

Hedging of long term zero-coupon bonds in a market model with reinvestment risk

David Stefanovits · Mario V. Wüthrich

Received: 28 June 2013 / Revised: 1 November 2013 / Accepted: 18 November 2013 /
Published online: 14 January 2014
© DAV / DGVFM 2013

Abstract We present a computational methodology to value and hedge long term zero-coupon bonds trading in short and medium term ones. For this purpose we develop a discrete time stochastic yield curve model with limited availability of maturity dates at a fixed time point and newly issued bonds at future time points. This involves reinvestment risk and there is no perfect hedging strategy available for long term liabilities. We calibrate the model to market data and describe optimal hedging strategies under a given risk tolerance. These considerations provide a natural extrapolation of the yield curve beyond the last liquid maturity date, and a framework which allows to value long term insurance liabilities, for instance, under Solvency 2. Moreover, we determine the optimal trading strategy replicating the liabilities under the given risk tolerance.

1 Introduction

This paper is motivated by the problem of valuing and hedging long term insurance liabilities. Solvency directives require to value liabilities in a market-consistent way. That is, insurance cash-flows should be optimally replicated by an appropriate investment strategy of liquid financial instruments. This replication might prove difficult when considering long term liability cash-flows such as life insurance contracts or pension liabilities. This is because the value of these cash-flows may depend on factors, such as mortality rates, which cannot be traded in a liquid market. Therefore the insurer is facing underwriting risk. Moreover, the value of these cash-flows may also depend on yields beyond the maturities available in the liquid bond market. In that case, securities with shorter times to maturity need to be rolled over and the insurer is also subject to reinvestment risk. Because of the

D. Stefanovits (✉) · M. V. Wüthrich
Department of Mathematics, RiskLab, ETH Zurich, 8092 Zurich, Switzerland
e-mail: david.stefanovits@math.ethz.ch

absence of liquid markets for these risks, the insurer cannot hedge the liabilities completely, or complete hedging may be highly capital inefficient. A possible approach is to value long term contracts as the minimal capital which must be invested in liquid financial instruments to cover the contract at an acceptable level of risk, see Hilli et al. [9]. In the literature this type of hedging technique is known as acceptable hedging. A mathematical presentation can be found in Artzner et al. [1].

In this paper we apply this valuation approach to a specific sub-problem of the one outlined above. We aim to value long term deterministic cash-flows with short and medium term zero-coupon bonds. This means that we only consider reinvestment risk. Our valuation methodology is not based on simple yield curve extrapolation using families of parametrized curves, such as the Svensson family (see e.g. Müller [14]), but rather on replication with actual investment strategies. We assume that we can trade short and medium term zero-coupon bonds in a liquid market. In practice, this is not entirely the case, since most of the medium term securities in the liquid government and corporate bond market are coupon bearing. Because coupon bonds are simply linear combinations of zero-coupon bonds with different maturities, our methodology can be modified for that case. However, the valuation algorithm becomes more complex from the computational point of view.

The first step of our work is to model the yield curve process appropriately. Classical stochastic term structure models, which assume that bonds with unlimited times to maturity are traded at each point in time, do not provide the right framework for our problem and do not seem appropriate for real world solvency modeling. In our set-up we can only trade bonds up to a certain time to maturity, and at each point in time a new bond is issued into the market with maturity date unavailable at the previous time points. This involves roll over and reinvestment risk in a natural way. To our knowledge there are only a few papers in the literature dealing with the problem of modeling bonds up to a finite time to maturity with newly issued bonds as time passes. See Sommer [19], and Dahl [4, 5]. In Sommer [19] a continuous time model with new bonds being issued continuously in time is considered, whereas in the continuous time setting of Dahl [5] new bonds are issued at a fixed set of time points. In Dahl [4] reinvestment risk is introduced in a discrete time non-recombining binomial model.

We work in a discrete time set-up and present a Heath-Jarrow-Morton type model with time-dependent volatility structure based on ideas developed in Ortega et al. [16], and Teichmann and Wüthrich [20]. In particular, we modify the model presented in [20] introducing the restriction of finitely many available times to maturity, and modeling newly issued zero-coupon bonds using an additional stochastic structure. In this way we obtain a market model with reinvestment risk. Except for the market price of risk, the calibration of the model to market data is straightforward. The model has two essential features for the problem of valuing long term liabilities: its calibration is consistent over a long period of time and it does not require nested simulations to generate yield curve scenarios. The model is presented in Sect. 2 and the calibration procedure in Sect. 5. The rest of the paper is organized as follows. In Sect. 3 we introduce trading strategies in this market model and prove formally that deterministic long term cash-flows are unattainable. Then,

we define the concept of acceptable hedging in the context of our model and formulate the corresponding dynamic stochastic optimization problem. In Sect. 4.1 we apply the Galerkin approximation technique which allows to efficiently solve the optimization problem numerically, see e.g. Koivu and Pennanen [13]. To apply this technique, classes of parametric trading rules need to be considered. In Sect. 4.2 we consider four classes and derive their properties. Finally, in Sect. 6 we apply the methodology to zero-coupon bond price time series derived from Swiss government bonds.

2 Bond market model with reinvestment risk

2.1 Bond market model with finitely many instruments

We propose a discrete-time bond market model. Choose a fixed grid size $\delta > 0$ and $d, n \in \mathbb{N}$ with $d < n$. Let $J = \{0, \delta, \dots, n\delta\}$ be the set of discrete time points with final time horizon $T = n\delta$. Let $D = \{1, \dots, d\}$ and assume that at each time point $t \in J$ zero-coupon bonds with times to maturity $j\delta$ for $j \in D$ are available at the financial market. Define the sets $J' = J \setminus \{0\} = \{\delta, \dots, n\delta\}$ and $D^- = D \setminus \{d\} = \{1, \dots, d-1\}$. Our aim is to model the d zero-coupon bond prices stochastically at each $t \in J'$ taking into account that the zero-coupon bonds with maturities $t + j\delta$ for $j \in D^-$ are available at the market at time $t - \delta$, whereas the one with maturity $t + d\delta$ is not.

Let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a filtered probability space with real-world probability measure \mathbb{P} and filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in J}$. We aim to model the d -dimensional \mathbb{F} -adapted yield curve process

$$(\mathbf{Y}_t)_{t \in J} = ((Y_{t,1}, \dots, Y_{t,d})')_{t \in J}.$$

Note that for $t \in J$ close to T the yield curve \mathbf{Y}_t refers to securities which may expire after the final time horizon T . The zero-coupon bond price process is defined by

$$(\mathbf{P}_t)_{t \in J} = ((P_{t,1}, \dots, P_{t,d})')_{t \in J},$$

where $P_{t,j} = \exp(-j\delta Y_{t,j})$ for $j \in D$. This price process is by assumption \mathbb{F} -adapted and observable at the market. Let $(\boldsymbol{\varepsilon}_t)_{t \in J'} = ((\varepsilon_{t,1}, \dots, \varepsilon_{t,d})')_{t \in J'}$ be \mathbb{F} -adapted, d -dimensional independent standard Gaussian innovations under \mathbb{P} , i.e. $\boldsymbol{\varepsilon}_t$ is $\mathcal{F}_{t-\delta}$ -measurable, independent of $\mathcal{F}_{t-\delta}$ and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, I_d)$ under \mathbb{P} for all $t \in J'$.

Define a probability measure \mathbb{P}^* equivalent to \mathbb{P} via the density process $(\zeta_t)_{t \in J'}$ given by

$$\zeta_t = \prod_{s=1}^{t/\delta} \exp\left(-\frac{1}{2} \|\boldsymbol{\mu}_{s\delta}\|^2 + \boldsymbol{\mu}'_{s\delta} \boldsymbol{\varepsilon}_{s\delta}\right),$$

where $(\boldsymbol{\mu}_t)_{t \in J'}$ is a d -dimensional previsible process called the market price of risk process. Let $(\boldsymbol{\varepsilon}_t^*)_{t \in J'}$ be \mathbb{F} -adapted, d -dimensional independent standard Gaussian innovations under \mathbb{P}^* . For $t \in J'$ we have under \mathbb{P}^*

$$\varepsilon_t - \mu_t \stackrel{(d)}{=} \varepsilon_t^*, \quad \text{given } \mathcal{F}_{t-\delta}. \quad (1)$$

For more details concerning \mathbb{P}^* we refer to Wüthrich and Merz [21], Sects. 2.3 and 2.4.

The bank account process is defined by $B_0 = 1$ and for $t \in J'$

$$B_t = \exp\left(\sum_{s=1}^{t/\delta} \delta Y_{(s-1)\delta,1}\right) = \prod_{s=1}^{t/\delta} P_{(s-1)\delta,1}^{-1}.$$

This process is previsible and describes the one period risk-free roll over, also called locally riskless investment (see Example 5.5 in Föllmer and Schied [7]). In order to obtain a model which is free of arbitrage we require under \mathbb{P}^* for all $t \in J'$ and $j \in D^-$

$$B_{t-\delta}^{-1} P_{t-\delta,j+1} = \mathbb{E}^*[B_t^{-1} P_{t,j} | \mathcal{F}_{t-\delta}]. \quad (2)$$

The existence of an equivalent probability measure $\mathbb{P}^* \sim \mathbb{P}$ such that the price process fulfills (2) rules out arbitrage in the resulting pricing model, we refer to Delbaen and Schachermayer [6] for the general version of the no-arbitrage condition. In our setting condition (2) is formulated for $j \in D^-$ because the bond with time to maturity $d\delta$ is not available at previous time points. As we discuss below, the extension of the model to bonds that are only available at later time points immediately leads to an incomplete market model with infinitely many equivalent measures which satisfy (2). Our aim is to price these bonds using dynamic hedging arguments with instruments available at the market.

2.2 Stochastic yield curve modeling

First, we formulate our model under \mathbb{P}^* considering only maturities in D^- . We construct a semi-parametric model based on ideas presented in Teichmann and Wüthrich [20]. Choose $t \in J'$. For given yield curve $Y_{t-\delta}$ we make the following model assumption for $j \in D^-$

$$j\delta Y_{t,j} = (j+1)\delta Y_{t-\delta,j+1} - \delta Y_{t-\delta,1} + \alpha_j(Y_{t-\delta,j+1}) + \sqrt{\delta} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1}) \lambda_{ij} \varepsilon_{t,i}^*, \quad (3)$$

where $h : \mathbb{R} \rightarrow \mathbb{R}$ is a real function which describes how yield volatilities scale with respect to yield levels. Set

$$\Lambda^- = (\Lambda_{ji}^-)_{i,j \in D^-} = (\lambda_{ij})_{i,j \in D^-} \in \mathbb{R}^{(d-1) \times (d-1)}.$$

Our explicit choice of h is discussed later on. The function h is called volatility scaling factor and Λ^- is called matrix of return directions. If we drop the last two terms on the right-hand side of (3), we obtain the no-arbitrage condition in case of a deterministic yield curve. The last term specifies uncertainty in the future yield curve. The drift term α_j is determined by the no-arbitrage condition (2). From [20], Lemma 3.1, we have for all $t \in J'$ and $j \in D^-$

$$\alpha_j(Y_{t-\delta,j+1}) = \frac{\delta}{2} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1})^2 \lambda_{ij}^2. \quad (4)$$

We introduce some notation to simplify (3) and (4). Let $\mathbf{e}_t^- = (e_{t,1}, \dots, e_{t,d-1})'$ and $\boldsymbol{\mu}_t^- = (\mu_{t,1}, \dots, \mu_{t,d-1})'$. Define the functions $\zeta^- : \mathbb{R}^d \rightarrow \mathbb{R}^{(d-1) \times (d-1)}$ and $\Sigma^- : \mathbb{R}^d \rightarrow \mathbb{R}^{(d-1) \times (d-1)}$ by

$$\mathbf{y} \mapsto \zeta^-(\mathbf{y}) = \text{diag}(h(y_2), \dots, h(y_d)), \quad \text{and} \quad \mathbf{y} \mapsto \Sigma^-(\mathbf{y}) = \zeta^-(\mathbf{y}) \Lambda^- (\Lambda^-)' \zeta^-(\mathbf{y}).$$

We consider the process $(\tilde{Y}_t^-)_{t \in J'} = ((\tilde{Y}_{t,1}^-, \dots, \tilde{Y}_{t,d-1}^-)')_{t \in J'}$ given by

$$\tilde{Y}_{t,j}^- = j\delta Y_{t,j} - (j+1)\delta Y_{t-\delta,j+1} = -\log \frac{P_{t,j}}{P_{t-\delta,j+1}}.$$

These describe the one period log-returns on the bonds. Under the real-world measure \mathbb{P} we obtain the following representation (see Sect. [Proof of Lemma 2.1](#) in “Appendix”).

