

CONSISTENT YIELD CURVE PREDICTION

BY

JOSEF TEICHMANN AND MARIO V. WÜTHRICH

ABSTRACT

We present an arbitrage-free non-parametric yield curve prediction model which takes the full discretized yield curve data as input state variable. Absence of arbitrage is a particularly important model feature for prediction models in case of highly correlated data as, for instance, interest rates. Furthermore, the model structure allows to separate constructing the daily yield curve from estimating its volatility structure and from calibrating the market prices of risk. The empirical part includes tests on modeling assumptions, out-of-sample back-testing and a comparison with the Vasiček (1977) short-rate model.

KEYWORDS

Yield curve prediction, no-arbitrage, Heath–Jarrow–Morton type model, volatility scaling factors and return directions.

1. INTRODUCTION

Future cash flows are valued using today's risk-free yield curve and its predicted future shapes. For this purpose, first today's yield curve needs to be constructed from appropriate financial instruments like government bonds, swap rates and corporate bonds, and, second, a stochastic dynamics for yield curve predictions needs to be specified. This specification is a delicate task because, in general, it involves time series of potentially infinite dimensional random vectors and/or random functions. We aim for long term prediction of yield curves as needed, for instance, in the insurance industry. Here, "long term" means predictions up to $t = 5$ years into the future, however, we are dealing with the whole yield curve with considering times to maturity up to $T = 30$ years.

We introduce notation to fix ideas: $t \geq 0$ denotes running time in years. For $T \geq t$, we denote by $P(t, T) > 0$ the price at time t of the (default-free) zero coupon bond (ZCB) that pays one unit of currency at maturity date T . The yield curve at (running) time t for maturity dates $T \geq t$ is then given by the continuously-compounded spot rate (yield) defined by

$$Y(t, T) = -\frac{1}{T-t} \log P(t, T). \quad (1.1)$$

1.1. Aim and scope

Model the stochastic evolution of yield curves $T \mapsto Y(t, T)$ for future dates $t \in (0, T)$ such that:

- i. the model is economically reasonable, e.g. rates are most likely positive and the model is free of arbitrage;
- ii. the model explains past yield curve observations, i.e. it is not rejected by back-testing;
- iii. the model allows to predict future yield curve evolution, i.e. it can be evaluated and updated online.

More specifically, we aim to describe a comparatively simple methodology to specify stochastic models for bond price processes $(P(t, T))_{0 \leq t \leq T}$ of the type

$$dP(t, T) = \Sigma(t, T) d\epsilon_t + \delta(t, T) dN_t + \mu(t, T) dt,$$

with predictable matrix-valued processes $\Sigma(\cdot, T)$, $\delta(\cdot, T)$ and a predictable scalar-valued process $\mu(\cdot, T)$. These predictable processes are representing volatility acting on a vector of Brownian motions $(\epsilon_t)_t$, a jump height structure acting on a vector of random measures $(N_t)_t$ and a drift component of the ZCB evolution under the constraint $P(T, T) = 1$, respectively. In this introduction, we choose the language of continuous time processes even though later on our analysis is only done for discretized versions of those equations or, in other words, we talk about all possible and reasonable discretizations of the previously specified continuous time stochastic process. Furthermore, we assume that there exists an equivalent (local) martingale measure for discounted ZCB price processes. Since we are interested in long term phenomena with t running up to five years, it is not a too strong restriction to assume $\delta = 0$, i.e. to consider a purely diffusive evolution (this requirement can easily be weakened by replacing the Brownian motion through a Lévy process). Then, heuristically speaking, the previous Brownian motion conditions mean that the stochastic activity of the ZCB price processes vanishes when $t \rightarrow T$ and, most importantly, that

$$\mu(t, T) = \Sigma(t, T)\lambda_t + r_t, \tag{1.2}$$

where $(\lambda_t)_t$ is an appropriately integrable and predictable vector-valued process (called market price of risk), and $(r_t)_t$ is the risk-free short-rate process used for discounting. Formulation (1.2) is essential since we expect ZCB prices for different maturities T to be highly correlated, henceforth, instantaneous stochastic covariance Σ will be close to singular. Therefore, generically *not* every drift $\mu(\cdot, T)$ can be written in the form (1.2), which however is necessary for the no-arbitrage condition.

