Term structure models in multistage stochastic programming: Estimation and approximation

Karl Frauendorfer and Michael Schürle

University of St. Gallen, Institute of Operations Research, St. Gallen, Switzerland

This paper investigates some common interest rate models for scenario generation in financial applications of stochastic optimization. We discuss conditions for the underlying distributions of state variables which preserve convexity of value functions in a multistage stochastic program. One- and multi-factor term structure models are estimated based on historical data for the Swiss Franc. An analysis of the dynamic behavior of interest rates generated with these models reveals several deficiencies which have an impact on the performance of investment policies derived from the stochastic program. While barycentric approximation is used here for the generation of scenario trees, these insights may be generalized to other discretization techniques as well.

Keywords: Stochastic programming, approximation, term structure models

AMS Subject classification: Primary 90C15; Secondary 91B28

1. Introduction

In most financial planning problems, decision makers must cope with various forms of uncertainty. For example, future interest rates are unknown, and variations in yields may have a strong impact on the value of a portfolio of fixed-income securities, the positions in a bank's balance sheet, etc. Traditionally, immunization techniques like duration matching are exploited to hedge against small changes in interest rates. The increasing volatility in financial markets has brought the shortcomings of such approaches to light as frequent rebalancings of portfolios are required which induce significant transaction costs. Moreover, in a number of applications, in particular from asset and liability management, there is also uncertainty with respect to the amount and timing of cash flows which cannot be taken into account sufficiently when calculating durations.

Stochastic multistage problems reflect the dynamics of random data and decisions more appropriately. Transactions may take place at discrete points in time over some horizon T. At each stage $t = 1, \ldots, T$, a decision has to be taken based on prevailing market conditions like observations of interest rates, asset prices, cash flows etc., and earlier decisions (i.e., an existing portfolio composition), but without knowing the future. The uncertainty of the problem is captured by constructing a set of scenarios that is representative for the entire universe of possible outcomes of the relevant risk factors. In most cases, scenarios are derived from assumptions on *stochastic processes* for the evolution of these factors which imply certain (joint) distributions.

A natural way to describe the dynamics of interest rates is to exploit *term structure models* that have been developed within the financial literature for the pricing of derivative instruments. For example, Mulvey [28] suggests the model of Brennan and Schwartz [3] which calculates the short and long spot rates based on a two-factor pair of diffusion equations. The remaining points on the yield curve are determined by making economic assumptions like no-arbitrage. Zenios [34] uses the model of Black, Derman, and Toy [2] to construct a binomial tree (lattice) whose paths describe the evolution of the short rate from which the term structure can be derived in each node. In general, the number of scenarios is too large to be considered in a single stochastic program. Hence, one has to choose a set of manageable size which can be determined, e.g., by sampling strategies (see e.g. [35]). When using Monte Carlo methods, the solution of the corresponding optimization problem may depend heavily on the selection of scenarios (for an analysis of the robustness of optimal values, see [13]).

As an alternative, we propose to exploit *approximation techniques* that replace the original distribution by a simpler one. Since the number of scenarios is reduced, the corresponding programs can be solved easily while they provide *exact* bounds to the original problems. This allows for an improvement of the approximation until a desired accuracy is achieved. Furthermore, we show under which conditions some of the most common term structure models can be combined with such an approach. Clearly, it is of utmost importance that the distributions implied by these models come as close as possible to the empirical distribution of observed interest rates. Therefore, we estimate the model parameters based on historical data for the Swiss Franc and investigate their statistical properties to assess if they are useful for the generation of interest rate scenarios. Finally, we conduct a case study to answer the question if the difficulties in developing stochastic optimization models are compensated by an improved performance compared to traditional approaches for a simple investment problem.

2. Multistage stochastic programming

2.1. Formal description

Within stochastic programming, the evolution of uncertain data over some planning horizon T can be formally described by a multi-dimensional stochastic process $(\omega_t, t = 1, \ldots, T)$ in discrete time on a common Borel space $(\Omega, \mathcal{B}^M)$ with compact $\Omega \subset \mathbb{R}^M$ (see [19–21]). Let P represent the (regular) joint probability measure of $\omega := (\omega_1, \ldots, \omega_T)$. The associated conditional measure with respect to ω_t is denoted $P_t(\cdot|\omega^{t-1})$ for $t = 1, \ldots, T$. For reasons of compactness, $\omega^t :=$

$(\omega_1, \ldots, \omega_t)$ represents the sequence of observations of $\omega_t \in \Omega_t \subset \mathbb{R}^{M_t}$ up to time t, where $\Omega_1 \times \ldots \times \Omega_T = \Omega$, $M_1 + \ldots + M_T = M$. Note that ω_0 denotes those data that are currently observed and, hence, deterministic.

At time $t = 0$, a decision $u_0 \in \mathbb{R}^{n_0}$ is made without knowing ω_t for the subsequent stages $t = 1, \ldots, T$. After ω_t was observed at time $t > 0$, the initial policy may be corrected by a new decision $u_t \in \mathbb{R}^{n_t}$ based on the known history of observations ω^t and decisions $u^t := (u_0, u_1, \ldots, u_t) \in \mathbb{R}^{n^t}$, $n^t = n_0 + \ldots + n_t$. In particular, u_t has to be independent of future outcomes $\omega_{t+1}, \ldots, \omega_T$. Therefore, the solution of the underlying stochastic optimization problem is a *recourse function* with the property

$$u(\omega) = (u_0, u_1(\omega^1), \ldots, u_T(\omega^T)) \in \mathbb{R}^n, \qquad n = n_0 + n_1 + \ldots + n_T,$$

known as *nonanticipativity.* The feasible set of the first decision u_0 is deterministic. For decisions at time $t = 1, \ldots, T$, it depends on previous decisions and observations up to t. Here, feasible decisions are given by a system of inequalities

$$\begin{array}{ll} f_0(u_0) \leq 0 & \\ f_t(u^t, \omega^t) \leq 0 & \quad t = 1, \ldots, T. \end{array} \tag{1}$$

The initial decision u_0 induces some (deterministic) costs $\rho_0(u_0)$. Again, in a subsequent stage t, the costs $\rho_t(u^t, \omega^t)$ are determined by the sequence of earlier decisions u^t and observations ω^t. It is assumed that $\rho_0(\cdot)$ and $\rho_t(\cdot, \cdot)$ are real-valued and $f_0(\cdot)$ and $f_t(\cdot, \cdot)$ are vector-valued functions defined on the corresponding Euclidian spaces, and that the feasible set given by (1) is convex, compact and non-empty for any ω. Furthermore, $\rho_t(\cdot, \cdot)$ are supposed to be convex in u^t for any random outcome ω^t. The objective is to find a nonanticipative recourse function $u(\cdot)$ that minimizes the expected value of the total costs $\rho_0(u_0) + \sum_{t=1}^T \rho_t(u^t, \omega^t)$ over the planning horizon and satisfies (1). The corresponding multistage stochastic program reads

$$\begin{array}{ll} \min & \left\{ \rho_0(u_0) + \int_\Omega \left[\sum_{t=1}^T \rho_t(u_0, u_1, \ldots, u_t, \omega_1, \ldots, \omega_t) \right] dP(\omega) \right\} \\ \text{s.t.} & f_0(u_0) \leq 0 \\ & f_t(u_0, u_1, \ldots, u_t, \omega_1, \ldots, \omega_t) \leq 0, \qquad t = 1, \ldots, T, \\ & u(\cdot) \text{ nonanticipative.} \end{array} \tag{2}$$

The meaning of the last (nonanticipativity) constraint is: Let ω^t and u^t satisfy $f_0(u_0) \leq 0, f_1(u^1, \omega^1) \leq 0, \ldots, f_t(u^t, \omega^t) \leq 0$, then there always exists some $u_{t+1}, \ldots, u_T$ for any $(\omega_{t+1}, \ldots, \omega_T)$, so that $u = (u_0, u_1, \ldots, u_T)$ is feasible with respect to (1). As a consequence, there always exists a feasible completion of the problem (2) for the remaining stages $t+1, \ldots, T$ independent of the realizations of ω_τ, $\tau = t+1, \ldots, T$, provided that the decisions u_τ in $\tau = 0, 1, \ldots, t$ are feasible. Note that this can be seen as a counterpart to relatively complete recourse in the two-stage case.