Lemma 2.1 *Let $t \in J'$. Under (1), model equation (3) and condition (4) can be formulated as*

$$\tilde{Y}_t^- = \delta \left[-Y_{t-\delta,1} \mathbf{1}^- + \frac{1}{2} \text{sp}(\Sigma^-(Y_{t-\delta})) \right] + \sqrt{\delta} \zeta^-(Y_{t-\delta}) \Lambda^- (\mathbf{e}_t^- - \boldsymbol{\mu}_t^-), \quad (5)$$

where $\mathbf{1}^- = (1, \dots, 1)' \in \mathbb{R}^{d-1}$ and $\text{sp}(\Sigma^-(\cdot)) = (\Sigma^-(\cdot)_{11}, \dots, \Sigma^-(\cdot)_{d-1,d-1})'$. Thus, \tilde{Y}_t^- is multivariate Gaussian distributed under \mathbb{P} , given $\mathcal{F}_{t-\delta}$, with conditional mean

$$\mathbb{E}[\tilde{Y}_t^- | \mathcal{F}_{t-\delta}] = \delta \left[-Y_{t-\delta,1} \mathbf{1}^- + \frac{1}{2} \text{sp}(\Sigma^-(Y_{t-\delta})) \right] - \sqrt{\delta} \zeta^-(Y_{t-\delta}) \Lambda^- \boldsymbol{\mu}_t^-,$$

and conditional covariance matrix

$$\text{Cov}[\tilde{Y}_t^- | \mathcal{F}_{t-\delta}] = \delta \Sigma^-(Y_{t-\delta}).$$

The lemma provides a vectorial representation of (3) and (4), and characterizes the multivariate distribution of the one period log-returns on the bonds.

For $t \in J'$ the bond which matures at time $t + d\delta$ is not traded at time $t - \delta$ and therefore the choice of the model for this security is not restricted by the no-arbitrage condition (2) under \mathbb{P}^* . For this reason we formulate a stochastic model for the process $(Y_{t,d})_{t \in J}$ directly under \mathbb{P} . We propose to couple equation (3) with the following stochastic representation for $t \in J'$ and a given yield curve $Y_{t-\delta}$

$$d\delta Y_{t,d} = (d-1)\delta Y_{t,d-1} + \delta\beta(Y_{t-\delta}) + \sqrt{\delta}h(Y_{t-\delta,d})\kappa\mathbf{e}_{t,d}, \quad (6)$$

where $\kappa \in \mathbb{R}$ is a constant and $\beta : \mathbb{R}^d \rightarrow \mathbb{R}$ is a real function which describes the slope at the long end of the yield curve. We call this function long end slope factor and choose it explicitly later on. The first two terms in (6) describe a linear continuation of the yield curve at the long end. The last term adds a stochastic part. This completes the formulation of our model assumptions.

To condense the model assumptions in one equation we introduce additional notation. Define the functions $\zeta : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ and $\Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by

$$\mathbf{y} \mapsto \zeta(\mathbf{y}) = \text{diag}(h(y_2), \dots, h(y_d), h(y_d)), \quad \text{and} \quad \mathbf{y} \mapsto \Sigma(\mathbf{y}) = \zeta(\mathbf{y}) \Lambda \Lambda' \zeta(\mathbf{y}), \quad (7)$$

where

$$\Lambda = \left[\begin{array}{ccc|c} & & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline \Lambda_{d-1,1}^- & \dots & \Lambda_{d-1,d-1}^- & \kappa \end{array} \right] \in \mathbb{R}^{d \times d}.$$

We also define the transformed yield curve process $\tilde{\mathbf{Y}}_t = (\tilde{\mathbf{Y}}_t^-, d\delta Y_{t,d} - d\delta Y_{t-\delta,d})'$ and $\tilde{\boldsymbol{\mu}}_t = (\boldsymbol{\mu}_t^-, 0)'$. We formulate our bond market model under the real-world measure \mathbb{P} as follows (see Sect. [Proof of Lemma 2.2](#) in “[Appendix](#)”).

Lemma 2.2 *Let $t \in J'$. Under (1), model Eqs. (3), (6) and condition (4) can be formulated as*

$$\begin{aligned} \tilde{\mathbf{Y}}_t = & \delta \left[-Y_{t-\delta,1} \mathbf{1} + \frac{1}{2} \text{sp}(\Sigma(\mathbf{Y}_{t-\delta})) + \left(\beta(\mathbf{Y}_{t-\delta}) - \frac{1}{2} (\zeta(\mathbf{Y}_{t-\delta}) \Lambda)_{dd}^2 \right) \mathbf{e}_d \right] \\ & + \sqrt{\delta} \zeta(\mathbf{Y}_{t-\delta}) \Lambda (\boldsymbol{\varepsilon}_t - \tilde{\boldsymbol{\mu}}_t), \end{aligned} \quad (8)$$

where $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^d$ and $\mathbf{e}_d = (0, \dots, 0, 1)' \in \mathbb{R}^d$. Thus, $\tilde{\mathbf{Y}}_t$ is multivariate Gaussian distributed under \mathbb{P} , given $\mathcal{F}_{t-\delta}$, with conditional mean

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{Y}}_t | \mathcal{F}_{t-\delta}] = & \delta \left[-Y_{t-\delta,1} \mathbf{1} + \frac{1}{2} \text{sp}(\Sigma(\mathbf{Y}_{t-\delta})) + \left(\beta(\mathbf{Y}_{t-\delta}) - \frac{1}{2} (\zeta(\mathbf{Y}_{t-\delta}) \Lambda)_{dd}^2 \right) \mathbf{e}_d \right] \\ & - \sqrt{\delta} \zeta(\mathbf{Y}_{t-\delta}) \Lambda \tilde{\boldsymbol{\mu}}_t, \end{aligned}$$

and conditional covariance matrix

$$\text{Cov}[\tilde{\mathbf{Y}}_t | \mathcal{F}_{t-\delta}] = \delta \Sigma(\mathbf{Y}_{t-\delta}).$$

Equation (8) shows that the market model is fully specified by h , β , Λ and $(\boldsymbol{\mu}_t^-)_{t \in J'}$. The process $(\mu_{t,d})_{t \in J'}$ is not determined by the market model. Therefore, there are infinitely many equivalent measures which satisfy (2) in this model, i.e. we have incompleteness. Note that only the conditional distribution given the information in the previous period is Gaussian. For general h , conditioning over multiple periods leads to more complicated distributions and there are no closed form expressions for the moments available.

3 Trading strategies and hedging

A trading strategy is a d -dimensional previsible stochastic process

$$\mathbf{X} = (\mathbf{X}_t)_{t \in J'} = ((X_{t,1}, \dots, X_{t,d})')_{t \in J'},$$

where $X_{t,j}$ denotes the amount of money invested from time $t - \delta$ to t in the zero-coupon bond with time to maturity $j\delta$. We define the \mathbb{F} -adapted return process $(\mathbf{R}_t)_{t \in J'} = ((R_{t,1}, \dots, R_{t,d})')_{t \in J'}$ by

$$R_{t,j} = \begin{cases} P_{t,j-1}P_{t-\delta,j}^{-1} = \exp\{j\delta Y_{t-\delta,j} - (j-1)\delta Y_{t,j-1}\} = \exp\{-\tilde{Y}_{t,j-1}\}, & \text{for } j \in D \setminus \{1\}, \\ P_{t-\delta,j}^{-1} = \exp\{\delta Y_{t-\delta,1}\}, & \text{for } j = 1. \end{cases}$$

This return denotes the value at time $t \in J'$ of one unit of cash invested at time $t - \delta$ in the security with time to maturity $j\delta$. We also introduce the \mathbb{F} -adapted process $(V_t^-(\mathbf{X}))_{t \in J'}$ defined by the value of \mathbf{X} at time $t \in J'$ before rebalancing, i.e. $V_t^-(\mathbf{X}) = \mathbf{X}_t' \mathbf{R}_t$. The \mathbb{F} -adapted value process $(V_t(\mathbf{X}))_{t \in J}$ for a trading strategy \mathbf{X} is defined by

$$V_t(\mathbf{X}) = \begin{cases} \mathbf{X}_{t+\delta}' \mathbf{1} & \text{for } t < T, \\ V_T^-(\mathbf{X}) & \text{for } t = T. \end{cases}$$

Trading strategies with no outflows or inflows of cash over time are called self-financing. Formally, \mathbf{X} is self-financing if for all $t \in J'$

$$V_t(\mathbf{X}) = V_t^-(\mathbf{X}). \quad (9)$$

For \mathbf{X} self-financing we identify $V_t^-(\mathbf{X})$ with $V_t(\mathbf{X})$ and we only use the latter notation. Note that linear combinations of self-financing trading strategies are self-financing.

An arbitrage opportunity is a self-financing trading strategy \mathbf{X} having initial value $V_0(\mathbf{X}) = 0$, terminal value $V_T(\mathbf{X}) \geq 0$, \mathbb{P} -a.s., and $\mathbb{P}[V_T(\mathbf{X}) > 0] > 0$. Such trading strategies are ruled out in our market model by the existence of an equivalent probability measure satisfying (2). A claim with maturity T is an \mathcal{F}_T -measurable random variable c_T . We think of c_T as a future liability which has to be covered at time T . A claim c_T is called attainable if there exists a self-financing trading strategy \mathbf{X}_{c_T} such that $V_T(\mathbf{X}_{c_T}) = c_T$, \mathbb{P} -a.s. An attainable claim can perfectly be replicated by investing $V_0(\mathbf{X}_{c_T})$ at time 0 and adjusting the portfolio holdings at times $t \in J'$ according to the trading strategy \mathbf{X}_{c_T} . In our market model we have the following.

Theorem 3.1 *The market model (8) is incomplete, i.e. there are claims with maturity T which are not attainable. In particular, the claim $c_T = 1$, which corresponds to a zero-coupon bond with time to maturity $T > d\delta$, is not attainable.*

The first statement follows from Föllmer and Schied [7], Theorem 5.38. The second statement is intuitively clear given our model assumptions. A formal proof is given in [Proof of Theorem 3.1](#) in “[Appendix](#)”. The unattainability of $c_T = 1$ for $T > d\delta$ is a desired property of our model. For instance, a life-time annuity with expected payments having more than 30 years to maturity cannot be hedged because there are no bonds available at the financial market with such long times to maturity. Bonds of shorter maturities need to be rolled over. This involves reinvestment risk because of the uncertainty in future yields.

In the literature there are several approaches to tackle the problem of hedging claims in incomplete financial markets. Super-replication is one example. A super-

replicating strategy X is a self-financing trading strategy such that $V_T(X) \geq c_T$, \mathbb{P} -a.s. Such X is often very expensive (see Example 3.4 below) and in some cases it does not even exist. For more details see Föllmer and Schied ([7], Sects. 7 and 8).

