Abstract

Application of Stochastic Programming and Stable Distributions to Asset Liability Management

by

Michael J. Grebeck

Multistage stochastic programming methods are applied to portfolio optimization in the context of asset liability management. These methods can incorporate rebalancing decisions and transaction costs to find optimal investment strategies over a time horizon. Additionally, assets and liabilities are considered in the same risk-reward optimization problem, thereby taking advantage of common risk factors.

The specific problem examined is that of a pension fund. The allocations are found among various asset classes that optimize a tradeoff between the risk and the expected final surplus wealth. A weighted average of the Conditional Value at Risk of the negative surplus wealth over the time horizon is used as the multiperiod measure of risk. This particular risk measure permits a formulation of the problem that has a convex, piecewise linear objective and linear constraints. A decomposition procedure in the solution method allows parallel implementation.

Uncertainty is represented through a scenario tree, resulting in a very large deterministic formulation of the stochastic program. The scenario generation procedures attempt to produce representative discrete distributions that will result in good decisions. The scenarios are generated from two multivariate time series models that incorporate volatility clustering: The first assumes the innovations are normal, and the second assumes the innovations are stable. Value at Risk

backtesting of the time series models rejects the normality assumption and shows the superiority of the stable assumption.

Efficient frontiers for the 2-stage problem are found under both distributional assumptions. Backtesting of the minimum risk portfolios is carried out to compare the performance of the 1-stage problem with the 2-stage recourse problem and the normal distribution with the stable distribution. By computing the risk of the realized surplus wealths resulting from the optimal allocations, it is shown that the 2-stage recourse problem outperforms the 1-stage problem in dynamic backtesting with transaction costs. Portfolio backtesting in a static setting without transaction costs provides a better comparison of the distributional assumptions and shows that the stable assumption produces a smaller realized risk.

Contents

List of Figures

List of Tables

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, these 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.

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 payment 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.

Both of these earlier frameworks are inadequate for ALM because they lack

the stochastic nature of interest rates and liabilities and the dynamic nature of investing. The two main tools that help capture these 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. Merton's classical consumption and investment problem in [16] is one successful application of stochastic control, and there are extensions of this problem to include a stochastic interest rate in [39]. The application of stochastic control to ALM has been limited so far, but an exception is surplus optimization for pension funds and life insurance in [54]: It is shown that an investor holds a combination of the riskless asset, the market portfolio, a hedge portfolio for the stochastic interest rate, and also a hedge portfolio for the liabilities.

An advantage of stochastic control over stochastic programming is that there is no need to approximate uncertainty with scenarios; however, there are many components of ALM, such as legal constraints, policy constraints, and transaction costs, that the stochastic control methods cannot handle. The remainder of this paper is devoted to stochastic programming techniques that can include such components in a direct manner.

A general definition of stochastic programming is mathematical optimization with uncertain parameters. These problems typically rely on uncertainty approximated by a set of scenarios, often in the form of a scenario tree. For example, a straightforward extension of dedication that includes scenarios for the liabilities, cashflows, and reinvestment rates results in a stochastic linear program in [20]. As another simple example, the scenario immunization technique in [14] finds a cheap portfolio that keeps the present value of the assets close to the present value of the liabilities under all scenarios. However, the dynamic nature of both of these

examples is limited because neither consider rebalancing decisions.

Multistage stochastic programming with recourse offers a general framework that can directly incorporate portfolio rebalancing and linear transaction costs. Two of the early successes in ALM, the Kusy-Ziemba model and the Russell-Yasuda Kasai model, illustrate the potential:

- The Kusy-Ziemba model in [31] is a 2-stage stochastic linear program that includes many various features: changing yield spreads over time, synchronization of cash flows by matching maturities of assets with expected cash outflows, simultaneous consideration of assets and liabilities to satisfy accounting principles and match liquidities, transaction costs, uncertainty in withdrawal claims and deposits, uncertainty in interest rates, and legal constraints. The stochastic program finds optimal decisions for asset allocations, deposit holdings, and borrowing throughout five years. Even with a very limited set of scenarios, the recourse model produces results superior to those of other models.

- The Kasai model found in [9, 11] is a 6-stage linear stochastic program for a Japanese insurance company. The model is designed to handle multiple investment and liability accounts while meeting regulations imposed by Japanese insurance laws. The goal of the stochastic program is to produce a high-income return to pay the interest on savings policies while also maximizing the wealth of the firm at a distant horizon. The model finds the best allocation strategy over time of the firm's funds to asset classes such as cash, loans, bonds, equities, and real estate. Implementation of this model proves to be a success over previous mean-variance methods used by the insurance company.

Stochastic programming is becoming more popular in finance as computing

power increases. While multistage stochastic programs with recourse can more adequately model dynamic and stochastic financial problems, they are much more of a computational challenge, and simplifications are sometimes needed to make the problems implementable. One popular approach in finance to ease the computational burden is to use decision rules, for instance as in [56]. Decision rules can greatly reduce the decision space but create non-convex optimization problems.

A scenario generation procedure is needed to construct the set of scenarios for the stochastic program that will result in good decisions. There are many different techniques in scenario generation, and most of them rely on a distributional assumption for a data set. Since there is empirical evidence of heavy tails, high peaks, skewness, long-range dependence, and volatility clustering in financial data, it is desirable to include these characteristics when constructing the scenarios. Special attention is given to time series models with stable distributions in this dissertation.

This dissertation is organized as follows:

- Chapter 2 reviews relevant material on risk and optimization, mostly in a single-stage context. Properties of the Conditional Value at Risk (CVaR) risk measure are covered, and its use in stochastic programming is illustrated.

- Chapter 3 discusses general multistage stochastic programs with recourse and sets up a program that is applied to pension fund data. It is a multi-objective program that maximizes the expected final surplus wealth and minimizes a multi-period risk measure while considering portfolio rebalancing and transaction costs.

- Chapter 4 describes the pension fund data and fits multivariate time series models that account for volatility clustering using the normal distribution

and a stable distribution. Value at Risk backtesting provides a comparison of the models.

- Chapter 5 solves the deterministic equivalent of the ALM problem in Chapter 3 with the scenarios generated from the time series model in Chapter 4. Portfolio backtesting is carried out to compare the 2-stage problem with the 1-stage problem and the normality assumption with the stable assumption.

Chapter 2

Risk and Optimization

The goal of risk-return optimization is to optimize the 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 are later used in a multistage problem in Chapter 3.

2.1 Risk Measures

The standard measure of risk for a portfolio of equities suggested by Markowitz in [38] is the *variance* of the return. A portfolio consists of weights $\omega = (\omega_1, ..., \omega_n)'$, such that $\omega_i \geq 0$ and $\sum_{i=1}^{n} \omega_i = 1$, in n assets with corresponding risky returns $R = (r_1, ..., r_n)'$. 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*:

$$\mathrm{E}\left([\omega'\mathrm{E}(R) - \omega' R]^+\right)^2.$$

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

$$\mathrm{E}\left([r^* - \omega' R]^+\right)^2.$$

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* of the portfolio

$$m_p = \mathrm{E}\left|\omega' R - \omega'\mathrm{E}(R)\right|,$$

or alternatively, the scale parameter of a stable distribution. Stable distributions are discussed in more detail in Chapter 4.

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 of $100\beta\%$ with $\beta \in (0, 1)$, 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 [53]. 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 $\Psi_L(x,\zeta)$ denote the cumulative distribution function for $L(x)$, i.e.,

$$\Psi_L(x,\zeta) = \mathrm{P}\left(L(x) \leq \zeta\right).$$

For a given decision x, the *Value at Risk* at the $100\beta\%$ confidence level is given by

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

While it is not widely used in finance, the Conditional Value at Risk (CVaR) has properties that make it a very logical alternative to VaR. These properties, referred to as *coherence*, are described in the next section.

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

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

For a given decision x, the *Conditional Value at Risk* at the $100\beta\%$ confidence level is the mean of the tail random variable $T_\beta(x)$ with distribution function (2.1):

$$\mathrm{CVaR}_\beta(x) = \mathrm{E}\left(T_\beta(x)\right).$$

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

$$\mathrm{E}\left(L(x)|L(x) \geq \mathrm{VaR}_\beta(x)\right) \leq \mathrm{CVaR}_\beta(x) \leq \mathrm{E}\left(L(x)|L(x) > \mathrm{VaR}_\beta(x)\right). \tag{2.2}$$

If there is no discontinuity in the distribution function of $L(x)$ at $\mathrm{VaR}_\beta(x)$, 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 [53], $\text{CVaR}_\beta(x)$ 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 when the distribution is described by a set of scenarios. This representation is referred to as Uryasev's formula and is reviewed in Section 2.4.

A discussion of several other tail risk measures can be found in [1] and [2].

2.2 Coherence

To help define a sensible risk measure, Artzner et al. introduce properties that are required of a *coherent* risk measure in [3]; 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 the individual VaRs. Lack of sub-additivity is very undesirable because diversification is not promoted. However, for the special class of elliptical distributions, VaR is sub-additive and coherent (see [6]).

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 $\rho : V \longrightarrow \mathbb{R}$. For two random variables $v, v' \in V$ that are thought of as losses, ρ is coherent if it is

i. sub-additive: $\rho(v + v') \leq \rho(v) + \rho(v')$,

ii. positive homogeneous: $\rho(\lambda v) = \lambda\rho(v), \quad \forall\lambda \geq 0$,

iii. translation invariant: $\rho(v + c) = \rho(v) + c, \quad \forall c \in \mathbb{R}$, and

iv. monotonous: $\rho(v) \geq 0, \quad \forall v \geq 0$.

Proof of the coherence of CVaR can be found in, for instance, [2, 47, 53]. The coherence of the set of random variables $\{L(x)\}$ can be stated as a function of x when $L(x)$ is linear:

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

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 $\text{CVaR}_\beta(x)$ is 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 [53] for a proof.

Note that sub-additivity and positive homogeneity guarantee 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. This dissertation also considers 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 have 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:

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

The solution to the above problem is easily obtained with Lagrangian techniques and can be found in [8]. As μ_0 is varied, the set of optimal portfolios trace out the mean-variance efficient frontier. If short selling is not 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 [58], is to model each asset with a multifactor equation:

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

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

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

It follows that 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 converts the mean-variance optimization problem to

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

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 mean-variance problems are quadratic optimization programs. As an alternative to mean-variance analysis, one can optimize the risk measures mentioned in Section 2.1. Also in [58], the author illustrates that a linear optimization problem can be constructed 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 produces 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 when minimizing VaR or CVaR. A very brief review follows; a more complete introduction to elliptical distributions and their portfolio implications is found in [6]. 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 $\tilde{\sigma}_p^2$, then

$$\rho\left(r_p - \mathrm{E}(r_p)\right) \leq \rho\left(\tilde{r}_p - \mathrm{E}(\tilde{r}_p)\right) \quad \Longleftrightarrow \quad \sigma_p^2 \leq \tilde{\sigma}_p^2.$$

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

$$\begin{array}{lll} \min\limits_{\omega} & \sigma_p^2 & \\ \text{s.t.} & r_p = \omega' R, & \\ & \mathrm{E}(r_p) = \mu_0, & \\ & \sum_{i=1}^n \omega_i = 1, & \end{array} \qquad \begin{array}{lll} \min\limits_{\omega} & \rho(r_p) & \\ \text{s.t.} & r_p = \omega' R, & \\ & \mathrm{E}(r_p) = \mu_0, & \\ & \sum_{i=1}^n \omega_i = 1, & \end{array}$$

where μ_0 is the desired return. Therefore, under this distributional assumption, minimization of VaR, CVaR, or variance will all 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 the variance is infinite and cannot be used as a risk measure. One natural solution is to use the scale parameter of the portfolio return. The scale parameter is just a generalization of the standard deviation of a normal distribution. Chapter 4 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 VaR and CVaR 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 and CVaR can both be achieved by the portfolio optimization problem

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

Details of stable portfolio theory are found in [49], and a comparison of allocations under the normality and stable assumptions is found in [48].

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 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 (see [57]). CVaR, on the other hand, has a representation that is easy to optimize for a set of scenarios both as a constraint and in an objective function. 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

$$\Gamma_\beta(x, \zeta) = \zeta + \frac{1}{1-\beta}\text{E}\left([L(x) - \zeta]^+\right). \tag{2.5}$$

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

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

and, in addition,

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

Equation (2.6) is referred to as *Uryasev's formula* in this dissertation. 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, the following optimization shortcut gives 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.,

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

Proofs of the above results are found in [53].

Similar to mean-variance efficient frontiers, [30] illustrates risk-reward analysis using CVaR as a risk measure. If $R(x)$ is a concave reward function and X is

convex, then

$$\min_{x\in X} \mathrm{CVaR}_\beta(x) \quad \text{subject to} \quad R(x) \geq \lambda, \tag{2.7}$$

$$\min_{x\in X} \mathrm{CVaR}_\beta(x) - \lambda R(x), \quad \text{and} \tag{2.8}$$

$$\min_{x\in X} -R(x) \quad \text{subject to} \quad \mathrm{CVaR}_\beta(x) \leq \lambda, \tag{2.9}$$

produce the same efficient frontiers as λ is varied. As is already shown, the optimal solution to (2.7) can be found by a jointly convex optimization problem. Similarly, problems (2.8) and (2.9) produce the same optimal solutions 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 [53]. If the CVaR at the $100\beta_i\%$ confidence level has a loss tolerance equal to λ_i, for $i = 1, ...I$, then

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

has the same optimal solution as

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

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

tree or sampling, equation (2.5) becomes

$$\tilde{\Gamma}_\beta(x,\zeta) = \zeta + \frac{1}{1-\beta}\sum_{s=1}^{S} p^s \left[L^s(x) - \zeta\right]^+,$$

where $L(x)$ takes the value $L^s(x)$ with probability p^s for $s = 1, ..., S$. 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.

2.4.2 1-Stage Portfolio Optimization

This subsection applies Uryasev's formula to risk-return analysis with CVaR and obtains a linear programming problem. Define

$$X = \left\{ \omega \in \mathbb{R}^n \,\middle|\, \sum_{j=1}^{n} \omega_j = 1, \omega_j \geq 0, j = 1, ..., n \right\},$$

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 $R = (r_1, ..., r_n)'$, and the negative return of the portfolio is given by

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

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

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

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 (2.7). If

the uncertainty in the return is given through the set of scenarios $\{R^1, ..., R^S\}$ where each $R^s \in \mathbb{R}^n$ occurs with probability p^s, Uryasev's optimization shortcut produces the equivalent problem

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

By introducing auxiliary variables y^s, $s = 1, ..., S$, a linear program results:

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

This program is used to compare hedging strategies for international asset allocation in [57]. In addition, the CVaR portfolio is compared with a portfolio minimizing the mean absolute deviation (MAD). 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.

