wealth is not strictly positive. By the translation invariance property of a coherent risk measure, it is more reasonable to look at the change in surplus wealth:

$$\text{CVaR}_{\beta}(-sw_2) = sw_1 + \text{CVaR}_{\beta}(-\Delta sw),$$

since sw_1 is a fixed initial condition. Therefore, minimizing the CVaR in the next time period has the effect of minimizing the CVaR of the change, but one cannot still make a direct comparison because the asset wealth also varies for the different assumptions over the horizon. The measure of realized risk, $\tilde{\rho}$, used in the comparison is the CVaR at 80% confidence level of the change in negative surplus wealth per dollar of asset wealth from the previous period. One can expect that minimizing ρ_1 and ρ_2 produces small values of $\tilde{\rho}$, but $\tilde{\rho}$ does not give a perfect comparison of risk because the resulting optimal allocations depend on the ratios of assets to liabilities, not just the asset wealths. Values of $\tilde{\rho}$ and

	$\tilde{\rho}$	final sw
1-stage Normal	0.0466	1177.29
1-stage Stable	0.0509	1077.64
2-stage Normal	0.0456	1209.22
2-stage Stable	0.0491	1217.92
Fixed-Mixed 0/40/60/0/0	0.0924	241.04
Fixed-Mixed 0/100/0/0/0	0.0776	-371.39

Table 5.8: Dynamic backtesting results.

the final surplus wealth are found in table (5.8). For comparison, this table also includes values for the fixed-mixed rule of 40% bonds and 60% equity, and the rule of 100% in bonds. Under both the normal and stable assumptions, the 2-stage recourse problem outperforms the 1-stage problem by both reducing $\tilde{\rho}$ and increasing the final surplus wealth. While the 2-stage problem under the stable assumption results in the highest final surplus wealth, the normal assumption gave lower values of $\tilde{\rho}$. The fixed-mixed rules were no comparison with the stochastic programs.

Figures (5.9-5.11) show the evolution of the asset wealths and liability value over the time horizon. One can see that minimizing CVaR does not look like a typical index tracking problem because the upside is not penalized. The asset wealths and the liability values are in table (6.5), and the optimal allocations can be found in the appendix. These tables also include the percent of asset wealth loss to transaction costs.

An additional comparison of the performance of the stable and normal distributions can be obtained by VaR backtesting similar to section 5.1.2. Future material on this issue will be provided in the lecture.

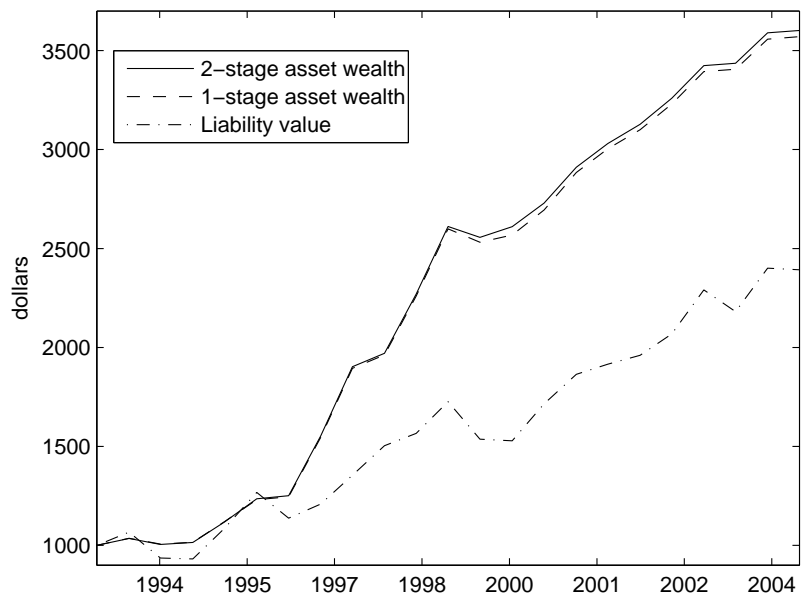


Figure 5.9: Dynamic backtesting: 1-stage versus 2-stage under the normal assumption.