In this paper we follow a more practical approach based on Artzner et al. [1], and Hilli et al. [9]. We define acceptable hedging strategies as those which cover the claim at maturity at an acceptable level of risk. Let \mathcal{A} denote a set of \mathcal{F}_T -measurable random variables. This set is called acceptance set and is interpreted as the set of all payoffs at maturity which are within our risk tolerance in hedging the claim c_T . For example, in Hilli et al. [9] acceptance sets of the form $\mathcal{A} = \{V \text{ is } \mathcal{F}_T\text{-measurable} \mid \rho(V - c_T) \leq 0\}$ for given risk measures ρ are considered. A self-financing trading strategy X is said to hedge c_T at acceptance level \mathcal{A} if $V_T(X) \in \mathcal{A}$, \mathbb{P} -a.s. The initial value of c_T at acceptance level \mathcal{A} is defined by

$$\mathcal{V}_0(\mathcal{A}) = \inf_{\substack{X \text{ self-financing} \\ V_T(X) \in \mathcal{A}, \mathbb{P}\text{-a.s.}}} V_0(X). \quad (10)$$

Remark 3.2 In [1] and [9] \mathcal{A} is defined with respect to the net terminal payoff after the claim has been paid, i.e. $V_T(X) - c_T$. We define \mathcal{A} with respect to the terminal payoff $V_T(X)$. Note that \mathcal{A} depends on the claim c_T because it should be chosen reasonably relative to c_T . In this way we obtain a more concise representation since the dependence from c_T is implicit in \mathcal{A} . For clarity we should write $\mathcal{A} = \mathcal{A}(c_T)$. We drop this explicit dependence in the notation, since later on we fix $c_T = 1$.

Remark 3.3 We only consider constraints for the value at the final time horizon T . In practice, we might be interested in additional value constraints at previous time points (e.g. at the end of each calendar year). Our methodology can also be applied for these constraints. For computational simplicity we do not consider them in this paper.

The rest of the paper exposes a method to solve (10) numerically for the claim $c_T = 1$ in the market model of Sect. 2. We define the set of feasible trading strategies by¹

$$\mathcal{D} = \{X \text{ is a self-financing trading strategy} \mid X_t \geq \mathbf{0}, \mathbb{P}\text{-a.s., for all } t \in J'\},$$

i.e. we restrict ourselves to long-only trading strategies. Note that \mathcal{D} is a convex cone. In this work we consider acceptance sets of the following form

$$\mathcal{A} = \{V \text{ is } \mathcal{F}_T\text{-measurable} \mid \mathbb{E}[f(V)] \leq \mathbf{0}\}, \quad (11)$$

where $f: \mathbb{R} \rightarrow \mathbb{R}^m$ is a continuous function with convex components. We consider such functions only on $\mathbb{R}_{\geq 0}$ because $V_T(X) \geq 0$, \mathbb{P} -a.s., for all $X \in \mathcal{D}$.

Example 3.4 Let $m = 1$, $f_0 > 0$ and define $f(v) = \max\{f_0 - v, 0\}$. This corresponds to the acceptance set $\mathcal{A} = \{V \mid V \geq f_0, \mathbb{P}\text{-a.s.}\}$. In particular, for $f_0 = 1$, this corresponds to super-replication of $c_T = 1$. Note that for all $t \in J'$ we have

¹ Throughout this paper for a vector \mathbf{v} we write $\mathbf{v} \leq \mathbf{0}$ meaning that all components of \mathbf{v} are lower or equal to zero. In the same sense we also write $\mathbf{v} \geq \mathbf{0}$.

$R_{t,1} \geq 1$, \mathbb{P} -a.s. Then, the risk-free roll over with sufficiently large initial capital provides a super-replicating strategy. Typically, such a super-replication is rather capital inefficient.

Example 3.5 Let $m = 1, f_0 > 0, u > 0, k \in \mathbb{N}$ and define $f(v) = \max\{f_0 - v, 0\}^k - u^k$. In this way we require the upper bound u^k to the lower k -th partial moment with target level f_0 . We interpret $u > 0$ as a risk tolerance parameter. Higher lower partial moments indicate higher shortfall risk. We use these acceptance sets in the numerical example of Sect. 6.

Summarizing we study the following optimization problem

$$\begin{aligned} & \text{minimize} && V_0(\mathbf{X}) \quad \text{over} \quad \mathbf{X} \in \mathcal{D} \\ & \text{subject to} && \mathbb{E}[f(V_T(\mathbf{X}))] \leq 0. \end{aligned} \quad (12)$$

This corresponds to (10) for the acceptance set (11). Solving (12) we achieve the following: a methodology to price and hedge non-tradable long term zero-coupon bonds with tradable ones under a certain risk tolerance and no short positions, where risk is caused by uncertainty in the interest rates on the unavoidable reinvestments.

4 Acceptable hedging and stochastic optimization

4.1 Galerkin approximation

Since the distribution of $(Y_t)_{t \in J}$ is continuous, \mathcal{D} is typically infinite dimensional and (12) cannot be solved analytically or using standard algorithms for convex optimization. For this reason we consider finite dimensional approximations. Instead of (12) it is more convenient for the application of certain numerical methods to consider the problem

$$\text{minimize} \quad V_0(\mathbf{X}) + \theta(\mathbb{E}[f(V_T(\mathbf{X}))]) \quad \text{over} \quad \mathbf{X} \in \mathcal{D}, \quad (13)$$

where $\theta : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ denotes the convex and increasing penalty function

$$\mathbf{v} \mapsto \theta(\mathbf{v}) = \sum_{i=1}^m \theta_{0i} \max\{v_i, 0\}^2,$$

for $\theta_{01}, \dots, \theta_{0m} > 0$. This is because defining parametric classes of trading strategies in \mathcal{D} is easier compared to $\mathcal{D} \cap \{\mathbf{X} | V_T(\mathbf{X}) \in \mathcal{A}, \mathbb{P}\text{-a.s.}\}$. The second term in (13) is positive whenever the inequality constraints in (12) are violated and zero otherwise. The solution of (13) approximates the one of (12) for large $\theta_{01}, \dots, \theta_{0m}$. The choice of smooth θ is convenient for the application of numerical optimization algorithms. The problem (13) is convex (see [Convexity of \(13\)](#) in “Appendix”).

Remark 4.1 An alternative approach to obtain numerical solutions for (12) is to follow Hilli et al. [9] and minimize $\mathbb{E}[f(V_T(\mathbf{X}))]$ for fixed values of the initial cost $V_0(\mathbf{X})$. Then, the optimal initial wealth is the value for which the minimum is zero.

Numerical methods to approximate infinite dimensional optimization problems such as (13) to finite dimensional ones have been discussed in the literature. A possible approach is the discretization of $(Y_t)_{t \in J}$, see e.g. Hilli and Pennanen [11] or Pflug [17]. This technique is referred to as scenario discretization. The main idea is to construct a finite set of scenarios for the yield curve. Such a discretization leads to a finite dimensional \mathcal{D} and a standard convex optimization problem that can be solved with conventional algorithms. A well known method to construct such scenarios is conditional sampling, where these are obtained by recursively sampling from the conditional distribution $Y_t | \mathcal{F}_{t-\delta}$, see Shapiro [18]. However, scenario discretization presents some serious computational drawbacks. The dimensionality of the resulting approximation typically increases exponentially as the number of trading periods n rises. We do not want to be constrained in the number of trading periods and therefore we do not consider scenario discretization.

Instead, we turn our attention to the Galerkin method, see Koivu and Pennanen [13], and Hilli et al. [10]. This is computationally more attractive and easier to implement. The idea is to look for optimal solutions not over the entire set \mathcal{D} , but over a subset of \mathcal{D} consisting of finite linear combinations of feasible trading strategies, called basis strategies. Let $\chi = (X^{(1)}, \dots, X^{(N)})' \in \mathcal{D}^N$ be N such basis strategies. The set of conical combinations $\{w' \chi | w \in \mathbb{R}_{\geq 0}^N\}$ is a finite dimensional subset of \mathcal{D} . Let $V_t(\chi) = (V_t(X^{(1)}), \dots, V_t(X^{(N)}))'$ for $t \in J$. The Galerkin approximation of (13) is given by

$$\text{minimize } w' V_0(\chi) + \theta(\mathbb{E}[f(w' V_T(\chi))]) \quad \text{over } w \in \mathbb{R}_{\geq 0}^N. \quad (14)$$

The convexity is preserved in (14), because it is simply the convex optimization problem (13) restricted to a finite dimensional convex subset of \mathcal{D} . This approximation leads to a standard finite dimensional convex minimization problem for which standard numerical algorithms can be used.

Observe that (14) is in some sense a one-period optimization problem. It depends directly only on $V_0(\chi)$ and $V_T(\chi)$. In contrast to scenario discretization, increasing n does not increase the complexity of the optimization directly. The dimension of (14) depends solely on N . The complexity of evaluating the term $\mathbb{E}[f(w' V_T(\chi))]$ typically depends on the number of trading periods n . This is because n influences the valuation of $V_T(\chi)$. Since in our market model we are not able to find closed form solutions for such expectations (except for very simple special cases) we need to evaluate them through Monte Carlo simulations, which are computationally more expensive if n is large. This is nevertheless easier and typically less time consuming compared to the evaluation in tree-based discretization. In particular, for the model of Sect. 2, we do not have to work with nested simulations. We only need to generate nd standard Gaussian random numbers in order to obtain one observation of $V_T(\chi)$. Because of the computational advantages, we take (14) together with (8) as our basic model to partially hedge and price long term zero-coupon bonds.

4.2 Parametric basis strategies

In this section we consider four parametric families of basis strategies which are in \mathcal{D} for every possible parameter choice. These are buy & hold (BH), fixed times to maturity (FTM), fixed proportions (FP) and target date fund (TDF) trading strategies. Let $V_0 > 0$ and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d)' \in \mathbb{R}_{\geq 0}^d$ so that $\boldsymbol{\pi}'\mathbf{1} = 1$. The BH, FTM and FP families are defined by $\mathbf{X}_\delta^{\text{BH}} = \mathbf{X}_\delta^{\text{FTM}} = \mathbf{X}_\delta^{\text{FP}} = V_0\boldsymbol{\pi}$ and for $t > \delta$

$$\begin{aligned}\mathbf{X}_t^{\text{BH}} &= \left(X_{t-\delta,2}^{\text{BH}} R_{t-\delta,2}, \dots, X_{t-\delta,d}^{\text{BH}} R_{t-\delta,d}, X_{t-\delta,1}^{\text{BH}} R_{t-\delta,1} \right)', \\ \mathbf{X}_t^{\text{FTM}} &= \left(X_{t-\delta,1}^{\text{FTM}} R_{t-\delta,1}, \dots, X_{t-\delta,d}^{\text{FTM}} R_{t-\delta,d} \right)', \\ \mathbf{X}_t^{\text{FP}} &= (\mathbf{R}'_{t-\delta} \mathbf{X}_{t-\delta}^{\text{FP}}) \boldsymbol{\pi}.\end{aligned}$$

At time 0 we invest V_0 in the d zero-coupon bonds according to the proportions $\boldsymbol{\pi}$. For BH trading strategies we do not rebalance at future time points. At each point in time we invest the cash obtained from the security that expires in the newly available zero-coupon bond with time to maturity $d\delta$. For FTM trading strategies we rebalance the portfolio at each point in time so that the value of the investment in each of the bonds is invested again in the security with the same time to maturity. For FP trading strategies we rebalance the portfolio at each point in time so that the total portfolio value is invested in each bond according to the proportions $\boldsymbol{\pi}$.

We define TDF trading strategies similarly to Bodie and Treussard [2]. Fix $d' \in D^-$ and divide the set of times to maturity into two disjoint subsets $\{1, \dots, d'\}$ and $\{d' + 1, \dots, d\}$. The bonds with shorter maturities $1, \dots, d'$ are referred to as lower duration and those with longer maturities as higher duration. Let $\boldsymbol{\eta} = (\eta_1, \dots, \eta_d) \in \mathbb{R}_{\geq 0}^d$ so that

$$\sum_{j=1}^{d'} \eta_j = \sum_{j=d'+1}^d \eta_j = 1.$$

Let $0 \leq p_\delta < p_T \leq 1$ be the initial and terminal proportions invested in the lower duration bonds. We increase this proportion linearly from p_δ to p_T to reflect the fact that the duration of the claim decreases as its expire date approaches. That is,

$$p_t = p_\delta + \frac{p_T - p_\delta}{T - \delta}(t - \delta).$$

The trading strategy is then defined by $\mathbf{X}_{\delta,j}^{\text{TDF}} = (p_\delta \eta_j \mathbf{1}_{1 \leq j \leq d'} + (1 - p_\delta) \eta_j \mathbf{1}_{d' < j \leq d}) V_0$ and for $t > \delta$

$$\mathbf{X}_{t,j}^{\text{TDF}} = (p_t \eta_j \mathbf{1}_{1 \leq j \leq d'} + (1 - p_t) \eta_j \mathbf{1}_{d' < j \leq d}) \mathbf{R}'_{t-\delta} \mathbf{X}_{t-\delta}^{\text{TDF}}, \quad j \in D.$$

This means that at each point in time the portfolio is rebalanced so that the proportion invested in low duration bonds is p_t , the one invested in high duration bonds is $1 - p_t$ and those invested within the low and high duration ones are given by $\boldsymbol{\eta}$. Note that these four families are, in general, not conical. Therefore, using trading

strategies in one of these as basis for (14), we optimize over strategies that do not belong to the family.