From statistical theory, it is well-known that model selection has to deal with a hierarchy of estimations from given time series: the best observable quantities are bond prices themselves leading to the initialization problem of constructing an appropriate yield curve for each day. This problem has been well studied and

is not dealt with in the present work, see for instance Nelson and Siegel (1987), Svensson (1994, 1995), Cairns (1998, 2004), Müller (2002), Brigo and Mercurio (2006), Filipović (2009).

The second best observable quantities are the components of the instantaneous covariance matrix Σ along the observed time series, which could — in principle — be fully identified from a continuous observation along the observation path in a compact time interval. Here, several approaches from classical statistical theory could be used incorporating local or stochastic volatilities. In the present paper, we tackle the problem of identifying the volatility structure for yield curve prediction by using a non-parametric approach which is based on ideas presented in Barone-Adesi *et al.* (1998) and Ortega *et al.* (2009), together with a parametric approach for the specification of the local volatility impact of yield levels based on the work of Deguillaume *et al.* (2013).

The most difficult part is the identification of the market price of risk $(\lambda_t)_t$, i.e. the deviation of μ from the risk-free short-rate process $(r_t)_t$ in terms of volatility Σ , see (1.2). These are good and bad news: it means that for pricing purposes (i.e. under the equivalent martingale measure) we are practically able to identify the model from time series data under fairly weak assumptions, since $(\lambda_t)_t$ is not needed in the equivalent martingale measure; but for prediction purposes, we have to deal with one quantity that is hard to measure, namely the market price of risk $(\lambda_t)_t$. We provide some results in this direction, but we do *not* provide any methodology there, except singling out the problem and separating it from determining the yield curve and determining the volatility structure. Notice, however, that determining Σ is *the crucial step* towards a measurement of μ since, due to no-arbitrage considerations, the difference $\mu - r$ has to lie in the range of Σ , see (1.2).

The presented Heath–Jarrow–Morton (HJM)-type approach, we refer to Heath *et al.* (1990, 1992), allows to separate the different tasks of initializing the yield curve, of estimating the volatility structure and of identifying the market price of risk. In particular, such models are from the very beginning on free of arbitrage. We consider this separation feature as the most important aspect of the present work.

Standard literature on prediction (e.g. on p. 139 in Brigo and Mercurio (2006) or the recent contribution of Audrino and Filipova (2009)) often directly starts with yield curves and applies statistical methods like principal component analysis (PCA) to time series of yield curves. In this context — by the seemingly innocent fact that drifts cannot be measured — drifts are often neglected or fixed in an almost arbitrary manner. This, however, leads to arbitrage since arbitrage-free models written in terms of yield curves *must* have a drift correction in the equivalent martingale measure, namely, exactly the HJM-type drift correction, which is due to the (non-linear) logarithmic transform between prices and yields, see (1.1). Because of the rather singular nature of the instantaneous covariance matrix, this drift correction will remain visible even in the real-world probability measure. One could argue that covariance matrices are rather invertible and hence any drift can be removed, but this argument does no longer hold true in

the highly correlated infinite-dimensional world of yield curve movements. On the other hand, adding the correct drift corrections to PCA approaches for yield curve evolutions restricts the PCA machinery substantially, since one cannot expect — by the very nature of the HJM equation — the stochastic evolution of yield curves to remain generically in a finite dimensional subspace. See for instance, Filipović and Teichmann (2004), where the general situation is clarified following ideas of Björk and Landén (2000) and Björk and Svensson (2001).

In contrast, factor models, we refer to Cairns (2004), Brigo and Mercurio (2006) and Filipović (2009), preserve by their very construction the no-arbitrage property, but there are undesirable interactions between the initial construction of the yield curve and specifying the parameter structure of the factor model. This can be circumvented by Hull–White extensions, see Hull and White (1990, 1994), in case of affine models, however, then fundamental time inhomogeneities are introduced into the model, which are not desirable features from statistical and from (re-)calibration points of view. Notice also that in our setting we *always* have to see factor models from the term structure perspective, i.e. in which the stochastic evolution of the term structure is described by a given factor model. In case, e.g. of the Vasiček model, see Section 4.3 below, this means that the volatility structure of the term structure of yields determines σ and κ (sic!) and that the rank of the covariance matrix is 1, which in turn strongly influences how a generic initial yield is changing instantaneously.