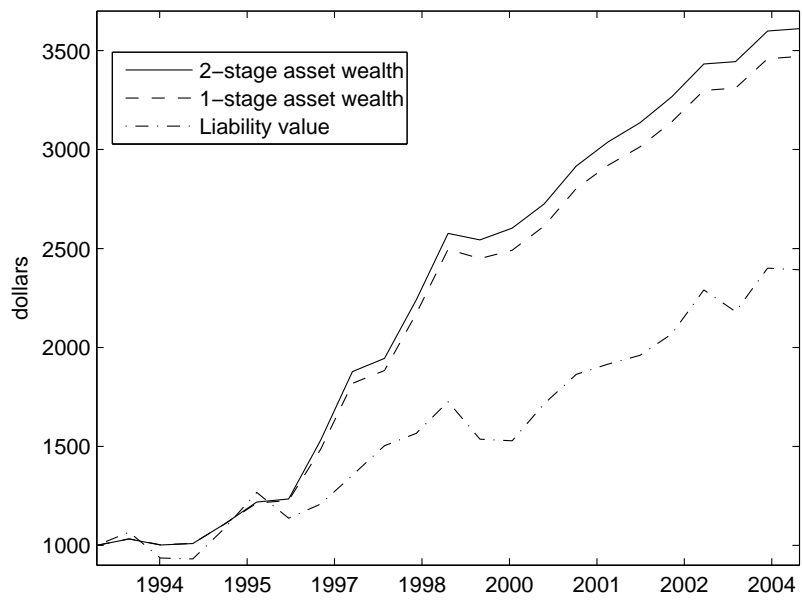


Figure 5.10: Dynamic backtesting: 1-stage versus 2-stage under the stable assumption.

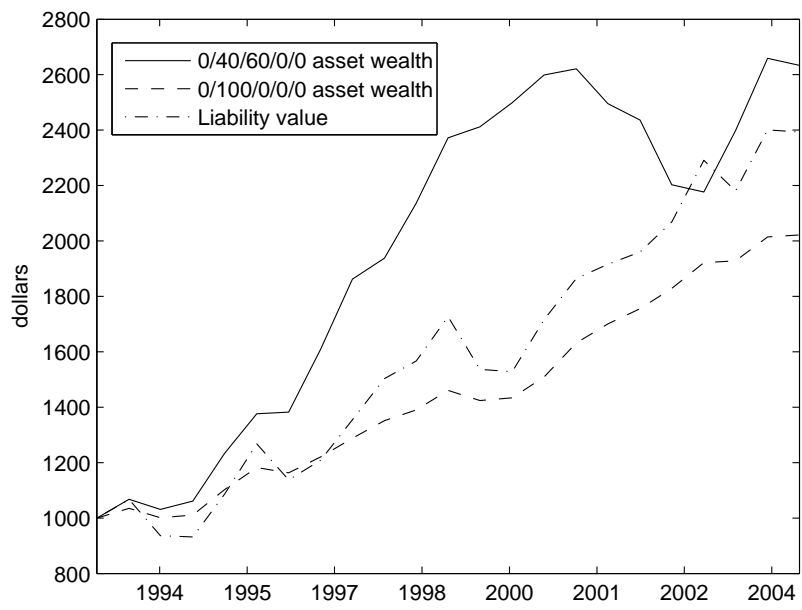


Figure 5.11: Dynamic backtesting: Fixed-mixed rules.

Chapter 6

Appendix - Tables for Empirical Analysis

					Optimal First Stage Allocations				
λ	$E(-sw_3)$	ρ_2	CVaR1	CVaR2	Cash	Bonds	Equities	Int. Eq.	Mortgages
0	72.18	399.46	319.44	479.48	0	0	0	100	0
0.10	71.68	389.07	319.44	458.70	0	0	0	100	0
0.20	70.82	384.04	319.44	448.63	0	0	0	100	0
0.25	64.49	364.44	299.09	429.80	0	10.6483	0	89.3517	0
0.30	42.92	306.50	232.28	380.71	0	48.3706	0	51.6294	0
0.35	32.63	284.80	207.23	362.36	0	64.9963	0	35.0037	0
0.40	25.61	273.09	194.71	351.47	0	75.1793	0	24.8207	0
0.45	19.75	265.12	186.68	343.57	0	82.9291	0	17.0709	0
0.50	15.08	259.96	182.16	337.76	0	88.1928	0	11.8072	0
0.60	6.72	253.12	177.31	328.92	0	95.7591	0	4.2409	0
0.75	-2.74	248.36	175.47	321.24	0	100	0	0	0
1.00	-20.23	246.13	175.47	316.79	0	100	0	0	0

Table 6.1: Efficient frontier under the normal assumption with $\beta = 0.95$.