2.5 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 the 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 addressed 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 [37]:

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

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 $w = (w_1, ..., w_T)$ where w_1 is a known deterministic wealth. It is shown in [21] that a weighted average of CVaR over the time horizon is coherent: If $\mathrm{CVaR}_\beta(-w_t)$ is the CVaR associated with the negative wealth $-w_t$, then a coherent risk measure is given by

$$\rho(w) = \rho(w_1, ..., w_T) = \sum_{t=2}^{T} \mu_t \mathrm{CVaR}_\beta(-w_t), \tag{2.10}$$

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

i. convex: $\rho(\lambda w + (1-\lambda)\tilde{w}) \leq \lambda\rho(w) + (1-\lambda)\rho(\tilde{w}), \quad \forall\lambda \in [0,1]$,

ii. positive homogeneous: $\rho(\lambda w) = \lambda\rho(w), \quad \forall\lambda \geq 0$,

iii. translation invariant: $\rho(w_1 + c, ..., w_T + c) = \rho(w) - c, \quad \forall c \in \mathbb{R}$, and

iv. monotonous: if $w_t \leq \tilde{w}_t$ a.s. for $t = 1, ..., T$, then $\rho(w) \geq \rho(\tilde{w})$.

When implementing the risk measure in equation (2.10), one can use Uryasev's optimization shortcut in a similar manner as the previous section: 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 are also 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.

Chapter 3

Formulation of the Stochastic Program

Stochastic programming offers a framework that can incorporate many of the characteristics of an ALM problem. This chapter first discusses a general setup for stochastic programs with recourse and then applies this framework to an ALM problem for a pension fund.

3.1 General Recourse Problems

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

$$\begin{aligned} \min_{x_2} \quad & q_2(x_1, x_2, \xi_1) + \mathrm{E}_{\xi_2}\left(Q_2(x_1, x_2, \xi_1, \xi_2) | \, \xi_1\right) \\ \text{s.t.} \quad & B_2(\xi_1)x_1 + A_2(\xi_1)x_2 = b_2(\xi_1), \\ & l_2(\xi_1) \leq x_2 \leq u_2(\xi_1), \end{aligned} \tag{3.1}$$

where

- $q_2(x_1, x_2, \xi_1)$ is the 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 the 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 (3.1) 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 $\mathrm{E}Q_1(x_1, \xi_1)$:

$$\begin{array}{ll} \min_{x_1} & q_1(x_1) + \mathrm{E}Q_1(x_1, \xi_1) \\ \text{s.t.} & A_1 x_1 = b_1, \\ & l_1 \leq x_1 \leq u_1. \end{array} \tag{3.2}$$

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, $t = 1, ..., T-1$, and the decisions vectors x_t, $t = 1, ..., T$. 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 $t = 1, ..., T$, be denoted by $\xi^t = \{\xi_j, j = 1, ..., t\}$, where each ξ_j is the uncertainty realized in stage j. Similarly, let the decisions up to and including stage t be denoted by $x^t = \{x_j, j = 1, ..., t\}$, where each x_j is the decision made for stage j. The first

stage problem is essentially the same as problem (3.2):

$$\begin{aligned} \min_{x_1} \quad & q_1(x_1) + \mathrm{E}_{\xi_1} Q_1(x^1, \xi^1) \\ \text{s.t.} \quad & A_1 x_1 = b_1, \\ & l_1 \leq x_1 \leq u_1, \end{aligned} \tag{3.3}$$

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

$$\begin{aligned} Q_t(x^t, \xi^t) = \quad \min_{x_{t+1}} \quad & q_{t+1}(x^{t+1}, \xi^t) + \mathrm{E}_{\xi_{t+1}} \left(Q_{t+1}(x^{t+1}, \xi^{t+1}) | \, \xi^t \right) \\ \text{s.t.} \quad & B_{t+1}(\xi^t) x_t + A_{t+1}(\xi^t) x_{t+1} = b_{t+1}(\xi^t), \\ & l_{t+1}(\xi^t) \leq x_{t+1} \leq u_{t+1}(\xi^t), \end{aligned} \tag{3.4}$$

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 (3.3-3.4) is a form of the multistage recourse program that is relevant to the ALM problem that is presented in Section 3.4. Other forms, such as those found in [20], allow the first constraint of (3.4) to depend on all decisions up to t:

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

but this type of constraint is not necessary in this dissertation.

3.2 Scenario Generation

To numerically solve the recourse problem (3.3-3.4), the distribution of $(\xi_1, ..., \xi_T)$ 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

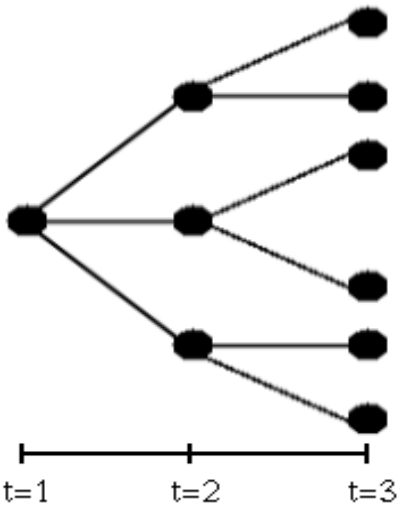

Figure 3.1: A balanced scenario tree.

that is used in the 2-stage ALM problem in Chapter 5. 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 are no additional allocation decisions made in 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 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 $\mathcal{I}_t = \{I_{t-1}+1, ..., I_t\}$, for $t = 2, ..., T+1$, with $I_1 = 1$. A scenario s, which is a path through the scenario tree, is then represented by the set of indices $(i_2, ..., i_{T+1})$ 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 $\sum_{i_{t+1} \in dec(i_t)} p(i_t, i_{t+1}) = 1$.

A topic of active research examines how to generate a 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 an importance sampling criterion are obtained at each node. These nodal values determine 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 [19]. As an example, the importance sampling criterion used in [15] 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. The work of Dupačová [20] presents a moment matching model for a two-dimensional random vector where the first and second random variables may represent the first and second stage uncertainties, 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, Dupačová suggests a weighted least squares minimization problem.

As an alternative to moment matching, the discretization technique of Pflug in [46] relies on the minimization of a transportation metric 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 approximation error, and Pflug 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 [10]. A second approach involving probability metrics is found in [17] and [25]: 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 uncertainty, 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 [20]: 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 paths. 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 [24].

3.3 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 determines realizations 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 (3.3-3.4) can then be written as

$$\begin{array}{ll} \min_{x_1} & q_1(x_1) + \sum_{i_2 \in \mathcal{I}_2} p(i_2) Q_{i_2}(x^1) \\ \text{s.t.} & A_1 x_1 = b_1, \\ & l_1 \leq x_1 \leq u_1, \end{array} \tag{3.5}$$

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

$$\begin{aligned} Q_{i_t}(x^{t-1}) = \quad & \min_{x_t} \quad q_{i_t}(x^t) + \textstyle\sum_{i_{t+1} \in dec(i_t)} p(i_t, i_{t+1}) Q_{i_{t+1}}(x^t) \\ & \text{s.t.} \quad B_{i_t} x_{t-1} + A_{i_t} x_t = b_{i_t}, \\ & \qquad l_{i_t} \leq x_t \leq u_{i_t}, \end{aligned} \tag{3.6}$$

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

The above (3.5-3.6) is the form of the recourse program that is relevant when the solution method for the ALM problem is discussed in Section 3.4.4; 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 is shown in the next section, the ALM problem has 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 function $q_{i_t}(\cdot)$ takes a linear form:

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

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

The deterministic equivalent for the linear program in arborescent form carefully considers the structure of the scenario tree:

$$\min \qquad \langle q_1, x_1 \rangle + \sum_{i_2 \in \mathcal{I}_2} p(i_2) \langle q_{i_2}, x_{i_2} \rangle + \cdots + \sum_{i_T \in \mathcal{I}_T} p(i_T) \langle q_{i_T}, x_{i_T} \rangle$$

subject to

$$
\begin{aligned}
& A_1 x_1 && && && = b_1, && \\
& B_{i_2} x_1 + A_{i_2} x_{i_2} && && && = b_{i_2}, && \forall i_2 \in \mathcal{I}_2, \\
& \qquad B_{i_3} x_{pred(i_3)} + A_{i_3} x_{i_3} && && && = b_{i_3}, && \forall i_3 \in \mathcal{I}_3, \\
& && && && \vdots && \\
& && && B_{i_T} x_{pred(i_T)} + A_{i_T} x_{i_T} && = b_{i_T}, && \forall i_T \in \mathcal{I}_T, \\
& l_{i_t} \le x_{i_t} \le u_{i_t}, \qquad \forall i_t \in \mathcal{I}_t, \qquad t = 1, ..., T. &&
\end{aligned}
\tag{3.7}
$$

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}, $i_t \in \mathcal{I}_t$, $t = 1, ..., T$, 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 $(i_2, ..., i_{T+1})$ is

$$
\begin{aligned}
& \min \qquad \langle q_1, x_1^s \rangle + \langle q_{i_2}, x_2^s \rangle + ... + \langle q_{i_T}, x_T^s \rangle \\
& \\
& \text{s.t.} \quad A_1 x_1^s && && && = b_1, \\
& \qquad B_{i_2} x_1^s + A_{i_2} x_2^s && && && = b_{i_2}, \\
& \qquad\qquad B_{i_3} x_2^s + A_{i_3} x_3^s && && && = b_{i_3}, \\
& \qquad\qquad\qquad && && && \vdots \\
& \qquad\qquad\qquad && && B_{i_T} x_{T-1}^s + A_{i_T} x_T^s && = b_{i_T},
\end{aligned}
\tag{3.8}
$$

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 $\tilde{s}$ with a common path up to and including t, $x_j^s = x_j^{\tilde{s}}$, for $j = 1, ..., t$, 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

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

subject to a set of constraints (3.8) for each s, the non-anticipatory constraints, and any upper and lower bound constraints on x_t^s. As [45] 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 are found in the work of Mulvey and Vladimirou [41–44]. See also [12], especially for parallel implementation.

Additional resources including solutions techniques for 2-stage and multistage linear stochastic programs with recourse are in [4], [12], and [20].

3.4 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 optimize

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 value be denoted by s_t and l_t, respectively. There are n assets available at each t 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 $t = 1, ..., T+1$. 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 t for $t = 2, ..., T+1$. 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 made at $t = 1$. For this reason, the CVaR of interest in stage t is written as $\text{CVaR}_\beta(-sw_{t+1})$ where sw_{t+1} is the surplus wealth at $t+1$. The ALM problem of interest is now written as:

$$\min \quad \lambda \left(\sum_{t=1}^{T} \mu_t \text{CVaR}_\beta(-sw_{t+1}) \right) - (1-\lambda)\text{E}(sw_{T+1}) \tag{3.9}$$

$$\text{s.t.} \quad \text{an initial wealth constraint,} \tag{3.10}$$

$$\text{balance of wealth constraints between stages, and} \tag{3.11}$$

$$\text{linear transaction costs.} \tag{3.12}$$

Other constraints may include bounds on positions invested in each asset, bounds on the total transaction costs in each stage, and bounds permitting short selling; however, these are not included in this dissertation.

The above problem (3.9-3.12) does not directly fit into the form (3.3-3.4), but the deterministic equivalent can be put into form (3.5-3.6) with the help of Uryasev's formula for CVaR. To begin, assume a scenario tree has already been

constructed for (s_t, l_t):

$$(s_t, l_t) = (s_{i_t}, l_{i_t}) \text{ with probability } p(i_t), \quad \forall i_t \in \mathcal{I}_t, \quad t = 1, ..., T+1. \tag{3.13}$$

The deterministic equivalent of the optimization problem determines 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 $i_t \in \mathcal{I}_t$, $t = 1, ..., T$. 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 in the nodes at t. Note that this corresponds to the surplus wealth at $t+1$ 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 $t+1 = 2, ..., T+1$, 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: The term $\text{CVaR}_\beta(-sw_{t+1})$ is replaced with

$$\zeta_t + \frac{1}{1-\beta} \sum_{i_{t+1} \in \mathcal{I}_{t+1}} p(i_{t+1}) \left[l_{i_{t+1}} - \langle s_{i_{t+1}}, a_{pred(i_{t+1})} \rangle - \zeta_t \right]^+,$$

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

$$h_{i_{t+1}}(\zeta_t, a_{pred(i_{t+1})}) = \left[l_{i_{t+1}} - \langle s_{i_{t+1}}, a_{pred(i_{t+1})} \rangle - \zeta_t \right]^+, \text{ and} \tag{3.14}$$

$$g_{i_{T+1}}(a_{pred(i_{T+1})}) = \langle s_{i_{T+1}}, a_{pred(i_{T+1})} \rangle - l_{i_{T+1}}. \tag{3.15}$$

The entire objective function is then

$$\begin{aligned} OBJ &= \lambda \sum_{t=1}^{T} \mu_t \zeta_t + \sum_{t=1}^{T} \left(\frac{\lambda \mu_t}{1-\beta} \sum_{i_{t+1} \in \mathcal{I}_{t+1}} p(i_{t+1}) h_{i_{t+1}}(\zeta_t, a_{pred(i_{t+1})}) \right) \\ &\quad -(1-\lambda) \sum_{i_{T+1} \in \mathcal{I}_{T+1}} p(i_{T+1}) g_{i_{T+1}}(a_{pred(i_{T+1})}). \end{aligned} \tag{3.16}$$

Two versions of the constraints for the main problem are given in the next two subsections: the first is without transaction costs, and the second is with transaction costs. The version with transaction costs requires introducing additional decision variables that account for the quantities of assets bought and sold because it is not enough to only keep track of the quantities held.

3.4.1 Constraints Without Transaction Costs

Recall that the quantities of assets held in each node is a n-dimensional vector, that is, for all $i_t \in \mathcal{I}_t$, $t = 1, ..., T$,

$$a_{i_t} = \begin{bmatrix} a_{i_t}^1 \\ \vdots \\ a_{i_t}^n \end{bmatrix}. \tag{3.17}$$

Now define the first stage decision variables, which are just the decision variables in node $i_1 (= 1)$, as

$$x_1 = \begin{bmatrix} \zeta \\ a_1 \end{bmatrix}, \quad \text{where} \quad \zeta = \begin{bmatrix} \zeta_1 \\ \vdots \\ \zeta_T \end{bmatrix}, \tag{3.18}$$

and define the decision variables at each node $i_t \in \mathcal{I}_t$, for $t = 2, ..., T$, as $x_{i_t} = a_{i_t}$.