Remark 4.2 Let X be a BH, FTM, FP or TDF trading strategy, X^{norm} have $V_0 = 1$ and otherwise be identical to X . Then, $X_t = V_0 X_t^{\text{norm}}$ for $t \in J'$.

For $j_1, \dots, j_n \in D$ we define the \mathbb{F} -adapted stochastic process $(B_t(j_1, \dots, j_n))_{t \in J}$ by $B_0(j_1, \dots, j_n) = 1$ and for $t \in J'$

$$B_t(j_1, \dots, j_n) = \prod_{s=1}^{t/\delta} R_{s\delta, j_s}. \quad (15)$$

This process corresponds to the value at time t of investing one unit of cash in the zero-coupon bond with time to maturity j_1 at time 0, rolling over at time δ in the one with time to maturity j_2 , and so on for j_3, \dots, j_n . In particular we have that $B_t(1, \dots, 1) = B_t$ is the classical bank account process. Note that we have $B_t(j_1, \dots, j_n) = B_t(j_1, \dots, j_s^*, 1, \dots, 1)$ for all $j_1, \dots, j_n \in D$ and $t \in J'$. Applying the definitions above recursively we have the following result (see [Proof of Lemma 4.3](#) in “[Appendix](#)”):

Lemma 4.3 Let X be a FTM, FP or TDF strategy. Then, we have for $t \in J'$

$$V_t(X) = \sum_{j_1, \dots, j_n=1}^d a_t(j_1, \dots, j_n, 1, \dots, 1) B_t(j_1, \dots, j_n, 1, \dots, 1),$$

where

$$a_t(j_1, \dots, j_n) = \begin{cases} V_0 \pi_{j_1} 1_{j_1 = \dots = j_n^*} & \text{if } X \text{ is FTM,} \\ V_0 \prod_{s=1}^{t/\delta} \pi_{j_s} & \text{if } X \text{ is FP,} \\ V_0 \prod_{s=1}^{t/\delta} (p_{s\delta} \eta_{j_s} 1_{1 \leq j_s \leq d'} + (1 - p_{s\delta}) \eta_{j_s} 1_{d' < j_s \leq d}) & \text{if } X \text{ is TDF.} \end{cases}$$

The above lemma states that the value processes for FTM, FP and TDF trading strategies are linear combinations of the random variables $B_t(\cdot, \dots, \cdot)$ with deterministic coefficients. These random variables depend only on the market model whereas the coefficients $a_t(\cdot, \dots, \cdot)$ depend only on the strategy parameters. Thus, we obtain a split in two terms, the market model (random) and the deterministic coefficients from the strategy which allows to evaluate the objective function in (14) more efficiently. A closed form expression in terms of $B_t(\cdot, \dots, \cdot)$ can also be worked out for the value process of BH trading strategies, but it does not have a simple form as for the other three families. On the distribution of $B_t(\cdot, \dots, \cdot)$ in our model we know the following (see [Proof of Lemma 4.4](#) in “[Appendix](#)”).

Lemma 4.4 Let $t \in J'$, $s = \frac{t}{\delta} \in \{1, \dots, n\}$ and $j_1, \dots, j_n \in D$. For $j_s > 1$, the random variable $\log B_t(j_1, \dots, j_n)$ is normally distributed under \mathbb{P} , given $\mathcal{F}_{t-\delta}$, with conditional mean

$$\log P_{t-\delta, 1} + \log B_{t-\delta}(j_1, \dots, j_n) + \sqrt{\delta}(\zeta(Y_{t-\delta}) \Lambda \tilde{\mu}_t)_{j_s-1} - \frac{\delta}{2} \text{sp}(\Sigma(Y_{t-\delta}))_{j_s-1},$$

and variance $\delta \text{sp}(\Sigma(\mathbf{Y}_{t-\delta}))_{j_s-1}$. For $j_s = 1$, $B_t(j_1, \dots, j_n)$ is previsible and given by $B_t(j_1, \dots, j_n) = B_{t-\delta}(j_1, \dots, j_n)P_{t-\delta,1}$.

The conditional expectation and variance of $B_t(\cdot, \dots, \cdot)$, assuming $j_s > 1$, are given by

$$\begin{aligned}\mathbb{E}[B_t(j_1, \dots, j_n) | \mathcal{F}_{t-\delta}] &= B_{t-\delta}(j_1, \dots, j_n)P_{t-\delta,1} \exp \left\{ \sqrt{\delta}(\zeta(\mathbf{Y}_{t-\delta})\Lambda \tilde{\boldsymbol{\mu}}_t)_{j_s-1} \right\}, \\ \text{Var}[B_t(j_1, \dots, j_n) | \mathcal{F}_{t-\delta}] &= \mathbb{E}[B_t(j_1, \dots, j_n) | \mathcal{F}_{t-\delta}]^2 \left(\exp \left\{ \delta \text{sp}(\Sigma(\mathbf{Y}_{t-\delta}))_{j_s-1} \right\} - 1 \right).\end{aligned}$$

Observe that the conditional distribution of $B_t(\cdot, \dots, \cdot)$, given $\mathcal{F}_{t-\delta}$, does not depend on β . If we condition over multiple periods we obtain distributions which are not log-normal and depend on β .

5 Calibration procedure

Our aim is to solve (14) numerically under the model of Sect. 2. This requires the calibration to market data. Once h and β have been chosen, it remains to calibrate Λ to real market data. Our explicit choice of h and β is discussed in Sect. 6. Assume that $\boldsymbol{\mu} \equiv \mathbf{0}$, i.e. $\boldsymbol{\mu}_t = \mathbf{0}$, \mathbb{P} -a.s., for all $t \in J'$. This implies that the real world measure \mathbb{P} satisfies condition (2). This assumption is discussed in more details in Sect. 6. Under this assumption, using Lemma 2.2 and Eq. (26) in “Appendix”, we have that $\tilde{\mathbf{Y}}_t$ is multivariate Gaussian distributed, given $\mathcal{F}_{t-\delta}$, with conditional mean

$$\delta \left[-Y_{t-\delta,1} \mathbf{1} + \frac{1}{2} \text{sp}(\Sigma(\mathbf{Y}_{t-\delta})) + \left(\beta(\mathbf{Y}_{t-\delta}) + \frac{1}{2} (\text{sp}(\Sigma(\mathbf{Y}_{t-\delta}))_d - \text{sp}(\Sigma(\mathbf{Y}_{t-\delta}))_{d-1}) \right) \mathbf{e}_d \right],$$

and conditional covariance matrix $\delta \Sigma(\mathbf{Y}_{t-\delta})$ for $t \in J'$. Therefore, we do not need to directly calibrate Λ to market data but only $\Sigma(\cdot)$ as defined in (7). In particular, we need to construct estimators for the elements of $\Lambda \Lambda'$. In this section we show that the calibration procedure presented in Teichmann and Wüthrich [20] can also be applied with slight modifications to our market model specified by the above conditional distribution. Let $(\mathbf{Y}_{k\delta})_{k=0,\dots,K}$ be observations of the yield curve and $(\tilde{\mathbf{Y}}_{k\delta})_{k=1,\dots,K}$ be the corresponding transformed observations. We introduce some additional notation to simplify the expressions below. For $k=0,\dots,K$ let $b_k = Y_{k\delta,\delta}$, $\beta_k = \beta(\mathbf{Y}_{k\delta})$, $\gamma_k = \beta_k - b_k$, $\zeta_k = \zeta(\mathbf{Y}_{k\delta})$, $\zeta_{k,j} = \text{sp}(\zeta_k)_j$ for $j \in D$ and $\Sigma_k = \Sigma(\mathbf{Y}_{k\delta})$. Define the following matrix based on the observations,

$$C_{(K)} = \frac{1}{\sqrt{K}} \left((\zeta_{k-1}^{-1} \tilde{\mathbf{Y}}_{k\delta})_j \right)_{\substack{j=1,\dots,d \\ k=1,\dots,K}} \in \mathbb{R}^{d \times K},$$

and the function $S_{(K)} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ by $\mathbf{y} \mapsto \zeta(\mathbf{y}) C_{(K)} C'_{(K)} \zeta(\mathbf{y})$. From the the proof of Theorem 3.4 in [20] we have for $\mathbf{y} \in \mathbb{R}^d$

$$\mathbb{E}[S_{(K)}(\mathbf{y})] = \delta \Sigma(\mathbf{y}) + \frac{\delta^2}{K} \sum_{k=0}^{K-1} \zeta(\mathbf{y}) \mathbb{E}[G_{h,\beta,\Sigma}(b_k \mathbf{1}, \mathbf{Y}_{k\delta})] \zeta(\mathbf{y}), \quad (16)$$

where $G_{h,\beta,\Sigma} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ has the form

$$(\mathbf{z}, \mathbf{y}) \mapsto \zeta(\mathbf{y})^{-1} \mathbf{g}_{\beta,\Sigma}(\mathbf{z}, \mathbf{y}) \mathbf{g}_{\beta,\Sigma}(\mathbf{z}, \mathbf{y})' \zeta(\mathbf{y})^{-1},$$

with $\mathbf{g}_{\beta,\Sigma} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ given by

$$(\mathbf{z}, \mathbf{y}) \mapsto -\mathbf{z} + \frac{1}{2} \text{sp}(\Sigma(\mathbf{y})) + \left(\beta(\mathbf{y}) - \frac{1}{2} \left(\text{sp}(\Sigma(\mathbf{y}))_d - \text{sp}(\Sigma(\mathbf{y}))_{d-1} \right) \right) \mathbf{e}_d.$$

Similar to (3.6) in [20] we get the crucial observation in (16) that $S_{(K)}(\mathbf{y})$ may serve as an estimator for $\delta \Sigma(\mathbf{y})$ with a bias going to zero as $\delta \rightarrow 0$.

