

Dynamic modelling and optimization of non-maturing accounts

Karl Frauendorfer Michael Schürle

Institute for Operations Research and Computational Finance
University of St. Gallen, Switzerland

Abstract

The risk management of non-maturing account positions in a bank's balance like savings deposits or certain types of loans is complicated by the embedded options that clients may exercise. In addition to the usual interest rate risk, there is also uncertainty in the timing and amount of future cash flows. Since the corresponding volume risk cannot directly be hedged, the account must be replicated by a portfolio of instruments with explicit maturities. This paper introduces a multistage stochastic programming model that determines an optimal replicating portfolio from scenarios for future outcomes of the relevant risk factors: Market rates, client rates and volume of the non-maturing account. The weights for the allocation of new tranches are frequently adjusted to latest observations of the latter. A case study based on data of a real deposit position demonstrates that the resulting dynamic portfolio provides a significantly higher margin at lower risk compared to a static benchmark.

1 Introduction

A significant portion of a typical bank's balance are so-called non-maturing accounts (NMAs). Their characteristic feature is that they have no specific contractual maturity, and individual clients can always add or withdraw investments or credits at no (or a negligible) penalty. On the other hand, the bank is allowed to adjust the customer rate any time as a matter of policy. Typical examples of NMAs include savings and sight deposits on the liability side of the balance as well as credit card loans or variable-rate mortgages as they are common in some European countries on the asset side.

Although the client rate is mostly adjusted in sympathy with the direction of changes in money and capital market yields, it does not completely depend on the latter. In practice, an adaption (often in discrete increments) follows only after larger variations in open-market rates and with some delay due to administrative and other costs involved. There might also be a (political) cap like in case of variable-rate mortgages in Switzerland where housing rents are indexed to an "official" mortgages rate.

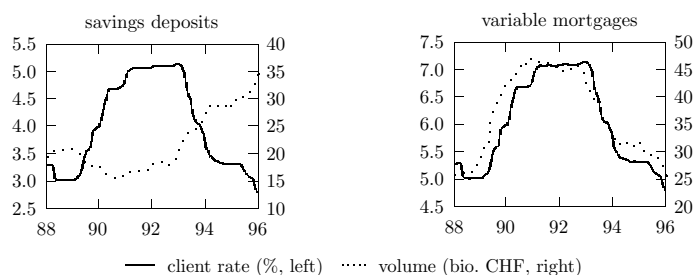


Figure 1: Correlation between client rate and volume during the early 1990s

1.1 Specific problems of non-maturing accounts

One can often observe that the volume of a NMA position fluctuates heavily as clients react to changes in the customer rate and the relative attractiveness of alternative investment or financing opportunities etc. For example, in the past a typical Swiss bank faced a significant increase in homeowners' demand for variable-rate mortgages during a period of high interest rates (see Figure 1) while refinancing them on the interbank market is particularly unfavorable. At the same time, depositors tend to substitute their savings by long-term investments to "lock up" the high level of yields (withdrawal).

When interest rates are low, clients switch to fix-rate mortgages in order to hedge themselves against a future rise (prepayment) while the deposit rate becomes more profitable compared to alternative short-term investments which attracts additional savings volume. Since the volume of the variable positions on both sides of the balance fluctuates "phase-delayed", mortgages (or other non-maturing loans) cannot be funded completely by deposits. Therefore, numerous Swiss banks suffered significant losses on their variable mortgage positions during the early 1990s when they had to refinance them at market rates around 10 % while the corresponding client rate never climbed over 7 % due to the cap.

In general, the incomplete adaption of the client rate to money and capital market rates may have a significant impact on the profitability and often result in unstable margins. Moreover, the embedded prepayment and withdrawal options induce a *volume risk*. Unlike the interest rate risk associated with fixed income instruments, the latter cannot be hedged directly since the volume is not traded in the market. Hence, conventional hedging techniques like duration matching can only be applied with additional assumptions about the amount and timing of (future) cash flows. A fix maturity profile must therefore be assigned to the NMA in order to transform the uncertain cash flows into (apparently) certain ones.

1.2 Implications for risk management

To obtain such a transformation, a *replicating portfolio* that mimics the payments of the variable position is constructed using traded standard instruments. Then, one is able to immunize (at least theoretically) the position against the

risk of changes in interest rates. The replicating portfolio also defines the transfer price at which the margin is split between the business unit retail that “acquires” the account and the treasury as the bank’s central unit where interest rate risk is managed. Obviously, an accurate determination of the replicating portfolio composition is of utmost importance for the risk exposure since an incorrect transformation may result in inefficient hedging decisions.

Given the significant amounts of NMAs in the balances of most retail banks, a proper management of the associated interest rate and volume risk therefore becomes a critical issue. Its importance has increased since the regulation of interest rate risk in the banking book plays a central role in the second pillar of Basel II (cf. Fiedler et al. [13]). For example, a number of banks regard savings accounts as short term liabilities (according to their legal maturity) and, thus, have assigned a short maturity to them in their interest obligation balance. When the term transformation is carried out by the treasury and variable positions are invested at longer terms to take advantage of higher rates in a normal yield curve, a term mismatch results. Although the latter will not directly be charged with capital, it leads to a “quasi capital adequacy” since the corresponding risk must not exceed 20 % of the bank’s Tier 1 and Tier 2 capital – otherwise the bank is disqualified as an “outlier” [1].

Similar problems are currently arising in the context of hedge accounting under IFRS. For example, a bank with large savings positions can hardly find “natural hedges” in its balance for the credit business due to the provisional non-approval of deposit accounts. Qualification for hedge accounting would require that the hedge effectiveness of non-maturing deposits could precisely be estimated.

Last but not least, pressure on the profit side has also increased after the introduction of alternative products such as money market funds or the appearance of direct banks in different countries that have achieved a large share in the savings market by aggressively offering attractive deposit rates. As a consequence, a number of traditional institutions have experienced a substantial contraction in their deposit volumes which had been an inexpensive funding opportunity so far. Even banks with stable volumes had to face smaller margins in the recent past due to historically low credit rates while the client cannot be charged a negative deposit rate [13].

Therefore, any methodology for the determination of the replicating portfolio not only has to hedge the inherent risk adequately but must also provide a sufficient margin that allows a bank to remain in the competition for “cheap” funding sources. In this sense, it is required to manage the “trade-off” between risk and return successfully. The fact that non-maturity products typically exhibit a complex dependency of the cash flows on the evolution of the yield curve demands for sophisticated modelling and decision making techniques.

At present time, the banking industry uses mostly static portfolios (i.e. with constant weights) whose composition is fitted to historical data. We illustrate in the next section that this approach is inefficient and holds a significant model risk. Based on an analysis of its shortcomings, we motivate why the replicating portfolio should be frequently readjusted to scenarios for the future risk factors:

Market rates, client rate and volume. The problem of the determination of a *dynamic replicating portfolio* is formulated in section 3 as a multistage stochastic program that can be solved with standard optimization algorithms. Models for the stochastic evolution of the risk factors are described in section 4 considering Swiss savings deposits as example. As discussed in section 5, specific attention must be paid to the generation of the scenarios that serve as input to the optimization problem to ensure consistence and robustness of the solution. Section 6 presents results of a case study for real Swiss savings deposit data that show the advantage of dynamic replication over the conventional static approach. The paper finishes with some remarks on the application to different products and future directions of research

2 Modelling non-maturing accounts

2.1 Shortcomings of static replication

A widely-used method in many European countries is the construction of the replicating portfolio by an analysis of the past evolution of the NMA position. This requires a set of historical data of client rates, volumes and market rates that may be used for reinvesting or financing. The portfolio composition is then determined under the condition that the cash flows of the (fixed income) instruments used for replication match those of the NMA (except for the margin) as close as possible, i.e., the average yield on the portfolio moves parallel to the client rate, and a drop in volume is compensated by maturing instruments. By minimization of the tracking error between the payments of the portfolio and the underlying NMA position, one aims at the minimization of the margin's volatility, and an estimate for its expectation may be obtained from the average of the sample period.

For the practical implementation, the resulting portfolio weights remain constant over time. Therefore, this method can be categorized as a *static* approach. Maturing instruments are always renewed at the same maturity except for volume changes which are compensated by buying additional or selling existing tranches of the specified instruments. Another possibility is the segmentation of the portfolio into a constant core volume and a volatile component. The core is itself subdivided into portions that are invested (or funded) in different time bands using medium and long term maturities, and maturing tranches are always reinvested in the same time band. The volatile part consists of overnight money and serves as a "buffer" for volume fluctuations. Since in both cases only a small portion of the replicating portfolio is renewed at a certain point in time, the portfolio return as an average of several market rates exhibits a low volatility and changes only slowly over time, similar to the client rate of a typical NMA.

An example for the former variant can be found in Table 1 for an asset (variable mortgages) and a liability product (deposit account, cf. Bühler [4]). The "optimal" portfolio composition in the static framework was determined by

begin end	<i>deposit account</i>			<i>variable mortgages</i>		
	Dec 92 Jun 99	Dec 92 Mar 96	Apr 96 Jun 99	Dec 92 Jun 99	Dec 92 Mar 96	Apr 96 Jun 99
6M	0.17		0.19			
1Y		0.11	0.81	0.22	0.07	0.51
2Y					0.48	
3Y						0.36
4Y		0.31				
5Y	0.83	0.58		0.78	0.45	0.13
margin [%]	3.83	4.25	1.50	0.31	2.71	1.68

Table 1: Characteristics of static replicating portfolios for two non-maturing accounts over a historical sample period and two subperiods [4]

minimization of the tracking error for a period of approx. 6.5 years (December 92 to June 99). For instance, if savings deposits had been replicated by a constant mix of 17 % six month and 83 % five year instruments, then – in this ex post analysis – a least volatile margin of 383 basis points (bp) would have been achieved during the sample period. Applying the same calculation for variable mortgages would have provided a rather small margin of 31 bp only which might not be sufficient to cover the administrative and other non-interest cost for holding the account.

In a second step, the author of this study divided the entire sample period into two subperiods (December 92 to March 96 and April 96 to June 99) and applied the same analysis to each of them. According to Table 1, this results in completely different portfolio compositions for both products. The observation that the weights depend heavily on the sample period and the corresponding margins are unstable implies a substantial *model risk* and has a direct impact on risk management and transfer pricing: The significant fluctuations in the portfolios would make hedging strategies inefficient, and the apparently low margin of mortgages from December 92 to June 99 might have led the bank management to the conclusion that the product is not profitable although an analysis for a different period could provide a completely different result.