					Optimal First Stage Allocations				
λ	$E(-sw_3)$	ρ_2	CVaR1	CVaR2	Cash	Bonds	Equities	Int. Eq.	Mortgages
0	70.47	409.60	321.91	497.29	0	0	0	100	0
0.10	70.08	401.18	321.91	480.45	0	0	0	100	0
0.20	69.36	396.97	321.91	472.03	0	0	0	100	0
0.25	68.92	395.45	321.91	468.99	0	0	0	100	0
0.30	57.09	365.74	288.65	442.84	0	20.8749	0	79.1251	0
0.35	43.60	337.37	256.10	418.64	0	44.3374	0	55.6626	0
0.40	34.79	322.43	239.12	405.73	0	58.8847	0	41.1153	0
0.45	27.67	312.73	228.49	396.97	0	69.8927	0	30.1073	0
0.50	21.83	306.25	221.75	390.75	0	78.1959	0	21.8041	0
0.60	13.85	299.59	216.06	383.12	0	87.1561	0	12.8439	0
0.75	2.97	294.23	212.78	375.68	0	95.7426	0	4.2574	0
1.00	-19.85	291.21	212.08	370.34	0	100	0	0	0

Table 6.2: Efficient frontier under the stable assumption with $\beta = 0.95$.

					Optimal First Stage Allocations					Trans.
λ	$E(-sw_3)$	ρ_2	CVaR1	CVaR2	Cash	Bonds	Equities	Int. Eq.	Mortgages	Costs
0	59.28	415.61	328.26	502.96	0	0	0	99.0050	0	0.9950
0.10	58.80	405.57	328.26	482.89	0	0	0	99.0050	0	0.9950
0.20	57.83	399.83	328.26	471.41	0	0	0	99.0050	0	0.9950
0.25	37.16	327.52	251.65	403.39	0	40.0000	0	59.4030	0	0.5970
0.30	34.81	321.78	247.18	396.37	0	42.6534	0	56.7496	0	0.5970
0.35	21.80	294.29	217.23	371.35	0	61.8604	0	37.5426	0	0.5970
0.40	13.93	281.12	203.63	358.61	0	72.4813	0	26.9217	0	0.5970
0.45	7.14	271.91	194.45	349.36	0	81.0294	0	18.3736	0	0.5970
0.50	2.07	266.31	189.48	343.14	0	86.5774	0	12.8256	0	0.5970
0.60	-6.25	259.47	184.36	334.58	0	94.1698	0	5.2331	0	0.5971
0.75	-15.27	254.86	181.98	327.75	0	99.4030	0	0	0	0.5970
1.00	-31.96	253.21	181.98	324.45	0	99.4030	0	0	0	0.5970

Table 6.3: Efficient frontier under the normal assumption with transaction costs and $\beta = 0.95$.

					Optimal First Stage Allocations				
λ	$E(-sw_3)$	ρ_2	CVaR1	CVaR2	Cash	Bonds	Equities	Int. Eq.	Mortgages
0	72.18	399.46	319.44	479.48	0	0	0	100	0
0.25	56.24	349.48	271.26	427.70	0	25.7090	0	74.2910	0
0.50	8.00	293.29	214.21	372.38	0	0	0	20.8588	79.1412
0.75	-40.48	333.07	247.64	418.49	74.2995	0	0	3.9844	21.7161
1.00	-53.29	353.66	253.02	454.30	84.9059	1.0005	0	2.4420	11.6516

Table 6.4: Wealth optimization under the normal assumption with no transaction costs and $\beta = 0.95$.