We use (16) to derive estimators for $\Lambda \Lambda' = (s_{ij})_{i,j \in D}$. We set $\mathbf{y} = \mathbf{1}$. For $i, j \in D^-$ we obtain the same estimators for s_{ij} as in [20] formulas (3.12)–(3.13). For the last diagonal element s_{dd} we have

$$\begin{aligned} \frac{\mathbb{E}[S_{(K)}(\mathbf{1})]_{dd}}{\delta \zeta(\mathbf{1})_{dd}^2} &= s_{dd} + \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[\left(\gamma_k + \frac{1}{2} \zeta_{k,d-1}^2 s_{d-1,d-1} \right)^2 \zeta_{k,d}^{\zeta-2} \right] \\ &= s_{dd} + \frac{\delta}{K} \sum_{k=0}^{K-1} \left(A_k^{(1)} + s_{d-1,d-1} A_k^{(2)} + \frac{1}{4} s_{d-1,d-1}^2 A_k^{(3)} \right), \end{aligned} \quad (17)$$

where we have used that $\zeta_{k,d-1} = \zeta_{k,d}$ for all $k = 0, \dots, K$ and set

$$A_k^{(1)} = \mathbb{E}[\gamma_k^2 \zeta_{k,d}^{\zeta-2}], \quad A_k^{(2)} = \mathbb{E}[\gamma_k] \quad \text{and} \quad A_k^{(3)} = \mathbb{E}[\zeta_{k,d}^2]. \quad (18)$$

For the elements $(s_{id})_{i=1,\dots,d-1}$ we have

$$\begin{aligned} \frac{\mathbb{E}[S_{(K)}(\mathbf{1})]_{id}}{\delta \zeta(\mathbf{1})_{ii} \zeta(\mathbf{1})_{dd}} &= s_{id} - \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[b_k \left(\gamma_k + \frac{1}{2} \zeta_{k,d-1}^2 s_{d-1,d-1} \right) \zeta_{k,i}^{\zeta-1} \zeta_{k,d}^{\zeta-1} \right] \\ &\quad + \frac{\delta}{K} \sum_{k=0}^{K-1} \mathbb{E} \left[\frac{1}{2} \zeta_{k,i}^2 s_{ii} \left(\gamma_k + \frac{1}{2} \zeta_{k,d-1}^2 s_{d-1,d-1} \right) \zeta_{k,i}^{\zeta-1} \zeta_{k,d}^{\zeta-1} \right] \\ &= s_{id} + \frac{\delta}{K} \sum_{k=0}^{K-1} \left(B_{k,i}^{(1)} + \frac{1}{2} s_{d-1,d-1} \left(B_{k,i}^{(2)} + \frac{1}{2} s_{ii} B_{k,i}^{(4)} \right) + \frac{1}{2} s_{ii} B_{k,i}^{(3)} \right), \end{aligned} \quad (19)$$

where we have used again that $\zeta_{k,d-1} = \zeta_{k,d}$ for all $k = 0, \dots, K$ and set

$$\begin{aligned} B_{k,i}^{(1)} &= -\mathbb{E}[b_k \gamma_k \zeta_{k,i}^{\zeta-1} \zeta_{k,d}^{\zeta-1}], \quad B_{k,i}^{(2)} = -\mathbb{E}[b_k \zeta_{k,i}^{\zeta-1} \zeta_{k,d}], \\ B_{k,i}^{(3)} &= \mathbb{E}[\gamma_k \zeta_{k,i}^{\zeta-1}] \quad \text{and} \quad B_{k,i}^{(4)} = \mathbb{E}[\zeta_{k,i} \zeta_{k,d}]. \end{aligned} \quad (20)$$

Note that $s_{di} = s_{id}$ for all $i \in D$. We can derive estimates for s_{id} in the following way. We replace the expectations (18) and (20) by the observations. For the diagonal elements $(s_{ii})_{i \in D^-}$ we can use the estimates derived in [20] formula (3.12). By

solving the linear Eqs. (17) and (19) for s_{id} we obtain the estimates. This is the approach presented in [20] and completes our estimation procedure.

6 Numerical example: Swiss government bonds

6.1 Choice of data, $h(\cdot)$ and $\beta(\cdot)$

In this section we consider CHF zero-coupon bond yields derived from Swiss government bonds. We work on a monthly grid, i.e. $\delta = 1/12$. The Swiss government issues bonds with times to maturity in

$$\{1/12, 1/6, 1/4, 1/2, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30\}.$$

However, securities with times to maturity of 15 years or more are not liquid. For this reason we consider only bonds with times to maturity up to 10 years and set $d = 120$. Note that medium and long term Swiss government bonds (i.e. those issued with times to maturity larger than one year) are coupon bearing securities. Therefore, the yield curve needs to be extrapolated from their market prices. For this purpose the Swiss National Bank uses the Svensson method, see Müller [14]. Bloomberg L.P. uses a similar procedure and provides these extrapolated rates for a history of more than 18 years and times to maturity in $\{j\delta | j \in D_{\text{obs}}\}$, where

$$D_{\text{obs}} = \{3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120\},$$

see Bloomberg tickers: F25603M Index, F25606M Index, F25601Y Index and so on. We choose these time series from 31.01.1995 to 31.12.2012 (i.e. $K = 215$) as our monthly yield curve observations. Therefore, we have observations

$$(Y_{k\delta,j})_{k=0,\dots,K,j \in D_{\text{obs}}}.$$

To obtain the yield curve for all times to maturity in D we set for $k \in \{0, \dots, K\}$, $j \in D \setminus D_{\text{obs}}$ and $j > 3$

$$Y_{k\delta,j} = \frac{j - d_l(j)}{d_u(j) - d_l(j)} Y_{k\delta,d_l(j)} + \frac{d_u(j) - j}{d_u(j) - d_l(j)} Y_{k\delta,d_u(j)},$$

where $d_l(j) = \max\{k \in D_{\text{obs}} | k \leq j\}$ and $d_u(j) = \min\{k \in D_{\text{obs}} | k \geq j\}$. This means we interpolate linearly between the two nearest known values. We continue the curve linearly at the short end to extrapolate data for $\delta j = 1/12$ and $1/6$. In this way we obtain observations $(Y_{k\delta})_{k=0,\dots,K}$ for the full yield curve. Some of these are shown in Fig. 1 below.

For the volatility scaling factor we choose the square root scaling $h(y) = \sqrt{y}$, which has been proved reasonable for yield curve modeling, see Guillaume et al. [8]. For the long end slope factor, which describes the steepness of the yield curve at the long end, we make the following intuitive choice

$$\beta(y) = \frac{1}{\delta} (d\delta y_d - (d-1)\delta y_{d-1}) = dy_d - (d-1)y_{d-1}.$$

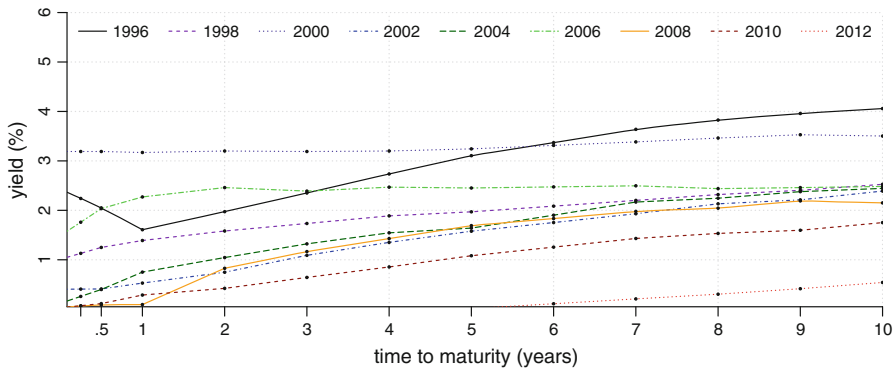


Fig. 1 CHF yield curve (values in percent) at the end of December for some of the years in the sample. We can observe that the yields at the end of 2012 are at historical lows with short term rates being near zero

6.2 Calibration results

In order to test if our model is reasonable and consistent, we split the sample of Sect. 6.1 in the middle and calibrate $\Lambda\Lambda'$ to the first half. We use the procedure described in Sect. 5 under the assumption $\mu \equiv \mathbf{0}$. This assumption is discussed below. For the calibration we use Teichmann and Wüthrich [20] formulas (3.12)–(3.13), and Eqs. (17) and (19). For these estimators we consider first and second order terms in δ . We only use the first half of the data for the calibration since $\Lambda\Lambda'$ is constant by model assumption, and we want to test if the model provides reasonable and consistent out-of-sample results. For this purpose we define the following process for $\vartheta \in \mathbb{R}^d$

$$(\phi_t(\vartheta))_{t \in J'} = \left(\frac{1}{\sqrt{\delta}} \vartheta' \zeta(Y_{t-\delta})^{-1} (\tilde{Y}_t - \psi_t) \right)_{t \in J'}, \quad (21)$$

where

$$\begin{aligned} \psi_t = & \delta \left[-Y_{t-\delta,1} \mathbf{1} + \frac{1}{2} \text{sp}(\zeta(Y_{t-\delta}) \Lambda \Lambda' \zeta(Y_{t-\delta})) + \beta(Y_{t-\delta}) \mathbf{e}_d \right. \\ & \left. + \frac{1}{2} \text{sp}(\zeta(Y_{t-\delta}) \Lambda \Lambda' \zeta(Y_{t-\delta}))_d \mathbf{e}_d - \frac{1}{2} \text{sp}(\zeta(Y_{t-\delta}) \Lambda \Lambda' \zeta(Y_{t-\delta}))_{d-1} \mathbf{e}_d \right], \end{aligned}$$

which depends on $\Lambda\Lambda'$. A direct consequence of Lemma 2.2 under the assumption $\mu^- \equiv \mathbf{0}$ is that $\psi_t = \mathbb{E}[\tilde{Y}_t | \mathcal{F}_{t-\delta}]$ for all $t \in J'$ and for any $\vartheta \in \mathbb{R}^d$

$$(\phi_t(\vartheta))_{t \in J'} \stackrel{iid}{\sim} \mathcal{N}(\mathbf{0}, \vartheta' \Lambda \Lambda' \vartheta). \quad (22)$$

We compute (21) for different values of ϑ using the observations $(Y_{k\delta})_{k=0,\dots,K}$ and the matrix $\Lambda\Lambda'$ calibrated to the first half of the sample. We do this for

$$\vartheta_1 = \frac{1}{\lfloor \frac{d}{3} \rfloor} \sum_{j=1}^{\lfloor \frac{d}{3} \rfloor} e_j, \quad \vartheta_2 = \frac{1}{\lfloor \frac{d}{3} \rfloor} \sum_{j=\lfloor \frac{d}{3} \rfloor+1}^{d-\lfloor \frac{d}{3} \rfloor} e_j, \quad \text{and} \quad \vartheta_3 = \frac{1}{\lfloor \frac{d}{3} \rfloor} \sum_{j=d-\lfloor \frac{d}{3} \rfloor+1}^d e_j, \quad (23)$$

which correspond to short, medium and long term parts of the yield curve. We obtain observations

$$(\phi_{k\delta}(\vartheta_i))_{k=0,\dots,K,i=1,2,3}.$$

If the model is reasonable then these sequences should look like i.i.d. normal random variables including the out-of-sample part, see (22). Figure 2 shows that the serial correlation is low also for small lags in all three cases as expected from the model. From Fig. 3 we can see that the observations of the sequences appear to be reasonable given the model over the full period. There are no particular changes visible in the second half of the sample which has not been used to calibrate the model. Table 1 provides some statistics. The observed standard deviations are close to the ones given by the model. Application of the Jarque–Bera test (see [12]) shows that for the medium and long term parts the normality of the series cannot be rejected at 5 % significance level, whereas for the short end normality is rejected at 1% level. The series also appear to be slightly negatively drifted. Given these results, we conclude that our model does a good job in explaining the yield curve development in the medium and long term parts, whereas for the short term part the model does not appear to work optimally. The assumption $\mu^- \equiv \mathbf{0}$ does not seem optimal either.

There is no clear procedure to estimate the market price of risk process μ . A heuristic approach is to extrapolate it from the sample applying linear filtering techniques. Lemma 2.2 implies for any $\vartheta \in \mathbb{R}^d$ and $t \in J'$ that $\vartheta' \Lambda \mu_t = \mathbb{E}[\phi_t(\vartheta) | \mathcal{F}_{t-\delta}]$. As discussed above the series $(\phi_{k\delta}(\vartheta_i))_{k=0,\dots,K}$ for $i = 1, 2, 3$ show very low serial correlation and we can see from Fig. 4 that a constant drift assumption is reasonable for these series. Hence, we make the following approximation for $t \in J'$ and $i = 1, 2, 3$

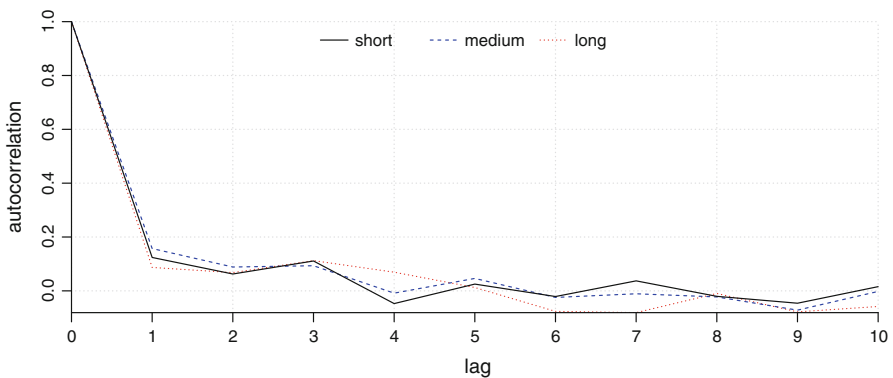


Fig. 2 Autocorrelation of the observed series $(\phi_{k\delta}(\cdot))_{k=0,\dots,K}$ for different lags. The three lines correspond to ϑ_1 (short), ϑ_2 (medium) and ϑ_3 (long). See also (21), (22) and (23)

$$\vartheta'_i \Lambda \boldsymbol{\mu}_t \approx \frac{1}{K+1} \sum_{k=0}^K \phi_{k\delta}(\vartheta_i).$$

We can use these approximations of the projections on the short, medium and long term parts to obtain an approximation for $\Lambda \boldsymbol{\mu}$, i.e. we set for $t \in J'$

$$\Lambda \boldsymbol{\mu}_t \approx \frac{\lfloor \frac{d}{3} \rfloor}{K+1} \sum_{k=0}^K (\phi_{k\delta}(\vartheta_1) \vartheta_1 + \phi_{k\delta}(\vartheta_2) \vartheta_2 + \phi_{k\delta}(\vartheta_3) \vartheta_3), \quad (24)$$

In this way we approximate $\boldsymbol{\mu}$ by a time-independent deterministic vector.