In practice, one would determine the (static) replicating portfolio once and adjust the weights and transfer prices only in longer intervals (e.g., yearly). If this results in a new portfolio composition, then instruments of the existing portfolio must be squared, and it is not clear which unit in the bank has to bear the losses that result more or less from the pure adjustment of a “calculation rule” only and not from a fundamental change in the market or client behavior. Moreover, the observation that the least volatile margin may (apparently) be not sufficient to cover the (non-interest) costs questions the usefulness of the (lowest) variance as an appropriate risk measure.

To overcome these obvious deficiencies of the conventional static replication approach, we suggest to abandon two basic assumptions made so far:

1. Instead of a constant portfolio composition and identical maturities (up to the next readjustment), the weights and also relevant maturities at which

maturing instruments and a volume change are (re-) invested or financed have to be recalculated frequently.

2. The replicating portfolio is not derived from a *single historical* scenario but from a *large number* of scenarios for the possible *future* outcomes of the relevant risk factors, i.e., market rates, client rate and volume.

The motivation of the first point is to exploit the latest information from the markets and client behavior whenever instruments must be allocated. The second point requires a quantitative model that describes the (joint) stochastic dynamics of market rates, client rates and volume over time as well as the dependencies between them for specific retail products.

2.2 Stochastic models

Given the practical relevance of NMAs, there has been a relatively small number of papers on their stochastic analysis. Selvaggio [44] analyzes the premiums that banks pay for the acquisition of demand deposit accounts (DDAs) and explains the volume as a function of interests and nominal income with some time lag plus a seasonal correction. This volume model is then combined with the short rate process of Cox, Ingersoll and Ross [5] (CIR) to calculate an option-adjusted spread and the net present value (NPV) of future payments associated with deposit liabilities. Risk can be quantified by a recalculation of the NPV after shifts in the underlying term structure which, in principle, allows the determination of a hedging policy. Unfortunately, data for DDA premiums required for calibration are not available for European markets, and the significance of nominal income as explanatory variable for the volume of individual banks is not fully plausible.

Hutchison and Pennacchi [26] employ a general contingent claims framework for valuation and calculate duration measures for demand deposits analytically. In contrast to these equilibrium-based models, Jarrow and van Deventer [30] (JvD) introduce an arbitrage-free modelling approach to obtain the value of NMAs which allows for consistence with the pricing and risk management of traded products. They also show how the risky position may be hedged by a replicating portfolio that results from investing and financing in fixed income securities.

These first arbitrage-free and equilibrium based models are often based on simplifying assumptions for the number and type of underlying risk factor processes and, in particular, the dependencies between client rate, volume and market rates to derive closed form solutions of the corresponding pricing equations. For example, the client rate in the original JvD model is a deterministic function of the short rate as the only source of uncertainty, and the aggregated volume depends only on the evolution of the term structure. To reflect the typical stickiness of client rates, O'Brien [37] models balances and deposit rates as autoregressive processes and also studies alternative specifications with asymmetric adjustments. Here, the CIR one-factor interest rate model is used again to describe the dynamics of market rates.

While these papers provide empirical evidence for the US market (see also Janosi et al. [29] for an investigation of the JvD approach), an application on European market calls for some modifications due to the different characteristics of retail products. For example, Laurent [36] extends the JvD model to the case that, as in the Benelux, fidelity premium rates are rewarded for deposits that have remained in the account for a certain period. Kalkbrener and Willing [32] propose a very general modelling approach with two factors for market rates which allows for more complex movements of the yield curve. A two-factor Vasicek-type model in the Heath, Jarrow and Morton [22] (HJM) framework is used for this purpose. The latter turned out to be superior to alternative specifications in an empirical analysis [23].

Because the main focus is a realistic development of interest rates over a long period of time, historical time series are used for calibration instead of fitting the model parameters to the current market prices of plain vanilla instruments as in derivative pricing. The client rate is described as a piecewise linear function with the short rate as stochastic argument only, and a (log-) normally distributed diffusion model is used for the volume with a third stochastic factor that may be correlated with the two factors of the market rate model. In this way, correlations between market rates and volume observed in the German market are taken into account. Monte Carlo simulation is then applied to compute the value and sensitivities of a NMA, the delta profile obtained from shifting (parts of) the yield curve is used for the construction of a replicating portfolio. Certainly this numerical approach requires some more computation time than analytically solvable models, but this slight drawback is offset by a more realistic description of risk factor dynamics. Due to its generality, any of its components – term structure, deposit rate and volume model – may be replaced by some alternative specification that reflects the characteristics of the relevant market and/or retail product more appropriately.

All of the latter models are motivated by the trend to mark-to-market risk management in banking and have been derived by extending some “classical” term structure models (that have originally been developed for the pricing of interest rate derivatives) by client rates and volume as additional risk factors. In principle, a replicating portfolio would be defined as the combination of fixed income securities that exhibits the same delta profile. Caps and floors may additionally be integrated if there is a significant vega risk. However, Monte Carlo based approaches become computationally more complex when the sensitivity is calculated with respect to shifts of *several* key interest rates. The existence of volume risk would also require the consideration of liquidity constraints. Unfortunately, little is reported about the practicability and efficiency of such replicating strategies in the long run.

2.3 Dynamic replication

Immunizing the actual NMA position against movements of the present term structure only is in some sense a myopic policy. Positions of today’s “perfect hedge” possibly have to be squared at a later point in time and, thus, large

transaction volumes might be cumulated. To overcome any potential inefficiencies arising from this “lack of foresight”, future responses to changes in the risk factors should also be taken into account in the determination of the current decision which can be achieved by *multistage stochastic programming* techniques: Given the (conditional) probability distributions of the future risk factors, those policies are determined which satisfy some optimality criterion and are feasible for all possible instances of the random data. For practical reasons that will be motivated later, the probability distributions are typically expressed by a set of scenarios.

In the context of NMA replication, this allows to address the two aspects motivated at the end of the previous section 2.1: Instead of fitting the weights of the replicating portfolio to a single historic scenario only, the optimal allocation of new tranches for a given planning horizon is calculated at all points in time along each scenario for future market rates, client rate and volume. “Optimal” can be specified in a number of ways, e.g., minimization of the tracking error, maximization of the margin subject to risk limits etc. Relevant constraints include liquidity, i.e., a drop in volume must be compensated by maturing tranches (or existing instruments must be sold which may affect the P&L and/or margin). Additional restrictions, e.g., for minimum margin, loss, portfolio composition, implementable strategies etc., may easily be integrated if desired.

Note that, in contrast to static replication, the weights at which new tranches are allocated will be readjusted from stage to stage along each scenario. There are numerous examples from portfolio management that dynamic strategies derived in this way are superior to static ones with respect to risk and return (for example, see Cariño et al. [6] for a comparison of multistage stochastic programming with single-period mean variance optimization, Consiglio and Zenios [8] for tracking fixed income indices, or Fleten et al. [14] for an analysis of dynamic and fixed mix portfolio models, to mention just a few).

Multistage stochastic programming was also used in our earlier models for the determination of profit-maximizing investment or cost-minimizing funding strategies, respectively, for positions with uncertain maturities. These are based on scenarios for the future evolution of interest rates and volume, client rates were explicitly not taken into account because they do not affect the optimal policy under this objective.

The motivation for this specification are the often insufficient profits for many retail products in the banking industry. There is evidence from various applications like savings deposits [17, 19] or variable mortgages [20] that dynamic policies obtained from these stochastic programming models provide higher and simultaneously less volatile margins than the static replicating portfolio approach although they were not constructed to “track” the non-maturing asset or liability exactly. The next section describes the formulation of a multistage stochastic program for the *dynamic replication* of NMAs. To this end, the client rate will also be included in the model, and investment or financing strategies are optimized relative to the latter as benchmark.

3 Formulation as multistage stochastic program

For simplicity, we restrict ourselves to a description of the model for the management of liability positions like deposit accounts. The formulation for asset products is equivalent and can easily be derived from this specification when investing is replaced by borrowing and vice versa. At this time, we do not make any assumptions on the stochastic evolution of the risk factors except that – in addition to the factors that drive the evolution of the term structure – there is at least one additional stochastic variable which affects solely the total volume of the relevant NMA position. This is motivated by the fact that – despite the high correlations between interest rates and volume that are observed for many retail products – the volume dynamics cannot be explained fully by interest rates. Potential candidates of such additional factors may be the return on stocks as investment opportunities competing with savings deposits, or simply a residual variable for non-systematic variations in volume.

3.1 Notation

Let D be the longest maturity used for the construction of the replicating portfolio. $\mathcal{D} = \{1, \dots, D\}$ denotes the set of dates where fixed-income securities held in the portfolio mature. Maturities of standard instruments that can be used for investing are given by the set $\mathcal{D}^S \subseteq \mathcal{D}$. The model has also the option to square existing instruments prior to maturity which would be required if a decline in volume cannot be compensated by maturing tranches. This is modelled as borrowing funds of the corresponding maturities.

Due to liquidity restrictions that apply for some markets, e.g., the interbank market for the Swiss Franc, the bid-ask spreads may increase if a certain transaction volume is exceeded. This might occur when a major bank places larger amounts particularly in longer maturities. Therefore, the transaction volume in each maturity is split into *several tranches* that are priced at different spreads. The number of possible tranches is given by I^d for maturity d , $\mathcal{I}^d := \{1, \dots, I^d\}$ is a corresponding index set and ℓ_i^d the maximum amount that can be traded in the i -tranche.

The joint evolution of random data (market rates, client rate and volume of the relevant position) is driven by a stochastic process $\omega := (\omega_t; t = 1, \dots, T)$ in discrete time. The latter is formally defined on a probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = \Omega_1 \times \dots \times \Omega_T$ is the sample space, $\mathcal{F}$ the σ -field of subsets on Ω and P a probability measure. The filtration $\mathcal{F}_t := \sigma\{\omega^t\}$ generated in Ω by the history $\omega^t := (\omega_1, \dots, \omega_t)$ of the stochastic process ω defines the information set available at time t and satisfies $\{\emptyset, \Omega\} \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_T$. The random vector $\omega_t := (\eta_t, \xi_t) \in \Omega_t^\eta \times \Omega_t^\xi =: \Omega_t \subseteq \mathbb{R}^{K+L}$ can be decomposed into two components: $\eta_t \in \Omega_t^\eta \subseteq \mathbb{R}^K$ is equivalent to the state variables of a K -factor term structure model and controls market rates, client rate and volume. $\xi_t \in \Omega_t^\xi \subseteq \mathbb{R}^L$ represents L additional (economic) factors that influence only the volume.