We insist that models should be free of arbitrage. This requirement is crucial when it comes to the prediction of highly correlated prices as it is the case for interest rates. Otherwise, it is possible to “artificially” shift profit-and-loss distributions. More precisely, if a prediction model admits arbitrage then implementing appropriate long–short arbitrage portfolios yields always positive profit-and-loss structures. In practice, adding such an arbitrage portfolio can then be used to shift profit-and-loss distributions of general portfolios, which is an undesired effect from the point of view of risk management and regulation, see Figure 17 and Section 4.4, below.

1.2. Organization of the paper

The remainder of the paper is organized as follows: in Section 2, we describe the discrete time setup for the (discretized) yield curve evolution. Moreover, we describe — for the sake of completeness — the no-arbitrage conditions within the given discrete time modeling setup. In Section 3, we describe the actual estimation procedure of the volatility structure and in Section 4, we present a concrete application with real market data.

2. DISCRETE TIME MODEL AND NO-ARBITRAGE

In this section, we describe the discrete time HJM-modeling setup, see Heath *et al.* (1990, 1992). Choose a fixed grid size $\Delta = 1/n$ for $n \in \mathbb{N}$. We consider

the discrete time points $t \in \Delta\mathbb{N}_0 = \{0, \Delta, 2\Delta, 3\Delta, \dots\}$ and the maturity dates $T \in t + \Delta\mathbb{N}$. For example, choice $n = 1$ corresponds to a yearly grid, $n = 4$ to a quarterly grid, $n = 12$ to a monthly grid, $n = 52$ to a weekly grid and $n = 250$ to a business day grid. The filtered probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ with real-world probability measure \mathbb{P} and (discrete time) filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in \Delta\mathbb{N}_0}$.

We assume that ZCBs exist at all time points $t \in \Delta\mathbb{N}_0$ for all maturity dates $T = t + m$ with times to maturity $m \in \Delta\mathbb{N}$. Then, we consider the discrete time yield curves

$$\mathbf{Y}_t = (Y(t, t + m))'_{m \in \Delta\mathbb{N}}$$

at time points $t \in \Delta\mathbb{N}_0$. Assume that $(\mathbf{Y}_t)_{t \in \Delta\mathbb{N}_0}$ is \mathbb{F} -adapted, that is, $(\mathbf{Y}_s)_{s \leq t}$ is observable at time t with respect to the information given by the σ -field \mathcal{F}_t . Our aim is to model and predict $(\mathbf{Y}_t)_{t \in \Delta\mathbb{N}_0}$. We assume that there exists an equivalent martingale measure $\mathbb{P}^* \sim \mathbb{P}$ with respect to the bank account numeraire discount $(B_t^{-1})_{t \in \Delta\mathbb{N}_0}$ and, in a first step, we describe $(\mathbf{Y}_t)_{t \in \Delta\mathbb{N}_0}$ directly under this equivalent martingale measure \mathbb{P}^* . Notice that here the bank account numeraire is a discrete time roll-over portfolio, we also refer to the next section.

Remark. The assumption that the yield curve is given at any moment $t \in \Delta\mathbb{N}_0$ for sufficiently many maturities is a very strong one. In practice, the yield curve is inter- and extrapolated every day from quite different traded instruments like coupon bearing bonds, swap rates, etc. This inter- and extrapolation allows for a lot of freedom, often parametric families are considered, e.g. the Nelson and Siegel (1987) or the Svensson (1994, 1995) families are used, but also non-parametric approaches such as splines are applied, see Björk (1998), Cairns (1998), Filipović (1999, 2000, 2009) and Christensen *et al.* (2007, 2009).