6.3 Pricing and replication for $T = 20$

Our aim is to solve (14) numerically for $T = 20$ (i.e. $n = 240$) and the stochastic model of Sect. 2 calibrated on the sample of Sect. 6.1. The first step is to simulate yield curve developments. We set $\boldsymbol{\mu} \equiv \mathbf{0}$. Alternatively the heuristic approach (24) might be used. The complexity of the pricing algorithm is not affected by the presence of $\boldsymbol{\mu}$. We calibrate the model on the full sample using the procedure of Sect. 6.2. Our market model provides two essential features for this application. First, as we verified in the previous section, the calibrated parameters are

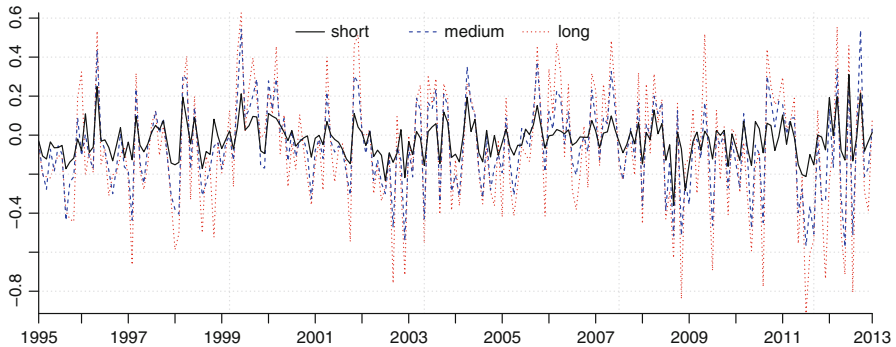


Fig. 3 Values of the observed series $(\phi_{k\delta}(\cdot))_{k=0,\dots,K}$ for ϑ_1 (short), ϑ_2 (medium) and ϑ_3 (long). See also (21), (22) and (23)

Table 1 This table presents the sample mean and standard deviation of the observed series $(\phi(\cdot)_{k\delta})_{k=0,\dots,K}$ in the three cases

	Sample mean	Sample std	Model std	Jarque–Bera test statistics
<i>Short</i>	−0.02	0.09	0.10	14.40
<i>Medium</i>	−0.06	0.22	0.22	0.18
<i>Long</i>	−0.09	0.31	0.30	1.10

It also presents the standard deviation implied by the model, i.e. $\sqrt{\vartheta' \Lambda \Lambda' \vartheta}$, and the Jarque–Bera test statistic for normality. The 5 and 1 % critical values of the test are 5.99 and 9.21, respectively

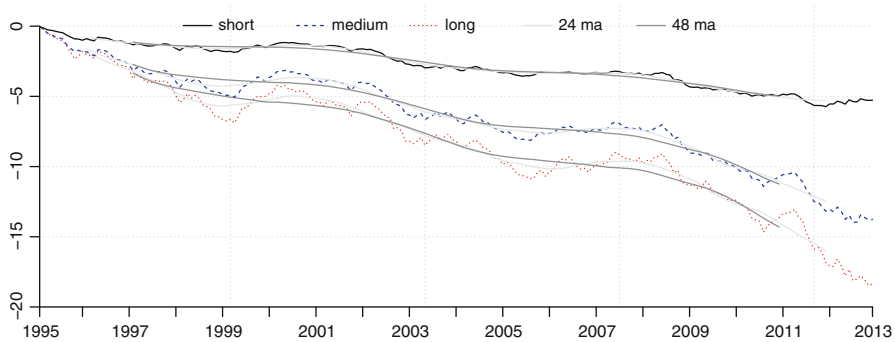


Fig. 4 Cumulated values over time of the observed series $(\phi_{k\delta}(\cdot))_{k=0,\dots,K}$ for ϑ_1 (short), ϑ_2 (medium) and ϑ_3 (long). See also (21), (22) and (23). We also plot the 24 and 48 months symmetric moving averages to smooth the series out. We can see that the drift is quite stable over time

appropriate over a long period of time. This is important since the calibrated model is being used to simulate yield paths for several years ahead. Secondly, generating yield curves in our model does not require nested simulations. This is crucial since we need to simulate with reasonable computational time monthly yield curve observations for a 20 years period several thousand times.

We choose the yield curve at 31.12.2012 as the starting point for our simulation. Observe that financial market conditions in 2012 resulted in extraordinary low yields for Swiss government bonds with flat and near zero short term rates. For the number of simulated developments (each one having 20 years of monthly yield curves simulations) we consider $M = 125, 250, 500, 1,000, 2,000$ and $4,000$.

We consider the acceptance set given in Example 3.5. We set $m = 1$, $f_0 = 1$ and $k = 2$. This means $f(v) = \max\{1 - v, 0\}^2 - u^2$, where we consider the values $u = 0.2, 0.15, 0.1, 0.05$ and 0.025 for the risk tolerance parameter. This corresponds to the acceptance set

$$\mathcal{A} = \{V \text{ is } \mathcal{F}_T\text{-measurable} \mid \mathbb{E}[\max\{1 - V, 0\}^2] \leq u^2\},$$

i.e. terminal values are acceptable if the shortfall risk is not greater than u^2 , where this is measured by the lower partial second moment of the loss. Hence, a lower value of u corresponds to a lower risk tolerance.

We define 10 basis strategies, i.e. $N = 10$. We set $V_0 = 1$ and $\pi_i = \vartheta_i$ for $i = 1, 2, 3$, where ϑ_i denote the weight vector defined in Sect. 6.2. Using these parameters we define 3 BH, FTM and FP strategies. Let $p_0 = 0$, $p_T = 1$, $d' = d - \lfloor \frac{d}{3} \rfloor$ and $\eta = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2 + \pi_3$. From this we define one TDF strategy. We introduce the abbreviations BH1, BH2, BH3, FTM1, FTM2, FTM3, FP1, FP2, FP3 and TDF to reference these 10 basis strategies, where the numbers correspond to $\pi_{1,2,3}$ respectively. We evaluate the terminal values of these strategies for each simulated yield curve development. This can be done very efficiently for FTM, FP and TDF strategies using Lemma 4.3.

Remark 6.1 The choice $V_0 = 1$ is irrelevant to (14). To see this, let $\chi = (X^{(1)}, \dots, X^{(N)})'$ be a N -dimensional vector of BH, FTM, FP or TDF trading strategies, χ^{norm} have $V_0(\chi^{\text{norm}}) = \mathbf{1}$ and otherwise be identical to χ . Then, Remark 4.2 implies

$$\{\mathbf{w}'\chi | \mathbf{w} \in \mathbb{R}_{\geq 0}^N\} = \left\{ \sum_{i=1}^N w_i V_0(X^{(i)})^{-1} X^{(i)} \mid \mathbf{w} \in \mathbb{R}_{\geq 0}^N \right\} = \{\mathbf{w}'\chi^{\text{norm}} | \mathbf{w} \in \mathbb{R}_{\geq 0}^N\}.$$

This implies that the solution of (14) for BH, FTM, FP or TDF basis strategies does not depend on $V_0(\chi)$.

Finally, we solve (14) using a quasi-Newton method, see Nocedal and Wright ([15], Sect. 8). The expectations in (14) are evaluated using the terminal values obtained from the simulations. In this example we consider only ten basis strategies and so the time needed to solve the optimization accounts only for a minimal part of the total computational time, which is almost entirely determined by the simulations. If we consider hundreds of basis strategies then the situation might be different.

Our numerical results for $T = 20$ are presented in Tables 2, 3 and Fig. 5. We have solved (14) for different values of M and u . First, we see that the resulting prices in Table 2 are stable as M increases for fixed values of u . This indicates convergence. Secondly, we observe that prices increase considerably with decreasing u . This makes sense since we require lower shortfall risk for the optimal replicating strategy. Reasonably the risk tolerance parameter plays a crucial role in determining the value of the long term zero-coupon bond. In Table 3 we report for fixed M some risk measures related to the probability distribution of the loss. We observe that all the risk measures considered decrease as u decreases. This is sound and allows shortfall risk to be managed with respect to several risk measures at the same time by selecting u . We also observe that the lower partial second moments are nearly equal to the values of u . Of course, this must be the case since we are using this measure for the optimization. From Fig. 5 we observe that the solution tend to be concentrated in 3–4 basis strategies and the concentration increases as u decreases. These are BH1, FTM1, FP1, which invest in short term securities, and TDF, which invests across all maturities and increases the proportion invested in short term securities as time passes. This concentration makes intuitively sense. In our example BH1, FTM1 and FP1 have low terminal value volatility which appears to be an attractive property for the optimization algorithm. The attractiveness of TDF might be explained as follows. Among our basis strategies, TDF most closely matches the duration over time of the long term zero-coupon bond.

Next we apply our pricing algorithm for $T = 11, 12, \dots, 20$ and derive an extension of the CHF yield curve for these times to maturity. For this purpose we cannot set one single value for u across all times to maturity since we are considering losses over different periods and this would reflect different risk tolerances across maturities. Therefore we set $u_{11} > 0$ and consider the square root of time scaling

$$u_T = u_{11} \sqrt{T - 10},$$

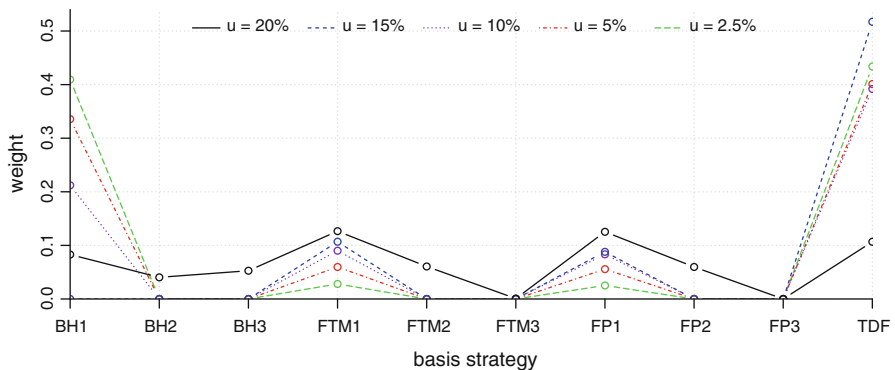
Table 2 Prices of a 20 years to maturity zero-coupon bond given by the approximation algorithm for different values of M, u

	$u = 20 \%$	$u = 15 \%$	$u = 10 \%$	$u = 5 \%$	$u = 2.5 \%$
$M = 125$	0.6490	0.7070	0.7703	0.8471	0.8940
$M = 250$	0.6572	0.7155	0.7788	0.8508	0.8920
$M = 500$	0.6575	0.7134	0.7811	0.8523	0.8975
$M = 1,000$	0.6564	0.7156	0.7796	0.8533	0.8990
$M = 2,000$	0.6554	0.7143	0.7789	0.8535	0.8992
$M = 4,000$	0.6558	0.7121	0.7773	0.8521	0.8964