The relevant stochastic coefficients in the optimization model derived from outcomes of ω^t at time $t = 1, \dots, T$ are:

$r_{i,t}^{d,+}(\eta^t)$	bid rate per period for investing in the i -th tranche ($i \in \mathcal{I}^d$) of maturity $d \in \mathcal{D}^S$
$r_{i,t}^{d,-}(\eta^t)$	ask rate per period for borrowing in the i -th tranche ($i \in \mathcal{I}^d$) of maturity $d \in \mathcal{D}^S$
$c_t(\eta^t)$	client rate paid per period for holding the deposit account
$v_t(\omega^t)$	volume of the relevant position.

The initial bid and ask rates $r_{i,0}^{d,+}$ and $r_{i,0}^{d,-}$ for $i \in \mathcal{I}^d$, $d \in \mathcal{D}^S$ as well as the client rate c_0 and volume v_0 at time $t = 0$ can be observed in the market and, hence, are deterministic. In the sequel, the dependency of the stochastic coefficients on ω^t or η^t will not be stressed in the notation for simplicity.

At each point in time $t = 0, \dots, T$, where T denotes the planning horizon, decisions are made on the transactions in each maturity for the allocation of maturing tranches (which are in general *not* renewed in the same maturity) plus the change in volume. This requires the following decision and state variables:

$x_{i,t}^{d,+}$	amount invested in the i -th tranche ($i \in \mathcal{I}^d$) of maturity $d \in \mathcal{D}^S$
$x_{i,t}^{d,-}$	amount financed in the i -th tranche ($i \in \mathcal{I}^d$) of maturity $d \in \mathcal{D}^S$
x_t^d	nominal amount maturing after $d \in \mathcal{D}$ periods
x_t^S	surplus in absolute terms (income from the replicating portfolio minus costs for holding the account)

A variable x_{-1}^d with negative time index represents the nominal volume in the initial portfolio with maturity $d \in \mathcal{D}$ that results from decisions in the past. The corresponding cash flow cf_{-1}^d received from these positions at date d are also known.

3.2 Specification of constraints

To ensure feasibility of investment and borrowing decisions, budget constraints must hold at each stage t that correct the nominal volume with maturity date $d \in \mathcal{D}$ by the corresponding transaction amounts:

$$\begin{aligned} x_t^d &= x_{t-1}^{d+1} + \sum_{i \in \mathcal{I}^d} x_{i,t}^{d,+} - \sum_{i \in \mathcal{I}^d} x_{i,t}^{d,-} & \forall d \in \mathcal{D}^S \\ x_t^d &= x_{t-1}^{d+1} & \forall d \in \mathcal{D} \setminus \mathcal{D}^S \end{aligned} \quad (1)$$

The next restriction controls that the sum of all positions in the portfolio matches the volume of the managed NMA at time t :

$$\sum_{d \in \mathcal{D}} x_t^d = v_t \quad (2)$$

Limits for the portion of the nominal volume in certain time buckets may optionally be defined. Let w_i^l and w_i^u be the lower and upper bound for the percentage

of the i -th bucket defined by the subset of maturity dates $\mathcal{D}_i^w \subseteq \mathcal{D}$, $i = 1, \dots, n$, where n is the number of time buckets for which such restrictions apply:

$$w_i^l \cdot v_t \leq \sum_{d \in \mathcal{D}_i^w} x_t^d \leq w_i^u \cdot v_t \quad i = 1, \dots, n \quad (3)$$

Generally the decision maker should not constrain the structure of the replicating portfolio too much since it should be a result of the optimization and not an input. However, these limits may be useful to ensure that a certain percentage of the portfolio matures in the short term for liquidity reasons. A lower bound of zero for all positions enforces that all netted exposures per maturity date have a positive sign, i.e., leverage using term transformation is not allowed. Corresponding bounds for absolute limits instead of percentages of the total volume can be defined analogously.

Without any further constraints, the model may decide to reduce the exposure in existing instruments and reinvest the money from squared positions in different maturities. Such activities can be restricted at time t to an amount equal to the sum of tranches maturing in $t, \dots, t + m - 1$:

$$\sum_{d \in \mathcal{D}^S} \sum_{i \in \mathcal{I}^d} x_{i,t}^{d,+} - \sum_{d=1}^m x_{t-1}^d \leq v_t - v_{t-1} \quad (4)$$

As a stronger limitation, short positions may be used only when a volume decline in t cannot be compensated by maturing tranches:

$$\sum_{d \in \mathcal{D}^S} \sum_{i \in \mathcal{I}^d} x_{i,t}^{d,-} \leq \max\{0, -v_t + v_{t-1}\} \quad (5)$$

Because the income from the replicating portfolio must be sufficient to cover the payments to clients and non-interest expenses α_0 for holding the deposit, we define the earnings surplus that results from transactions up to time t as

$$x_t^S = \sum_{\tau=0}^{\min\{t, D-1\}} \sum_{\substack{d \in \mathcal{D}^S \\ d > \tau}} \sum_{i \in \mathcal{I}^d} \left(r_{i,t-\tau}^{d,+} \cdot x_{i,t-\tau}^{d,+} - r_{i,t-\tau}^{d,-} \cdot x_{i,t-\tau}^{d,-} \right) + cf_{-1}^{t+2} - (c_t + \alpha_0) \cdot v_t \quad (6)$$

The first line contains the interest cash flows from positions in the replicating portfolio that have not yet matured in time t . Here, it is assumed that a corresponding fraction of the (yearly) coupon is paid at each point in time, i.e., the first line divided by the volume represents the average yield of the replicating portfolio. Note that the first cash flow resulting from a transaction in t is already taken into account in the constraint of the same stage although it actually accrues one period later. Because D is the longest available maturity, only instruments from transactions after stage $t - D$ contribute to the earnings in the first term of (6) according to this definition. The formulation of the optimization model can be extended by a large number of additional constraints for feasible

portfolio compositions, implementable transactions, minimum margin (surplus divided by volume) etc. Since these are not relevant for an understanding of the model, we do not continue the discussion here.

3.3 Complete optimization model

There is a large controversy among theorists as well as practitioners about appropriate risk measures (e.g, see the recent compilation by Szegö [47]). Volatility measured by the variance has often been criticized in this respect since it “penalizes” also above-average returns. As motivated in section 2.1, the replicating portfolio with the lowest variance might generate a margin that is insufficient to cover the total costs (including non-interest expenses). Therefore, it cannot be seen as the “risk-minimal” portfolio if its implementation leads to a sure loss. Based on these considerations, the suggested objective of the multistage stochastic program for dynamic replication of a NMA is to minimize the expected downside deviation of not meeting the overall costs $(c_t + \alpha_0)$ for all stages $t = 0, \dots, T$, as a “benchmark” for the earnings:

$$\begin{aligned}
& \min \int_{\Omega} \sum_{t=0}^T x_t^M dP(\omega) \\
& \text{s.t. mandatory constr. (1), (2), (6)} \quad t = 0, \dots, T; \text{ a.s.} \\
& \quad \text{optional constr. (3), (4), (5)} \quad t = 0, \dots, T; \text{ a.s.} \\
& \quad x_t^M \geq -x_t^S \quad t = 0, \dots, T; \text{ a.s.} \tag{7} \\
& \quad 0 \leq x_{i,t}^{d,+} \leq \ell_i^{d,+} \quad \mathcal{F}_t\text{-meas.} \quad t = 0, \dots, T; \forall i \in \mathcal{I}^d; \forall d \in \mathcal{D}^S; \text{ a.s.} \\
& \quad 0 \leq x_{i,t}^{d,-} \leq \ell_i^{d,-} \quad \mathcal{F}_t\text{-meas.} \quad t = 0, \dots, T; \forall i \in \mathcal{I}^d; \forall d \in \mathcal{D}^S; \text{ a.s.} \\
& \quad x_t^d \in \mathbb{R}; \quad \mathcal{F}_t\text{-meas.} \quad t = 0, \dots, T; \forall d \in \mathcal{D}; \text{ a.s.} \\
& \quad x_t^S \in \mathbb{R}; x_t^M \geq 0 \quad \mathcal{F}_t\text{-meas.} \quad t = 0, \dots, T; \text{ a.s.}
\end{aligned}$$

The (non-negative) auxiliary variable x_t^M ensures that only a negative surplus at stage t enters the objective and, hence, earnings with positive sign will not be minimized. All decisions and state variables in the multistage stochastic program (7) for $t > 0$ are stochastic since they depend on observations of the random data process ω^t up to time t . Therefore, they are adapted to the filtration $\mathcal{F}_t$ that specifies the information structure, i.e., they are taken only with respect to the information available at this time (*nonanticipativity*). All constraints must hold *almost surely (a.s.)*, i.e., for all $\omega \in \Omega$ except for sets with zero probability.

The numerical difficulty in the solution of the optimization problem above lies in the integration of the objective function with respect to the probability measure P . The costs in the objective at each stage $t > 0$ depend on a decision that results from the solution of a multistage stochastic program for the remaining stages $t + 1, \dots, T$. Because the objective function is not given explicitly, the integration in (7) cannot be performed analytically, and numerical methods are required.

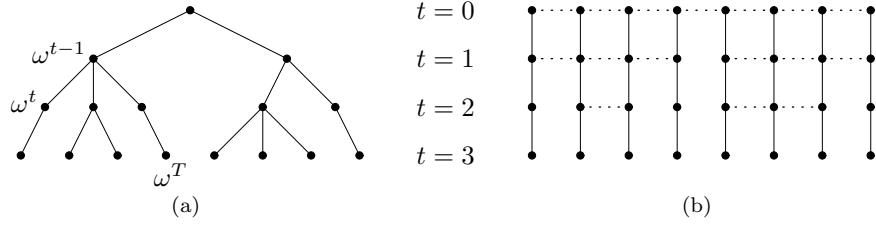


Figure 2: Scenario tree (a) and nonanticipativity constraints (b)

A common approach to make a stochastic program computationally tractable is the substitution of the original (continuous) distributions by discrete ones. To this end, the vector stochastic process ω in $(\Omega, \mathcal{F}, P)$ is approximated by a finite sample of its paths that may be represented conveniently as a *scenario tree* (see Figure 2 (a)), and each scenario corresponds to a trajectory of the process $\omega^T = (\omega_1, \dots, \omega_T)$ at the horizon T . The generation of the sample data paths is a crucial step for the solution of the problem and will be discussed separately in a later section. As a result of the approximation, the integration in the objective is replaced by probability weighted summations over the finite number of scenarios, and the constraints are duplicated for each single path. To ensure *nonanticipativity* of decisions, additional constraints link those decision and state variables that share a common history up to a certain node of the scenario tree (see Figure 2 (b)).