2.2. Solution techniques

If the random vector ω is continuously distributed, it can be seen from (2) that to evaluate the objective function for a given policy u, one has to calculate a M-dimensional integral with respect to the measure P describing the distribution of ω. In general, there is no analytical solution to (2). Therefore, numerical methods must be exploited that usually require much computational effort which increases rapidly with the desired accuracy and/or the dimension M of Ω.

A common approach is to choose a set $\mathcal{A}$ of *scenarios* that are considered as representative for the evolution of ω_t. Since some of these scenarios share a common history up to some stage t, their relations may be described by a tree. Let $\mathcal{A}_t(\omega^{t-1})$ denote the set of finitely many outcomes for ω_t conditioned on ω^{t-1} and $q_t(\omega_t|\omega^{t-1})$ the corresponding conditional probability of $\omega_t \in \mathcal{A}_t$. Then, a scenario tree and its path probabilities are formally characterized as

$$\mathcal{A} = \{\omega^T \in \Omega | \omega_t \in \mathcal{A}_t(\omega^{t-1}) \, \forall t > 1\}$$
$$q(\omega^T) = \prod_{t=1}^{T} q_t(\omega_t|\omega^{t-1}).$$

In other words, the original (continuous) joint probability measure P is replaced by Q, the joint probability measure of $\mathcal{A}$. The associated conditional probability measure is denoted by $Q_t(\cdot|\omega^{t-1})$. By approximating the original continuous distribution with a discrete one, the problem of evaluating the objective function for a given policy u reduces to the calculation of a sum, which can easily be performed, instead of solving an integral.

Intuitively, one can imagine that the accuracy of the solution depends on the choice and number of scenarios. Clearly, the problem size and, hence, the numerical effort grows rapidly with an increasing number. Therefore, a critical issue is to determine a "good" selection of scenarios whose size still remains manageable. For example, one may rely on Monte Carlo methods combined with powerful variance reduction techniques like importance sampling (see e.g. [11]) or the *expected value of perfect information* (see [8,12]). The latter allows to identify those scenarios that have a stronger impact on the solution of the stochastic optimization problem than others. In such simulation-based approaches, one obtains probabilistic error bounds (i.e., a confidence interval for the objective value) to quantify the accuracy of the solution (see e.g. [24]). Unfortunately, Monte Carlo methods often suffer from a low convergence speed until the desired quality of the solution is achieved, particularly in the multistage case.

On the other hand, there are often certain properties like convexity of value functions associated with the stochastic optimization problem that can be exploited (see [1,25]). In approximation techniques, the elements of the scenario tree $\mathcal{A}$ are determined in a way that an *exact* upper or lower bound to the original problem may be derived under certain conditions (see [14–17,19–21]). Typically, these bounds are linear or bilinear functions that can be easily integrated. A detailed analysis of the solution also allows for refinements of the discretization which can be controlled in a way that the increase in the number of scenarios is moderate. However, in the multistage case, one faces the *curse of dimensionality*, meaning that the scenario tree grows exponentially with the number of stages under consideration and the dimension of ω_t. Therefore, it is of utmost importance that the stochasticity of the problem can be described by a low-dimensional distribution in order to consider a sufficiently large planning horizon.

From statistical analysis, it is well known that the evolvement of interest rates can be modeled appropriately with only 2 or 3 risk factors (see [27]). Hence, the use of approximation schemes seems to be a natural choice. In the next section, the barycentric approximation technique is introduced which allows to determine exact bounds for problems of type (2) where the distributions may be conditioned on the past history.

3. Barycentric approximation

3.1. Saddle property of value functions

From now on, we concentrate on the case where the random vectors ω_t, $t = 1, \ldots, T$, can be decomposed into two components $\eta_t \in \Theta_t \subset \mathbb{R}^{K_t}$ and $\xi_t \in \Xi_t \subset \mathbb{R}^{L_t}$ with $\Omega_t = \Theta_t \times \Xi_t$, $M_t = K_t + L_t$. This helps to distinguish those random data that affect the objective from those that influence the constraints. If such a decomposition is not obvious, one may augment the probability space (for details, see [19]). For the sake of simplicity, we introduce the function

$$g_t(u^t, \omega^t) := \begin{cases} \rho_t(u^t, \eta^t) & \text{if } f_t(u^t, \xi^t) \le 0, \\ +\infty & \text{otherwise.} \end{cases}$$

which quantifies the costs in $t = 1, \ldots, T$ induced by the sequence of decisions u^t. Analogously, we define $g_0(u_0) := \rho_0(u_0)$, if $f_0(u_0) \le 0$, and $+\infty$ otherwise in the first stage. It is useful to consider the dynamic representation of the multistage program (2) for the following analysis, given in terms of *value functions* (with $\phi_{T+1}(\cdot) := 0$):

$$\phi_t(u^{t-1}, \omega^t) := \min\Big\{ g_t(u^t, \omega^t) + \int_{\Omega_{t+1}} \phi_{t+1}(u^t, \omega^{t+1}) dP_{t+1}(\omega_{t+1}|\omega^t) \big| u_t \ge 0 \Big\} \quad (3)$$

for $t = 1, \ldots, T$, whereas in the first stage, the problem is

$$\min\Big\{ g_0(u_0) + \int_{\Omega_1} \phi_1(u_0, \omega_1) dP(\omega_1) \big| u_0 \ge 0 \Big\}. \quad (4)$$

To ensure that the problems (3) and (4) can be solved, the following conditions for $t = 1, \ldots, T$ are sufficient: (i) $\Theta_t \times \Xi_t$ are compact, convex and cover the support of $\omega_t = (\eta_t, \xi_t)$, (ii) $\rho_t(\cdot, \cdot)$ are continuous saddle functions, which are concave in η_t, and (iii) the constraints can be written in the form $f_t(u^t, \xi^t) = d_t(u_t) - e_t(u^{t-1}, \xi^t)$ with $d_t(\cdot)$ convex and $e_t(\cdot, \cdot)$ linear-affine (see Proposition 2.1 in [21]).

Barycentric approximation was originally introduced in [19] in the context of two-stage stochastic programming and can be seen as a generalization of the inequalities due to Jensen and Edmundson-Madansky (see e.g. [1,25] and the references herein) that allow for bounding the expectation of a saddle function. From the representation (3), it becomes obvious that a multistage problem can be seen as a series of nested two-stage programs. This motivates the extension to the multistage case.

Assumption (i) implies that the problem of the last stage T is convex. According to (ii), the cost function g_T is a lower closed saddle function. Since $\phi_{T+1} = 0$, it follows immediately that the value function $\phi_T(u^{T-1}, \omega^T)$ in T is also a saddle function which is convex in (u^{T-1}, ξ^T) and concave in η^T. To apply the concept of barycentric approximation to the multistage case, it is necessary that the saddle property of the value function in T is inherited to the remaining stages $T-1, \ldots, 1$. When calculating the expectation of the value function, the probability measure P_{t+1} depends on ω^t. As a consequence, the saddle property is not preserved in general due to the integration with respect to $P_{t+1}(\omega_{t+1}|\omega^t)$.