The initial wealth constraint (3.10) can now be expressed as a stage 1 constraint. If $a_0 \in \mathbb{R}^n$ contains the current investments, the total wealth in node i_1 is $\langle s_{i_1}, a_0 \rangle$. The initial wealth constraint just states that the wealth before reallocation is equal to the wealth after reallocation. That is,

$$\begin{gathered} A_1 x_1 = b_1, \quad \text{where} \\ A_1 = \left[\begin{array}{cc} 0_{1\times T} & s'_{i_1} \end{array} \right], \quad \text{and} \quad b_1 = \langle s_{i_1}, a_0 \rangle. \end{gathered} \tag{3.19}$$

Since there is no short selling of the assets, and there is no restriction on each ζ_i is Uryasev's formula, the bounds on the first stage decision variables are

$$x_1 \geq l_t = \left[\begin{array}{c} -\infty_{T\times 1} \\ 0_{n\times 1} \end{array} \right]. \tag{3.20}$$

Slightly different than the initial wealth constraint, the balance of wealth constraint for each $i_2 \in \mathcal{I}_2$ is written as

$$\begin{gathered} B_{i_2} x_1 + A_{i_2} x_{i_2} = b_{i_2}, \quad \text{where} \\ B_{i_2} = \left[\begin{array}{cc} 0_{1\times T} & -s'_{i_2} \end{array} \right], \quad A_{i_2} = [s'_{i_2}], \quad \text{and} \quad b_{i_2} = 0, \end{gathered} \tag{3.21}$$

and for each remaining node $i_t \in \mathcal{I}_t$, $t = 3, ..., T$, this constraint is

$$\begin{gathered} B_{i_t} x_{pred(i_t)} + A_{i_t} x_{i_t} = b_{i_t} \quad \text{where} \\ -B_{i_t} = A_{i_t} = [s'_{i_t}], \quad \text{and} \quad b_{i_t} = 0. \end{gathered} \tag{3.22}$$

The no short selling constraint in each nodes $i_t \in \mathcal{I}_t$, $t = 2, ..., T$, gives the bound

$$x_{i_t} \geq l_{i_t} = 0_{n\times 1}. \tag{3.23}$$

3.4.2 Constraints With Transaction Costs

The transaction cost to buy is determined by the positive vector $TCB = (TCB^1, ..., TCB^n)' \in \mathbb{R}^n$, and the transaction cost to sell is determined by the positive vector $TCS = (TCS^1, ..., TCS^n) \in \mathbb{R}^n$. Each element TCB^k gives the fraction of transaction wealth lost when asset k is bought. For example, if $TCB^k = .005$ and \$1000 worth of asset k is bought, then the total wealth lost in the transaction is \$5. Additional decision variables need to be included in each node that keep track of the amounts bought and sold of each asset. These are the nonnegative variables $buy_{i_t} \in \mathbb{R}^n$ and $sell_{i_t} \in \mathbb{R}^n$. The first stage decision variables in this case are

$$x_1 = \begin{bmatrix} \zeta \\ a_1 \\ buy_1 \\ sell_1 \end{bmatrix}, \tag{3.24}$$

and the decision variables in the remaining nodes are

$$x_{i_t} = \begin{bmatrix} a_{i_t} \\ buy_{i_t} \\ sell_{i_t} \end{bmatrix}, \tag{3.25}$$

for all $i_t \in \mathcal{I}_t$, $t = 2, ..., T$. The bounds become

$$x_1 \geq l_1 = \begin{bmatrix} -\infty_{T\times 1} \\ 0_{3n\times 1} \end{bmatrix}, \quad \text{and} \quad x_{i_t} \geq l_{i_t} = 0_{3n\times 1}. \tag{3.26}$$

There are now two types of constraints to consider: The first keeps track of the quantities bought and sold, and the second keeps track of the wealth. The first type states that the quantity of each asset held is the quantity previously held

plus the amount bought and less the amount sold. Let $a_0 \in \mathbb{R}^n$ denote the current quantities of investment again, then in the first stage the equation for asset k is

$$a_1^k = a_0^k + buy_1^k - sell_1^k,$$

and now in matrix form for all assets,

$$\left[\begin{array}{cccc} 0_{n\times T} & \mathrm{I}_n & -\mathrm{I}_n & \mathrm{I}_n \end{array} \right] x_1 = a_0,$$

where I_n is the $n \times n$ identity matrix. The second type of constraint states that the wealth after portfolio reallocation is equal to the wealth before reallocation minus the total transaction costs. In the first stage, that is

$$\langle s_{i_1}, a_1 \rangle = \langle s_{i_1}, a_0 \rangle - \langle s_{i_1} * TCB, buy_1 \rangle - \langle s_{i_1} * TCS, sell_1 \rangle,$$

where $*$ is element by element matrix multiplication:

$$s_{i_1} * TCB = \left[\begin{array}{c} s_{i_1}^1 TCB^1 \\ \vdots \\ s_{i_1}^n TCB^n \end{array} \right].$$

Now combining both types of constraints gives

$$A_1 x_1 = b_1 \quad \text{where}$$

$$A_1 = \left[\begin{array}{cccc} 0_{n\times T} & \mathrm{I}_n & -\mathrm{I}_n & \mathrm{I}_n \\ 0_{1\times T} & s_{i_1}' & (TCB * s_{i_1})' & (TCS * s_{i_1})' \end{array} \right], \quad \text{and} \tag{3.27}$$

$$b_1 = \left[\begin{array}{c} a_0 \\ \langle s_{i_1}, a_0 \rangle \end{array} \right].$$

For each node $i_2 \in \mathcal{I}_2$, the constraints on quantities become

$$\left[\begin{array}{cccc} 0_{n\times T} & 0_n & 0_n & -\mathrm{I}_n \end{array}\right] x_1 + \left[\begin{array}{ccc} \mathrm{I}_n & -\mathrm{I}_n & \mathrm{I}_n \end{array}\right] x_{i_2} = 0_{n\times 1},$$

and the wealth constraint becomes

$$\langle s_{i_2}, a_{i_2}\rangle = \langle s_{i_2}, a_1\rangle - \langle s_{i_2} * TCB, buy_{i_2}\rangle - \langle s_{i_2} * TCS, sell_{i_2}\rangle.$$

Combining the previous two gives

$$\begin{gathered} B_{i_2}x_1 + A_{i_2}x_{i_2} = b_{i_2}, \quad \text{where} \\ B_{i_2} = \left[\begin{array}{cccc} 0_{n\times T} & 0_n & 0_n & -\mathrm{I}_n \\ 0_{1\times T} & 0_{1\times n} & 0_{1\times n} & -s'_{i_2} \end{array}\right], \\ A_{i_2} = \left[\begin{array}{ccc} \mathrm{I}_n & -\mathrm{I}_n & \mathrm{I}_n \\ s'_{i_2} & (TCB * s_{i_2})' & (TCS * s_{i_2})' \end{array}\right], \quad \text{and} \\ b_{i_2} = [0_{(n+1)\times 1}]. \end{gathered} \tag{3.28}$$

For the remaining nodes $i_t \in \mathcal{I}_t$, $t = 3, ..., T$, the equations are easily modified to

$$\begin{gathered} B_{i_t}x_{pred(i_t)} + A_{i_t}x_{i_t} = b_{i_t}, \quad \text{where} \\ B_{i_t} = \left[\begin{array}{ccc} 0_n & 0_n & -\mathrm{I}_n \\ 0_{1\times n} & 0_{1\times n} & -s'_{i_t} \end{array}\right], \\ A_{i_t} = \left[\begin{array}{ccc} \mathrm{I}_n & -\mathrm{I}_n & \mathrm{I}_n \\ s'_{i_t} & (TCB * s_{i_t})' & (TCS * s_{i_t})' \end{array}\right], \quad \text{and} \\ b_{i_t} = [0_{(n+1)\times 1}]. \end{gathered} \tag{3.29}$$

3.4.3 Recursive Formulation of the Deterministic Equivalent

The problem is now to minimize the piecewise linear, convex objective function given by equation (3.16) subject to constraints (3.19-3.23) in the case of no transaction costs, or subject to constraints (3.26-3.29) in the case of transaction costs. In either case, the constraints are linear and fit exactly into the form of the constraints in problem (3.5-3.6). Additionally, the piecewise linear objective function can be converted into a fully linear objective by introducing auxiliary decision variables. The deterministic equivalent of the T-stage ALM problem can then be posed precisely as problem (3.7), and a good linear solver could potentially be used. The drawback of this setup is that there is one auxiliary variable introduced for each scenario in each stage. For a huge number of scenarios, this creates a huge number of decision variables, and even if the problem is linear, a solver might be inadequate because of memory and time considerations. For this reason, a solution algorithm for the piecewise linear objective is discussed in the next subsection; but first, the deterministic equivalent of the ALM problem needs to be rewritten into the recursive formulation of (3.5-3.6).

To put the objective function into the recursive form, note that the sums can be expressed in terms of the transition probabilities: for instance,

$$\begin{aligned}&\textstyle\sum_{i_{t+1}\in\mathcal{I}_{t+1}} p(i_{t+1})h_{i_{t+1}}(\zeta_t, a_{pred(i_{t+1})})\\ &\quad = \textstyle\sum_{i_2\in\mathcal{I}_2}\sum_{i_3\in dec(i_2)}\cdots\sum_{i_{t+1}\in dec(i_t)} p(i_2)p(i_2,i_3)\cdots p(i_t,i_{t+1})h_{i_{t+1}}(\zeta_t, a_{i_t}).\end{aligned}$$

Define the terms

$$q_1(\zeta_1, a_1) = \lambda\sum_{t=1}^{T}\mu_t\zeta_t + \frac{\lambda\mu_1}{1-\beta}\sum_{i_2\in\mathcal{I}_2} p(i_2)h_{i_2}(\zeta_1, a_1), \tag{3.30}$$

$$q_{i_t}(\zeta_t, a_{i_t}) = \frac{\lambda\mu_t}{1-\beta} \sum_{i_{t+1}\in dec(i_t)} p(i_t, i_{t+1}) h_{i_{t+1}}(\zeta_t, a_{i_t}), \quad t = 2, .., T-1, \tag{3.31}$$

$$q_{i_T}(\zeta_T, a_{i_T}) = \sum_{i_{T+1}\in dec(i_T)} p(i_T, i_{T+1}) \left(\frac{\lambda\mu_T}{1-\beta} h_{i_{T+1}}(\zeta_T, a_{i_T}) - (1-\lambda) g_{i_{T+1}}(a_{i_T}) \right). \tag{3.32}$$

Then after rearranging terms, and moving the minimizations, the restated problem is

$$\begin{array}{ll} \min_{x_1} & q_1(\zeta_1, a_1) + \sum_{i_2\in\mathcal{I}_2} p(i_2) Q_{i_2}(x_1) \\ \text{s.t.} & A_1 x_1 = b_1 \\ & x_1 \geq l_1. \end{array} \tag{3.33}$$

The Q_{i_t}, for $t = 2, ..., T$, are defined recursively as

$$\begin{array}{rll} Q_{i_t}(\zeta_t, ..., \zeta_T, x_{t-1}) = & \min_{x_t} & q_{i_t}(\zeta_t, a_t) \\ & & \quad + \sum_{i_{t+1}\in dec(i_t)} p(i_t, i_{t+1}) Q_{i_{t+1}}(\zeta_{t+1}, ..., \zeta_T, x_t) \\ & \text{s.t.} & B_{i_t} x_{t-1} + A_{i_t} x_t = b_{i_t}, \\ & & x_t \geq l_{i_t}, \end{array} \tag{3.34}$$

with $Q_{i_{T+1}} = 0$. Note that it is redundant to write $Q_{i_2}(\zeta_2, ..., \zeta_T, x_1)$ since ζ is contained in x_1, so one can just write $Q_{i_2}(x_1)$ as in (3.33).

3.4.4 Solution Algorithm for the 2-Stage Problem

In the ALM problem (3.9-3.12), both the risk and reward are included in the objective function. While it might be more desirable to minimize the risk subject to a reward constraint, or maximize the reward subject to a risk constraint, including both the risk and reward in the objective allows a solution method that is easily computed in parallel. If the risk or reward is included as a constraint, there is a single constraint over all nodes at $t = 3$ in the 2-stage problem, and the algorithm

of this section can no longer be implemented.

This solution algorithm for a piecewise linear 2-stage problem is developed in detail in [27], and a sketch of this algorithm is repeated here. This algorithm uses a modification of Bender's outer linearization. The basic idea is to linearize the objective function at various points and find the point that yields the minimum value that lies above all the objective linearizations. Linearization points are sequentially added until the minimum value above all linearizations coincides with the actual objective value at the corresponding point. This point is the optimal solution.

The 2-stage ALM problem fits into the following form:

$$\begin{array}{ll} \min_{x_1 \geq 0} & q_1(x_1) + \sum_{i_2 \in I_2} p(i_2) Q_{i_2}(x_1) \\ \text{s.t.} & A_1 x_1 = b_1, \end{array} \tag{3.35}$$

with

$$\begin{array}{rll} Q_{i_2}(x_1) = & \min_{x_2 \geq 0} & q_{i_2}(x_1, x_2) \\ & \text{s.t.} & A_{i_2} x_1 + B_{i_2} x_2 = b_{i_2}. \end{array} \tag{3.36}$$

This form is basically the same as (3.33-3.34) since the bounds on the decision variables can be changed by separating each ζ_i into a positive and negative part (this only slightly modifies the constraint matrixes). The functions q_1 and q_{i_2} are essentially the same as those given in the previous section, only the dependence is written differently to make the following easier to present.

The solution algorithm makes use of subgradients since the objective function is piecewise linear. Let $q_1'(x_1^0)$ be a column vector that denotes a subgradient of q_1 at the point x_1^0, and let $Q_{i_2}'(x_1^0)$ similarly denote a subgradient of Q_{i_2} at x_1^0. For simplicity, let $Q_1(x_1) = \sum_{i_2} p(i_2) Q_{i_2}(x_1)$, which implies a subgradient of Q_1 at x_1^0 is $Q_1'(x_1^0) = \sum_{i_2} p(i_2) Q_{i_2}'(x_1^0)$. Note that q_1 and each q_{i_2} are just piecewise linear

functions, so it is easy to calculate a subgradient for each of them. Calculating a subgradient for the optimal value of each minimization problem Q_{i_2} is more difficult and is shown shortly.

The first stage algorithm is:

(A.0) Select an initial point x_1^0 satisfying $A_1 x_1^0 = b_1$ and solve

$$\begin{array}{ll} \min_{(\theta, x_1)} & \theta \\ \text{s.t.} & A_1 x_1 = b_1, \\ & q_1(x_1^0) + Q_1(x_1^0) + \langle q_1'(x_1^0) + Q_1'(x_1^0), x_1 - x_1^0 \rangle \leq \theta. \end{array}$$

Let the optimal solution to this problem be at (θ^1, x_1^1).