The resulting deterministic equivalent problem (DEQ) is a large-scale linear program that can be solved efficiently with the simplex or interior point methods. Alternatively, structural properties of the DEQ can be exploited by specialized decomposition methods [2, 41, 42, 43], but a detailed discussion of solution algorithms is beyond the scope of this paper (see the textbook by Birge and Louveaux [3] or the compilation by Wallace and Ziemba [48] for an introduction to currently available algorithms). The fact that the scenario tree grows exponentially with the number of stages leads to a corresponding growth in the size of the optimization problem which imposes some practical restrictions on the planning horizon T due to the available computational capacities. However, the significant improvement in hardware and algorithmic research in the recent years has made stochastic programming problems now solvable within reasonable time on an ordinary personal computer.

4 Models for the evolution of risk factors

In this section, we give some examples of models for the stochastic evolution of the relevant risk factors. The latter are used in combination with the stochastic programming model (7) for the management of savings deposits in Switzerland. Note that the methodology itself is not restricted to certain specifications or modelling frameworks. For instance, while we use a term structure model to

describe the dynamics of the yield curve, other approaches like econometric models or random fields may also serve well (e.g., see James and Webber [28] for a good introduction to traditional and latest term structure modelling approaches). The deposit rate and volume model can also be replaced easily for retail products with different characteristics.

4.1 The market rate model and its calibration

The market rate model describes the evolution of those interest rates that are relevant for the reinvestment of the deposit position. In an earlier work [18], we compared one- and multi-factor versions of the classical term structure models of Vasicek [49] and Cox, Ingersoll and Ross [5] (CIR) where different types of diffusion processes are used to model the state variables. A specific diffusion implies a certain (conditional) distribution of the factors, their number determines also the possible shapes of the yield curve and if interest rates of different maturities can change independently from each other. One-factor model imply a perfect correlation between rates of all maturities which clearly contradicts empirical observations. Moreover, the set of possible term structures is restricted to uniformly rising, falling or unimodally humped yield curves while a U-shaped curve can only be represented by multi-factor models.

According to our study, an extension of the Vasicek model with $K = 2$ factors provides the best description for the evolution of Swiss market rates. The dynamics of these two factors are described by the stochastic differential equations

$$\begin{aligned} d\eta_{1t} &= \kappa_1(\theta - \eta_{1t}) dt + \sigma_1 dz_{1t} \\ d\eta_{2t} &= -\kappa_2\eta_{2t} dt + \sigma_2 dz_{2t}. \end{aligned} \tag{8}$$

Both state variables exhibit mean reversion around long term means of θ and zero, respectively. The parameters κ_i , $i = 1, 2$, control the speed at which the factors revert to these levels while σ_i quantifies the instantaneous volatility of the i -th process. Under the assumption that the two Wiener processes z_1 and z_2 are uncorrelated and the factors sum up to the instantaneous short rate, the fundamental equation for the price B of a pure discount bond with maturity d at time t has some affine structure:

$$-\ln B(d; \eta_{1t}, \eta_{2t}) = \sum_{i=1}^2 [-a_i(d) + b_i(d) \cdot \eta_{it}] \tag{9}$$

where

$$\begin{aligned} a_i(d) &= \frac{(b_i(d) - d)(\kappa_i\phi_i - \sigma_i/2)}{\kappa_i^2} - \frac{\sigma_i^2 b_i^2(d)}{4\kappa_i} \\ b_i(d) &= (1 - e^{-\kappa_i d})/\kappa_i \end{aligned}$$

and $\phi_1 = \kappa_1\theta_1 + \lambda_1\sigma_1$, $\phi_2 = \lambda_2\sigma_2$. The parameters λ_1 and λ_2 are called the “market prices of risk” with respect to both factors and control the shape of the term structure. Figure 3 illustrates that the first factor closely tracks the 5

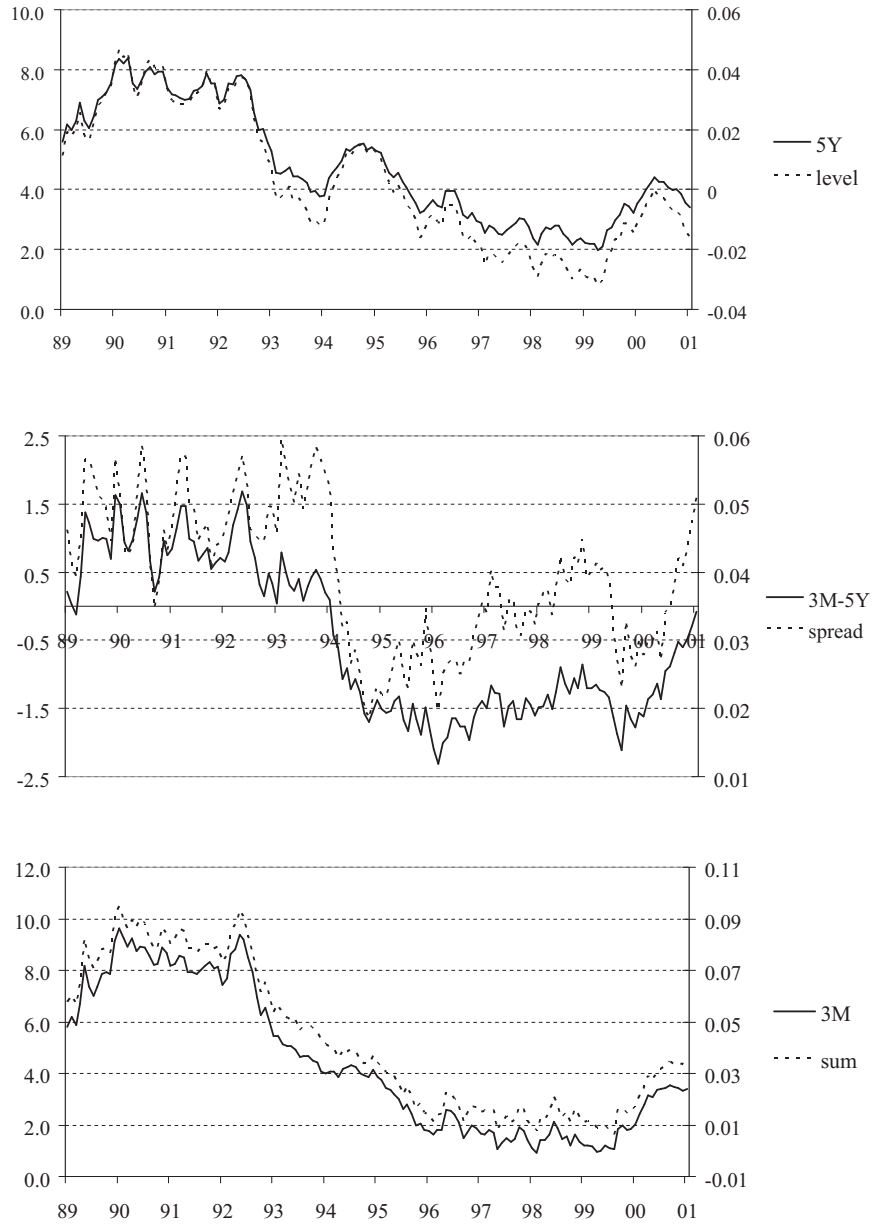


Figure 3: Interpretation of the state variables of the 2F-Vasicek model: The first factor moves parallel to the 5Y rate (top), the second factor is highly correlated with the spread between short and long rates (middle). Therefore, η_1 and η_2 may be seen as “level” and “spread” factors. By construction, their sum $\eta_1 + \eta_2$ is equivalent to the instantaneous short rate of a single-factor model (below).

year rate and the second factor the difference between the 3 month and the 5 year rate. This motivates their interpretation as “level” and “spread” factors.

If the values η_{1t}, η_{2t} at time t are known, then the factors at time $s > t$ are normally distributed with expectations and variances

$$E(\eta_{1s} | \eta_{1t}) = \theta + (\eta_{1t} - \theta)e^{-\kappa_1(s-t)} \quad E(\eta_{2s} | \eta_{2t}) = \eta_{2t}e^{-\kappa_2(s-t)} \quad (10)$$

$$\text{Var}(\eta_{is} | \eta_{it}) = \frac{\sigma_i^2}{2\kappa_i}(1 - e^{-2\kappa_i(s-t)}), \quad i = 1, 2. \quad (11)$$

Since our intention is a realistic description of the interest rate evolution for a planning horizon of several years, historical data are used for the calibration where the sample period covers a complete cycle, i.e., includes high and low level of rates, normal and inverse yield curves etc. We do explicitly not fit the model to currently observed option prices only as common in derivative valuation because this may result in unstable parameter estimates over time. Obviously, this would not be suitable for scenario generation with a long planning horizon.

The specific estimation method is a modification of the maximum likelihood approach used by Chen and Scott [7] for the calibration of one- and multi-factor CIR models. Given N observations of each state variable, the sum of logarithms of its conditional density that will enter the log likelihood function is

$$\mathcal{L}(\eta_{i1}, \dots, \eta_{iN}) = \sum_{t=1}^N \ln \left[\frac{1}{\sqrt{2\pi V_{it}}} \exp \left(-\frac{(\eta_{it} - E_{it})^2}{2V_{it}} \right) \right], \quad i = 1, 2, \quad (12)$$

with the expectations and variances

$$E_{it} := E(\eta_{it} | \eta_{i,t-1}), \quad V_{it} := \text{Var}(\eta_{it} | \eta_{i,t-1}), \quad t = 2, \dots, N.$$

according to (10) and (11). The first observation cannot be conditioned on a previous one, therefore the expectations $E_{11} := \theta$, $E_{21} := 0$ and variances $V_{i1} := \sigma_i^2/(2\kappa_i)$, $i = 1, 2$, of the unconditional distribution are used for $t = 1$.