However, if the distribution functions are of the form $\widetilde{P}_{t+1}(\omega_{t+1} + H_{t+1}(\omega^t))$, where H_{t+1} is a linear mapping, then it can be shown that the relevant expectation functionals $(E_{t+1}\phi_{t+1})(u^t, \omega^{t+1})$ are continuous saddle functions. In other words, if the distribution function of P_{t+1} depends linearly on the past, the value function $\phi_t(u^{t-1}, \eta^t, \xi^t)$ is convex in η^t and concave in ξ^t. Under this condition, barycentric approximation provides *exact* upper and lower bounds for the multistage problem. As we shall see in 4.1, this has a strong impact on the choice of the term structure model we use to capture the uncertainty of interest rates and the distribution assumption associated with it.

To ensure that the stochastic program is solvable in the multistage case, beside assumptions $(i) - (iii)$ continuity of the value functions in $t = 1, \ldots, T$ is required. This is preserved if there exists some point $\hat{u}_t$ depending on (u^{t-1}, ξ^t) for which the Slater condition $f_t(u^{t-1}, \hat{u}_t, \xi^t) < 0$ holds. Note that this is a stronger assumption than nonanticipativity as introduced in 2.1. The latter states that the feasible region in t is non-empty. Now, we suppose in addition that it contains inner points (see [20] for details).

3.2. Barycentric scenario trees

As mentioned above, approximation techniques are based on discretizations of continuous distributions for the underlying stochastic factors. From the inequalities due to Edmundson-Madansky and Jensen, one can derive upper and lower bounds using extreme points and expected values of the distribution which are used to construct the scenario trees. Barycentric approximation is a generalization of these concepts. Instead of expected values, the bounds are based on so-called generalized barycenters (see [19–21]). This allows the approximation of problems where uncertainty occurs in the objective as well as in the constraints.

In particular, it is assumed that Θ_t and Ξ_t are regular simplices in $\mathbb{R}^{K_t}$ and $\mathbb{R}^{L_t}$ that cover the support of (η_t, ξ_t). Both may depend on the history of observations (η^{t-1}, ξ^{t-1}) although this is not stressed in the notation for simplicity. When $P_t(\cdot|\eta^{t-1}, \xi^{t-1})$ has unbounded support, one has to ensure that $P_t(\Theta_t \times \Xi_t|\eta^{t-1}, \xi^{t-1}) \geq 1 - \epsilon$ for some sufficiently small $\epsilon > 0$ and substitute P_t by its normalized truncation. Let the vertices of Θ_t and Ξ_t be denoted by a_{ν_t}, $\nu_t = 0, 1, \ldots, K_t$, and b_{μ_t}, $\mu_t = 0, 1, \ldots, L_t$. The probability measure P_t induces mass distributions $\mathcal{M}_{\nu_t}$ with the associated generalized barycenters ξ_{ν_t} on the L_t-dimensional simplices $a_{\nu_t} \times \Xi_t$. As for $\nu_t = 0, 1, \ldots, K_t$ the mass distributions $\mathcal{M}_{\nu_t}$ add up to a probability distribution, a discrete probability measure Q_t^l is derived when probabilities $\mathcal{M}_{\nu_t}(\{a_{\nu_t}\} \times \Xi_t)$ are assigned to points (a_{ν_t}, ξ_{ν_t}).

Analogously, P_t induces mass distributions $\mathcal{M}_{\mu_t}$ with associated generalized barycenters η_{μ_t} on the K_t-dimensional simplices $\Theta_t \times \{b_{\mu_t}\}$. Again, the mass distributions $\mathcal{M}_{\mu_t}$ add up to a probability distribution for $\mu_t = 0, 1, \ldots, L_t$, yielding a discrete probability measure Q_t^u on $\Theta_t \times \Xi_t$ by assigning probability $\mathcal{M}_{\mu_t}(\Theta_t \times \{b_{\mu_t}\})$ to point $(\eta_{\mu_t}, b_{\mu_t})$. Note that the support of these discrete measures is given by $\operatorname{supp} Q_t^l = \{(a_{\nu_t}, \xi_{\nu_t})|\nu_t = 0, 1, \ldots, K_t\}$ and $\operatorname{supp} Q_t^u = \{(\eta_{\mu_t}, b_{\mu_t})|\mu_t = 0, 1, \ldots, L_t\}$. The corresponding probabilities can be obtained from the barycentric weights with respect to the simplices Θ_t and Ξ_t (for details, see [19]).

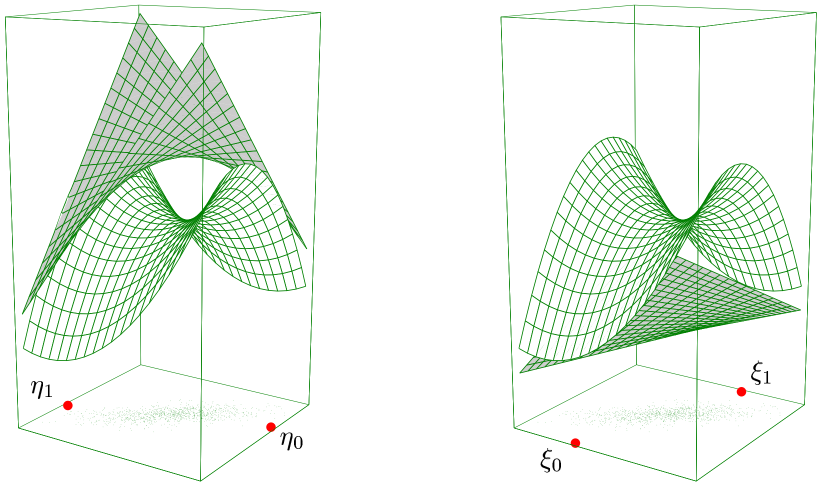
Using these discretizations, one finally gets two scenario trees from $\mathcal{A}_t^l(\omega^{t-1}) = \operatorname{supp} Q_t^l(\cdot|\omega^{t-1})$ and $\mathcal{A}_t^u(\omega^{t-1}) = \operatorname{supp} Q_t^u(\cdot|\omega^{t-1})$. These trees provide the desired bounds which can be seen as follows: Substituting Q_t^l and Q_t^u for P_t in (3) yields the corresponding value functions for $t = 1, \ldots, T$ (with $\psi_{T+1}(\cdot) = \Psi_{T+1}(\cdot) := 0$):

$$\psi_t(u^{t-1}, \omega^t) = \min_{u_t} \Big\{ g_t(u^t, \omega^t) + \int_{\Omega_{t+1}} \phi_{t+1}(u^t, \omega^{t+1}) dQ_{t+1}^l(\omega_{t+1}|\omega^t) | u^t \geq 0 \Big\},$$
$$\Psi_t(u^{t-1}, \omega^t) = \min_{u_t} \Big\{ g_t(u^t, \omega^t) + \int_{\Omega_{t+1}} \phi_{t+1}(u^t, \omega^{t+1}) dQ_{t+1}^u(\omega_{t+1}|\omega^t) | u^t \geq 0 \Big\}.$$

It is shown in [20] that $\psi_t(u^{t-1}, \omega^t) \leq \phi_t(u^{t-1}, \omega^t) \leq \Psi_t(u^{t-1}, \omega^t)$. Therefore, replacing P_t by Q_t^l in the original stochastic program yields a lower and substituting it by Q_t^u yields an upper bound to the original problem (see Figure 1).