(A.1) If $q_1(x_1^1) + Q_1(x_1^1) \leq \theta^1$, then the optimal solution to (3.35) is at x_1^1. If not, add an additional constraint and solve the linear problem

$$\begin{array}{ll} \min_{(\theta, x_1)} & \theta \\ \text{s.t.} & A_1 x_1 = b_1, \\ & q_1(x_1^0) + Q_1(x_1^0) + \langle q_1'(x_1^0) + Q_1'(x_1^0), x_1 - x_1^0 \rangle \leq \theta, \\ & q_1(x_1^1) + Q_1(x_1^1) + \langle q_1'(x_1^1) + Q_1'(x_1^1), x_1 - x_1^1 \rangle \leq \theta. \end{array}$$

Let the optimal solution to this problem be at (θ^2, x_1^2).

$$\vdots$$

(A.m) If $q_1(x_1^m) + Q_1(x_1^m) \leq \theta^m$, then the optimal solution to (3.35) is at x_1^m. If

not, add an additional constraint and solve the linear problem

$$\begin{aligned}
\min_{(\theta,x_1)} \quad & \theta \\
\text{s.t.} \quad & A_1 x_1 = b_1, \\
& q_1(x_1^0) + Q_1(x_1^0) + \langle q_1'(x_1^0) + Q_1'(x_1^0), x_1 - x_1^0 \rangle \leq \theta, \\
& \vdots \\
& q_1(x_1^m) + Q_1(x_1^m) + \langle q_1'(x_1^m) + Q_1'(x_1^m), x_1 - x_1^m \rangle \leq \theta.
\end{aligned}$$

Let the optimal solution to this problem be at $(\theta^{m+1}, x_1^{m+1})$.

$$\vdots$$

In order to find Q_1 and Q_1' at the given point x_1^i in the above, one has to compute $Q_{i_2}(x_1^i)$ and $Q_{i_2}'(x_1^i)$ for all $i_2 \in I_2$. A subgradient of q_{i_2} at (x_1^i, x_2^0) can be separated into two parts: A part corresponding to the first stage variables $q_{i_2}^{x_1}(x_1^i, x_2^0)$ and a part corresponding to the second stage variables $q_{i_2}^{x_2}(x_1^i, x_2^0)$. The value of $Q_{i_2}(x_1^i)$ is then found through the following algorithm:

(B.0) Select an initial point x_2^0 satisfying $A_{i_2} x_1^i + B_{i_2} x_2^0 = b_{i_2}$ and solve

$$\begin{aligned}
\min_{(\theta,x_2)} \quad & \theta \\
\text{s.t.} \quad & A_{i_2} x_1^i + B_{i_2} x_2 = b_{i_2}, \\
& q_{i_2}(x_1^i, x_2^0) + \langle q_{i_2}^{x_2}(x_1^i, x_2^0), x_2 - x_2^0 \rangle \leq \theta.
\end{aligned}$$

Let the optimal solution to this problem be at (θ^1, x_2^1).

(B.1) If $q_{i_2}(x_1^i, x_2^1) \leq \theta^1$, then the optimal solution is at x_2^1 and $Q_{i_2}(x_1^i) = \theta^1$. If

not, add an additional constraint and solve the linear problem

$$\begin{aligned} \min_{(\theta,x_2)} \quad & \theta \\ \text{s.t.} \quad & A_{i_2}x_1^i + B_{i_2}x_2 = b_{i_2}, \\ & q_{i_2}(x_1^i, x_2^0) + \langle q_{i_2}^{x_2}(x_1^i, x_2^0), x_2 - x_2^0 \rangle \leq \theta, \\ & q_{i_2}(x_1^i, x_2^1) + \langle q_{i_2}^{x_2}(x_1^i, x_2^1), x_2 - x_2^1 \rangle \leq \theta. \end{aligned}$$

Let the optimal solution to this problem be at (θ^2, x_2^2).

$$\vdots$$

(B.m) If $q_{i_2}(x_1^i, x_2^m) \leq \theta^m$, then the optimal solution is at x_2^m and $Q_{i_2}(x_1^i) = \theta^m$. If not, add an additional constraint and solve the linear problem

$$\begin{aligned} \min_{(\theta,x_2)} \quad & \theta \\ \text{s.t.} \quad & A_{i_2}x_1^i + B_{i_2}x_2 = b_{i_2}, \\ & q_{i_2}(x_1^i, x_2^0) + \langle q_{i_2}^{x_2}(x_1^i, x_2^0), x_2 - x_2^0 \rangle \leq \theta, \\ & \vdots \\ & q_{i_2}(x_1^i, x_2^m) + \langle q_{i_2}^{x_2}(x_1^i, x_2^m), x_2 - x_2^m \rangle \leq \theta. \end{aligned} \tag{3.37}$$

Let the optimal solution to this problem be at $(\theta^{m+1}, x_2^{m+1})$.

$$\vdots$$

To obtain a subgradient of Q_{i_2}, assume that the procedure is stopped immediately after step (**B.m**), so $Q_{i_2}(x_1^i) = \theta^{m+1}$, where θ^{m+1} is found by solving problem (3.37). Let the lagrange multipliers of (3.37) be denoted by the column vector π_{i_2}, then

$$Q'_{i_2}(x_1^i) = \begin{bmatrix} A'_{i_2} & q_{i_2}^{x_1}(x_1^i, x_2^0) & q_{i_2}^{x_1}(x_1^i, x_2^1) & \cdots & q_{i_2}^{x_1}(x_1^i, x_2^n) \end{bmatrix} \pi_{i_2}$$

This algorithm is easily implemented in parallel. Given a first stage decision x_1^i, the second stage procedure (B) for $Q'_{i_2}(x_1^i)$ can be computed on different processors corresponding to different nodes i_2.

Chapter 4

Time Series Methods

For the ALM problem of the previous chapter, scenarios are generated by calibrating and simulating a time series model. This chapter first describes the multivariate data set and then fits a time series model under the normal and stable distributional assumptions.

There are two major approaches in modeling multivariate data:

- 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 univariate GARCH models with stable distributions to each time series and then capture the dependence with a time-varying copula. This approach was attempted for the ALM data, but it produced poor scenarios: Estimation resulted in an integrated GARCH model and an explosion of sample paths in simulation.

This chapter only discusses the first approach for which scenarios are generated and the optimization problem is solved in the next chapter. A good illustration

of the second approach is found in [26].

4.1 Data Description and VAR modeling

The T-stage ALM problem of Section 3.4 is now 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 [55]. This same reference also provides the typical asset classes invested in by pension funds: cash, bonds, equities, real estate, international stocks, international bonds, mortgages, GIC's and annuities, and private equities. Table (4.1) contains the benchmarks used for the asset classes in this dissertation. It was difficult to obtain benchmarks for all the asset classes just listed, but the major three (bonds, equities, and international equities) are included. Given the historical data for the liability index l_τ and asset indexes s^i_τ, $i = 1, ..., 5$, one can construct a multivariate scenario tree. This is achieved by fitting a multivariate time series model to the return vector:

$$R_\tau = \begin{bmatrix} r^1_\tau \\ r^2_\tau \\ \vdots \\ r^6_\tau \end{bmatrix} = \begin{bmatrix} l_\tau / l_{\tau-1} - 1 \\ s^1_\tau / s^1_{\tau-1} - 1 \\ \vdots \\ s^5_\tau / s^5_{\tau-1} - 1 \end{bmatrix}.$$

Once a time series model is calibrated, it is simple to generate sample paths for the returns and then convert the returns back to index values. It is important to note that in this chapter 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 coincide;

	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 4.1: Benchmarks for the pension fund asset classes.

however, there are usually many smaller time periods between stages. In one application in this dissertation, a time series model is fit to monthly data, but a stage covers a 6 month period.

Figure (4.1) 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, but the length of the data set proved to be very short when trying to also incorporate a multivariate stable distribution. Ideally, one would have at least 1000 points of daily data, but daily data is not available for all the benchmarks in Table (4.1).

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 means. The means of the univariate return series are assumed to follow

$$m_\tau = \lambda_m m_{\tau-1} + (1 - \lambda_m) r_{\tau-1},$$

where m_τ is the vector of means and λ_m is a fixed parameter. The new return series of interest is

$$\tilde{R}_\tau = R_\tau - m_\tau,$$

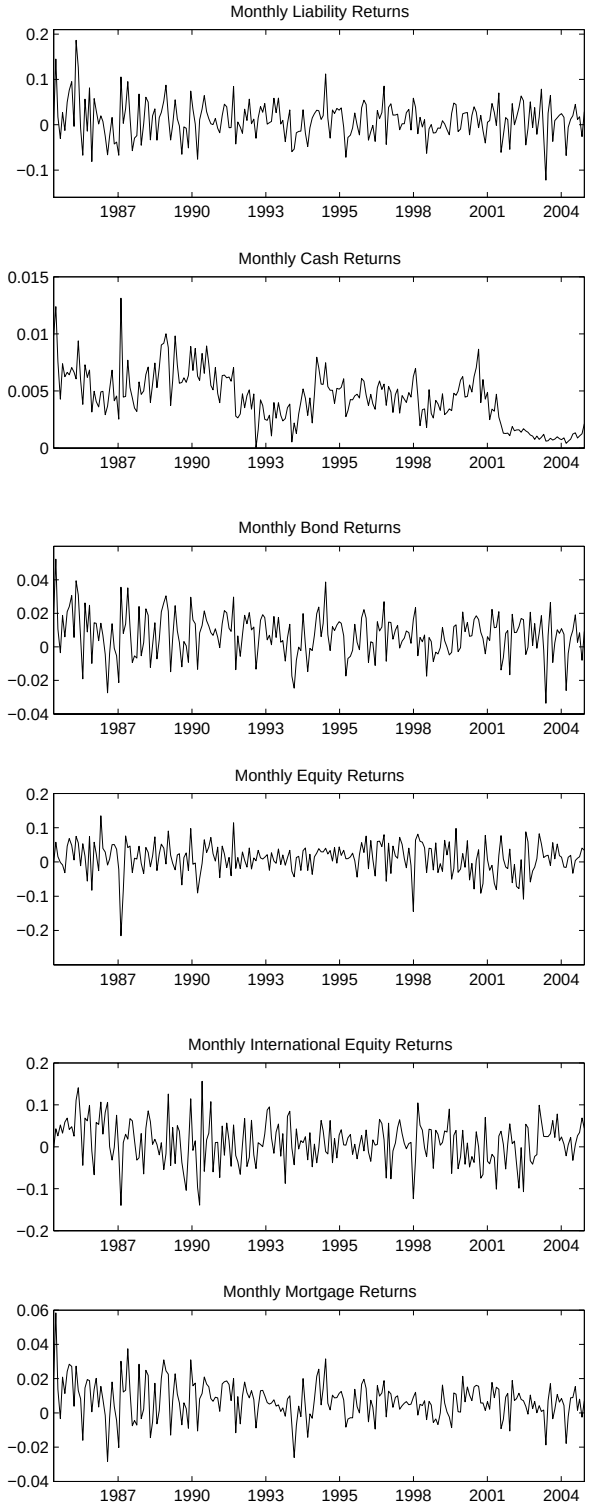

Figure 4.1: Monthly returns R_τ from April 1985 to December 2004.

and as the next step, a vector autoregressive (VAR) model is calibrated to $\tilde{R}_\tau$.

The VAR model has had much success in modeling financial and economic data. The general VAR(p) model for $\tilde{R}_\tau$ is

$$\tilde{R}_\tau = \Pi_1 \tilde{R}_{\tau-1} + ... + \Pi_p \tilde{R}_{\tau-p} + E_\tau,$$

where the innovations process $E_\tau = (e_\tau^1, ..., e_\tau^6)'$ 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 [59]. For the data at hand, the Akaike Information Criteria (AIC) indicates that the VAR equation of order 1 is optimal, and all the backtesting in the next chapter is carried out under the assumption of a VAR(1) model. More generally, one may fit a multivariate autoregressive moving average (ARMA) model such as in [7]; however, multivariate financial data typically indicates only an autoregressive component, so it is reasonable to restrict the model to VAR. Another extension of the above VAR model that additionally includes economic regime changes and long term equilibria in an ALM context can be found in [5].

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

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

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

and the standardized residuals $\hat{\Sigma}^{-1/2}\hat{E}_\tau$ are plotted in Figure (4.2). The usual assumption is that the innovations are Gaussian, in which case the standardized residuals should be i.i.d. 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 distribution is found in Figure (4.3).

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

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

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. When the optimization program is solved for the minimum

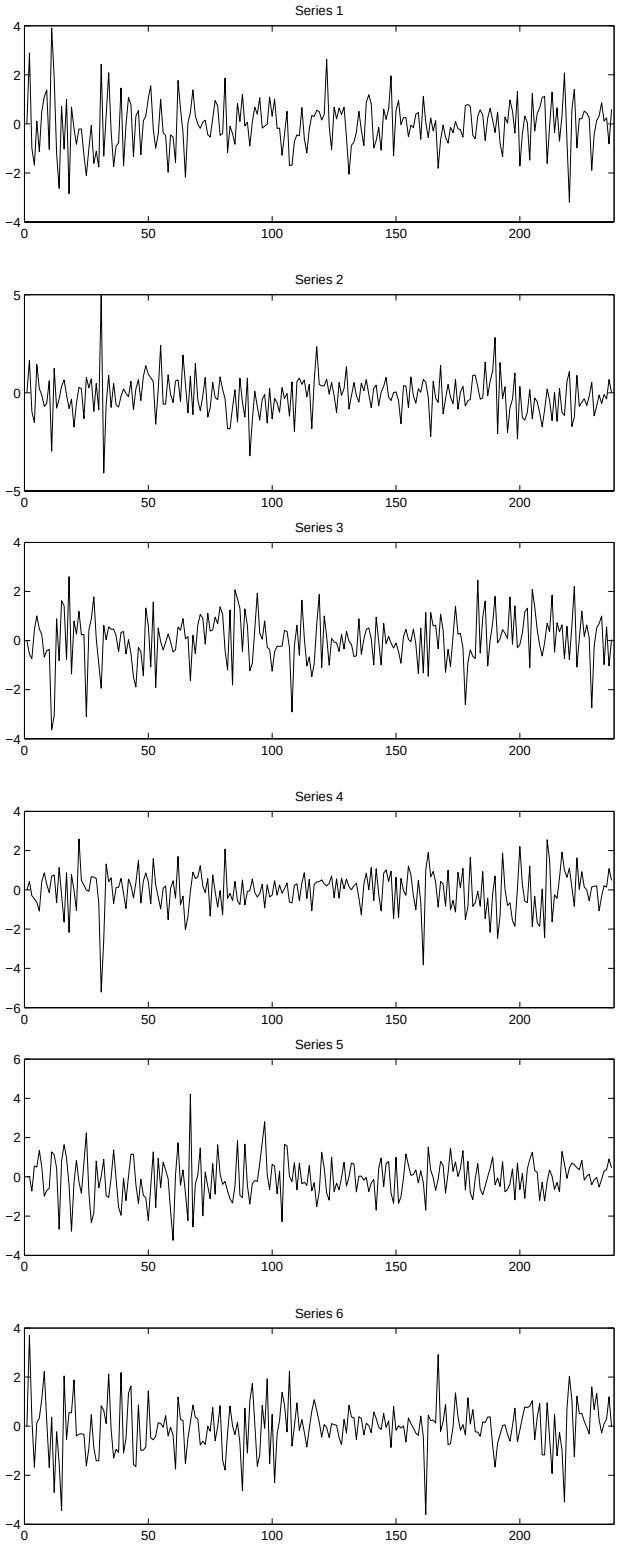

Figure 4.2: Standardized residuals $\hat{\Sigma}^{-1/2}\hat{E}_\tau$ of the VAR(1) model for R_τ.