Because the factors η_1 and η_2 are not directly observable, a change of variables is applied to construct the likelihood function from the joint density of logarithms of discount bond prices which are linear functions of the state variables. If there is no market for discount bonds, their (hypothetical) prices can easily be derived from observed par rate curves [10]. To exploit information from different segments of the term structure, discount bonds of four different maturities $d_1, \dots, d_4$ are included in the estimation (for example, the use of two bonds only would make the calibration sensitive to the choice of maturities, see also Pearson and Sun [38]). Since this exceeds the original number of factors, additional stochastic variables φ_1 and φ_2 are introduced that control the measurement errors, i.e., the differences between observed discount bond prices and those obtained from the model. Given a historical time series of N observations of the four discount bonds, the model factors and measurement errors at time $t = 1, \dots, N$ can be obtained from the linear system of equations:

$$-\ln B(d_j; \eta_{1t}, \eta_{2t}) = \sum_{i=1}^2 [-a_i(d_j) + b_i(d_j) \cdot \eta_{it}] + u_{jt}, \quad j = 1, \dots, 4, \quad (13)$$

with

$$u_{1t} := 0, \quad u_{2t} := \varphi_{1t}, \quad u_{3t} := -(\varphi_{1t} + \varphi_{2t}), \quad u_{4t} := \varphi_{2t},$$

i.e., the price for the shortest maturity is observed without bias and the errors for the remaining maturities sum up to zero. This structure is motivated primarily by tractability in deriving the likelihood function. The evolution of the error variables is modelled by two first-order autoregressive processes

$$\varphi_{jt} = \rho_j \cdot \varphi_{j,t-1} + \varepsilon_{jt}, \quad t = 1, \dots, N; j = 1, 2,$$

with $\varphi_{10} = \varphi_{20} := 0$. Because all stochastic variables are normally distributed, the log likelihood function for the calibration with discount bond prices becomes

$$\sum_{i=1}^2 \mathcal{L}(\eta_{i1}, \dots, \eta_{iN}) - N \ln |J| - N \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^N \varepsilon_t' \Sigma^{-1} \varepsilon_t, \quad (14)$$

where Σ is the covariance matrix of $\varepsilon_t := (\varepsilon_{1t}, \varepsilon_{2t})'$. The matrix J is the Jacobian of the linear transformation from latent factors η_{1t}, η_{2t} to observed price logarithms $[-\ln P(d_j; \eta_{1t}, \eta_{2t})]$ due to the change of variables [27, p. 85], and the elements of J are functions of $b_i(d_j)$.

The maximum of function (14) can be found using standard algorithms for non-linear optimization. Because the likelihood function is extremely flat, analytical derivatives must be exploited. We use the BFGS-algorithm in the implementation of Press et al. [39], combined with an extensive grid search over the parameter space for promising starting values. Typically this requires some minutes on a personal computer for sample periods of 120 months and more. Note that other specifications of the stochastic processes than (8) may increase the numerical efforts significantly. For example, the densities in a CIR-type model are non-central χ^2 which would require complex calculations of modified Bessel functions of the first kind and their derivatives with respect to the (non-integer) order.

4.2 Evolution of deposit rate and volume

The client rate represents the “benchmark” for the surplus constraint (6) in the optimization model and has direct impact on the composition of the replicating portfolio. Therefore, particular attention must be paid to the specification of its evolution and dependency on interest rates. In the case of Swiss savings accounts, one can observe that the deposit rate remains constant for longer time periods. After larger changes in the level of market rates, it is then adjusted by the bank with some delay at discrete increments of 25 or 50 basis points. The reason for these “frictions” are the associated costs (marketing, operational etc.) for such corrections.

Often an asymmetric response to market rate changes can be observed, i.e., deposit rates move more rapidly when market rates drop than they rise (see Figure 4). Moreover, the deposit rate does not exceed a certain level even if

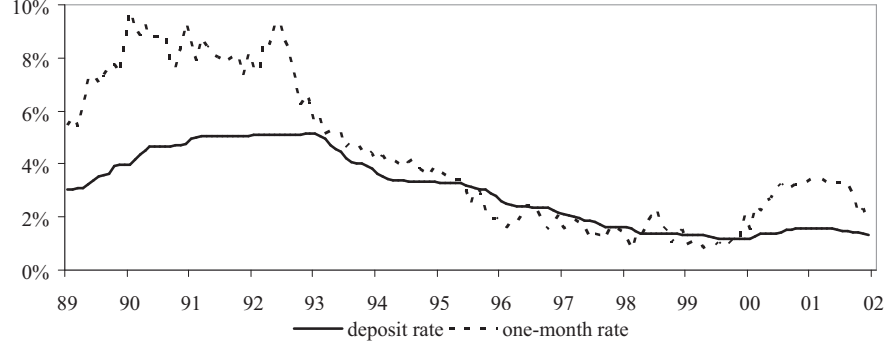


Figure 4: The deposit rate often exhibits a cap or asymmetric adjustments.

market rates are still increasing. This results from the political cap on variable mortgage rates in Switzerland since both are typically adjusted at the same pace. As a consequence, ordinary least square regression cannot be used for the estimation. The fact that the dependent variable remains unchanged while market rates fluctuate without restriction contradicts the assumption of a normally distributed error term and will lead to biased and inconsistent estimators (cf. Judge et al. [31, p. 770]).

We propose to use the previous value of the deposit rate as well as the current and past observations of the level factor as explanatory variables for the model. Assume that the possible values for the (discrete) deposit rate increments are given by $\delta_0 < \dots < \delta_n$ (including the value 0 for an unchanged rate). The *actual* change at time t depends on the outcome of a (latent) control variable

$$c_t^* = \beta_0 c_{t-1} + \beta_1 \eta_{1t} + \dots + \beta_{m+1} \eta_{1,t-m} \quad (15)$$

and follows the rule

$$\Delta c_t = \begin{cases} \delta_0 & \text{if } c_t^* \leq \gamma_1 \\ \delta_1 & \text{if } \gamma_1 < c_t^* \leq \gamma_2 \\ \vdots & \\ \delta_{1-1} & \text{if } \gamma_{n-1} < c_t^* \leq \gamma_n \\ \delta_n & \text{if } \gamma_n < c_t^*, \end{cases} \quad (16)$$

where $\gamma_1 < \dots < \gamma_n$ represent some threshold values. Note that an asymmetric adjustment of the deposit rate is reflected by different thresholds for positive and negative increments of the same size.

The deposit rate model given by (15) and (16) can be calibrated as follows: For simplicity, define the vectors

$$\beta := (\beta_0, \dots, \beta_{m+1})' \quad \text{and} \quad \psi_t := (c_{t-1}, \eta_{1t}, \dots, \eta_{1,t-m})'.$$

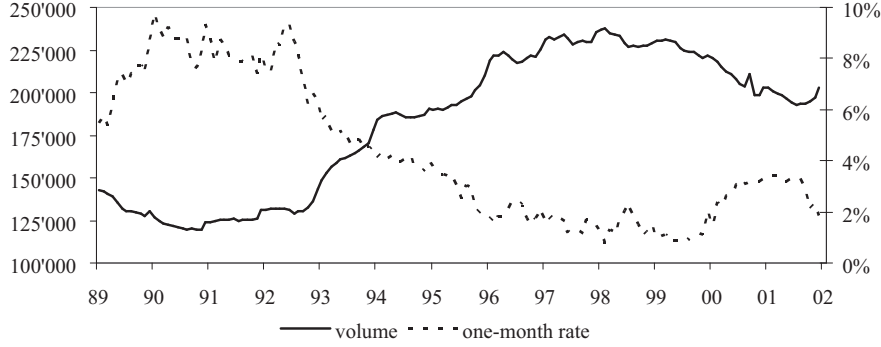


Figure 5: Aggregated savings volume (mio CHF) of all banks in Switzerland compared to the one-month rate (source: reports of the Swiss National Bank)

Let $\epsilon_t := c_t^* - \beta' \psi_t$ be the residuum for the control variable with standard deviation σ_ϵ . Then, parameter estimates are obtained by maximizing the likelihood function

$$\begin{aligned}
 & \sum_{t=1}^N \left\{ I_0(\Delta c_t) \cdot \ln \Phi\left(\frac{\gamma_1 - \beta' \psi_t}{\sigma_\epsilon}\right) + \right. \\
 & \quad I_1(\Delta c_t) \cdot \ln \left[\Phi\left(\frac{\gamma_2 - \beta' \psi_t}{\sigma_\epsilon}\right) - \Phi\left(\frac{\gamma_1 - \beta' \psi_t}{\sigma_\epsilon}\right) \right] + \dots + \\
 & \quad I_{n-1}(\Delta c_t) \cdot \ln \left[\Phi\left(\frac{\gamma_n - \beta' \psi_t}{\sigma_\epsilon}\right) - \Phi\left(\frac{\gamma_{n-1} - \beta' \psi_t}{\sigma_\epsilon}\right) \right] + \\
 & \quad \left. I_n(\Delta c_t) \cdot \ln \left[1 - \Phi\left(\frac{\gamma_n - \beta' \psi_t}{\sigma_\epsilon}\right) \right] \right\}
 \end{aligned} \tag{17}$$

where Φ is the standard normal cumulative distribution function and the indicator function $I_i(\Delta c_t)$ takes on value 1 if $\Delta c_t = \delta_i$ and 0 otherwise. The optimum of the likelihood function (17) can be found, e.g., by Newton's method.

A dependency on interest rates (or the relevant stochastic factors that drive the evolution of the term structure, respectively) must also be taken into account in a volume model. Figure 5 illustrates that the volume of savings deposits is low when market rates are above average and vice versa as clients transfer money from or to other investment opportunities. Therefore, we model relative changes in the deposit volume v_t over time by

$$\ln v_t = \ln v_{t-1} + e_0 + e_1 t + e_2 \eta_{1t} + e_3 \eta_{2t} + \xi_t. \tag{18}$$

The constant e_0 and the time component $e_1 t$ reflect that the total deposit volume exhibits a positive trend. The factors η_1 and η_2 of the term structure model are included as explanatory variables to reflect the discussed dependency on

market rates. An additional stochastic factor ξ which is uncorrelated with the market rate model factors takes into account that the latter do not fully explain the observed evolution of the balance. Given time series of η_1 and η_2 from the calibration of the term structure model factors, equation (18) can easily be estimated by ordinary least squares regression. The volatility σ_ξ of the residuum factor ξ is then derived from the standard error.