		Asset Wealth					
Date	Liability Value	1-stage		2-stage		Fixed-mixed	
		Normal	Stable	Normal	Stable	0/40/60/0/0	0/100/0/0/0
7/93	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00
1/94	1067.19	1036.08	1033.60	1034.93	1032.71	1067.69	1034.72
7/94	936.20	1006.58	1003.17	1005.07	1002.12	1031.23	1000.94
1/95	932.01	1015.83	1009.89	1014.77	1009.76	1061.09	1010.78
7/95	1085.82	1119.68	1106.91	1118.45	1108.27	1233.43	1102.13
1/96	1267.52	1230.57	1213.50	1235.10	1219.35	1376.50	1182.06
7/96	1137.78	1245.16	1227.06	1250.82	1234.14	1382.09	1163.18
1/97	1208.93	1545.82	1486.14	1552.85	1532.14	1609.50	1220.62
7/97	1354.74	1894.36	1818.87	1902.97	1877.59	1862.24	1288.37
1/98	1503.39	1961.80	1883.63	1970.72	1944.43	1937.84	1351.51
7/98	1566.35	2259.70	2169.65	2269.97	2239.69	2136.26	1389.74
1/99	1726.53	2599.19	2495.61	2611.01	2576.17	2371.80	1460.65
7/99	1536.38	2531.37	2448.79	2555.91	2543.31	2411.78	1424.34
1/00	1528.22	2568.31	2490.92	2610.18	2603.17	2498.60	1433.68
7/00	1715.38	2694.08	2612.77	2729.57	2724.48	2599.06	1509.31
1/01	1864.70	2884.14	2803.19	2910.55	2916.84	2621.32	1631.85
7/01	1915.56	3003.98	2920.14	3030.85	3038.17	2495.37	1700.87
1/02	1960.23	3100.08	3013.55	3127.81	3135.36	2436.33	1755.28
7/02	2068.61	3230.29	3140.12	3259.18	3267.05	2202.77	1829.00
1/03	2291.06	3393.44	3298.72	3423.79	3432.06	2176.37	1921.38
7/03	2180.78	3405.28	3310.23	3435.73	3444.03	2398.13	1928.08
1/04	2400.67	3558.11	3458.80	3589.93	3598.60	2659.32	2014.61
7/04	2392.77	3570.05	3470.40	3601.98	3610.68	2633.80	2021.38

Table 6.5: Dynamic backtesting: Realized liability value and asset wealths for the optimal allocations with $\beta = 0.80$.

Date	Cash	Bonds	Equities	Intern. Equities	Mortgages	Transaction Costs
(initial)	0	40	60	0	0	
7/93	0	89.9043	4.9691	4.5791	0	0.5476
1/94	0	89.7865	0.1678	9.9954	0	0.0503
7/94	0	85.6719	9.9092	4.3211	0	0.0979
1/95	0	89.7313	10.2284	0	0	0.0403
7/95	0	58.9767	40.7269	0	0	0.2964
1/96	0	0	99.4273	0	0	0.5727
7/96	0	0	100	0	0	0
1/97	0	0	100	0	0	0
7/97	0	0	100	0	0	0
1/98	0	0	100	0	0	0
7/98	0	0	100	0	0	0
1/99	0	88.4698	10.6410	0	0	0.8891
7/99	0	82.3947	17.5437	0	0	0.0616
1/00	0	81.7418	18.2582	0	0	0
7/00	0	91.9907	7.9093	0	0	0.1000
1/01	0	99.9294	0	0	0	0.0706
7/01	0	100	0	0	0	0
1/02	0	100	0	0	0	0
7/02	0	100	0	0	0	0
1/03	0	100	0	0	0	0
7/03	0	100	0	0	0	0
1/04	0	100	0	0	0	0

Table 6.6: Dynamic backtesting: Allocations (as a percent of asset wealth) for the 1-stage optimization problem under the normal assumption with $\beta = 0.80$.

Date	Cash	Bonds	Equities	Intern. Equities	Mortgages	Transaction Costs
(initial)	0	40	60	0	0	
7/93	0	91.4944	3.7318	4.2140	0	0.5599
1/94	0	90.9029	0	9.0523	0	0.0448
7/94	0	86.5537	9.8023	3.5455	0	0.0985
1/95	0	89.8535	10.1135	0	0	0.0330
7/95	0	50.4015	49.2155	0	0	0.3830
1/96	0	0	99.5129	0	0	0.4871
7/96	0	0	100	0	0	0
1/97	0	0	100	0	0	0
7/97	0	0	100	0	0	0
1/98	0	0	100	0	0	0
7/98	0	0	100	0	0	0
1/99	0	82.2585	16.9148	0	0	0.8267
7/99	0	67.1027	32.7497	0	0	0.1477
1/00	0	66.1382	33.8618	0	0	0
7/00	0	89.5188	10.2507	0	0	0.2305
1/01	0	99.9081	0	0	0	0.0919
7/01	0	100	0	0	0	0
1/02	0	100	0	0	0	0
7/02	0	100	0	0	0	0
1/03	0	100	0	0	0	0
7/03	0	100	0	0	0	0
1/04	0	100	0	0	0	0