The approximation obtained from Q_t^l and Q_t^u can be improved by partitioning the simplices into sub-simplices. In case that these sub-simplices become arbitrary small, weak convergence of the discrete measures to P_t and, hence, the convergence of ψ_t and Ψ_t to the value function ϕ_t is guaranteed. However, since each partitioning increases the size of the corresponding scenario trees, the process of refinements must be controlled carefully. The advantageous feature

Figure 1. Upper and lower bilinear approximations of the value function



of the approximation technique is that – under the assumptions for the convexity of ϕ_t – the accuracy of the discretization is quantifiable by the difference $\epsilon_t(\omega_t) = \Psi_t(u^{t-1}, \omega^t) - \psi_t(u^{t-1}, \omega^t)$ between the upper and lower bound. Hence, one has to identify those nodes of the scenario tree where the largest approximation error $\epsilon_t(\cdot)$ is observed. Note that if $\epsilon_t(\cdot) = 0$ for some node, the approximation of ϕ_t is exact there, and a (further) partitioning of the associated simplex does not improve the accuracy of the solution. For details on efficient refinement techniques, it is referred to [22].

3.3. Application: A simplified investment model

As an example for stochastic optimization in financial planning, we introduce a simplified version of the model by Forrest, Frauendorfer, and Schürle [18]. The latter has been successfully applied to the management of savings accounts by a major Swiss bank. Let $\mathcal{D} = \{1, 2, \ldots, D\}$ denote a set of maturity dates for the assets hold. Investment opportunities are given by a set of some standard maturities $\mathcal{D}^S \subset \mathcal{D}$ that are traded in the market. The amount invested in maturity $d \in \mathcal{D}^S$ is denoted by x_t^d, and φ_t^d is the corresponding sum of discounted interest payments received per unit. It is assumed that φ_t^d is determined by the realizations of a K-dimensional process $(\eta_t, t = 1, \ldots, T)$ driving the stochastic evolution of the yield curve. The relationship between η_t and φ_t^d also incorporates transaction costs and the discount mechanism. Uncertainty of the total portfolio volume is taken into account by a stochastic volume change $(\xi_t, t = 1, \ldots, T)$. P is the joint probability measure of (η, ξ) associated with time $t = 1, \ldots, T$, where $\eta = (\eta_1, \ldots, \eta_T)$, $\xi = (\xi_1, \ldots, \xi_T)$. The objective is to maximize the income from all investments over the planning horizon T:

$$
\begin{array}{llll}
\max \displaystyle\int \sum_{t=0}^{T} \sum_{d\in\mathcal{D}^S} \varphi_t^d(\eta_t)\cdot x_t^d \, dP(\eta,\xi) & & & \\
\text{(1a)} & v_t^d - v_{t-1}^{d+1} - x_t^d = 0 & t=0,\dots,T; \forall d \in \mathcal{D}^S & \\
\text{(1b)} & v_t^d - v_{t-1}^{d+1} \qquad = 0 & t=0,\dots,T; \forall d \notin \mathcal{D}^S & (5) \\
\text{(2)} & v_t - \sum_{d\in\mathcal{D}} v_t^d \quad = 0 & t=0,\dots,T & \\
\text{(3)} & v_t - v_{t-1} \qquad = \xi_t & t=1,\dots,T & \\
\text{(4)} & x_t^d \geq 0,\ v_t^d, v_t \in \mathbb{R} \text{ nonanticipative} & t=0,\dots,T; \forall d &
\end{array}
$$

The sum of assets v_t^d at t maturing after d periods is determined by (1a) and (1b) where v_{-1}^d with negative time index specifies the initial portfolio composition from decisions in the past. Constraint (2) requires that the savings volume v_t equals the total sum of investments v_t^d at all points in time. Equation (3) characterizes the change in the total volume. Finally, the nonanticipativity constraint (4) requires that two scenarios which share the same history up to time t must have identical decision and state variables up to t. Obviously, in the stochastic program (5) the risk factor η_t appears only in the objective while the volume change ξ_t affects solely the right-hand side of the constraints. In particular, the coefficients on the left-hand side are deterministic. According to section 3.1, this preserves the saddle property of the corresponding value functions at each time t if the distribution function of $P_t(\cdot|\eta_{t-1},\xi_{t-1})$ depends linearly on the past. It remains to discuss under which conditions this is given for a class of common term structure models.

4. Stochastic interest rate models

4.1. Evolution of the term structure

Clearly, beside the structural properties postulated above, it is required that the stochastic process we use to model the dynamics of η provides an appropriate description for the evolution of interest rates. For example, it should reflect that interest rates tend to fluctuate around their long-term mean *(mean reversion)*. Moreover, the dynamics of the entire yield curve should be explained by a small number of factors. These characteristics are incorporated in most of the common *term structure models* that have been developed in the financial literature since the seminal work of Vasicek [31] and Cox, Ingersoll, and Ross [10]. In the simplest case, it is assumed that there is a single state variable which is associated with the instantaneous rate r (the yield for an infinitely short holding period). Its dynamics are usually described by a stochastic differential equation of the form

$$dr = \kappa(\theta - r)dt + \sigma r^{\delta} dz. \tag{6}$$

The parameter κ defines the speed of adjustment of the continuous-time process from its current level towards the long-term mean θ. The source of uncertainty is represented by the increment of a Wiener process dz. The instantaneous volatility σr^δ depends on the level of the state variable if $\delta > 0$. In many empirical studies, a discrete-time approximation of the process (6) is investigated, and the instantaneous rate is replaced by the yield for a short maturity like one month. Chan et al. [6] for US data or Leithner [26] for the Swiss market claim that the process should indeed reflect such kind of heteroskedasticity.

Since the instantaneous rate is the only risk factor (i.e., $r = \eta$), yields of all maturities depend exclusively on r. Ito's Lemma allows the derivation of the process for the price of a discount bond from (6). By construction of a hedge portfolio out of two such instruments with different maturities, one obtains a partial differential equation for the term structure using a no-arbitrage argument (see [31] for details). In case that $\delta \in \{0, 0.5\}$, this equation can be solved analytically which results in a formula for the spot rate $R(d)$ with maturity d

$$R(d) \equiv R(d, r) = \frac{1}{d}\left[-\ln A(d) + B(d)r\right]. \tag{7}$$

This postulates that spot rates are derived from the risk factor by a linear transformation. The functions $A(\cdot)$ and $B(\cdot)$ depend on the parameters of the stochastic process (6) and some additional parameter λ known as the *market price of risk*. Both are not stated here explicitly but may be found in the literature. The case $\delta = 0$ is equivalent to Vasicek's model [31]. Here, the distribution of the factor η_s at time s conditioned on a prior realization η_t is normal with mean $\theta + (\eta_t - \theta)e^{-\kappa(s-t)}$ and variance $\frac{\sigma^2}{2\kappa}(1 - e^{-2(s-t)})$. Note that the latter is independent of η_t. According to (7), spot rates for all maturities are also normally distributed. $\delta = 0.5$ yields the model of Cox, Ingersoll, and Ross [10] where the conditional distribution of the factor is non-central χ^2 and its volatility is proportional to the square-root of its level. In the remainder, we restrict ourselves to these two distributions due to the tractability of the corresponding models.