5 Scenario generation

The models introduced in the previous section imply that the stochastic factors $\omega_t := (\eta_{1t}, \eta_{2t}, \xi_t)$ at $t = 1, \dots, T$ are multivariate normally distributed with conditional expectation

$$\mu(\omega_t | \omega^{t-1}) = \begin{pmatrix} \theta + (\eta_{1,t-1} - \theta) \cdot e^{-\kappa_1 \cdot h} \\ \eta_{2,t-1} \cdot e^{-\kappa_2 \cdot h} \\ 0 \end{pmatrix} \quad (19)$$

and covariance matrix

$$\Sigma(\omega_t | \omega^{t-1}) = \begin{pmatrix} \frac{\sigma_1^2}{2\kappa_1}(1 - e^{-2\kappa_1 \cdot h}) & 0 & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}(1 - e^{-2\kappa_2 \cdot h}) & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix}. \quad (20)$$

The parameter h is the period length between two successive stages of the multi-period optimization problem (7). We have already motivated at the end of section 3 that these continuous distributions must be approximated by discrete ones to make the multistage program computationally tractable. Moreover, the approximation is a critical element for the solution with respect to accuracy and robustness. Loosely speaking, the substitution of the complete universe of possible outcomes by a relatively small sample should lead to the same solution as it would be in the (non-tractable) case of the original distributions. Therefore, an adequate representation of the underlying random data is of utmost importance.

At first glance, Monte Carlo simulation might be an obvious approach to generate finite sets of samples of the given distributions. Unfortunately, we must inevitably draw large numbers of points to obtain a reasonable accuracy (in the sense of small confidence intervals and robustness of the solution with respect to a “contamination” of the sample, e.g., by using a different initial value for the random number generator). While large set of samples can be acceptable for two-stage problems, the number of required scenarios would become too large for a multistage stochastic program due to the exponential growth in problem size with the number of stages. Therefore, any Monte Carlo based sampling procedure must be combined with a variance reduction technique like importance sampling [9] for example.

A special variant of the latter that appears promising for multistage stochastic programming problems uses the *expected value of perfect information* (EVPI)

[11] to identify scenarios with higher impact on the solution. The EVPI describes the loss in the objective value due to uncertainty and thus measures the amount a decision maker would pay in return for complete information. Local EVPIs are calculated for each node of the scenario tree from the difference in the objective values of the original solution and a corresponding problem where the nonanticipativity constraints for this node and all of its successors are relaxed. Starting with a tree generated by Monte Carlo, scenarios are then added at nodes with a large EVPI or possibly removed where the value is close to zero. By this means, a sequence of scenario trees is generated and corresponding (large-scale) optimization problems are solved until a stopping criterion is fulfilled.

An alternative to sampling procedures, which provide probabilistic error bounds only, are bound-based approximations. As an example, the *barycentric approximation* proposed by Frauendorfer [15, 16] deliberately constructs the scenario trees so that exact upper and lower bounds to the stochastic program with the original distributions are derived. The analysis of the discretization error – defined as the difference between the two bounds – identifies those nodes where the approximation must be improved by generating additional scenarios until the accuracy is sufficient. Again, this might require the solution of a sequence of increasingly larger optimization problems although experience from our earlier model for income maximization of variable products [18, 20] where we used this technique implies that the iteration can be stopped already after calculation of the first lower and upper bound. A drawback is that the method requires certain structural properties of the stochastic processes and the formulation of the optimization problem such as deterministic left-hand-sides of the constraints which are not given for the dynamic replication model (7).

For practical reasons, a robust solution should preferably be available already after solution of a single optimization problem to reduce the overall computation time. Thus the initial scenario tree must be “accurate enough” to provide such a result. Another approach first suggested by Høyland and Wallace [24] and Dupačová et al. [12] builds the tree in a way so that some statistical properties of the data process are preserved, e.g., expectations, variances, correlations (also over time) etc. This is achieved by matching the moments of the discrete approximated distributions to those of the original (continuous) ones.

As an illustration, consider a random variable ω for which values of certain moments or expectations of continuous functions $\mu_k = \int_{\Omega} u_k(\omega) dP$, $k = 1, \dots, m$, are given. Then, a modest number of sample points ω^s with nonnegative probabilities p_s , $s = 1, \dots, S$, is determined so that the probabilities sum up to one and the moment values are matched as well as possible. To this end, a measure of distance, e.g., the square norm, is minimized with the outcomes and probabilities as decision variables where different weights γ_k , $k = 1, \dots, m$, may be used to account for the relevance of the k -th moment:

$$\min \sum_{k=1}^m \gamma_k \left(\sum_{s=1}^S p_s u_k(\omega^s) - \mu_k \right)^2 \quad (21)$$

subject to

$$\begin{aligned} \sum_{s=1}^S p_s &= 1, \\ p_s &\geq 0, \quad s = 1, \dots, S. \end{aligned}$$

Kouwenberg [35] presents a case study where this type of approach is compared with Monte Carlo sampling: His stochastic programming model with randomly sampled scenario trees performs poorly, and the solutions exhibit extremely large buying and selling amounts. Fitting the tree to the expectations and covariances of the original distributions stabilizes the optimal decisions and greatly improves the performance which underlines the importance of choosing an appropriate scenario generation method.

However, the optimization problem (21) is highly non-linear, and its solution in each node of the scenario tree can make the generation of the latter extremely time-consuming. Some authors try to overcome this obstacle by restricting themselves to the first two moments only [35], or by using heuristics to identify a local optimum as a “sufficiently good” solution of the moment matching problem [24, 25].

As an alternative, a representative selection of scenarios may be derived by a discrete distribution that converges to the desired one by refinements. Siede [45] approximates the multivariate normal distributions defined by (19) and (20) by a multinomial distribution which reveals the same expectations and covariance matrix after a transformation. This procedure can roughly be outlined as follows (for a k -dimensional normal distribution): Consider a multinomial distribution $\mathcal{N}$ with parameters

$$(l; \underbrace{\frac{1}{k+1}, \dots, \frac{1}{k+1}}_{k+1}),$$

i.e., the probability distribution of a $(k+1)$ -dimensional random vector $(X_0, \dots, X_k)$ with nonnegative integer elements that satisfies $X_0 + \dots + X_k = l$. In the first step, an orthogonal transformation $\mathcal{T}$ is applied to the support of $\mathcal{N}$ which will remove the correlations between the components and sets the expectations $E(\mathcal{T} \cdot n_i)$, $n_i \in \mathcal{N}$, for $i = 1, \dots, k$ to zero. A second transformation “stretches” the variances to one. The first two moments are now equivalent to those of a standard normal distribution. Denote the resulting random vector by $y \in \mathbb{R}^k$. The distribution of y converges to the k -dimensional standard normal distribution for $l \rightarrow \infty$. Thus by the transformation

$$\hat{y} := \mu + Ly$$

a random variable $\hat{y}$ can be constructed with expectation μ and covariance matrix $\Sigma = LL'$ (i.e., L is the Cholesky decomposition of Σ). It converges for $l \rightarrow \infty$ to the k -dimensional multivariate normal distribution with expectation μ and covariance matrix Σ .

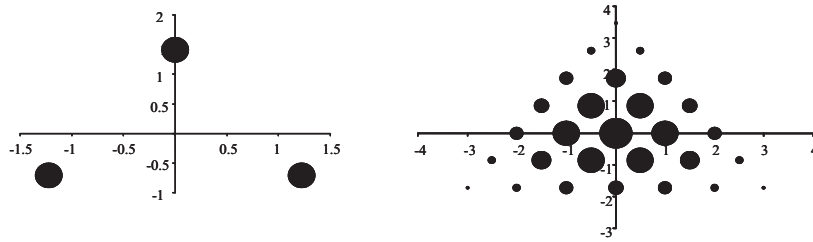


Figure 6: Impact of the parameter l on the sample size

As a consequence, for a given multivariate normal distribution a discrete approximation $y(n)$ can be constructed which is based on the multinomial distribution and has $(l+1) \cdot \dots \cdot (l+k)/k!$ sample points. The influence of l on sample size is illustrated in Figure 6 for the approximation of a 2-dimensional standard normal distribution. Note the different scaling of the left ($l=1$) and right ($l=6$) diagram: The larger the parameter, the more extreme (but less likely) scenarios are generated. The probability of a point is represented by its diameter.

The advantage of this approach is that the time required for the generation of the scenario tree becomes negligibly short. We also compared discretizations by multinomial distributions with approximations from exact bounds for benchmark problems that satisfy the properties required for the latter: The obtained solution was practically identical to those derived from the bounded problems with low discretization error after some refinement iterations. Thus we can be confident that the approximation is sufficiently accurate.

6 Application to savings deposits

The ability of the multistage stochastic programming model (7) to replicate a NMA position was tested in a case study and compared to the “conventional” static approach. Client rate and volume data of a real Swiss savings deposit account were provided by a bank, the investigation started in January 1989 and ended in December 2001. This case study period is characterized by an increase in interest rates shortly after the beginning. The yield curve was inverse on a high level during the first half of the 1990s, then interest rates dropped sharply while the volume increased for most of the remaining time. The evolution of the relevant interest rates, deposit rate and volume for the case study period is shown in Figure 7.