Table 3 $T = 20$ and $M = 4,000$. Shortfall risk measures of the optimal portfolio for different values of u

	LSM1	LSM2	VaR95	VaR99	ES95	ES99
$u = 20 \%$	0.1541	0.2004	0.3254	0.3338	0.3314	0.3393
$u = 15 \%$	0.1087	0.1508	0.2605	0.2678	0.2656	0.2738
$u = 10 \%$	0.0675	0.1015	0.1913	0.2011	0.1978	0.2066
$u = 5 \%$	0.0295	0.0514	0.1123	0.1241	0.1201	0.1304
$u = 2.5 \%$	0.0127	0.0262	0.0653	0.0778	0.0739	0.0853

LSM1 and LSM2 are the lower standardized moments given by $\mathbb{E}[\max\{1 - V, 0\}^k]^{\frac{1}{k}}$ for $k = 1, 2$ respectively. VaR and ES are the Value-At-Risk and Expected Shortfall measures for 95 and 99 % confidence levels. All four measures are computed applying sample estimators on the simulated terminal values of the optimal portfolio

**Fig. 5** Optimal portfolio weights assigned to the 10 basis strategies for $M = 4,000$ and different values of u

for $T = 11, 12, \dots, 20$. This rule is well known in the literature to scale volatility estimates and relies on the rationale of one-period returns being i.i.d. over time. In non-i.i.d. environments, such as in our model, this rule produces estimates which are reasonable on average but show too large volatility fluctuations, see Christoffersen et al. [3] for more details. In Fig. 6 we show the extension. This extension has a

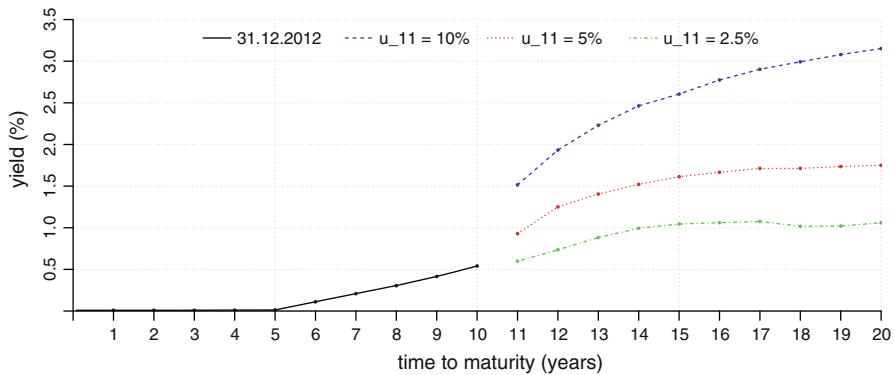


Fig. 6 $M = 4,000$. Extension of the CHF yield curve at 31.12.2012 for different values of u_{11} . Note that the market yields are at historical lows

positive spread compared to the market yield curve which becomes smaller as u_{11} decreases. This makes intuitively sense since the derived yields are not risk-free. Because of reinvestment risk we can only replicate those securities up to a certain risk tolerance specified by $u_{11} > 0$. We also observe that the steepness and concavity of the extension matches the market yield curve better as u_{11} decreases. Thus, we obtain a natural extension of the yield curve beyond the last liquidly traded time to maturity. This extrapolation is obtained by searching explicitly for replicating strategies up to the risk tolerance.

7 Conclusions

In this paper we have developed a computationally efficient algorithm to extrapolate the yield curve at the long end by explicitly considering replication strategies. Our procedure is based on a stochastic yield curve model with reinvestment risk which can be effectively calibrated to market data. The hedging strategy is constructed by taking an optimal linear combination of trading rules specified in advance according to a certain risk tolerance. This involves reinvestment risk, which cannot be completely eliminated, and the lower this risk the smaller the spread between the extrapolated yields and the market yields. Thus, we obtain a natural extension of the yield curve beyond the last liquid time to maturity and moreover, we also obtain the hedging strategy which replicates this value.

Appendix: Proofs

Proof of Lemma 2.1

Let $t \in J'$ and $j \in D^-$. We set $\mathbf{y} = \mathbf{Y}_{t-\delta} \in \mathbb{R}^d$ for this proof. For the drift term (4) we have

$$\alpha_j(y_{j+1}) = \frac{\delta}{2} h(y_{j+1})^2 \sum_{i=1}^{d-1} \lambda_{ij}^2 = \frac{\delta}{2} \zeta^-(\mathbf{y})_{jj}^2 \text{sp}(\Lambda^-(\Lambda^-)')_j = \frac{\delta}{2} \text{sp}(\Sigma^-(\mathbf{y}))_j.$$

Using (1) we rewrite the model assumption (3) under the real world measure \mathbb{P} and obtain

$$\begin{aligned} \tilde{Y}_{t,j}^- &= -\delta Y_{t-\delta,1} + \alpha_j(Y_{t-\delta,j+1}) + \sqrt{\delta} \sum_{i=1}^{d-1} h(Y_{t-\delta,j+1}) \lambda_{ij} (\varepsilon_{t,i} - \mu_{t,i}) \\ &= -\delta Y_{t-\delta,1} + \frac{\delta}{2} \text{sp}(\Sigma^-(\mathbf{y}))_j + \sqrt{\delta} (\zeta^-(\mathbf{y}) \Lambda^-(\mathbf{e}_t^- - \mu_t^-))_j. \end{aligned}$$

The statement on the conditional normality of \tilde{Y}_t follows directly from (5). \square

Proof of Lemma 2.2

Let $t \in J'$ and set $\mathbf{y} = \mathbf{Y}_{t-\delta}$ for this proof. For the components $(\tilde{Y}_{t,j})_{j \in D^-}$ Eq. (8) is equivalent to (5). Therefore, we only need to consider the random variable $\tilde{Y}_{t,d}$. From the definitions of $\zeta(\cdot)$ and Λ observe that $\zeta(\cdot)_{dd} = \zeta(\cdot)_{d-1,d-1}$, $\Lambda_{dd} = \kappa$ and $\Lambda_{d,j} = \Lambda_{d-1,j} = \lambda_{j,d-1}$ for $j \in D^-$. Hence, for $\tilde{\mu}^- = (\mu^-, 0)$ we have

$$(\zeta(\mathbf{y}) \Lambda(\mathbf{e}_t - \tilde{\mu}_t))_d = \zeta^-(\mathbf{y}) \Lambda^-(\mathbf{e}_t^- - \mu_t^-)_{d-1} + h(y_d) \kappa \varepsilon_{t,d}, \quad (25)$$

and

$$\text{sp}(\Sigma(\mathbf{y}))_d = \zeta(\mathbf{y})_{dd}^2 \sum_{i=1}^{d-1} \lambda_{i,d-1}^2 + \zeta(\mathbf{y})_{dd}^2 \kappa^2 = \text{sp}(\Sigma^-(\mathbf{y}))_{d-1} + (\zeta(\mathbf{y}) \Lambda)_{dd}^2. \quad (26)$$

Using model assumptions (3) and (6) we obtain

$$\begin{aligned} \tilde{Y}_{t,d} &= d\delta Y_{t,d} - d\delta y_d \\ &= \delta \left(-y_1 + \frac{1}{2} \text{sp}(\Sigma^-(\mathbf{y}))_{d-1} + \beta(\mathbf{y}) \right) + \sqrt{\delta} ((\zeta^-(\mathbf{y}) \Lambda^-(\mathbf{e}_t^- - \mu_t^-))_{d-1} + h(y_d) \kappa \varepsilon_{t,d}) \\ &= \delta \left(-y_1 + \frac{1}{2} \text{sp}(\Sigma(\mathbf{y}))_d + \beta(\mathbf{y}) - \frac{1}{2} (\zeta(\mathbf{y}) \Lambda)_{dd}^2 \right) + \sqrt{\delta} (\zeta(\mathbf{y}) \Lambda(\mathbf{e}_t - \tilde{\mu}_t))_d. \end{aligned}$$

This proves the stochastic representation (8) for the last component under the real-world measure \mathbb{P} . The statements on the conditional distribution follow directly from (8). \square

Proof of Theorem 3.1

We only have to prove the second part. The statement is intuitively clear because $\varepsilon_{t,d}$ and $\varepsilon_{t,j}$ are independent, given $\mathcal{F}_{t-\delta}$, for all $t \in J'$ and $j \in D^-$. We prove the statement indirectly. Assume that the claim $c_T = 1$ is attainable. Let \mathbf{X} be a replicating strategy for c_T . This means that \mathbf{X} is self-financing and $V_T(\mathbf{X}) = 1$, \mathbb{P} -a.s. The no-arbitrage condition (2) and the self-financing condition (9) imply for all $t \in J'$

$$\begin{aligned}
\mathbb{E}^*[B_t^{-1}V_t(\mathbf{X})|\mathcal{F}_{t-\delta}] &= B_t^{-1}\mathbb{E}^*[V_t^-(\mathbf{X})|\mathcal{F}_{t-\delta}] = B_t^{-1}\mathbb{E}^*[\mathbf{X}'_t\mathbf{R}_t|\mathcal{F}_{t-\delta}] \\
&= B_t^{-1}X_{t,1}P_{t-\delta,1}^{-1} + B_t^{-1}\sum_{j=2}^d X_{t,j}P_{t-\delta,j}^{-1}\mathbb{E}^*[P_{t,j-1}|\mathcal{F}_{t-\delta}] \\
&= B_{t-\delta}^{-1}P_{t-\delta,1}X_{t,1}P_{t-\delta,1}^{-1} + B_t^{-1}\sum_{j=2}^d X_{t,j}P_{t-\delta,j}^{-1}\frac{B_{t-\delta}^{-1}}{B_t^{-1}}P_{t-\delta,j} \\
&= B_{t-\delta}^{-1}\sum_{j=1}^d X_{t,j} = B_{t-\delta}^{-1}V_{t-\delta}(\mathbf{X}),
\end{aligned}$$

i.e. the discounted value process $(B_t^{-1}V_t(\mathbf{X}))_{t \in J}$ is a $(\mathbb{P}^*, \mathbb{F})$ -martingale. The value of the replicating strategy at time $T - d\delta$ is then given by (see explanation below)

$$\begin{aligned}
B_{T-d\delta}^{-1}V_{T-d\delta}(\mathbf{X}) &= \mathbb{E}^*[B_T^{-1}V_T(\mathbf{X})|\mathcal{F}_{T-d\delta}] = \mathbb{E}^*[B_T^{-1}|\mathcal{F}_{T-d\delta}] = \mathbb{E}^*[B_{T-\delta}^{-1}P_{T-\delta,1}|\mathcal{F}_{T-d\delta}] \\
&= \mathbb{E}^*[\mathbb{E}^*[B_{T-\delta}^{-1}P_{T-\delta,1}|\mathcal{F}_{T-2\delta}]\mathcal{F}_{T-d\delta}] = \mathbb{E}^*[B_{T-2\delta}^{-1}P_{T-2\delta,2}|\mathcal{F}_{T-d\delta}] \\
&= \dots = B_{T-d\delta}^{-1}P_{T-d\delta,d}.
\end{aligned}$$