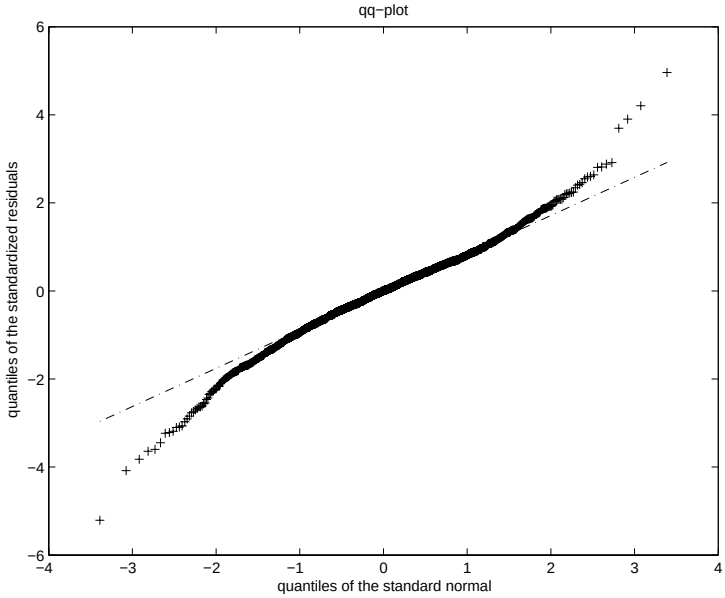

Figure 4.3: QQ-plot of the standard normal versus the standardized residuals $\hat{\Sigma}^{-1/2}\hat{E}_\tau$.

risk portfolio, one could expect a large allocation in the bonds to offset the risk in the liabilities. As is shown by the numerical results in the next chapter, this is indeed the case.

4.2 Stable Distributions

When looking at the distributions of financial time series, the Gaussian assumption often provides a poor fit to the data. As far back as the 1960's, the works of Mandelbrot [33–36] and Fama [22] proposed the more general stable laws as an alternative to the special case of Gaussian laws for the returns of various financial assets. In addition to the ability to capture heavy tails, skewness, and peakedness, stable distributions have many other desirable properties:

Stability Property Each stable distribution has an index of stability α that can act as an overall parameter in inference and decision making.

Additive Property Any linear combination of independent stable random variables with given α is a stable random variable with the same α. This property is advantageous when looking at portfolios of assets with stable returns.

Domain of Attraction It is possible to determine if a random variable is in the domain of attraction of a stable law by looking at the decay of the tails. The normalized sum of independent random variables in the domain of attraction of a stable law has properties similar to those of the stable law.

4.2.1 Univariate Stable

A random variable z has a stable distribution if for any $a > 0$ and $b > 0$ there exists $c > 0$ and $d \in \mathbb{R}$ such that

$$az_1 + bz_2 \stackrel{d}{=} cz + d,$$

where z_1 and z_2 are independent copies of z. Stable distributions are represented by four parameters, α, β, μ, and σ, and are denoted by $S_\alpha(\sigma, \beta, \mu)$. When the index of stability, $\alpha \in (0, 2]$, is small, the distribution has a high peak and heavy tails. The skewness parameter, β, determines if the distribution is skewed to the left ($\beta < 0$) or the right ($\beta > 0$). The scale parameter, σ, generalizes the notion of standard deviation, and the variation, σ^α, generalizes the notion of variance. If $\alpha = 2$ and $\beta = 0$ the stable distribution is the Gaussian distribution.

For $\alpha \in (0, 2)$ the heavy tails can be described by

$$\mathrm{P}(z > \lambda) \sim c_1\lambda^{-\alpha}, \qquad \mathrm{P}(z < -\lambda) \sim c_2\lambda^{-\alpha}, \qquad \forall\lambda > 0,$$

where c_1 and c_2 are constants. In this case, the p-th absolute moment of z, $\mathrm{E}|z|^p = \int_0^\infty \mathrm{P}(|z|^p > \lambda)d\lambda$, is finite if and only if $p < \alpha$. If $\alpha = 2$, all absolute

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

Table 4.2: Univariate ML estimates of the tail index and scale parameter for each residual series $\hat{e}^i$.

moments are finite. Models of financial data typically assume $\alpha \in (1, 2]$, so it is possible to discuss expected returns. In general, there is no closed form density or distribution function for stable random variables (see [23]). However, the characteristic function, $\Phi_z(\theta) = \mathrm{E}\left(\exp(iz\theta)\right)$, is given by

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

Densities can be computed through the characteristic function by fast Fourier transform methods, and parameters can by estimated by maximum likelihood methods. Also, see [28] for another interesting approach in estimating α.

Returning to the ALM data, 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 $\hat{\sigma}^i_\alpha$ from each univariate series $\hat{e}^i$ is given in Table (4.2). The estimation is 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 both the estimated normal density and the estimated stable density in Figures (4.4) and (4.5). 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.

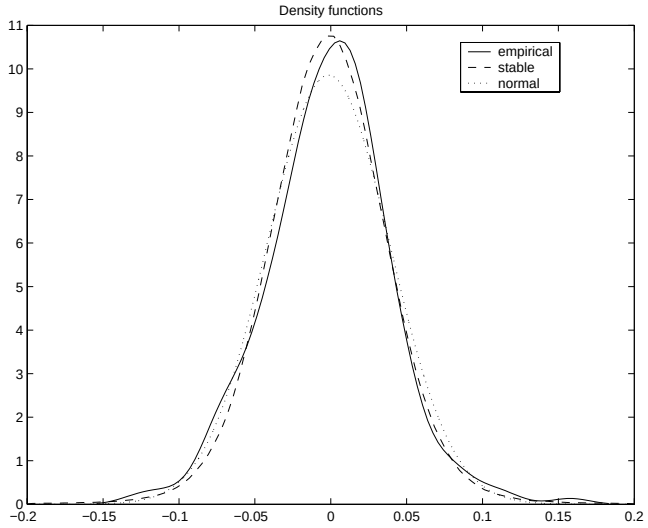

Figure 4.4: Density functions for the residuals of the liability return series.

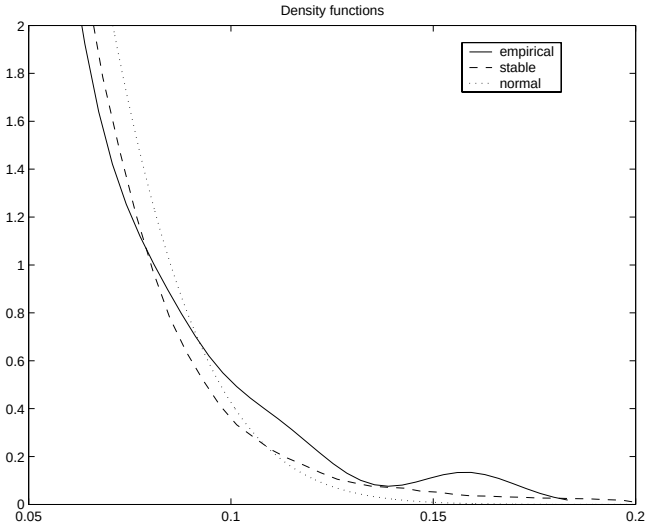

Figure 4.5: Right tail of the density functions for the residuals of the liability return series.

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

Table 4.3: Comparison of KD under the normality and stable assumptions for each residual series $\hat{e}^i$.

	$\hat{e}^1$	$\hat{e}^2$	$\hat{e}^3$	$\hat{e}^4$	$\hat{e}^5$	$\hat{e}^6$
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 4.4: Comparison of AD under the normality and stable assumptions for each residual series $\hat{e}^i$.

Two goodness-of-fit measures are employed to compare the normal fit and the stable fit of the univariate series. The first is the Kolmogorov distance (KD) between the empirical distribution function $F_e(x)$ and the estimated distribution function $\hat{F}(x)$:

$$KD = \sup_{x \in \mathbb{R}} \left| F_e(x) - \hat{F}(x) \right|.$$

The KD is considered a robust measure that emphasizes the deviation of the estimated distribution from the empirical distribution about the median. The second goodness-of-fit measure emphasizes the deviation at the tails and is called the Anderson-Darling (AD) statistic:

$$AD = \sup_{x \in \mathbb{R}} \frac{\left| F_e(x) - \hat{F}(x) \right|}{\sqrt{\hat{F}(x)\left(1 - \hat{F}(x)\right)}}.$$

The KD and AD of the normal and stable estimated distributions for each of the series can be found in Tables (4.3) and (4.4). The normal fit slightly outperforms the stable fit twice under the KD measure, but the stable fit is clearly superior under the AD measure.

4.2.2 Multivariate Stable and Sub-Gaussian

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

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

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

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

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

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

See [49] and [29] for details on estimating the index of stability, spectral measure, and scale parameter for a general stable random vector. A simple estimate of the index of stability $\hat{\alpha}$ can be obtained from the univariate estimates $\hat{\alpha}^i$ of each

component z^i of Z:

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \hat{\alpha}^i. \tag{4.1}$$

Alternatively, since all one-dimensional projections $u'Z$, $u \in \mathbb{R}^n$, are α-stable random variables, a better estimate might be obtained by averaging the univariate estimates of many one-dimensional projections.

The special case of a sub-Gaussian distribution for a random vector Z is both stable and elliptical with characteristic function

$$\Phi_Z(\theta) = \exp\left\{-\left(\theta' Q \theta\right)^{\alpha/2} + i\theta'\mu\right\},$$

where $Q = \left(q_{ij}^2\right)$ is called the dispersion matrix. If $\alpha = 2$, this reduces to the multivariate Gaussian distribution, and if $\alpha < 2$, the variance is infinite. Each term q_{ij}^2 is the co-dispersion between components z^i and z^j of Z, and it is defined in terms of the $S\alpha S$ vector $\tilde{Z} = Z - \mu$:

$$q_{ij}^2 = \left[\tilde{z}^i, \tilde{z}^j\right]_\alpha \|\tilde{z}^j\|_\alpha^{2-\alpha}.$$

The sub-Gaussian random vector offers an advantage in portfolio optimization because it is in the class of elliptical distributions. For instance, if the portfolio returns are modeled with a sub-Gaussian vector, the VaR of a portfolio with weights ω is just $k\sqrt{\omega' Q \omega}$, for a constant k. This means that minimizing VaR can be achieved by minimizing the quadratic function $\omega' Q \omega$. A second advantage of a sub-Gaussian random vector is that it is straightforward to simulate: One just needs to be able to simulate a multivariate Gaussian $G \sim N(0, 2Q)$ and an independent, positive, stable random variable $s \sim S_{\alpha/2}\left(\left(\cos(\pi\alpha/4)\right)^{2/\alpha}, 1, 0\right)$, and use

$$\tilde{Z} \stackrel{d}{=} \sqrt{s} G. \tag{4.2}$$

G is referred to as the governing Gaussian vector.

A method of estimation for Q based on the moments $\mathrm{E}\,|\tilde{z}^i|^p$ and $\mathrm{E}\left(\tilde{z}^i(\tilde{z}^j)^{\langle q-1\rangle}\right)$, for a $p \in (0,\alpha)$ and $q \in (1,\alpha)$, is found in [32]. Given a multivariate data series $\{Z_\tau, \tau = 1, ..., \tau_m\}$, the moment estimator of Q is computed from the centered data $\tilde{Z}_\tau = Z_\tau - \mu$ by

$$\hat{q}_{jj}^2 = \left(A(p)\frac{1}{\tau_m}\sum_{\tau=1}^{\tau_m} \left|\tilde{z}_\tau^j\right|^p \right)^{2/p}, \tag{4.3}$$

$$\hat{q}_{ij}^2 = \hat{q}_{jj}^{2-q} A(q)\frac{1}{\tau_m}\sum_{\tau=1}^{\tau_m} \tilde{z}_\tau^i \left(\tilde{z}_\tau^j\right)^{\langle q-1\rangle}, \tag{4.4}$$

where

$$A(p) = \frac{\Gamma(1-p/2)\sqrt{\pi}}{2^p\Gamma(1-p/\alpha)\Gamma((p+1)/2)}. \tag{4.5}$$

The authors of the above reference note that the rates of convergence for $\hat{q}_{jj}$ and $\hat{q}_{ij}$ are fastest for p and q as small as possible. They suggest to take $q = 1$ and describe a selection method for p. Alternatively, for financial data it is common to take $p = \alpha/3$.

A sub-Gaussian distribution can be fit to the residuals $\hat{E}_\tau$ of the ALM data. First, α is estimated using equation (4.1) and the univariate estimates from Table (4.2), yielding $\hat{\alpha} = 1.8705$. Assuming the residuals have zero mean, the moment estimator of Q in equations (4.3-4.5) 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 (4.2). The moment estimate for Q given by the above equations is not symmetric, but a symmetric estimate is given by $\hat{Q} = \left((\hat{q}_{ij}^2 + \hat{q}_{ji}^2)/2\right)$. The standardized residuals $\hat{Q}^{-1/2}\hat{E}_\tau$ are

also computed and plotted in Figure (4.6). 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 (4.7). This plot appears closer to linear than the QQ-plot with the standard normal in Figure (4.3), which indicates the sub-Gaussian distribution provides a better fit than the multivariate normal; however, neither of these capture the time-varying nature of the innovations.

4.3 Exponentially Weighted Moving Average Models

To account for the volatility clustering, three types of models are implemented: The first model assumes the innovations have a time-varying Gaussian distribution, and the second two models assume the innovations have a time-varying sub-Gaussian distribution.

4.3.1 Normal EWMA

Given a multivariate data set $\{E_\tau, \tau = 1, ..., \tau_m\}$ with zero mean, the sample estimate of the covariance is just

$$\hat{\Sigma} = \frac{1}{\tau_m - 1} \sum_{\tau=1}^{\tau_m} E_\tau E_\tau'.$$

Note that there is equal weight applied to each observation in the data set. To construct 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

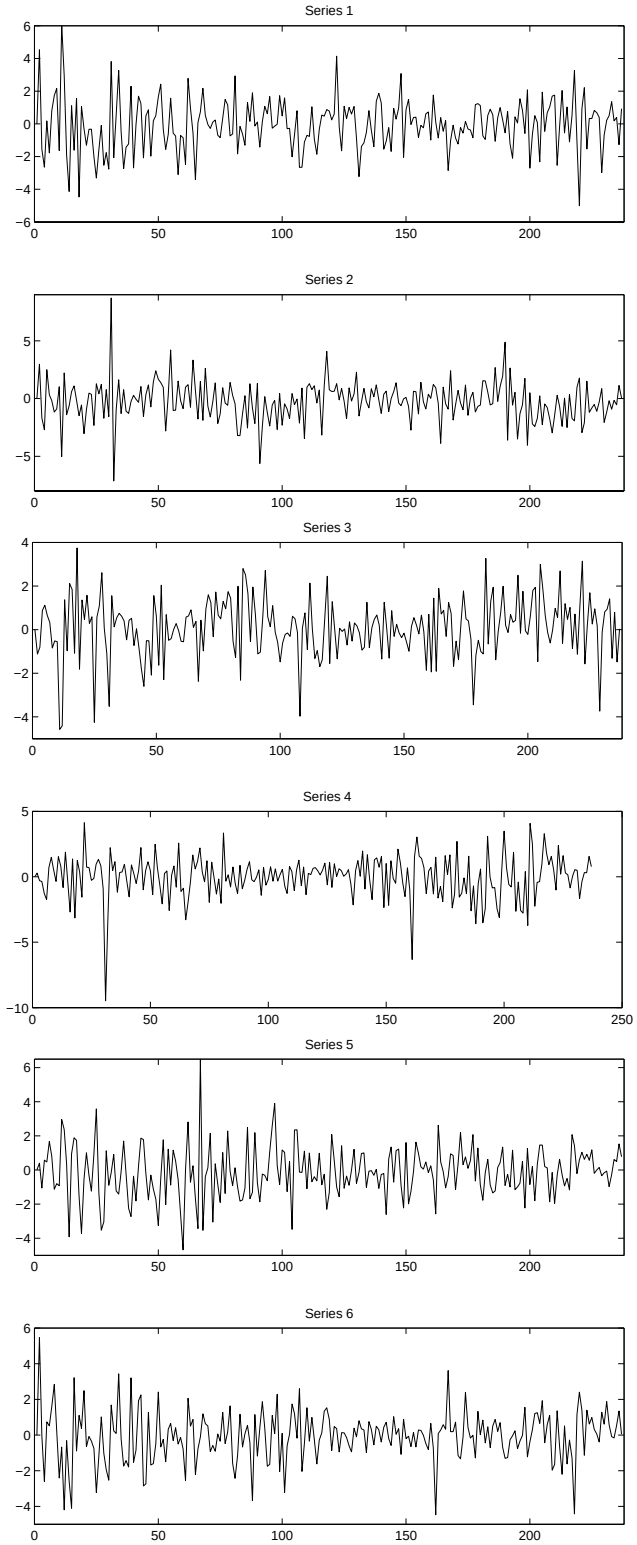

Figure 4.6: Standardized residuals $\hat{Q}^{-1/2}\hat{E}_\tau$ of the VAR(1) model for R_τ.