Assume the initial yield curve $\mathbf{Y}_0 = (Y(0, m))_{m \in \Delta\mathbb{N}}$ at time $t = 0$ is given. For $t, m \in \Delta\mathbb{N}$, we make the following model assumptions: assume there exist deterministic functions $\alpha_\Delta(\cdot, \cdot, \cdot)$ and $\mathbf{v}_\Delta(\cdot, \cdot, \cdot)$ such that the yield curve has the following stochastic representation:

$$m Y(t, t + m) = (m + \Delta) Y(t - \Delta, t + m) - \Delta Y(t - \Delta, t) + \alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t - \Delta}) + \mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t - \Delta}) \boldsymbol{\varepsilon}_t^*, \quad (2.1)$$

where the innovations $\boldsymbol{\varepsilon}_t^*$ are \mathcal{F}_t -measurable, independent of $\mathcal{F}_{t-\Delta}$, centered and have unit variance under \mathbb{P}^* . In general, the innovations $\boldsymbol{\varepsilon}_t^*$ are multivariate random vectors and the last product in (2.1) needs to be understood in the inner product sense. Notice that — in contrast to the introduction — we do not model bond prices directly but rather (absolute) yields, to fulfill the requirements $P(t, T) > 0$ and $P(T, T) = 1$.

Remark. The first two terms on the right-hand side of (2.1) exactly correspond to the no-arbitrage condition in a deterministic interest-rate model (see (2.2)

in Filipović (2009)). The fourth term on the right-hand side of (2.1) described by $\mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) \boldsymbol{\epsilon}_t^*$ adds the stochastic part to the future yield curve development. Finally, the third term $\alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta})$ will be recognized as the HJM term that makes the stochastic model free of arbitrage. This term is going to be analyzed in detail in Lemma 2.1 below. This approach allows us to separate conceptually the task of estimating volatilities, i.e. estimating \mathbf{v}_Δ , and estimating the market price of risk, i.e. the difference between the two measures \mathbb{P} and \mathbb{P}^* .

Assumption (2.1) implies for the price of the ZCB at time t with time to maturity m

$$P(t, t + m) = \frac{P(t - \Delta, t + m)}{P(t - \Delta, t)} \exp \left\{ -\alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) - \mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) \boldsymbol{\epsilon}_t^* \right\}.$$

In order to determine the HJM term $\alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta})$, we define the discrete time bank account value for an initial investment of 1 as follows: $B_0 = 1$ and for $t \in \Delta\mathbb{N}$

$$B_t = \prod_{s=0}^{t/\Delta-1} P(\Delta s, \Delta(s + 1))^{-1} = \exp \left\{ \Delta \sum_{s=0}^{t/\Delta-1} Y(\Delta s, \Delta(s + 1)) \right\} > 0.$$

The process $\mathbf{B} = (B_t)_{t \in \Delta\mathbb{N}_0}$ considers the riskless roll-over of an initial investment 1 into the (discrete time) bank account with grid size Δ . Note that \mathbf{B} is previsible, i.e. B_t is $\mathcal{F}_{t-\Delta}$ -measurable for all $t \in \Delta\mathbb{N}$.

Absence of arbitrage is now expressed in terms of the following $(\mathbb{P}^*, \mathbb{F})$ -martingale property (under the assumption that all the conditional expectations exist). We require for all $t, m \in \Delta\mathbb{N}$

$$\mathbb{E}^* \left[B_t^{-1} P(t, t + m) \mid \mathcal{F}_{t-\Delta} \right] \stackrel{!}{=} B_{t-\Delta}^{-1} P(t - \Delta, t + m). \tag{2.2}$$

Such a martingale property is sufficient and necessary (under integrability assumptions) for the absence of arbitrage due to the fundamental theorem of asset pricing (FTAP) derived in Delbaen and Schachermayer (1994). For notational convenience, we set $\mathbb{E}_t^*[\cdot] = \mathbb{E}^*[\cdot \mid \mathcal{F}_t]$ for $t \in \Delta\mathbb{N}_0$. No-arbitrage condition (2.2) immediately provides the following lemma, see also Heath *et al.* (1990, 1992) and Filipović and Zabczyk (2002).

Lemma 2.1. *Under the above assumptions, the absence of arbitrage condition (2.2) implies*

$$\alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) = \log \mathbb{E}_{t-\Delta}^* \left[\exp \left\{ -\mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) \boldsymbol{\epsilon}_t^* \right\} \right].$$

This solves item (i) of the aim and scope list. The proof is provided in the appendix.

Remark. Notice that we do not explicitly assume that $T \mapsto P(t, T)$ is decreasing in maturity T , even though we believe that this should be the case with high probability. The functional dependence of \mathbf{v}_Δ on the term structure \mathbf{Y} controls the probability of scenarios where the term structure of bond prices is not decreasing in maturity T . In our case, this is expressed through the choice of the scale function (3.7), below.