Table 6.7: Dynamic backtesting: First stage allocations (as a percent of asset wealth) for the 2-stage optimization problem under the normal assumption with $\beta = 0.80$.

Date	1-stage		2-stage			
	Normal	Stable	Normal		Stable	
	CVaR	CVaR	CVaR1	CVaR2	CVaR1	CVaR2
7/93	139.97	118.33	139.99	248.58	118.35	211.53
1/94	165.24	165.42	166.51	286.72	166.45	285.05
7/94	13.25	24.93	14.91	77.33	26.09	95.20
1/95	-21.85	-2.31	-20.81	25.29	-2.22	52.64
7/95	90.97	112.84	92.62	202.52	111.41	229.59
1/96	165.43	199.19	159.34	217.74	192.03	262.59
7/96	17.94	53.73	11.81	89.41	46.62	139.86
1/97	-294.58	-201.09	-302.18	-335.87	-253.20	-259.12
7/97	-530.24	-415.32	-539.84	-622.72	-481.01	-531.86
1/98	-343.84	-223.30	-353.58	-327.22	-290.62	-217.16
7/98	-603.38	-471.40	-614.36	-624.62	-545.81	-522.57
1/99	-682.94	-569.73	-693.82	-572.10	-652.59	-544.71
7/99	-914.55	-825.38	-937.53	-938.14	-921.85	-929.62
1/00	-971.35	-876.49	-1009.61	-1015.10	-987.07	-971.63
7/00	-839.05	-731.40	-871.71	-770.64	-847.11	-725.54
1/01	-911.81	-782.59	-939.38	-834.34	-904.92	-756.63
7/01	-986.08	-850.41	-1014.32	-934.61	-975.67	-848.00
1/02	-1053.78	-948.54	-1082.97	-1014.99	-1077.36	-991.05
7/02	-1032.56	-886.39	-1063.62	-965.99	-1023.88	-879.87
1/03	-899.44	-761.41	-931.94	-768.19	-904.66	-706.25
7/03	-1013.32	-881.03	-1046.02	-906.55	-1024.45	-858.39
1/04	-933.25	-832.85	-967.31	-802.10	-982.10	-820.57

Table 6.8: Dynamic backtesting: Optimal values of $CVaR_{0.80}$ for scenarios generated under the normal and stable assumptions.

Bibliography

- [1] Carlo Acerbi and Dirk Tasche. Expected shortfall: A natural coherent alternative to value at risk. <http://www.gloriamundi.org>, May 2001.
- [2] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, July 2002.
- [3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, July 1999.
- [4] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer, New York, 1997.
- [5] Brendan O. Bradley and Murad S. Taqqu. Financial risk and heavy tails. In Svetlozar T. Rachev, editor, *Handbook of heavy tailed distributions in finance*, Handbooks In Finance. Elsevier Science, The Netherlands, 2003.
- [6] John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey, 1997.
- [7] David R. Carino, David H. Myers, and William T. Ziemba. Concepts, technical issues, and uses of the russell-yasuda kasai financial planning model. *Operations Research*, 46(4):450–462, July-August 1998.