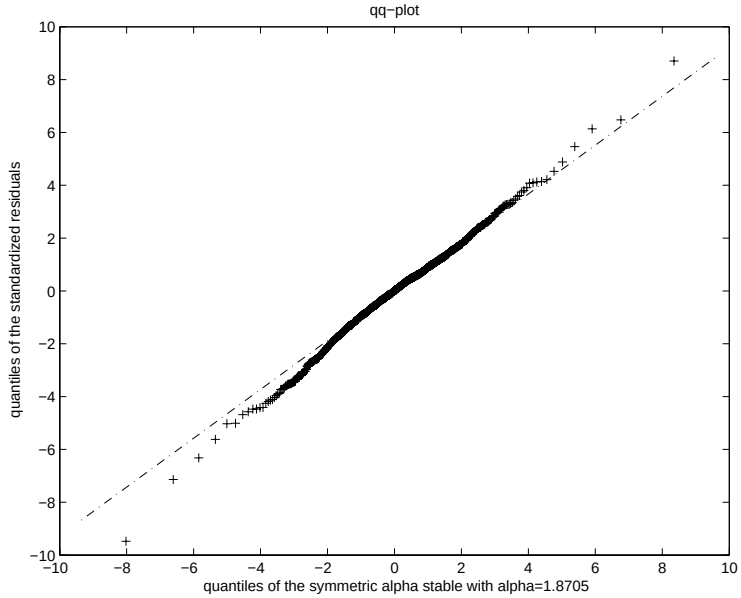

Figure 4.7: QQ-plot of the symmetric stable with $\alpha = 1.8705$ versus the standardized residuals $\hat{Q}^{-1/2}\hat{E}_\tau$.

decay exponentially from the most recent observation:

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

where $0 < \lambda_e < 1$ and the weights are chosen such that they sum to one for an infinite series. This estimate can also be written in the recursive form

$$\hat{\Sigma}_{\tau|\tau-1} = (1-\lambda_e)E_{\tau-1}E'_{\tau-1} + \lambda_e\hat{\Sigma}_{\tau-1|\tau-2}, \tag{4.6}$$

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 taken in this dissertation is to estimate $\Sigma_{0|-1}$ from the sample covariance of the initial 10% of the data set.

RiskMetrics [40] provides an estimation technique for λ_e based on the root mean squared prediction error (RMSE) of $(e_\tau^i)^2$:

$$RMSE_2^i(\lambda_e) = \sqrt{\frac{1}{\tau_m}\sum_{\tau=1}^{\tau_m}\left((e_\tau^i)^2 - \hat{\sigma}^2_{\tau|\tau-1,ii}(\lambda_e)\right)^2}, \tag{4.7}$$

where $\hat{\sigma}^2_{\tau|\tau-1,ii}(\lambda_e)$ is a diagonal component of $\hat{\Sigma}_{\tau|\tau-1}$ in equation (4.6). Since the data series is assumed to have zero mean, $\mathrm{E}_{\tau-1}\left(e_\tau^i\right)^2 = \sigma^2_{\tau|\tau-1,ii}$, so the prediction error of $(e_\tau^i)^2$ is the difference of terms that is squared in equation (4.7). A single optimal estimate λ_e^* for the decay factor is computed from the RMSE of each univariate series through the formula

$$\lambda_e^* = \sum_{i=1}^{n} \phi_i \lambda_i^*, \tag{4.8}$$

where

$$\lambda_i^* = \underset{\lambda}{\operatorname{argmin}}\ RMSE_2^i(\lambda), \qquad \theta_i = \frac{RMSE_2^i(\lambda_i^*)}{\sum_{k=1}^n RMSE_2^k(\lambda_k^*)}, \qquad \phi_i = \frac{\theta_i^{-1}}{\sum_{k=1}^n \theta_k^{-1}}. \tag{4.9}$$

Using this technique, RiskMetrics recommends typical parameter values of $\lambda_e = 0.94$ for daily data and $\lambda_e = 0.97$ for monthly data. Based on the experience of this author, reducing λ_e removes more of the volatility clustering, but the distribution of the standardized residuals becomes more peaked with heavier tails.

4.3.2 Stable EWMA

The ideas of the previous subsection are extended to the sub-Gaussian case in [32] by allowing a time-varying dispersion matrix. Similar to the Gaussian case, exponential weights are applied to the moment estimators in equations (4.3-4.4),

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

Table 4.5: Comparison of the optimal decay factors λ_e^* under the normality and stable assumptions using the selection criterion based on $RMSE_p^i$.

yielding the equations:

$$\hat{q}^p_{\tau|\tau-1,jj} = (1-\lambda_e)\left|e^j_{\tau-1}\right|^p A(p) + \lambda_e \hat{q}^p_{\tau-1|\tau-2,jj} \tag{4.10}$$

$$B_{\tau|\tau-1,ij} = (1-\lambda_e) e^i_{\tau-1}\left(e^j_{\tau-1}\right)^{\langle q-1\rangle} A(q) + \lambda_e B_{\tau-1|\tau-2,ij} \tag{4.11}$$

$$\hat{q}^2_{\tau|\tau-1,ij} = B_{\tau|\tau-1,ij}\hat{q}^{2-q}_{\tau|\tau-1,jj}, \qquad i \neq j, \tag{4.12}$$

and the symmetric estimator for the time-varying dispersion matrix is given by $\hat{Q}_{\tau|\tau-1} = \left(\hat{q}^2_{\tau|\tau-1,ij} + \hat{q}^2_{\tau|\tau-1,ji}\right)/2$. 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 $|e^i_\tau|^p$. They note that $\mathrm{E}_{\tau-1}\left(|e^i_\tau|^p\right) = q^p_{\tau|\tau-1,ii}/A(p)$ and suggest to minimize the following RMSE error for each univariate series:

$$RMSE^i_p(\lambda_e) = \sqrt{\frac{1}{\tau_m}\sum_{\tau=1}^{\tau_m}\left(A(p)\left|e^i_\tau\right|^p - \hat{q}^p_{\tau|\tau-1,ii}(\lambda_e)\right)^2}.$$

The single optimal decay factor λ_e^* is then found by replacing $RMSE_2^i$ with $RMSE_p^i$ in equations (4.8-4.9). Using the VAR(1) residuals of the ALM data, this technique is applied in both the Gaussian and sub-Gaussian cases with $p = \alpha/3$. A grid for λ is constructed with increments of 0.001, and $RMSE_p^i(\lambda)$ is minimized over this grid. In both cases, a value of $\lambda_e = 0.95$ for equations (4.10-4.12) is found to be appropriate. The exact values of λ_e^* are found in Table (4.5).

There are difficulties in implementing the SEWMA model for the ALM residuals: 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, 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 from the SEWMA model are not inputted into the ALM optimization problem.

4.3.3 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 the governing Gaussian distribution and the scale parameters of the individual univariate series. First, one needs the following result: if

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

and g and s are independent, then

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

See [52], and for an application in finance, [51]. If the governing Gaussian distribution G_τ has a time-varying covariance matrix $\Sigma_{\tau|\tau-1} = \left(\sigma^2_{\tau|\tau-1,ij}\right)$ 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 result and equation (4.2) suggest a way to

model E_τ with a time-varying distribution similar to the sub-Gaussian:

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

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

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

When generating a sample for E_τ, the samples of s_τ^i, $i = 1, ..., n$, are taken from the same random seed so that the above equations are close to the sub-Gaussian representation in equation (4.2) 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

$$\hat{q}^{p_i}_{\tau|\tau-1,i} = (1-\lambda_e)\left|e^j_{\tau-1}\right|^{p_i} A(p_i) + \lambda_e \hat{q}^{p_i}_{\tau-1|\tau-2,i},$$

which is similar to equation (4.10), and it is reasonable to take $p_i = \alpha_i/3$ for financial series. To obtain the estimate for the covariance of the governing Gaus-

sian, 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

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

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

4.4 VaR Backtesting

The forecasting performances of the EWMA and SSEWMA models are examined by comparing the predicted VaRs with the observed returns as done in [50]. From the definition of VaR, the null hypothesis to test is

$$\mathrm{P}\left(r_\tau < -\mathrm{VaR}_\beta(\tau)\right) = 1-\beta, \tag{4.13}$$

for a return series $\{r_\tau\}$. This hypothesis is tested for each ALM return series $r^i = \{r_\tau^i, \tau = 1, ..., 237\}$, $i = 1, ..., 6$, 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 $\overline{\mathrm{VaR}}_\beta(\tau)$, for $\tau = 101, ..., 237$, be the estimate of $\mathrm{VaR}_\beta(\tau)$ from a

model calibrated to $\{r_{\tilde{\tau}}, \tilde{\tau} = \tau - 100, ..., \tau - 1\}$. If equation (4.13) holds, then

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

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

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

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

$$\sum_{k=1}^{\mathcal{X}} \begin{pmatrix} 137 \\ k \end{pmatrix} (1-\beta)^k \beta^{137-k} \leq \delta/2, \tag{4.14}$$

or,

$$\sum_{k=1}^{\mathcal{X}} \begin{pmatrix} 137 \\ k \end{pmatrix} (1-\beta)^k \beta^{137-k} \geq 1 - \delta/2. \tag{4.15}$$

The number of exceedings and the corresponding p-values for each ALM return series are contained in Tables (4.6-4.7). The conclusions are:

- At 1% level of significance, neither the EWMA or the SSEWMA model is rejected for any value of β.
- At 5% level of significance, the EWMA model is rejected three times for $\beta = 0.99$ and once for $\beta = 0.95$ while the SSEWMA model is never rejected. 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-

	Exceedings and p-values					
β	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 4.6: Number of VaR_β exceedings in 137 data points with corresponding p-values under the normality assumption.

	Exceedings and p-values					
β	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 4.7: Number of VaR_β exceedings in 137 data points with corresponding p-values under the stable assumption ($\alpha = 1.8$).

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.

Chapter 5

Results for the 2-Stage Problem

The ALM optimization problem of Chapter 3 is now solved using scenarios generated from the time series models of Chapter 4. First, efficient frontiers are developed for the 2-stage problem with scenarios from the EWMA and SSEWMA models, and post-optimality 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 normality 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 12 month stage leaves too few data points in the backtesting, so the decision is made to shorten the stage to cover a 6 month period. In addition to giving more points for comparison in the backtesting, time series models typically generate more reliable scenarios over shorter horizons.

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 do not allow much larger scenario trees to be considered. The first stage scenarios are created by simulating 10^4 sample paths of the time series model out to 6 months, and the second stage scenarios are created by simulating another 10^3 sample paths out an additional 6 months for each of the first stage scenarios. Scenario reduction and bundling using the methods of probability metrics are also attempted in order to created a better set of first stage scenarios, but these methods cannot 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 tails of the return distributions better than the normality assumption.

For the efficient frontiers and at the start date of the dynamic 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 has 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.

When transaction costs are considered, 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 transaction costs, as a percent of wealth traded, is obtained from data on mutual funds in [13]. Given in that report, the median trading cost (TC) is 0.70% of fund assets per year:

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

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

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

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

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

Combining equations yields

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

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 $i = 1, ..., 5$, are used in the optimization problem.

All computations are carried out in Matlab on a Linux cluster of 10 nodes, each with 2 processors and at least 2 GB of RAM. The parallel implementation of the solution algorithm is run on either four or five processors. Solution times

typically vary between a couple hours and one day. Since incorporating transaction costs nearly triples the number of decision variables, solving the problem with transaction costs is more time-consuming than without.

5.1 Efficient Frontiers

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

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

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 CVaR confidence level of 95% is used in this section to emphasize the differences between the normality and stable assumptions. For the remainder of this dissertation, 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, $\text{E}(sw_3)$, the efficient frontier is obtained by varying λ in the minimization objective: $\lambda\rho_2 - (1-\lambda)\text{E}(sw_3)$. Figure (5.1) contains three different efficient frontiers:

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

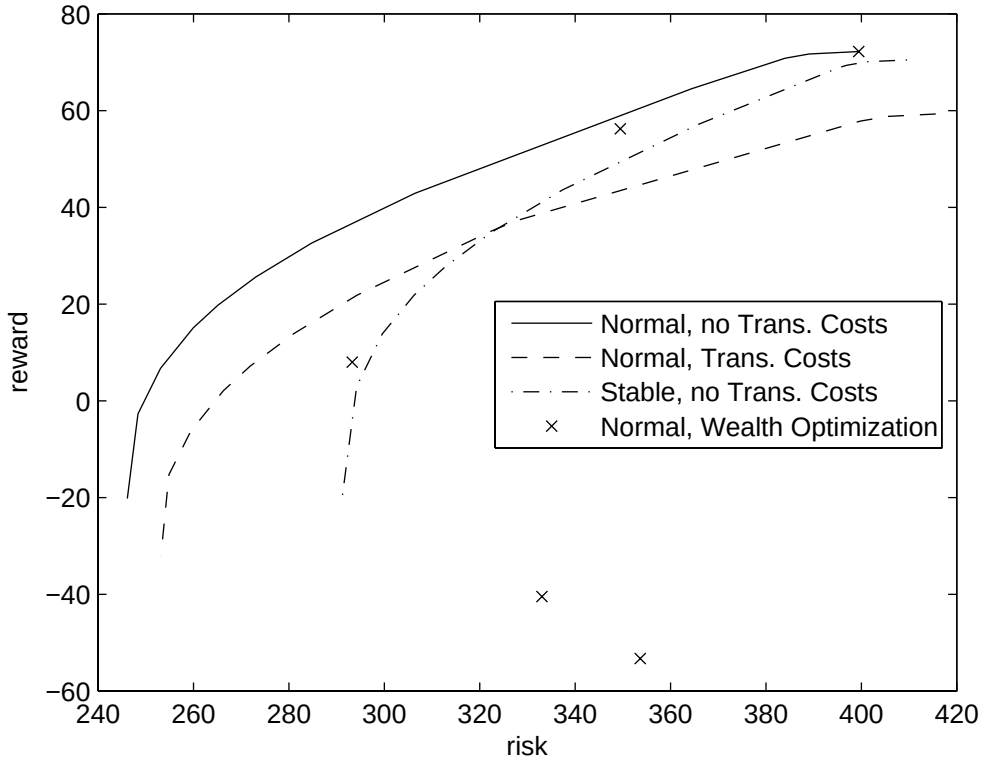

Figure 5.1: Efficient frontiers under the normality and stable assumptions for $\beta = 0.95$.