Money market instruments and swaps with maturities of 1, 2, 3, 4, 5, 7 and 10 years were allowed for the composition of the replicating portfolio. Because the Swiss interbank market exhibits a limited liquidity in particular at the long end of the yield curve, the spreads in each maturity increase with the transaction volume. The bid/ask spreads for maturities up to 5 years were set to 10 bp for initial tranches of 200 mio Francs (except 5 bp for the maturity one year) and

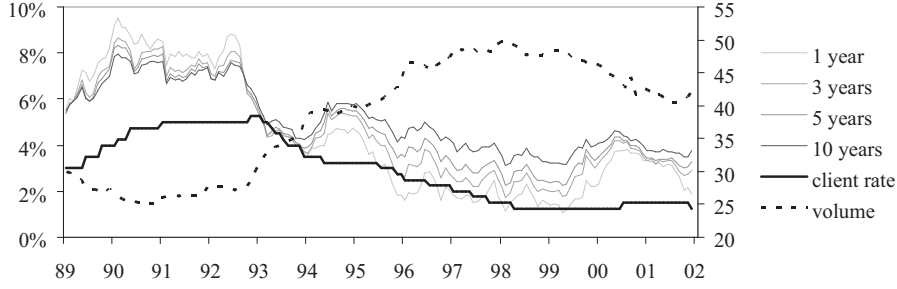


Figure 7: Market rates, client rate and volume used as input for the case study

increased by 1 bp for each additional tranche of 100 mio. In addition, absolute upper limits of 200 mio were applied to the amounts in 7 and 10 years. Short sales (squaring positions) were only permitted if a drop in volume could not be compensated by maturing tranches. The objective of the optimization model was to minimize the downside of not achieving a margin of 200 bp that is considered as a reasonable target for the income. The planning horizon was set to 7 years (one stage equals one year).

The static benchmark strategy was also calculated by the bank using a tracking error minimization over historical data and was frequently readjusted. The rule applied to the allocation of new positions worked as follows: Maturing tranches were always reinvested at their original maturity. Any change in volume is invested or financed at fixed weights, where negative amounts in case of a volume drop are set against maturing tranches. The initial portfolio for the investigation of both approaches was constructed under the assumption that a static mix of 2 year and 5 year tranches had been applied before, i.e., for a total volume of 30 bio an amount of 875 mio matures in each of the first 24 months and 250 mio in each of the following 36 months.

Starting with the first observations of market rates, deposit rate and volume in January 1989, an optimal investment policy for the maturing 875 mio was obtained from the optimization model and implemented by an update of the positions in the existing portfolio. This new portfolio was then used for the next optimization in February, together with the new interest rate curve, a new deposit rate and volume, and so on. After processing the complete 13 year period, the resulting margins defined as the income from the replicating portfolio minus the client rate were calculated for each month.

The corresponding evolutions for both approaches are shown in Figure 8. It can be seen that the dynamically managed portfolio mostly outperforms the static one. The fact that there is only a small difference between both methods at the beginning of the study can be attributed to the following reasons: First, it took five years until the last tranche from the initial portfolio is matured, i.e., a large portion of the replicating portfolio stems from old positions during the first years. Second, the market environment at the beginning was also “comfortable” for the static approach because the cap on the deposit rate allowed for large

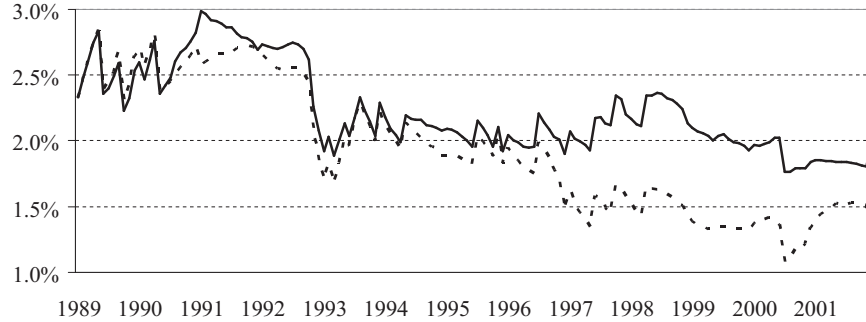


Figure 8: Margin evolution: Dynamic (solid) vs. static replication (dashed)

margins with any possible strategy. After the drop in market rates, the margin of the static portfolio reduces continually. On the other hand, the stochastic programming model achieves the specified target well for the remaining time.

Table 2 summarizes the relevant key data: Over the entire 13 year period, the average margin could be improved by 30 bp at a lower standard deviation using the dynamic replication. This shows that the latter can manage the trade-off between risk and return more efficiently. Since the volatility of the margin is reduced, the portfolio obtained from the stochastic optimization model may be viewed as the better replication of the uncertain position. For a better assessment of the given performance numbers, “difference to 3M portfolio” relates to the additional income in comparison with a portfolio of three continuously renewed 3 month tranches.

Figure 9 displays the composition of both portfolios over time. The stochastic programming model uses to some extent also longer maturities that are not considered by the static replication. However, a “dynamic management” does not result in larger fluctuations of the time buckets than a static rule as one might suspect. The dynamic portfolio exhibits a higher duration, indicating that the “true” maturity of the savings deposit position is approx. 0.5 years longer than implied by the static replication. One might argue that the higher margin results from an extension of the duration only. An analysis of potential static portfolios (with constant weights) where the average maturity was extended by half a year reveals that this would provide at most a gain of 10 bp at a larger volatility. Therefore, we can conclude that the extended margin here can mainly be attributed to the added value of dynamic management.

Finally, the number of financing activities refers to cases where a drop in volume could not be compensated by maturing tranches and, thus, existing positions had to be squared. Despite the longer duration, this occurs less frequent with the dynamic model than in the static approach. This can be explained by the fact that the volume risk is reflected more appropriately because the corresponding stochastic model also takes the correlations with interest rates into account.

	dynamic	static
mean margin	2.23 %	1.93 %
standard deviation	0.32 %	0.49 %
difference to 3M portfolio	92 bp	62 bp
avg. portfolio maturity	2.37 yrs	1.81 yrs
no. financing activities	14	26

Table 2: Case study results (including transaction costs)

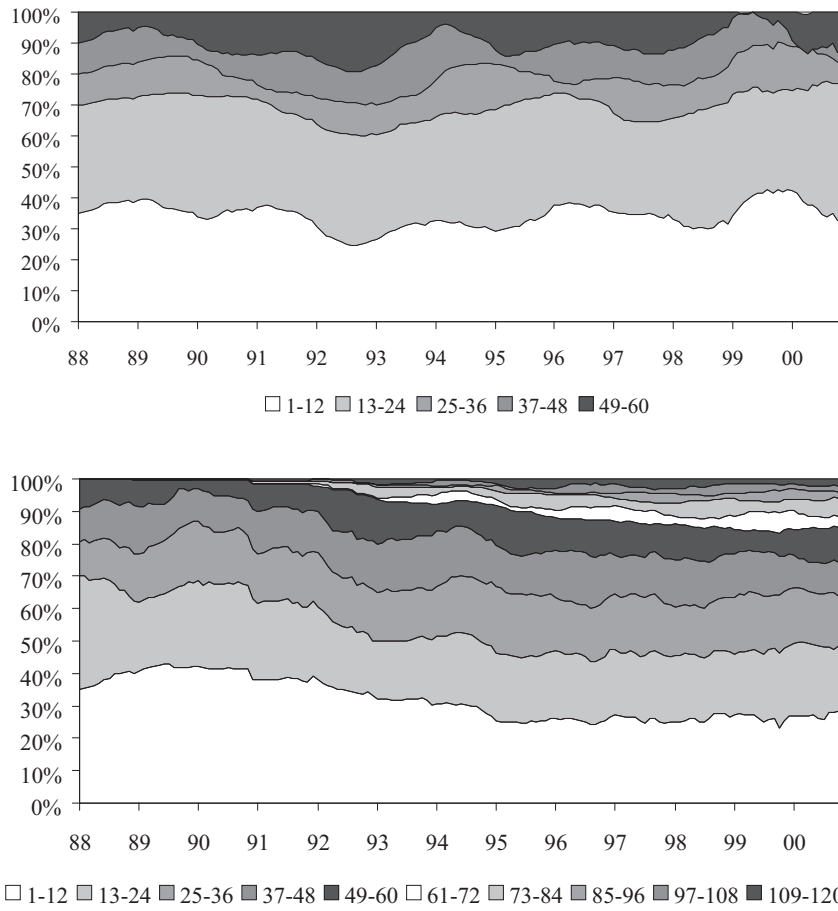


Figure 9: Portfolio compositions over time: Static (top) vs. dynamic (below)

7 Summary

The problem of calculating a replicating portfolio for non-maturing account (NMA) positions was formulated as a multistage stochastic program that can be solved by standard optimization algorithms. As opposed to the conventional approach where the portfolio weights are adjusted once by minimizing the tracking error over a historical sample period, the model is applied anytime when positions must be reinvested or financed due to maturing tranches or a change in volume. Instead of a single historic scenario, the decision about the allocation of new tranches is made with respect to a large set of scenarios for the relevant risk factors. Furthermore, future transactions subject to various constraints and their impact on today's portfolio are also taken into account.

Compared to the static approach, the dynamic replicating portfolio turned out to be more efficient since it provides a higher margin at reduced volatility. The selection of a different margin target in the downside minimization certainly has some impact on the realized earnings and volatility. In this sense, the model allows also for an analysis and optimization of the tradeoff between risk and return. Figure 10 shows the impact of different targets on the expected margins that are calculated from the portfolios along each path of the scenario tree. Too low targets can lead to inefficient solutions as there are portfolios that provide a higher margin at lower volatility. A decision maker with the goal of minimizing margin volatility will implement the strategy associated with the lowest standard deviation (which is not obtained always for the same target). If the expected margin of a certain decision is regarded as insufficient, the risk of higher targets may be analyzed ex ante. Likewise the influence of tighter or more relaxed constraints may be assessed.

The optimization approach is very flexible in respect of alternative risk measures as objective functions or additional restrictions. Although we have implemented some alternative formulations of the optimization problem (7), we did not give any examples here because such extensions are straightforward. Due to the fact that the resulting deterministic equivalent problem is of large size (all decision variables and constraints must be duplicated for each scenario), it is beneficial if the model can be formulated as a *linear* or other type of convex program. A number of popular risk measures fulfills this requirement, e.g., mean-absolute deviation [33] or conditional value-at-risk (CVaR) [40] can be formulated as linear models, minimization of variance in a multistage context can be solved efficiently by suitable optimization algorithms [46] or approximated by piecewise linear functions [34]. On the other hand, using value-at-risk (VaR) as objective leads to nonconvex optimization problems [21] that are extremely hard to solve.