Obviously, the assumption of only one risk factor is a strong simplification of the real world since it implies perfect correlation between the interest rates of different maturities which cannot be observed in real markets. In particular, one-factor models allow to reflect only a limited set of possible shapes for the term structure (i.e., the yield curve may be uniformly rising or falling, or it may be humped) with a fixed long end. One may overcome these weaknesses by introducing additional state variables. If these K factors are orthogonal, sum up to the instantaneous rate, and their evolution can be described by the stochastic differential equations $d\eta_i = \kappa_i(\theta_i - \eta_i)dt + \sigma_i \eta_i^\delta dz_i$, $i = 1, \ldots, K$, it can be shown that the term structure equation has some affine structure (see [7]):

$$R(d) \equiv R(d, \eta_1, \ldots, \eta_K) = \frac{1}{d}\sum_{i=1}^{K}\left[-\ln A_i(d) + B_i(d)\eta_i\right]. \tag{8}$$

Again, A_i and B_i depend on the process parameters $\kappa_i, \theta_i, \sigma_i$ and the market price of risk λ_i and have different functional forms for $\delta \in \{0, 0.5\}$. It is argued in 3.1 that to ensure the saddle property of the value function when using these models for scenario generation within a stochastic program, the distribution of the risk factors $\eta_s = (\eta_{1s}, \ldots, \eta_{Ks})'$ at time s conditioned on their realizations at t has to depend linearly on the past. Assume that $z = (z_1, \ldots, z_K)'$ is a random vector of a K-dimensional standard normal distribution, i.e., $z_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, K$. Clearly, for $\eta_s = z$ the saddle property is inherited since the distribution is independent of the past. However, in case of the normally distributed model introduced earlier, the expected values and variances of η_{is} are given by

$$\mu_{is} = \theta_i + (\eta_{it} - \theta_i)e^{-\kappa(s-t)} \quad \text{and} \quad \sigma_{is}^2 = \frac{\sigma_i^2}{2\kappa_i}(1 - e^{-2\kappa_i(s-t)}).$$

$\mu_s = (\mu_{1s}, \ldots, \mu_{Ks})'$ is the corresponding vector of means and $\Sigma_s = \text{diag}\,(\sigma_{1s}^2, \ldots, \sigma_{Ks}^2)$ the covariance matrix, respectively. With Δ_s obtained from $\Sigma_s = \Delta_s \Delta_s'$, we can use the transformation $\eta_s = \mu_s + \Delta_s z$. As μ_s is linear in η_t and Δ_s is independent of η_t (or earlier observations), the Vasicek model as well as its extension by additional normally distributed factors preserve the desired convexity. On the other hand, for a χ^2-distributed model the condition of a linear dependency is not fulfilled. The reason may be seen in the square-root term in the process equation that induces some non-linearity. As a consequence, we are not able to determine the upper and lower bounds for the value function by means of the barycentric approximation technique in the multistage case.

In the next step, we have to investigate how well term structure models reproduce the evolution of interest rates from a statistical point of view. It should be emphasized that these models have been developed mainly for an arbitrage-free pricing of securities under simplifying assumptions for reasons of analytical tractability. In principle, the restrictions implied by the absence of arbitrage may be so strong that they prevent a realistic modeling of the term structure dynamics (see e.g. [9] for a discussion of these issues). Therefore, the models must be examined by means of suitable econometric methods for the market to which the stochastic optimization model is applied. In particular, we face two questions for a comparison of alternative models: (a) How many factors are required to reflect the observed correlation structure and (b) which distribution (i.e., value of δ in the process equations for the risk factors) is appropriate?

4.2. Estimation and empirical results

Among the models introduced in section 4.1, only the normally distributed ones allow to inherit the saddle property of value functions in a multistage stochastic program. On the other hand, earlier results in the literature imply that the χ^2-distributed models can be expected to perform better empirically. In case of the Swiss Franc, Leithner [26] finds that the square-root process provides a

Table 1
Parameter estimates (standard errors in parenthesis)

	i	κ_i	θ_i	σ_i	λ_i	likelihood
$\mathcal{N}$-1F	–	0.4613 (0.1024)	0.04267 (0.006256)	0.01590 (0.00441)	0.5218 (0.08707)	3229.602
χ^2-1F	–	0.6078 (0.1030)	0.04409 (0.006300)	0.07863 (0.00441)	-0.1682 (0.08771)	3222.577
$\mathcal{N}$-2F	1	2.3174 (0.2172)	0.02705 (0.007316)	0.02406 (0.001447)	-0.1288 (0.2873)	3206.667
	2	0.1990 (0.02530)	0.01705 (0.007315)	0.01398 (0.001078)	0.3458 (0.2090)	
χ^2-2F	1	1.1990 (0.1981)	0.03376 (0.004847)	0.1283 (0.008453)	0.1041 (0.1957)	3118.973
	2	0.06404 (0.09603)	0.01154 (0.01475)	0.06091 (0.005316)	-0.2249 (0.0929)	

good description for the dynamics of short-term rates. Hence, although one of the models with linear dependency of the distributions on prior observations will be used for scenario generation, we estimate the χ^2-distributed models as a benchmark to assess the error that is made by ignoring a possible heteroskedasticity of interest rates.

In order to compare models with different distributions and up to $K \leq 3$ factors, we implement the maximum likelihood approach due to Chen and Scott [7] for the estimation. The advantageous feature of this procedure is that each observation covers interest rates of four different maturities. In order to derive the unobservable factors from market data, $4 - K$ measurement errors following a first-order autoregressive process are introduced as additional variables for reasons of tractability (for details, see [7]). In contrast to the investigations mentioned above, the inclusion of cross-sectional information from the yield curve allows to identify the market price of risk λ_i. For the estimation, 167 monthly observations of Euromarket and swap rates for the Swiss Franc from January 1983 until November 1996 with maturities of 3 and 6 months as well as 2 and 5 years are used. The Euro- and swap market are by far more liquid than the government bond market in Switzerland, in particular for the short end of the term structure. Results for the one- and two-factor models are given in Table 1 while the estimation of the three-factor models failed due to numerical problems. It turned out that the log-likelihood function was too flat in case of both distributions, and the BFGS-algorithm [29] we used for the optimization terminated after a few iterations because of almost singularity of the Hessian matrices.

The optimal values of the likelihood function quantify how well the models fit to the data from the sample, indicating that the Vasicek model slightly outperforms the square-root approach of Cox, Ingersoll, and Ross. In case of two factors, a drop in the function values can be observed, in particular for the χ^2-distributed model. For both distributions, models with a different number of factors represent a trade-off between state variables and measurement errors. Apparently, neither the normal nor the χ^2-distribution capture the dynamics of the second factor well since the ability of the models to fit the empirical distribution of the observations is reduced when one measurement error is replaced by an additional state variable. Furthermore, Vuong's likelihood ratio test for non-nested hypotheses [33] indicates that the models with one χ^2-distributed and two normally distributed factors are equivalent when tested against each other. Compared to the Vasicek model, both can only be rejected at a low significance level of 10 %. As expected, the χ^2-distributed two-factor model is clearly rejected at the 1 % level when compared to all other models under investigation.

However, the bad performance of the two-factor models is surprising since Chen and Scott obtained the best fit for the χ^2-distributed two-factor model (the normally distributed case was not considered) using US Treasury bills and bonds. An analysis of the time series of state variables shows that the second factor moves parallel to the longest rate in their study as well as in our investigation. This implies that the second state variable reflects information from the long end of the term structure. While (default-free) government bonds with maturities up to 30 years are traded at high volumes in the US, maturities of five years or more are less liquid in the Swiss interbank market since participants avoid a possibly increased counterparty risk associated with such long-term instruments. Therefore, a possible bias in the data for long maturities may be another explanation for the poor fit of the multifactor models (see [30]). Unfortunately, traded volumes are not available which might be used as weights for an adjustment of data to account for liquidity.

4.3. Tests of term structure models

The observation that the normally distributed models are superior to their χ^2-distributed counterparts contradicts earlier results from the literature as the former do not reflect a relation between the level of interest rates and their volatility. As mentioned, these investigations are based on a discrete-time approximation of the stochastic processes for the short rate whereas the approach we used in our analysis also takes cross-sectional information from yields of different maturities into account. Apparently, the Vasicek model and its extension by a second factor perform better in fitting the entire yield curve.

For an assessment how well the models capture the shape of the term structure, we examine the yield errors implied by the estimates. These errors are the differences between the model spot rates $R(\cdot)$ given the parameter estimates plus the recovered series of factors and the spot rates $R^m(\cdot)$ observed on the market for the corresponding maturity. Note that the three month rate is always observed without measurement error due to the construction of the estimation approach.