- [8] Yair Censor and Stavros A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 1997.
- [9] John M.R. Chalmers, Roger M. Edelen, and Gregory B. Kadlec. Transaction-cost expenditures and the relative performance of mutual funds. Technical report, The Wharton School, University of Pennsylvania, November 1999.
- [10] M.A.H. Dempster and R.T. Thompson. Evpi-based importance sampling solution procedures for multistage stochastic linear programmes on parallel mimd architectures. *Annals of Operations Research*, 90:161–184, 1999.
- [11] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming stochastic programming: An approach using probability metrics. *Mathematical Programming*, 94:493–511, 2003.
- [12] Jitka Dupačová, Giorgio Consigli, and Stein W. Wallace. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100:25–53, 2000.
- [13] Jitka Dupačová, Jan Hurt, and Josef Štěpán. *Stochastic Modeling in Economics and Finance*. Kluwer Academic Publishers, Netherlands, 2002.
- [14] A. Eichhorn and W. Rmisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.
- [15] P. Embrechts, A. McNeil, and D. Straumann. *Correlation and Dependence in Risk Management: Properties and Pitfalls*. 1999. In: *Risk management: value at risk and beyond*, ed. Dempster, M.
- [16] Nicole Gröwe-Kuska, Holger Heitsch, and Werner Römisch. Scenario reduction and scenario tree construction for power management problems. Technical report, Institute of Mathematics, Humboldt-University Berlin.

- [17] Holger Heitsch and Werner Römisch. Generation of multivariate scenario trees to model stochasticity in power management. Technical report, Institute of Mathematics, Humboldt-University Berlin.
- [18] Pavlo Krokmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk*, 4(2), 2002.
- [19] Fabio Lamantia, Sergio Ortobelli, and Svetlozar Rachev. Value at risk with stable distributed returns. Available at <http://www.pstat.ucsb.edu/research>.
- [20] C. D. Maranas, I. P. Androulakis, C. A. Floudas, A. J. Berger, and J. M. Mulvey. Solving long-term financial planning problems via global optimization. *Journal of Economic Dynamics and Control*, 21:1405–1425, 1997.
- [21] J.P. Morgan. Riskmetrics - Technical Document. Technical report, December 1996.
- [22] John Mulvey and Hercules Vladimirou. Stochastic optimization models for investment planning. *Annals of Operations Research*, 20:187–217, 1989.
- [23] John Mulvey and Hercules Vladimirou. Applying the progressive hedging algorithm to stochastic generalized networks. *Annals of Operations Research*, 31:399–424, 1991.
- [24] John Mulvey and Hercules Vladimirou. Solving multistage stochastic networks: An application of scenario aggregation. *Networks*, 21(6):619–643, 1991.
- [25] John Mulvey and Hercules Vladimirou. Stochastic network programming for financial planning problems. *Management Science*, 38(11):1642–1664, Nov. 1992.

- [26] Soren S. Nielsen and Stravos A. Zenios. Solving multistage stochastic network programs on massively parallel computers. *Mathematical Programming*, 73:227–250, 1996.
- [27] Sergio Ortobelli, Isabella Huber, Svetlozar Rachev, and Eduardo Schwartz. Portfolio choice theory with non-gaussian distributed returns.
- [28] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271, 2001.
- [29] Georg Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In Stanislav Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer Academic Publishers, 2002.
- [30] Svetlozar Rachev and Stefan Mittnik. *Stable Paretian Models in Finance*. Series in Financial Economics and Quantitative Analysis. John Wiley & Sons Ltd., 2000.
- [31] Svetlozar Rachev, Eduardo Schwartz, and Irina Khindanova. Stable modeling of market and credit value at risk. Technical report, University of California, Santa Barbara, October 2001. Available at <http://www.pstat.ucsb.edu/research/papers>.
- [32] Svetlozar T. Rachev, Eduardo S. Schwartz, and Irina Khindanova. Stable modeling of market and credit value at risk. In Svetlozar T. Rachev, editor, *Handbook of heavy tailed distributions in finance*, Handbooks In Finance. Elsevier Science, The Netherlands, 2003.
- [33] R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

- [34] Ronald J. Ryan and Frank J. Fabozzi. The pension crisis revealed. *The Journal of Investing*, Fall, 2005.
- [35] B Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland Elsevier, New York, 1983.
- [36] Nikolas Topaloglou, Hercules Vladimirov, and Stavros A. Zenios. Cvar models with selective hedging for international asset allocation. *Journal of Banking and Finance*, 26:1535–1561, 2002.
- [37] Stavros A. Zenios. Lecture notes: Mathematical modeling and its application in finance. <http://www.hermes.ucy.ac.cy/zenios/teaching/399.001/index.html>.
- [38] Eric Zivot and Jianhui Wang. *Modeling Financial Time Series with S-Plus*. Springer, New York, 2003.