Analogously, the multistage stochastic programming approach is not restricted to certain models for the evolution of risk factors. Arbitrary specifications can be chosen to describe the adjustment of client rates or dependency of the volume on interest rates to take certain characteristics of the relevant product into account. The models for interest rates, client rate and volume are estimated individually and, hence, different sample periods can be used for each

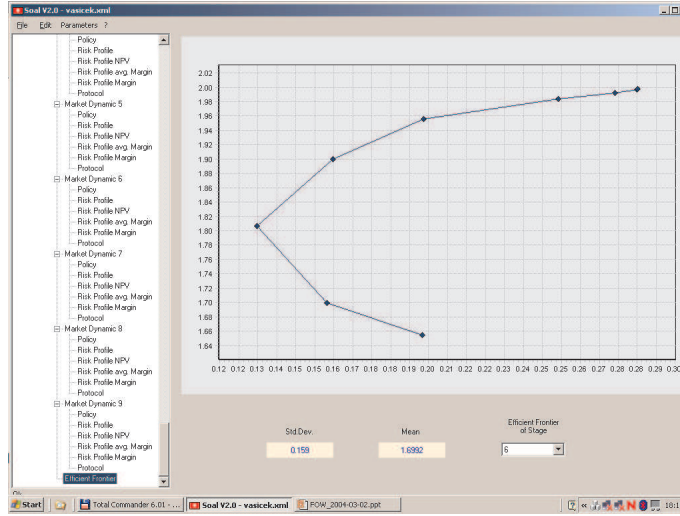


Figure 10: Impact of the margin target on the optimal solution. The expected margins (y-axis) and their standard deviation (x-axis) are calculated from the replicating portfolios along the paths of the scenario tree.

of them. This can be useful for products where only short time series of the bank-specific client rate and/or volume are available while long interest rate histories for the calibration of the term structure model can easily be obtained from public sources. According to our experience with the factor model described in section 4.1, the data should cover at least one interest cycle to ensure that the parameter for the mean reversion level is not biased by the selection of a specific sample period.

The dynamic optimization framework represents also a substantial difference to the stochastic models reviewed in section 2.2. The latter primarily calculate the value of a NMA position, a replicating portfolio can then be derived as the combination of instruments with known prices that has similar characteristics, i.e., sensitivity with respect to changes in the *current* yield curve. By consideration of the *future* transactions, the multistage stochastic programming model takes possible corrections of the initial hedge also into account. One can expect that this helps reduce (costly) transactions at later points in time and might also lead to a more efficient replication of the NMA. However, a comparison of our stochastic programming model with the contributions from the financial literature remains subject to further research.

References

- [1] Basel Committee on Banking Supervision (2004). *Principles for the management and supervision of interest rate risk*. Bank for International Settlements, Basel.
- [2] J.R. Birge, C.J. Donohue, D.F. Holmes, and O.G. Svintsitski (1996). A parallel implementation of the nested decomposition algorithm for multi-stage stochastic linear programs, *Mathematical Programming* 75, pp. 327–352.
- [3] J.R. Birge and F. Louveaux (1997), *Introduction to Stochastic Programming*. Springer, New York.
- [4] A. Bühler (2000). Modellierung von Bodensatzprodukten. *Schweizer Bank* 15, pp. 44–45.
- [5] J.C. Cox, J.E. Ingersoll and S.A. Ross (1985). A theory of the term structure of interest rates, *Econometrica* 53, pp. 385–407.
- [6] D.R. Cariño, T. Kent, D.H. Myers, C. Stacy, M. Sylvanus, A.L. Turner, K. Watanabe, W.T. Ziemba (1994). The Russell-Yasuda Kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming. *Interfaces* 24, pp. 29–49.
- [7] R.-R. Chen and L. Scott (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income* 3, pp. 14–31
- [8] A. Consiglio and S.A. Zenios (2001). Integrated simulation and optimization models for tracking international fixed income indices. *Mathematical Programming (Ser. B)* 89, pp. 311–339.
- [9] G.B. Dantzig and G. Infanger (1993). Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research* 45, pp. 59–76.
- [10] M. Deacon and A. Derry (1994). *Estimating the Term Structure of Interest Rates*. Working paper, Bank of England, London.
- [11] M.A.H. Dempster (1981). The expected value of perfect information in the optimal evolution of stochastic problems. In M. Arato, D. Vermes and A.V. Balakrishnan, eds.: *Stochastic Differential Systems*. Springer (Lecture Notes in Information and Control Vol. 36), Berlin, pp. 25–40.
- [12] J. Dupačová, G. Consigli and S.W. Wallace (2000). Scenarios for Multi-stage Stochastic Programs. *Annals of Operations Research* 100, pp. 25–53.
- [13] R. Fiedler, K. Frauendorfer and M. Schürle (2004). Dynamic management of core deposits. *Banking and Finance* 3, pp. 19–23.

- [14] S.-E. Fleten, K. Høyland and S.W. Wallace (2002). The performance of stochastic dynamic and fixed mix portfolio models. *European Journal of Operational Research* 140, pp. 37–49.
- [15] K. Frauendorfer (1994). Multistage stochastic programming: Error analysis for the convex case. *Mathematical Methods of Operations Research* 39, pp. 93–122.
- [16] K. Frauendorfer (1996). Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming (Ser. B)* 75, pp. 277–293.
- [17] B. Forrest, K. Frauendorfer and M. Schürle (1998). A stochastic optimization model for the investment of savings account deposits. In P. Kischka, H.-W. Lorenz, U. Derigs, W. Domschke, P. Kleinschmidt and R. Möhring, eds.: *Operations Research Proceedings 1997*, Springer, Berlin, pp. 382–387.
- [18] K. Frauendorfer and M. Schürle (2001). Term structure models in multistage stochastic programming: Estimation and approximation. *Annals of Operations Research* 100, pp. 189–209.
- [19] K. Frauendorfer and M. Schürle (2003). Management of non-maturing deposits by multistage stochastic programming. *European Journal of Operational Research* 151, pp. 602–616.
- [20] K. Frauendorfer and M. Schürle (2005). Refinancing mortgages in Switzerland. In S.W. Wallace and W.T. Ziemba, eds.: *Applications of Stochastic Programming*, MPS-SIAM Series in Optimization, Philadelphia, pp. 445–469.
- [21] A.A. Gaivoronski and G. Pflug (2005). Value at risk in portfolio optimization: Properties and computational approach. *Journal of Risk* 7, pp. 1–31.
- [22] D. Heath, R. Jarrow and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica* 60, pp. 77–105.
- [23] F. Heitmann and S. Trautmann (1995). *Gaussian multi-factor interest rate models: Theory, estimation, and implications for option pricing*. Working paper, Johannes Gutenberg University Mainz.
- [24] K. Høyland and S.W. Wallace (2001). Generating Scenario Trees for Multistage Decision Problems. *Management Science* 47, pp. 295–307.
- [25] K. Høyland, M. Kaut and S.W. Wallace (2001). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications* 24, pp. 169–185.
- [26] D.E. Hutchison and G.G. Pennacchi (1996). Measuring rents and interest rate risk in imperfect financial markets: The case of retail bank deposits. *Journal of Financial and Quantitative Analysis* 31, pp. 399–417.

- [27] M.D. Intriligator, R.G. Bodkin and C. Hsiao (1996). *Econometric Models, Techniques, and Applications*. Prentice-Hall, Upper Saddle River NJ.
- [28] J. James and N. Webber (2000). *Interest Rate Modelling*. Wiley, Chichester.
- [29] T. Janosi, R. Jarrow and F. Zullo (1999). An Empirical Analysis of the Jarrow-van Deventer Model for Valuing Non-Maturity Demand Deposits. *Journal of Derivatives* 7, pp. 8–31.
- [30] R.A. Jarrow and D.R. van Deventer (1998). The arbitrage-free valuation and hedging of demand deposits and credit card loans. *Journal of Banking and Finance* 22, pp. 249–272.
- [31] G.G. Judge, R.C. Hill, W. Griffiths, H. Lütkepohl, and T.-C. Lee (1987). *Introduction to the Theory and Practice of Econometrics*. Wiley, New York.
- [32] M. Kalkbrenner and J. Willing (2004). Risk management of non-maturing liabilities. *Journal of Banking and Finance* 28, pp. 1547–1568.
- [33] H. Konno and H. Yamazaki (1991). Mean-absolute deviation portfolio optimization model and its application to Tokyo stock market. *Management Science* 37, pp. 519–531.
- [34] H. Konno and K. Suzuki (1992). A fast algorithm for solving large scale mean-variance models by compact factorization of covariance matrices. *Journal of the Operations Research Society of Japan* 35, pp. 93–104.
- [35] R. Kouwenberg (2001). Scenario generation and stochastic programming models for asset and liability management. *European Journal of Operational Research* 134, pp. 279–292.
- [36] M.-P. Laurent (2004). *Non-maturity deposits with a fidelity premium*. Working paper, Solvay Business School, Université Libre de Bruxelles.
- [37] J.M. O’Brien (2000). *Estimating the value and interest rate risk of interest-bearing transactions deposits*. Board of Governors of the Federal Reserve System, Washington D.C.
- [38] N.D. Pearson and T.-S. Sun (1994). Exploiting the conditional density in estimating the term structure: An application of the Cox, Ingersoll, and Ross model. *Journal of Finance* 49, pp. 1279–1304.
- [39] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992). *Numerical Recipes in C* (2nd ed.). Cambridge University Press.
- [40] R.T. Rockafellar and Stanislav Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, pp. 1443–1471.

- [41] C.H. Rosa and A. Ruszczyński (1996). On augmented Lagrangian decomposition methods for multistage stochastic programs. *Annals of Operations Research* 64, pp. 289–309.
- [42] A. Ruszczyński (1993). Parallel decomposition of multistage stochastic programming problems. *Mathematical Programming* 58, pp. 201–228.
- [43] A. Ruszczyński (1997). Decomposition methods in stochastic programming. *Mathematical Programming* 79, pp. 333–353.
- [44] R. Selvaggio (1996). Using the OAS methodology to value and hedge commercial bank retail demand deposit premiums. In F. Fabozzi and A. Konishi, eds.: *The Handbook of Asset/Liability Management*, Irwin, pp. 363–373.
- [45] H. Siede (2000). *Multi-Period Portfolio Optimization*. PhD thesis, University of St. Gallen.
- [46] M. Steinbach (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review* 43, pp. 31–85.
- [47] G. Szegö, ed. (2004). *Risk Measures for the 21st Century*. Wiley, New York.
- [48] S.W. Wallace and W.T. Ziemba, eds. (2005). *Applications of Stochastic Programming*. MPS-SIAM Series on Optimization, Philadelphia.
- [49] O. Vasicek (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, pp. 177–188.