Table 2
Statistics of differences between implied and observed spot rates (in basis points, d in years)

one-factor models

d	normal distribution			χ^2-distribution		
	mean	std.dev.	max.	mean	std.dev.	max.
0.5	4.98	16.12	49.01	4.87	16.90	49.62
2.0	17.33	40.61	111.02	16.81	40.53	113.57
5.0	14.08	43.57	136.02	13.35	43.15	134.82

two-factor models

d	normal distribution			χ^2-distribution		
	mean	std.dev.	max.	mean	std.dev.	max.
0.5	-0.33	13.20	31.88	-3.52	16.19	39.90
2.0	0.18	14.93	37.79	-13.53	27.11	74.30
5.0	-0.04	6.88	17.35	5.76	12.20	31.80

According to Table 2, both one-factor models produce higher long rates than observed in reality. The highest deviation has an absolute value of 1.3 percent for the 5 year maturity which is economically significant. On the other hand, the errors implied by the two-factor models are less severe. Particularly the normal distribution fits well to the observed yield curves from the sample.

To assess how close the models come to the distribution of observed spot rates, the empirical moments from the underlying sample are compared to the theoretical moments for the steady state distribution of the factors. Cox, Ingersoll, and Ross [10] show that this is the Γ-distribution under the assumption of a square-root process for the state variables. In general, from the expected value $\mathrm{E}(\eta_i) = \theta_i$ of the stationary distribution, the *unconditional* variance $\mathrm{Var}(\eta_i) = 0.5\sigma_i\theta_i^{2\delta}/\kappa_i$ of the i-th factor and using equation (8), one obtains the mean and standard deviation for any maturity d implied by the models. The corresponding values for selected maturities can be found in Table 3. Compared to the descriptive statistics for the historical sample, the mean implied by all models is too low. The three month rate is underestimated by at least 50 basis points (BP) while the difference between the observed and the theoretical mean of the five year rate is reduced to 9 BP or 16 BP for the models of Cox, Ingersoll, and Ross and Vasicek, respectively. This can be explained by the fact that one-factor models generate steeper yield curves than occur in reality, compensating the effect that the probability of high values of the state variables is too small. Spot rate volatility is always underestimated by one- and overestimated by two-factor models, but the normally distributed two-factor model comes closest to the empirical values.

These conclusions are based on the unconditional distribution of the state

Table 3
Means and standard deviations of the steady state distribution (in %, d in years).

normal distribution

d	1 factor		2 factors		historic sample	
	E $(R(d))$	$\sigma\,(R(d))$	E $(R(d))$	$\sigma\,(R(d))$	$\overline{R}^m(d)$	$\sigma\,(R^m(d))$
0.25	4.36	1.56	4.44	2.32	5.02	2.09
0.5	4.45	1.48	4.47	2.21	5.03	2.00
2.0	4.87	1.08	4.72	1.84	5.16	1.53
5.0	5.31	0.65	5.13	1.41	5.47	1.05

χ^2-distribution

d	1 factor		2 factors		historic sample	
	E $(R(d))$	$\sigma\,(R(d))$	E $(R(d))$	$\sigma\,(R(d))$	$\overline{R}^m(d)$	$\sigma\,(R^m(d))$
0.25	4.50	1.42	4.52	2.27	5.02	2.09
0.5	4.58	1.34	4.52	2.21	5.03	2.00
2.0	4.96	0.99	4.64	2.22	5.16	1.53
5.0	5.38	0.60	5.11	2.75	5.47	1.05

variables. However, the evolution of the term structure in the course of time is determined by the *conditional* distribution. Therefore, further insights into the properties of the models can be obtained if we analyze the dynamic behavior of the implied spot rates. To this end, we conduct a series of Monte Carlo experiments similar to Vetzal [32] where for each of the four models with the parameter estimates from above, the process of η_i, $i = 1, \ldots, K$, is simulated. Using the term structure equation (8), this yields a series of corresponding spot rates for various maturities traded in the market. The simulation is repeated 1'000 times with 167 observations in each run, starting from the values of the state variables that are derived from the first observation in the historical sample. Then, the descriptive statistics are calculated and compared to the corresponding historic values.

Again, we can only briefly outline the results (details and further tests can be found in [30]). As before, the means and quantiles for short-term rates are too low. For example, the 90 %-quantile of the three month rate is 8.41 % in the historical sample, whereas the corresponding value generated with the Vasicek model is only 5.83 % in the simulation. With a second normally distributed factor, it can be increased to 6.87 %, and for both χ^2-distributed models, it lies between these values. Hence, we can conclude that all models underestimate the probability of high rates. For the five year maturity, the one-factor models also produce lower quantiles as in the sample, but the normally distributed two-factor model meets the historic 90 %-quantile fairly well. Furthermore, it provides the best description of spot rate volatility, although the distribution of the factors is homoskedastic. Regarding higher moments, the skewness of 0.51 for the three month rate is better met by the χ^2-distribution with 0.37 for one and 0.42 for two factors. All models seem to overestimate the observed kurtosis of 2.27 slightly, but the simulated estimates are within only one standard deviation from the empirical value.

Turning our attention to *changes* in spot rates, both one factor models meet the variance of 0.43 for the first differences in the historical sample almost exact while the normally and χ^2-distributed model overestimate it with 0.64 and 0.59, respectively. The normal distribution cannot reflect the high kurtosis of 4.30 for the changes in observed spot rates but the χ^2-distribution with 3.25 and 3.37, do also not achieve a similar magnitude. In case of the five year rate, the χ^2-distributed two-factor model reaches the largest kurtosis of 4.02 compared to an observed value of 5.91. An explanation for the thicker tails in the distribution of the differences is that Swiss interest rates are relatively stable over a long time and exhibit rare but relatively large changes which does not seem to be captured by the models. The negative skewness of -0.24 and -0.43 in the changes of the observed three month and five year rates indicates that downside movements are of larger magnitude than upside shifts while the corresponding statistics are close to zero for all simulations.

To summarize, all models show certain deficiencies when compared to the empirical properties of observed yields, in particular with respect to higher moments. Overall, the model with two-normally distributed state variables comes closest to the empirical distribution of Swiss interest rates. However, the fact that the models do not generate enough high rates may have a strong impact on the performance of investment policies derived from the stochastic program (5). This can be investigated if the decisions based on scenarios generated with different term structure models are compared.

4.4. Performance of investment policies

In order to test the suitability of scenarios generated with the term structure models under consideration for the determination of investment policies, we use 82 monthly Swiss yield curves from March 1989 to December 1995, covering a complete cycle of high and low rates, and the corresponding volume of a real savings account. These data were provided by a major Swiss bank for the conduction of a case study before an extended version of the stochastic optimization model described here was applied in practice. The latter also allows for short sales and generates interest rate scenarios by means of a three-factor model where the state variables follow Brownian motions in discrete time. Clearly, this specification does not reflect mean-reversion of interest rates and, hence, may be seen as more appropriate for short-term horizon. Despite this deficiency, results of the case study indicate that the stochastic optimization model clearly outperforms static investment policies for planning periods of several years when investment decisions are based on a sensitivity analysis which takes extreme movements of the term structure into account (see [18] for details).

For each month out of the sample period, we generate scenarios for future interest rates based on the parameters in Table 1 and initial values η_0 for the conditional distributions derived from the current yield curve. Although the volume fluctuates in the data set, it is assumed to be deterministic in the stochastic program (i.e., $\xi_1 = \ldots = \xi_T = 0$) since we restrict our analysis to models for the uncertainty in interest rates. The initial volume v_0 is 30 billion CHF. Investment opportunities consist of money market and swap positions with maturities of 1, 2, 3, 4, 5, 7, and 10 years. In order to take liquidity restrictions into account that come close to conditions on the Swiss market, upper limits are set to 500 million CHF for maturities up to 5 years and 200 million CHF above that.