Table (4.2).

The optimal allocations, as percents of the initial wealth, are found in Tables (5.1-5.3). 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

					Optimal First Stage Allocations				
λ	E(sw_3)	ρ_2	CVaR($-sw_2$)	CVaR($-sw_3$)	Cash	Bonds	Eq.	Int. Eq.	Mort.
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 5.1: Efficient frontier under the normality assumption with $\beta = 0.95$.

					Optimal First Stage Allocations				
λ	E(sw_3)	ρ_2	CVaR($-sw_2$)	CVaR($-sw_3$)	Cash	Bonds	Eq.	Int. Eq.	Mort.
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 5.2: Efficient frontier under the stable assumption with $\beta = 0.95$.

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 normality assumption, are also included in Figure (5.1). The optimal allocations, found in Table (5.4), 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

					Optimal First Stage Allocations				
λ	E(sw_3)	ρ_2	CVaR($-sw_2$)	CVaR($-sw_3$)	Cash	Bonds	Eq.	Int. Eq.	Mort.
0	59.28	415.61	328.26	502.96	0	0	0	99.0050	0
0.10	58.80	405.57	328.26	482.89	0	0	0	99.0050	0
0.20	57.83	399.83	328.26	471.41	0	0	0	99.0050	0
0.25	37.16	327.52	251.65	403.39	0	40.0000	0	59.4030	0
0.30	34.81	321.78	247.18	396.37	0	42.6534	0	56.7496	0
0.35	21.80	294.29	217.23	371.35	0	61.8604	0	37.5426	0
0.40	13.93	281.12	203.63	358.61	0	72.4813	0	26.9217	0
0.45	7.14	271.91	194.45	349.36	0	81.0294	0	18.3736	0
0.50	2.07	266.31	189.48	343.14	0	86.5774	0	12.8256	0
0.60	-6.25	259.47	184.36	334.58	0	94.1698	0	5.2331	0
0.75	-15.27	254.86	181.98	327.75	0	99.4030	0	0	0
1.00	-31.96	253.21	181.98	324.45	0	99.4030	0	0	0

Table 5.3: Efficient frontier under the normality assumption with transaction costs and $\beta = 0.95$.

					Optimal First Stage Allocations				
λ	E(sw_3)	ρ_2	CVaR($-sw_2$)	CVaR($-sw_3$)	Cash	Bonds	Eq.	Int. Eq.	Mort.
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 5.4: Wealth optimization under the normality assumption with no transaction costs and $\beta = 0.95$.

and E(sw_3) because of the linearity of the problem.

5.2 Post-optimality Analysis

A short review of the contamination technique for post-optimality analysis is found in the work of Dupačová et al. [18]. The basic technique 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. These authors consider problems of the form

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

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

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

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

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

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

$$\begin{aligned}(1-\psi)\phi(P_1) + \psi\phi(P_2) &\leq \phi(P_\psi) \\ &\leq \min\left\{(1-\psi)\phi(P_1) + \psi F(x_1(P_1), P_2), (1-\psi)F(x(P_2), P_1) + \psi\phi(P_2)\right\},\end{aligned} \tag{5.1}$$

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 normality 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

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

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 (5.1-5.2), the optimal allocations under both distributional assumptions 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 (5.1) 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 normality and stable assumptions 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.

5.3 Portfolio Backtesting

The results from two rounds of backtesting are now 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 is 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.

5.3.1 Dynamic Backtesting: 1-Stage versus 2-Stage

This subsection 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 the optimal allocation that minimizes

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

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 normality 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 6 month period beginning in July, 1993. It is again assumed that the pension fund is initially fully funded with 40% of its wealth in the bond index and 60% of its 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 normality 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 produces only 22 values of the surplus wealth for comparison, the CVaR confidence level is reduced to 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 wealths that are used as the initial conditions in the optimization problems vary over the time horizon and are different for the various 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:

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

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 still cannot make a direct comparison because the initial asset wealths also vary for the different assumptions over the horizon. The measure of realized risk, $\tilde{\rho}$, used in the comparison is the CVaR with $\beta = 0.80$ 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

	$\tilde{\rho}$	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 5.5: Dynamic backtesting results.

of assets to liabilities, not just the asset wealths. Values of $\tilde{\rho}$ and the final surplus wealth are found in Table (5.5). For comparison, this table also includes values for the fixed-mixed rule of 40% bonds and 60% equities, and the rule of 100% bonds. Under both the normality 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 normality assumption gives lower values of $\tilde{\rho}$. The fixed-mixed rules are no comparison with the stochastic programs.

Figures (5.2-5.4) 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 (5.6), and the optimal allocations are in Tables (5.7-5.10). These tables also include the percents of asset wealth lost to transaction costs.

An additional comparison of the performance of the stable and normal distributions is obtained by VaR backtesting similar to Section 4.4. Under the optimal allocations, the forecasted VaRs of the surplus wealth are compared with the observed surplus wealths. The number of exceedings of $\text{VaR}_{0.80}$ has a binomial distribution, $\text{Bin}(22, .2)$, and the rejection rule in equations (4.14-4.15) is easily modified to this situation. While it is very difficult to reject with only 22 points,

the resulting p-values of the test can be compared: In both the 1-stage and 2-stage problems, the normality assumption has 5 exceedings (p-value=0.5429) and the stable assumption has 4 exceedings (p-value=0.9142). This indicates that the stable assumption does a better job in forecasting the VaR of the surplus wealth.

In this round of backtesting, the expectation is that running the 2-stage problem on a rolling horizon should be less risky than the 1-stage problem, as $\tilde{\rho}$ indicates. Besides looking at scenarios farther into the future, the 2-stage problem considers the loss of wealth in rebalancing in the future, which can be costly for a significant change in allocations. If the distributions are independent over time, and if there are no transaction costs, the major advantages of the 2-stage model are lost.

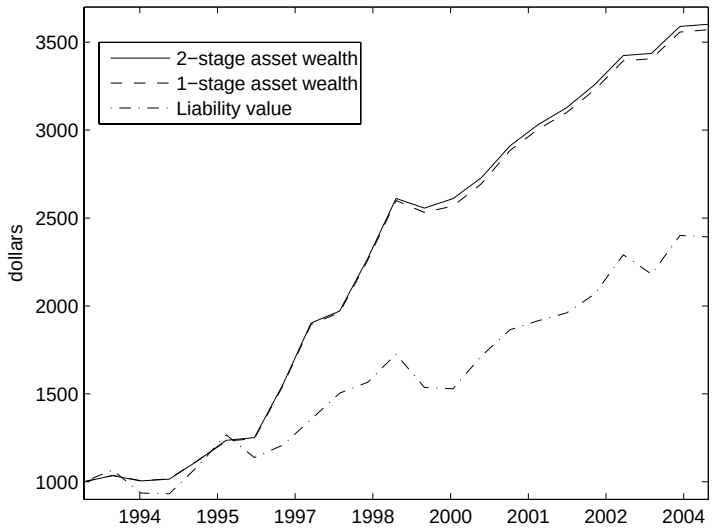

Figure 5.2: Dynamic backtesting: 1-stage versus 2-stage under the normality assumption.

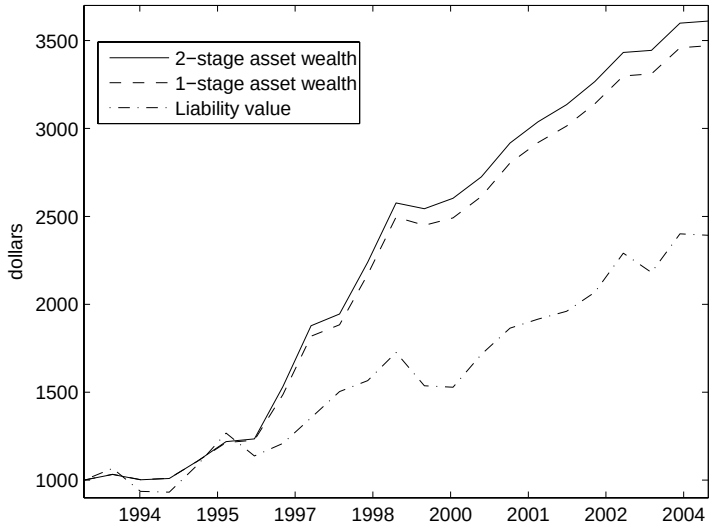

Figure 5.3: Dynamic backtesting: 1-stage versus 2-stage under the stable assumption.

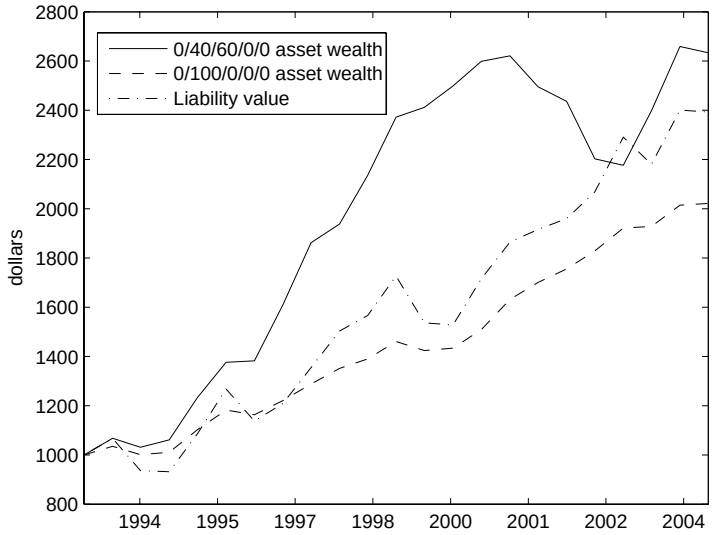

Figure 5.4: Dynamic backtesting: Fixed-mixed rules.

		Asset Wealth					
Date	Liability	1-stage		2-stage		Fixed-mixed	
	Value	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 5.6: Dynamic backtesting: Realized liability values 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 5.7: Dynamic backtesting: Allocations (as a percent of asset wealth) for the 1-stage optimization problem under the normality assumption with $\beta = 0.80$.

Date	Cash	Bonds	Equities	Intern. Equities	Mortgages	Transaction Costs
(initial)	0	40	60	0	0	
7/93	0	93.8328	0.5981	4.9780	0	0.5911
1/94	0	92.7472	0	7.2347	0	0.0181
7/94	0	88.2931	5.1188	6.5367	0	0.0514
1/95	0	94.6421	5.2968	0	0	0.0611
7/95	0	62.4417	37.2428	0	0	0.3155
1/96	0	1.3738	98.0320	0	0	0.5942
7/96	0	14.9073	84.9563	0	0	0.1364
1/97	0	0	99.8715	0	0	0.1285
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	79.3499	19.8526	0	0	0.7975
7/99	0	78.3104	21.6842	0	0	0.0054
1/00	0	85.5057	14.4138	0	0	0.0806
7/00	0	93.7601	6.1600	0	0	0.0798
1/01	0	99.9451	0	0	0	0.0549
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 5.8: Dynamic backtesting: Allocations (as a percent of asset wealth) for the 1-stage optimization problem under the stable 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 5.9: Dynamic backtesting: First stage allocations (as a percent of asset wealth) for the 2-stage optimization problem under the normality assumption with $\beta = 0.80$.

Date	Cash	Bonds	Equities	Intern. Equities	Mortgages	Transaction Costs
(initial)	0	40	60	0	0	
7/93	0	94.9591	0	4.4438	0	0.5970
1/94	0	93.0800	0	6.8995	0	0.0205
7/94	0	88.2264	6.2331	5.4779	0	0.0626
1/95	0	93.5050	6.4439	0	0	0.0511
7/95	0	56.2840	43.3517	0	0	0.3643
1/96	0	0	99.4541	0	0	0.5459
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	71.8792	27.3984	0	0	0.7224
7/99	0	64.1376	35.7941	0	0	0.0683
1/00	0	77.5415	22.3131	0	0	0.1454
7/00	0	92.3947	7.4606	0	0	0.1447
1/01	0	99.9334	0	0	0	0.0666
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 5.10: Dynamic backtesting: First stage allocations (as a percent of asset wealth) for the 2-stage optimization problem under the stable assumption with $\beta = 0.80$.

5.3.2 Static Backtesting: Stable versus Normal

By changing the initial conditions and removing the transaction costs, a better comparison is obtained for the distributional assumptions. To eliminate the difficulties of the previous subsection in measuring the risk of the realized surplus wealths, it is assumed that the pension fund is fully funded ($sw_1 = 0$) with an initial asset wealth and liability value of \$1000 for every run of the optimization problem. Transaction costs are not included because the optimal allocations would depend on the allocations from the prior 6 month period, which would not have a total asset wealth of \$1000. With these assumptions, one can directly compare the CVaR of the realized surplus wealths, denoted by $\tilde{\rho}_1$, and not use $\tilde{\rho}$.

This static backtesting is carried out under both distributional assumptions for the minimum risk portfolios. The time series models are again fit to a moving window of 100 data points that is shifted by 6 points each instance. A confidence level of 80% is used in ρ_1, ρ_2, and $\tilde{\rho}_1$, and the tail index for the SSEWMA model is again assumed to be 1.8. The results are found in Table (5.11), and they indicate that the stable assumption does a slightly better job of minimizing $\tilde{\rho}_1$. The optimal allocations and the resulting surplus wealth for each run of the optimization problem are in Tables (5.12-5.15). In 12 out of the 22 runs of the optimization problem, the normality and stable assumptions result in the same optimal allocations, which left only 10 points for comparison. While the stable assumption performs better here, more backtesting should be carried out on additional data sets to obtain a more reliable comparison.

As observed, one should expect that the 1-stage problem results in a lower value of $\tilde{\rho}_1$ in this round of backtesting. Since the initial conditions of the optimization problems are reset for every 6 month period, the second stage of the 2-stage problem no longer matters. The 1-stage problem directly minimizes ρ_1, which

	$\tilde{\rho}_1$	$s\bar{w}$
1-stage Normal	69.29	19.16
1-stage Stable	69.03	17.11
2-stage Normal	71.84	21.13
2-stage Stable	70.62	20.25

Table 5.11: Static backtesting results.

is the risk measure that is compared, while the 2-stage problem typically has a larger first stage CVaR because it also considers the second stage CVaR.