In the first step we use the martingale property. The second step follows from $\mathbb{P}^* \sim \mathbb{P}$ and $V_T(\mathbf{X}) = 1$, \mathbb{P} -a.s. We then use the tower property of conditional expectations and the no-arbitrage condition (2) iteratively. This result is intuitively clear: at time $T - d\delta$ holding the portfolio \mathbf{X} or investing $P(T - d\delta, T)$ in the zero-coupon bond with time to maturity $d\delta$ are both going to generate the same payoff of one unit of cash at maturity. Because of no-arbitrage the value of both strategies at time $T - d\delta$ must be the same. In this proof we set $\mathbf{y} = \mathbf{Y}_{T-(d+1)\delta}$, $\tilde{\boldsymbol{\mu}} = (\boldsymbol{\mu}^-, 0) = (\boldsymbol{\mu}_{T-d\delta}^-, 0)$ and $\mathbf{x} = \mathbf{X}_{T-d\delta}$. Note that these random variables are $\mathcal{F}_{T-(d+1)\delta}$ -measurable. Using (8), (25) and (26) we have

$$V_{T-d\delta}(\mathbf{X}) = P_{T-d\delta,d} = \exp\{-d\delta Y_{T-d\delta,T}\} = \exp\{\delta(dy_d - \beta(\mathbf{y}))\}k_{d-1}^-E_{d-1}^-E_d,$$

where for $j \in D^-$ we set

$$\begin{aligned}
k_j^- &= \exp\left\{\delta\left(y_1 - \frac{1}{2}\text{sp}(\Sigma^-(\mathbf{y}))_j\right) + \sqrt{\delta}(\zeta^-(\mathbf{y})\Lambda^-\boldsymbol{\mu}^-)_j\right\}, \\
E_j^- &= \exp\{-\sqrt{\delta}(\zeta^-(\mathbf{y})\Lambda^-\boldsymbol{\varepsilon}_t^-)_j\}, E_d = \exp\{-\sqrt{\delta}h(y_d)\kappa\varepsilon_{t,d}\}.
\end{aligned}$$

Define $\mathbf{k}^- = (k_1^-, \dots, k_{d-1}^-)'$ and $\mathbf{E}^- = (E_1^-, \dots, E_{d-1}^-)'$. Note that \mathbf{k}^- is $\mathcal{F}_{T-(d+1)\delta}$ -measurable, and \mathbf{E}^- and E_d are $\mathcal{F}_{T-d\delta}$ -measurable. Note that \mathbf{k}^- , \mathbf{E}^- , E_d have strictly positive components, \mathbb{P} -a.s. For all $j \in D^-$ the conditional distribution of $\log E_j^-$ is normal with zero mean and variance $\delta\text{sp}(\Sigma^-(\mathbf{y}))_j$. The conditional distribution of $\log E_d$ is also normal with zero mean and variance $\delta h(y_d)^2 \kappa^2$. On the other hand, using the self-financing condition (9) and the model Eq. (5) we obtain

$$\begin{aligned}
V_{T-d\delta}(\mathbf{X}) &= \mathbf{x}'\mathbf{R}_{T-d\delta} = x_1 \exp\{\delta y_1\} + \sum_{j=1}^{d-1} x_{j+1} \exp\{(j+1)\delta y_{j+1}\} \exp\{-j\delta Y_{T-d\delta,j}\} \\
&= x_1 \exp\{\delta y_1\} + \sum_{j=1}^{d-1} x_{j+1} k_j^- E_j^-.
\end{aligned}$$

From the model assumption we know that the components of ε_t are independent and therefore E_d is independent of E^- . This independence and the two expressions derived above for $V_{T-d\delta}(\mathbf{X})$ imply

$$\begin{aligned}
0 &= \exp\{-\delta(dy_d - \beta(\mathbf{y}))\} \sum_{j=1}^{d-1} x_{j+1} k_j^- \text{Cov}[E_j^-, E_d | \mathcal{F}_{T-(d+1)\delta}] \\
&= \exp\{-\delta(dy_d - \beta(\mathbf{y}))\} \text{Cov}[V_{T-d\delta}(\mathbf{X}) - x_1 \exp\{\delta y_1\}, E_d | \mathcal{F}_{T-(d+1)\delta}] \\
&= \exp\{-\delta(dy_d - \beta(\mathbf{y}))\} \text{Cov}[V_{T-d\delta}(\mathbf{X}), E_d | \mathcal{F}_{T-(d+1)\delta}] \\
&= k_{d-1}^- \text{Cov}[E_{d-1}^-, E_d | \mathcal{F}_{T-(d+1)\delta}] \\
&= k_{d-1}^- (\mathbb{E}[E_{d-1}^- E_d^2 | \mathcal{F}_{T-(d+1)\delta}] - \mathbb{E}[E_{d-1}^- E_d | \mathcal{F}_{T-(d+1)\delta}] \mathbb{E}[E_d | \mathcal{F}_{T-(d+1)\delta}]) \\
&= k_{d-1}^- \mathbb{E}[E_{d-1}^- | \mathcal{F}_{T-(d+1)\delta}] (\mathbb{E}[E_d^2 | \mathcal{F}_{T-(d+1)\delta}] - \mathbb{E}[E_d | \mathcal{F}_{T-(d+1)\delta}]^2) \\
&= k_{d-1}^- \mathbb{E}[E_{d-1}^- | \mathcal{F}_{T-(d+1)\delta}] \text{Var}[E_d | \mathcal{F}_{T-(d+1)\delta}] \\
&= k_{d-1}^- \exp\left\{\frac{1}{2} \delta \text{sp}(\Sigma^-(\mathbf{y}))_{d-1}\right\} (\exp\{\delta h(y_d)^2 \kappa^2\} - 1) \exp\{\delta h(y_d)^2 \kappa^2\} > 0,
\end{aligned}$$

where the right-hand side of the above equation is strictly larger than zero, given $\mathcal{F}_{T-(d+1)\delta}$ (assuming $\kappa \neq 0$ and $h(y_d) \neq 0$). This is a contradiction and proves the claim. \square

Convexity of (13)

The convexity of \mathcal{D} is clear. We consider the objective function in (13). Let $\mathbf{X}, \mathbf{Y} \in \mathcal{D}$ and $s \in [0, 1]$. We estimate (see explanation below)

$$\begin{aligned}
&V_0(s\mathbf{X} + (1-s)\mathbf{Y}) + \theta(\mathbb{E}[\mathbf{f}(V_T(s\mathbf{X} + (1-s)\mathbf{Y}))]) \\
&= sV_0(\mathbf{X}) + (1-s)V_0(\mathbf{Y}) + \theta(\mathbb{E}[\mathbf{f}(sV_T(\mathbf{X}) + (1-s)V_T(\mathbf{Y}))]) \\
&\leq sV_0(\mathbf{X}) + (1-s)V_0(\mathbf{Y}) + \theta(\mathbb{E}[s\mathbf{f}(V_T(\mathbf{X})) + (1-s)\mathbf{f}(V_T(\mathbf{Y}))]) \\
&= sV_0(\mathbf{X}) + (1-s)V_0(\mathbf{Y}) + \theta(s\mathbb{E}[\mathbf{f}(V_T(\mathbf{X}))] + (1-s)\mathbb{E}[\mathbf{f}(V_T(\mathbf{Y}))]) \\
&\leq s(V_0(\mathbf{X}) + \theta(\mathbb{E}[\mathbf{f}(V_T(\mathbf{X}))])) + (1-s)(V_0(\mathbf{Y}) + \theta(\mathbb{E}[\mathbf{f}(V_T(\mathbf{Y}))])),
\end{aligned}$$

where in the first equality we use the linearity of $V_0(\cdot)$ and $V_T(\cdot)$. In the second step we use the convexity of \mathbf{f} , the monotonicity of the expected value and the fact that θ is increasing. Finally in the fourth step the convexity of θ is used. This proves the convexity of the objective function. \square

Proof of Lemma 4.3

The given identity follows by straightforward application of the definitions. We show it only for TDF strategies and for the other strategies proceed similarly. Let $t \in J'$ and set $\mathbf{X} = \mathbf{X}^{\text{TDF}}$. Since the strategies considered are all self-financing we have $V_t(\mathbf{X}) = V_t^-(\mathbf{X}) = \mathbf{X}'_t \mathbf{R}_t$ for all $t \in J'$. Iterating the definition of TDF strategies we obtain

$$\begin{aligned} V_t(\mathbf{X}) &= \mathbf{X}'_t \mathbf{R}_t = \sum_{j,k=1}^d (p_t \pi_k 1_{1 \leq k \leq d'} + (1 - p_t) \pi_k 1_{d' < k \leq d}) X_{t-\delta,k} R_{t-\delta,k} R_{t,j} = \dots \\ &= V_0 \sum_{j_1, \dots, j_{\frac{t}{\delta}}=1}^d B_t(j_1, \dots, j_{\frac{t}{\delta}}, 1, \dots, 1) \prod_{s=1}^{\frac{t}{\delta}} (p_t \eta_{j_s} 1_{1 \leq j_s \leq d'} + (1 - p_t) \eta_{j_s} 1_{d' < j_s \leq d}). \end{aligned}$$

Proof of Lemma 4.4

Let $t \in J'$, $s = \frac{t}{\delta}$ and $j_1, \dots, j_n \in D$. For this proof we set $\mathbf{y} = \mathbf{Y}_{(s-1)\delta} = \mathbf{Y}_{t-\delta}$. The statement for $j_s = 1$ follows directly from (15). Assuming $j_s > 1$ and using (15) we have

$$B_t(j_1, \dots, j_n) = B_{t-\delta}(j_1, \dots, j_n) R_{t,j_s} = B_{t-\delta}(j_1, \dots, j_n) \exp\{-\tilde{Y}_{s\delta, j_s-1}\}.$$

Then, using Lemma 2.2, we have

$$\frac{B_t(j_1, \dots, j_n)}{B_{t-\delta}(j_1, \dots, j_n)} = P_{t-\delta,1} \exp\left\{-\frac{\delta}{2} \text{sp}(\Sigma(\mathbf{y}))_{j_s-1} + \sqrt{\delta}(\zeta(\mathbf{y})\Lambda(\tilde{\boldsymbol{\mu}}_t - \boldsymbol{\varepsilon}_t))_{j_s-1}\right\}.$$

Taking the logarithm of the right-hand side proves the statement. \square

References

1. Artzner P, Delbaen F, Koch-Medina P (2009) Risk measures and efficient use of capital. *Astin Bull* 39(1):203–228
2. Bodie Z, Treussard J (2007) Making investment choices as simple as possible, but not simpler. *Financial Analyst J* 63(3):42–47
3. Christoffersen PF, Diebold FX, Schuermann T (1998) Horizon problems and extreme events in financial risk management. *Econ Policy Rev* 1998:109–118
4. Dahl M (2007) A discrete-time model for reinvestment risk in bond markets. *Astin Bull* 37(2):235–264
5. Dahl M (2009) A continuous-time model for reinvestment risk in bond markets. *Quant Finance* 9(4):451–464
6. Delbaen F, Schachermayer W (1994) A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300:463–520
7. Föllmer H, Schied A (2002) *Stochastic Finance*. de Gruyter
8. Guillaume de N, Rebonato R, Pogudin A (2012) The nature of the dependence of the magnitude of rate moves on the level of rates: a universal relationship (Preprint)
9. Hilli P, Koivu M, Pennanen T (2011) Cash-flow based valuation of pension liabilities. *Eur Actuarial J* 1(2):329–343

10. Hilli P, Koivu M, Pennanen T (2011) Optimal construction of a fund of funds. *Eur Actuarial J* 1(2):345–359
11. Hilli P, Pennanen T (2008) Numerical study of discretizations of multistage stochastic programs. *Kybernetika* 44(2):185–204
12. Jarque CM, Bera AK (1987) A test for normality of observations and regression residuals. *Int Stat Rev* 55(2):163–172
13. Koivu M, Pennanen T (2010) Galerkin methods in dynamic stochastic programming. *Optimization* 59(3):339–354
14. Müller R (2002) Zur Berechnung der Obligationen Renditen im Statistischen Monatsheft der SNB. *Swiss National Bank Quartalsheft* 2002(2):64–73
15. Nocedal J, Wright SJ (2006) Numerical optimization. Springer, New York
16. Ortega JP, Pullirsch R, Teichmann J, Wergieluk J (2009) A dynamic approach for scenario generation in risk management (Preprint on arXiv)
17. Pflug G (2001) Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math Program* 89:251–271
18. Shapiro A (2003) Inference of statistical bounds for multistage stochastic programming problems. *Math Methods Oper Res* 58:57–68
19. Sommer D (1997) Pricing and hedging of contingent claims in term structure models with exogenous issuing of new bonds. *Eur Financial Manag* 3:269–292
20. Teichmann J, Wüthrich MV (2013) Consistent yield curve prediction. ETH Zurich (Preprint)
21. Wüthrich MV, Merz M (2013) Financial modeling, actuarial valuation and solvency. Springer, Berlin