After each optimization, the portfolio of assets is rebalanced by implementing the first-stage decision. This results in some kind of roll-over planning. As a benchmark, we use the allocation rule specified in [18] that was previously used by the bank for the management of their savings accounts. According to this policy, maturing funds are always renewed at their original maturity. Whenever the total savings volume increases or decreases due to changes in customer demand, funds are bought or sold at fixed proportions (1, 2, and 5 years at 35 %, 35 %, 30 %). These weights are determined through minimizing the tracking error, i.e., the difference between the customer rate plus margin and the return on the reinvested portfolio over a historic sample period (*replicating portfolio*).

It is noted that the sample period in our test and the data used to calibrate the term structure models overlap partially (the reason for this setting is that the bank did not reveal the time period covered by the case study to us to prevent that we anticipate forthcoming interest rates in the decision making). This might limit the validity of the results to a certain degree because information was used for the calibration that was unavailable at the time of decisions. However, our primary goal is to compare investment policies derived from the stochastic optimization model with a static allocation rule. Since the weights of the replicating portfolio provided by the bank are also determined using these data, we have to accept this deficiency in order to compare our results with the benchmark. Moreover, money market and swap rates for maturities over one year are not available for earlier periods in case of the Swiss Franc. As a consequence, the number of observations used for the estimation of the term structure models would have been insufficient if we restricted the calibration to data not contained in the sample period.

Table 4 shows the performance of investment policies derived from the stochastic program for the one- and two-factor models. The average margin (i.e., the income on reinvested savings deposits minus the customer rate) and its standard deviation over the entire sample period are compared to the corresponding values for the replicating portfolio. When the Vasicek model ($\mathcal{N}$-1F) is used for scenario generation, the stochastic program (2) can be approximated at high accuracy without any refinements. Unfortunately, the discretization error

Table 4
Performance of investment policies compared to a replicating portfolio

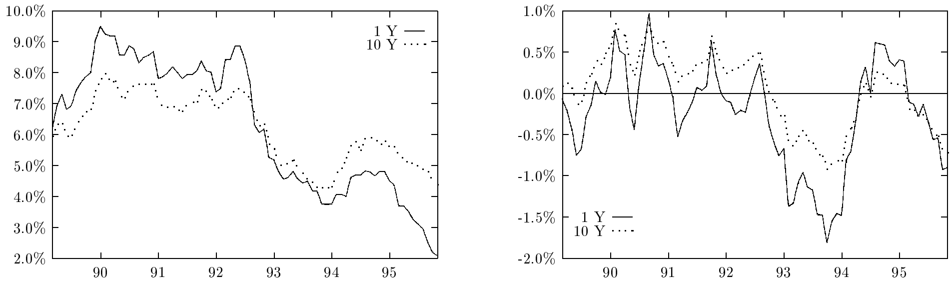
model	approx	#ref.	margin [%]	improv. [BP]	std. dev. [%]	improv. [%]
$\mathcal{N}$-1F	L/U	0	2.58	18.0	0.25	0.44
$\mathcal{N}$-2F	L	0	2.474	7.5	0.246	0.450
$\mathcal{N}$-2F	L	1	2.476	7.7	0.252	0.444
$\mathcal{N}$-2F	U	0	2.603	20.4	0.214	0.482
$\mathcal{N}$-2F	U	1	2.553	15.4	0.215	0.481
Brown.	U	0	2.444	4.5	0.609	0.087

increases with the second state variable in $\mathcal{N}$-2F. As a consequence, the decisions of both approximations may not coincide. This deficiency remains in some cases even after one partition of the $\times$- simplex corresponding to the root nodes of the scenario trees, and further refinements would increase the problem size seriously. Therefore, the test series are carried out twice, where decisions are either based on the lower (L) or upper (U) approximated problems.

The planning horizon is set to $T = 8$ which results in $h \cdot (K+1)^T = 2 \cdot 3^8 = 13122$ scenarios for the two-factor models ($K = 2$) with $h = 2$ partitions in the root node (one refinement). Larger problems with a higher number of stages and/or partitions cannot be solved with our computational resources. More efficient refinement algorithms which analyze the scenario trees for the nodes with the highest approximation error $\epsilon_t(\cdot)$ to achieve a better accuracy are currently under investigation and have not yet been implemented in the optimization model. Using the one-factor Vasicek model allows for an extension of the planning horizon, but results for $T = 10$ and $T = 12$ are not presented here since they are equivalent to the eight-period problem. As an additional benchmark, we use a three-factor model with Brownian motions. These reflect shift, tilt, and humped movements of the yield curve but do not incorporate mean reversion. Here, the planning horizon was reduced by one stage due to the higher dimension.

The scenarios generated with the Vasicek model yield an increase in the margin of 18 BP over the static approach. In addition, the volatility is reduced significantly by more than 0.4 %. On the other hand, the Brownian motion model shows no significant improvement in the margin or in its volatility at all. This must be seen as a justification that the interest rate model should reflect the mean reversion property. In case of the two-factor models, we gain 20 BP using the upper approximation. However, one partition in the root node reduces this value towards the average performance of the lower approximated problem of 8 BP. It can be expected to drop further since the lower approximation is the more stable one according to our experience. This raises the question why the two-factor model exhibits a worse performance compared to the Vasicek model although it provides a more accurate description of interest rate behavior.

Figure 2. Interest rates (left) and differences in forecasts (right) in the case study



An analysis of portfolio compositions in the course of time (see [30] for details) reveals that the two-factor model invests a larger proportion in short-term money than the Vasicek model during the first half of the sample period which is characterized by an inverse term structure on a high level. A possible explanation for the different investment policies is that the 12 month forecasts of the 10 year rate generated with $\mathcal{N}$-2F are 0.37 % larger on average while the predicted one year rates are equivalent to those of the Vasicek model (see Figure 2 (right) for the differences in the forecasts between the two- and the one-factor model over time). As a consequence, the former exploits the high short rates instead of securing the current above-average level of yields by a long-term investment. For the second half, $\mathcal{N}$-2F shows a stronger tendency towards long maturities when the term structure becomes normal after the drop in interest rates. The expected one year rate after 12 months is 0.49 %, the ten year rate 0.28 % lower on average then the Vasicek forecasts. In this case, the two-factor model takes advantage of the positive spread between long- and short-term yields.

However, in the situation of below-average interest rates, an increase can be expected due to mean reversion which implies a short-term investment. We found in section 4.3 that all models underestimate the probability of high values for the state variables, but the Vasicek model generates spot rate curves that are steeper than in reality, suggesting high long-term rates in the next period. Since the state variables correspond to both ends of the yield curve for the two-factor model, the spread between long and short rates remains moderate in the scenarios. Hence, the optimization model invests in the long maturities today because the expected increase in yields does not compensate the lower income for a short-term investment at the beginning. On the other hand, the Vasicek model tends to wait for a rise in the rates for long maturities and invests in them later as an analysis of the second-stage decisions shows. At least for the sample period considered in the case study, this turns out to be the better policy in many situations. In other words, the good result for the Vasicek model can be explained by the fact that two deficiencies eliminate each other: an underestimation of the probability of high rates and an overestimation of the steepness of the yield curve.

5. Conclusions and outlook

In this paper, we introduced barycentric approximation as a solution technique for stochastic multistage programs which allows to determine exact bounds under certain conditions for the structure of the optimization problem and the distributions of random data. In particular, it is required that the distribution functions for the risk factors depend linearly on the past. Among the most popular term structure models, the latter holds for the well-known Vasicek model and its extension by additional (normally distributed) state variables. We also conducted an empirical comparison of models with alternative distributions and number of factors using historic Swiss interest rates. In contrast to earlier results in the literature, our tests indicate that the normally distributed models are not inferior to others that allow to reflect heteroskedasticity in their state variables. This observation is of particular importance since it preserves the convexity of value functions in the multistage case on which the approximation is based.