Sim Num	Cash	Bonds	Equities	Intern. Equities	Mortgages	Surplus Wealth
1	0	94.2996	0	5.7004	0	-27.13
2	0	88.8915	0	11.1085	0	95.28
3	0	77.9172	19.7553	2.3275	0	19.10
4	0	95.3882	4.6118	0	0	-69.11
5	0	41.6561	58.3439	0	0	-52.28
6	0	0	100	0	0	120.04
7	0	15.0808	84.9192	0	0	149.96
8	0	0	100	0	0	104.85
9	0	0	100	0	0	-74.12
10	0	0	100	0	0	109.97
11	0	0	100	0	0	47.97
12	0	86.0574	13.9426	0	0	95.02
13	0	48.3285	51.6715	0	0	37.37
14	0	65.4060	34.5940	0	0	-76.88
15	0	93.0775	6.9225	0	0	-14.22
16	0	100	0	0	0	15.02
17	0	100	0	0	0	8.67
18	0	100	0	0	0	-13.29
19	0	100	0	0	0	-57.03
20	0	100	0	0	0	51.62
21	0	100	0	0	0	-55.95
22	0	100	0	0	0	6.65

Table 5.12: Static backtesting: Allocations for the 1-stage problem, and the surplus wealths resulting from those allocations, under the normality assumption with $\beta = 0.80$.

Sim Num	Cash	Bonds	Equities	Intern. Equities	Mortgages	Surplus Wealth
1	0	95.0270	0	4.9730	0	-27.81
2	0	91.5586	0	8.4414	0	94.03
3	0	78.2348	16.6446	5.1206	0	16.29
4	0	96.0098	3.9902	0	0	-69.86
5	0	49.5305	50.4695	0	0	-58.02
6	0	0	100	0	0	120.04
7	0	41.2263	58.7737	0	0	99.74
8	0	0	100	0	0	104.85
9	0	0	100	0	0	-74.12
10	0	0	100	0	0	109.97
11	0	0	100	0	0	47.97
12	0	73.8613	26.1387	0	0	103.55
13	0	40.6498	59.3502	0	0	41.16
14	0	78.6774	21.3226	0	0	-74.14
15	0	94.1093	5.8907	0	0	-12.98
16	0	100	0	0	0	15.02
17	0	100	0	0	0	8.67
18	0	100	0	0	0	-13.29
19	0	100	0	0	0	-57.03
20	0	100	0	0	0	51.62
21	0	100	0	0	0	-55.95
22	0	100	0	0	0	6.65

Table 5.13: Static backtesting: Allocations for the 1-stage problem, and the surplus wealths resulting from those allocations, under the stable assumption with $\beta = 0.80$.

Sim Num	Cash	Bonds	Equities	Intern. Equities	Mortgages	Surplus Wealth
1	0	95.3620	0	4.6380	0	-28.12
2	0	89.2761	0	10.7239	0	95.10
3	0	78.4857	20.5315	0.9828	0	20.22
4	0	99.5282	0.4718	0	0	-74.09
5	0	32.4868	67.5132	0	0	-45.59
6	0	0	100	0	0	120.04
7	0	0	100	0	0	178.93
8	0	0	100	0	0	104.85
9	0	0	100	0	0	-74.12
10	0	0	100	0	0	109.97
11	0	0	100	0	0	47.97
12	0	78.8588	21.1412	0	0	100.05
13	0	15.3084	84.6916	0	0	53.67
14	0	40.1295	59.8705	0	0	-82.12
15	0	90.2678	9.7322	0	0	-17.62
16	0	100	0	0	0	15.02
17	0	100	0	0	0	8.67
18	0	100	0	0	0	-13.29
19	0	100	0	0	0	-57.03
20	0	100	0	0	0	51.62
21	0	100	0	0	0	-55.95
22	0	100	0	0	0	6.65

Table 5.14: Static backtesting: First stage allocations for the 2-stage problem, and the surplus wealths resulting from those allocations, under the normality assumption with $\beta = 0.80$.

Sim Num	Cash	Bonds	Equities	Intern. Equities	Mortgages	Surplus Wealth
1	0	95.6634	0	4.3366	0	-28.40
2	0	91.8795	0	8.1205	0	93.88
3	0	78.7285	18.2053	3.0661	0	18.12
4	0	100	0	0	0	-74.66
5	0	42.5472	57.4528	0	0	-52.93
6	0	0	100	0	0	120.04
7	0	16.0125	83.9875	0	0	148.17
8	0	0	100	0	0	104.85
9	0	0	100	0	0	-74.12
10	0	0	100	0	0	109.97
11	0	0	100	0	0	47.97
12	0	63.4409	36.5591	0	0	110.83
13	0	9.6377	90.3623	0	0	56.47
14	0	66.5203	33.4797	0	0	-76.65
15	0	93.4359	6.5641	0	0	-13.79
16	0	100	0	0	0	15.02
17	0	100	0	0	0	8.67
18	0	100	0	0	0	-13.29
19	0	100	0	0	0	-57.03
20	0	100	0	0	0	51.62
21	0	100	0	0	0	-55.95
22	0	100	0	0	0	6.65

Table 5.15: Static backtesting: First stage allocations for the 2-stage problem, and the surplus wealths resulting from those allocations, under the stable assumption with $\beta = 0.80$.

Chapter 6

Conclusions and Future Work

An ALM problem for a pension fund is designed to find the optimal allocations among various asset classes while considering the liability obligation. A multistage stochastic program with recourse optimizes the tradeoff between the expected final surplus wealth and a multi-period risk measure. The major advantages of the recourse program are that it considers portfolio reallocation with linear transaction costs and models uncertainty through scenarios. A weighted average of the CVaR of the negative surplus wealth at each stage is used as the risk measure because it permits a deterministic equivalent form that has a piecewise linear objective with linear constraints. The scenarios are generated from time series models calibrated to monthly data representative of a pension fund. Two different multivariate distributional assumptions are made for the innovations of a VAR(1) model. Both assumptions use an EWMA process for the innovations in an attempt to capture the volatility clustering, but the first uses the normal distribution while the second uses a distribution similar to the sub-Gaussian.

As one would expect, dynamic portfolio backtesting indicates that the 2-stage recourse problem performs better than the 1-stage problem. The 2-stage problem considers the loss of wealth when the portfolio is reallocated and results in a higher

final surplus wealth and smaller realized risk than the 1-stage problem.

The monthly data provides a short time series and fitting a time-varying multivariate stable distribution proves to be difficult. While the VaR backtesting of the time series models for the returns favors the stable distribution, it is not clear which distributional assumption is superior in the portfolio backtesting. In dynamic portfolio backtesting, the normality assumption results in a smaller value of realized risk, but in static backtesting, the stable assumption performs better. It is desirable to conduct additional backtesting of the ALM program on a more extensive data set to obtain a better comparison.

The multistage program is much more of a computational challenge than the more common single-stage programs. Huge numbers of scenarios are generated to obtain reliable results for the comparisons in the backtesting, and extending beyond the 2-stage program requires a great reduction in the number of scenarios to solve the program in any reasonable amount of time. The techniques involving probability metrics should be examined in more detail as they were found to have a computational issue of their own for a large number of scenarios.

Bibliography

[1] Carlo Acerbi and Dirk Tasche. Expected Shortfall: A Natural Coherent Alternative to Value at Risk. Available at 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, Jul 2002.

[3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228, Jul 1999.

[4] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer, New York, 1997.

[5] Guus C. E. Boender, Paul van Aalst, and Fred Heemskerk. Modelling and Management of Assets and Liabilities of Pension Plans in the Netherlands. In John M. Mulvey and William T. Ziemba, editors, *Worldwide Asset and Liability Modeling*. Cambridge University Press, 1999.

[6] 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.

[7] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, second edition, 2002.

[8] John Y. Campbell, Andrew W. Lo, and A. Craig MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey, 1997.

[9] David R. Carino, Terry Kent, David H. Myers, Celine Stacy, Mike Sylvanus, Andrew L. Turner, Kanji Watanabe, and William T. Ziemba. The Russell-Yasuda Kasai Model: An Asset/Liability Model for a Japanese Insurance Company using Multistage Stochastic Programming. In John M. Mulvey and William T. Ziemba, editors, *Worldwide Asset and Liability Modeling*. Cambridge University Press, 1999.

[10] 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, Jul-Aug 1998.

[11] David R. Carino and William T. Ziemba. Formulation of the Russell-Yasuda Kasai Financial Planning Model. *Operations Research*, 46(4):433–449, Jul-Aug 1998.

[12] Yair Censor and Stavros A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 1997.

[13] 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, Nov 1999.

[14] Ron S. Dembo. Scenario Immunization. In Stavros A. Zenios, editor, *Financial Optimization*. Cambridge University Press, United Kingdom, 1999.

[15] M.A.H. Dempster and R.T. Thompson. EVPI-based Importance Sampling Solution Procedures for Multistage Stochastic Linear Programs on Parallel MIMD Architectures. *Annals of Operations Research*, 90:161–184, 1999.

[16] Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton, New Jersey, second edition, 1996.

[17] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario Reduction in Stochastic Programming: An Approach Using Probability Metrics. *Mathematical Programming*, 94:493–511, 2003.

[18] Jitka Dupačová, Marida Bertocchi, and Vittorio Moriggia. Postoptimality for Scenario Based Financial Planning Models with an Application to Bond Portfolio Management. In John M. Mulvey and William T. Ziemba, editors, *Worldwide Asset and Liability Modeling*. Cambridge University Press, 1999.

[19] Jitka Dupačová, Giorgio Consigli, and Stein W. Wallace. Scenarios for Multistage Stochastic Programs. *Annals of Operations Research*, 100:25–53, 2000.

[20] Jitka Dupačová, Jan Hurt, and Josef Štěpán. *Stochastic Modeling in Economics and Finance*. Kluwer Academic Publishers, Netherlands, 2002.

[21] A. Eichhorn and W. Römisch. Polyhedral Risk Measures in Stochastic Programming. *SIAM Journal on Optimization*, 16:69–95, 2005.

[22] Eugene Fama. The Behavior of Stock Market Prices. *Journal of Business*, 38:34–105, 1965.

[23] William Feller. *An Introduction to Probability Theory and Its Applications, Volume 2*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, second edition, 1971.

[24] 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.

[25] Holger Heitsch and Werner Römisch. Generation of Multivariate Scenario Trees to Model Stochasticity in Power Management. Technical report, Institute of Mathematics, Humboldt-University Berlin.

[26] Beat Huggler. Modelling Hedge Fund Returns. Technical report, Swiss Federal Institute of Technology of Zurich, University of Zurich, Dec 2004.

[27] Milen Ivanov and Jivko Stoyanov. Two Stage Portfolio Optimisation: Modification of the L-shaped Method. Technical report, Finalytica, Jun 2003.

[28] S. Rao Jammalamadaka and Srikanth K. Iyer. A Simple Estimate of the Index of Stability for Symmetric Stable Distributions. *Stochastic Modelling and Applications*, 6(1):9–19, 2003.

[29] Tomasz J. Kozubowski, Anna K. Panorska, and Svetlozar T. Rachev. Statistical Issues in Modeling Multivariate Stable Portfolios. In Svetlozar T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, Handbooks in Finance. Elsevier Science, The Netherlands, 2003.

[30] Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints. *The Journal of Risk*, 4(2):11–27, 2002.

[31] M. I. Kusy and W. T. Ziemba. A Bank Asset and Liability Management Model. *Operations Research*, 34(3):356–376, May-Jun 1986.

[32] Fabio Lamantia, Sergio Ortobelli, and Svetlozar Rachev. Value at Risk with Stable Distributed Returns. Technical report, University of California, Santa Barbara, 2004. Available at http://www.pstat.ucsb.edu/research.

[33] Benoit Mandelbrot. Paretian Distributions and Income Maximization. *Quarterly Journal of Economics*, 76:57–85, 1962.

[34] Benoit Mandelbrot. New Methods in Statistical Economics. *Journal of Political Economy*, 71:421–440, 1963.

[35] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *Journal of Business*, 36:394–419, 1963.

[36] Benoit Mandelbrot. The Variation of Some Other Speculative Prices. *Journal of Business*, 40:393–413, 1967.

[37] 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.

[38] Harry M. Markowitz. Portfolio Selection. *Journal of Finance*, 7:77–91, Mar 1952.

[39] Richard C. Merton. An Intertemporal Capital Asset Pricing Model. *Econometrica*, 41(5):867–887, Sept 1973.

[40] J.P. Morgan. Riskmetrics - Technical Document. Technical report, Dec 1996.

[41] John Mulvey and Hercules Vladimirou. Stochastic Optimization Models for Investment Planning. *Annals of Operations Research*, 20:187–217, 1989.

[42] John Mulvey and Hercules Vladimirou. Applying the Progressive Hedging Algorithm to Stochastic Generalized Networks. *Annals of Operations Research*, 31:399–424, 1991.

[43] John Mulvey and Hercules Vladimirou. Solving Multistage Stochastic Networks: An Application of Scenario Aggregation. *Networks*, 21(6):619–643, 1991.

[44] John Mulvey and Hercules Vladimirou. Stochastic Network Programming for Financial Planning Problems. *Management Science*, 38(11):1642–1664, Nov 1992.

[45] Soren S. Nielsen and Stravos A. Zenios. Solving Multistage Stochastic Network Programs on Massively Parallel Computers. *Mathematical Programming*, 73:227–250, 1996.

[46] G. Ch. Pflug. Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization. *Mathematical Programming*, 89:251–271, 2001.

[47] 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.

[48] Svetlozar Rachev, Isabella Huber, and Sergio Ortobelli. Portfolio Choice Theory with Heavy Tailed Distributions. Technical report, University of California, Santa Barbara, 2004. Available at http://www.pstat.ucsb.edu /research/papers.

[49] Svetlozar Rachev and Stefan Mittnik. *Stable Paretian Models in Finance*. Series in Financial Economics and Quantitative Analysis. John Wiley & Sons, 2000.

[50] Svetlozar Rachev, Eduardo Schwartz, and Irina Khindanova. Stable Modeling of Market and Credit Value at Risk. Technical report, University of California, Santa Barbara, Oct 2001. Available at http://www.pstat.ucsb.edu/research/papers.

[51] 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.

[52] J. S. Rao and James B. Robertson. A Characterization of Gaussian Spaces. *Journal of Multivariate Analysis*, 10(4):599–602, Dec 1980.

[53] R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

[54] Markus Rudolf and William T. Ziemba. Intertemporal Surplus Management. *Journal of Economic Dynamics & Control*, 28:975–990, 2004.

[55] Ronald J. Ryan and Frank J. Fabozzi. The Pension Crisis Revealed. *The Journal of Investing*, Fall, 2005.

[56] Yesim Tokat, Svetlozar T. Rachev, and Eduardo S. Schwartz. Asset and Liability Management: A Review and Some New Results in the Presence of Heavy Tails. In Svetlozar T. Rachev, editor, *Handbook of Heavy Tailed Dis-*

tributions in Finance, Handbooks in Finance. Elsevier Science, The Netherlands, 2003.

[57] Nikolas Topaloglou, Hercules Vladimirou, and Stavros A. Zenios. CVaR Models with Selective Hedging for International Asset Allocation. *Journal of Banking and Finance*, 26:1535–1561, 2002.

[58] Stavros A. Zenios. Lecture Notes: Mathematical Modeling and Its Application in Finance. Available at http://www.hermes.ucy.ac.cy /zenios/teaching/399.001/index.html.

[59] Eric Zivot and Jianhui Wang. *Modeling Financial Time Series with S-Plus*. Springer, New York, 2003.