The normally distributed models are used for scenario generation within a stochastic optimization problem for the reinvestment of savings accounts. Compared to a static approach, investment policies obtained from the stochastic program yield a higher margin at lower risk if the underlying model reflects mean reversion of interest rates. However, the results imply that some defects of the term structure models have an impact on the implemented investment decisions. For example, one-factor models generate yield curves that are steeper than observed in the Swiss market, and all models underestimate the probability of high rates. Alternative processes for the evolution of state variables may provide a more appropriate description for interest rate dynamics. Unfortunately, an analytical solution of the corresponding partial differential equation for the term structure does not exist in the general case.

Since we found that the investigated term structure models have a distinct bias, an improvement of interest rate scenarios might be achieved if the parameters are adjusted to reduce this distortion. For example, one could fit the models to the observed term structure by introducing time-dependent parameters according to the "arbitrage-free approach" (see e.g. Hull and White [23]). However, the estimation of such models is not covered by the procedure we used here. It remains subject to further investigation whether they show a higher dynamic accuracy. In particular, it must be clarified if a model with time-dependent parameters which has to be frequently calibrated is capable of providing an accurate description for the evolution of interest rates over a long-term horizon.

Empirical investigations for Swiss [4] as well as for US data [5] imply that the dynamics of interest rates cannot be explained when the model parameters are estimated from the current term structure under the arbitrage-free approach, at least not with a one-factor model. Therefore, it is possible that such kind of calibration provides more or less an "interpolation technique" for the observed yield curve rather than an appropriate characterization of its dynamics. In contrast to this, our estimation approach is focused on the time series of interest rates to obtain an explanation for the evolution of the term structure over a period of several years. Finally, we restricted ourselves to an examination of the uncertainty in interest rates although the future volume is also stochastic (e.g., due to withdrawals of savings accounts by customers or prepayments of non-fixed mortgages). The behavior of cash flows strongly depends on the particular problem and must be modeled differently for each application. A better understanding of the interaction between interest rates and volume is expected to yield an additional improvement.

Acknowledgements

Research for this paper was supported by the Science Foundation of the *University of St. Gallen*, Grant No. 121'417, and by the *Swiss National Science Foundation*, Grant No. 21-39'575.93. We thank the anonymous referees for many useful comments that helped improve the exposition.

References

[1] J.R. Birge and R.J.-B. Wets. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, *Mathematical Programming Study*, 27 (1986) 54-102.
[2] F. Black, E. Derman, and W. Toy. A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts Journal*, 46 (1990) 33-39.
[3] M.J. Brennan and E.S. Schwartz, An equilibrium model of bond pricing and a test of market efficiency, *Journal of Financial and Quantitative Analysis*, 17 (1982) 301-329.
[4] A. Bühler, *Einfaktormodelle der Fristenstruktur der Zinssätze*, Haupt, 1995.
[5] E. Canabarro, Comparing the dynamic accuracy of yield-curve-based interest rate contingent claim pricing models, *Journal of Financial Engineering*, 2 (1993) 365-401.
[6] K.C. Chan, G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, 47 (1992) 1209-1227.
[7] R.-R. Chen and L. Scott, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income*, 3 (1993) 14-31.
[8] Z. Chen, M.A.H. Dempster, and N. Hicks-Pedrón, Towards sequential sampling algorithms for dynamic portfolio management, in: *Operational Tools in the Management of Financial Risks*, ed. C. Zopounidis, Kluwer, 1998, pp. 197-211.
[9] R. Cont, Modeling term structure dynamics: an infinite dimensional approach, *International Journal of Theoretical and Applied Finance*, 3 (2000), to appear.
[10] J.C. Cox, J.E. Ingersoll, and S.A. Ross, A theory of the term structure of interest rates, *Econometrica*, 53 (1985) 385-407.
[11] G.B. Dantzig and G. Infanger, Multi-stage stochastic linear programs for portfolio optimization, *Annals of Operations Research*, 45 (1993) 59-76.
[12] M.A.H. Dempster, The expected value of perfect information in the optimal evolution of stochastic problems, in: *Stochastic Differential Systems*, eds. M. Arato, D. Vermes, and A.V. Balakrishnan, Springer, 1981, pp. 25-40.
[13] J. Dupačová, M. Bertocchi, and V. Moriggia, Postoptimality for scenario based financial planning models with an application to bond portfolio management, in: *World Wide Asset*

and Liability Modeling, eds. W.T. Ziemba and J.M. Mulvey, Cambridge University Press, 1998, pp. 263-285.

[14] N.C.P. Edirisinghe, New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming, *Operations Research*, 44 (1996) 909-922.

[15] N.C.P. Edirisinghe and W.T. Ziemba, Tight bounds for stochastic convex programs, *Operations Research*, 40 (1992) 660-677.

[16] N.C.P. Edirisinghe and W.T. Ziemba, Bounding the expectation of a saddle function with application to stochastic programming, *Mathematics of Operations Research*, 19 (1994) 314-340.

[17] N.C.P. Edirisinghe and W.T. Ziemba, Bounds for two-stage stochastic programs with fixed recourse, *Mathematics of Operations Research*, 19 (1994) 292-313.

[18] B. Forrest, K. Frauendorfer, and M. Schürle, A stochastic optimization model for the investment of savings account deposits, in: *Operations Research Proceedings 1997*, eds. P. Kischka, H.-W. Lorenz, U. Derigs, W. Domschke, P. Kleinschmidt, and R. Möhring, Springer, 1998, pp. 382-387.

[19] K. Frauendorfer, *Stochastic Two-Stage Programming*, Springer, 1992.

[20] K. Frauendorfer, Multistage stochastic programming: Error analysis for the convex case, *Mathematical Methods of Operations Research*, 39 (1994) 93-122.

[21] K. Frauendorfer, Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming (Series B)*, 75 (1996) 277-293.

[22] K. Frauendorfer and C. Marohn, Refinement issues in stochastic multistage linear programming, in: *Stochastic Programming Methods and Technical Applications*, eds. K. Marti and P. Kall, Springer, 1998, pp. 305-328.

[23] J. Hull and A. White, Pricing interest rate derivative securities, *Review of Financial Studies*, 3 (1990) 573-592.

[24] G. Infanger, *Planning under Uncertainty*, Boyd & Fraser, Danvers, 1994.

[25] P. Kall, A. Ruszczyński, and K. Frauendorfer, Approximation techniques in stochastic programming, in: *Numerical Techniques for Stochastic Optimization*, eds. Y. Ermoliev and R.J.-B. Wets, Springer, 1988, pp. 33-64.

[26] S. Leithner, *Valuation and Risk Management of Interest Rate Derivative Securities*, Haupt, 1992.

[27] R. Litterman and J. Scheinkman, Common factors affecting bond returns, *Journal of Fixed Income*, 1 (1991) 54-61.

[28] J.M. Mulvey, Financial planning via multi-stage stochastic programs, in: *Mathematical Programming: State of the Art 1994*, eds. J.R. Birge and K.G. Murty, University of Michigan, 1994, pp. 151-171.

[29] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 2nd ed., 1992.

[30] M. Schürle, *Zinsmodelle in der stochastischen Optimierung*, Haupt, 1998.

[31] O. Vasicek, An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5 (1977) 177-188.

[32] K.R. Vetzal, Stochastic volatility, movements in short term interest rates, and bond option values, *Journal of Banking and Finance*, 21 (1997) 169-196.

[33] Q.H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, 57 (1989) 307-333.

[34] S.A. Zenios, A model for portfolio management with mortgage-backed securities, *Annals of Operations Research*, 43 (1993) 337-356.

[35] S.A. Zenios and M.S. Shtilman, Constructing optimal samples from a binomial lattice, *Journal of Information and Optimization Sciences*, 14 (1993) 125-147.