3. MODELING ASPECTS AND CALIBRATION

We now discuss the explicit modeling choices $\mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta})$ and $\boldsymbol{\epsilon}_t^*$ under \mathbb{P}^* as well as the description of the change of measure $\mathbb{P}^* \sim \mathbb{P}$. Then, the pricing model w.r.t. \mathbb{P}^* and the prediction model w.r.t. \mathbb{P} are fully specified through Lemma 2.1.

3.1. Data and explicit model choice with respect to \mathbb{P}^*

Assume we study a finite set $\mathcal{M} \subset \Delta\mathbb{N}$ of times to maturity. We specify below necessary properties of \mathcal{M} for yield curve prediction. In view of (2.1), we define for these time to maturity choices and $t \in \Delta\mathbb{N}$ the (finite dimensional) random vectors

$$\begin{aligned} \Upsilon_t &= (\Upsilon_{t,m})'_{m \in \mathcal{M}} = (m Y(t, t+m) - (m + \Delta) Y(t - \Delta, t+m))'_{m \in \mathcal{M}} \\ &= \left(-\log \frac{P(t, t+m)}{P(t - \Delta, t+m)} \right)'_{m \in \mathcal{M}}. \end{aligned}$$

We set dimension $d = |\mathcal{M}|$ and for $\boldsymbol{\epsilon}_t^* | \mathcal{F}_{t-\Delta}$, we choose a d -dimensional standard Gaussian distribution with independent components under the equivalent martingale measure \mathbb{P}^* .

Remark. We are aware that the choice of multivariate Gaussian innovations $\boldsymbol{\epsilon}_t^*$ is only a first step towards more realistic innovation processes. However, we believe that already in this model, with suitably chosen estimations of the instantaneous covariance structure, the results are quite convincing — additionally chosen jump structures might even improve the situation. The independence assumption with respect to the equivalent martingale measure is an additional strong assumption which could be weakened. We make these choices here to have analytical tractability for sensitivity analysis and back-testing.

We re-scale the instantaneous variance term with the grid size Δ and we assume that at time t it only depends on the last observation $\mathbf{Y}_{t-\Delta}$: define $\mathbf{v}_\Delta(\cdot, \cdot, \cdot)$ by

$$\mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) = \sqrt{\Delta} \boldsymbol{\sigma}(t, m, \mathbf{Y}_{t-\Delta}),$$

where the function $\sigma(\cdot, \cdot, \cdot)$ does not (directly) depend on the grid size Δ . Lemma 2.1 implies for these choices for the HJM term

$$\begin{aligned} \alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) &= \log \mathbb{E}_{t-\Delta}^* \left[\exp \left\{ -\sqrt{\Delta} \sigma(t, m, \mathbf{Y}_{t-\Delta}) \mathbf{e}_t^* \right\} \right] \\ &= \frac{\Delta}{2} \|\sigma(t, m, \mathbf{Y}_{t-\Delta})\|^2. \end{aligned}$$

From (2.1), we then obtain for $t \in \Delta\mathbb{N}$ and $m \in \mathcal{M}$ under equivalent martingale measure \mathbb{P}^*

$$\Upsilon_{t,m} = \Delta \left[-Y(t - \Delta, t) + \frac{1}{2} \|\sigma(t, m, \mathbf{Y}_{t-\Delta})\|^2 \right] + \sqrt{\Delta} \sigma(t, m, \mathbf{Y}_{t-\Delta}) \mathbf{e}_t^*. \tag{3.1}$$

Note that $(\Upsilon_t)_{t \in \Delta\mathbb{N}}$ is a d -dimensional process, therefore, we need a d -dimensional Gaussian random vector $\mathbf{e}_t^*|_{\mathcal{F}_{t-\Delta}}$ for obtaining sufficiently rich stochastic structures. Next, we specify explicitly the d -dimensional vector-valued function $\sigma(\cdot, \cdot, \cdot)$. We proceed similar to Ortega *et al.* (2009) and we directly model volatilities and return directions: we choose (i) linear maps ζ which describe the *volatility scaling factors*; and (ii) a matrix $\Lambda = [\lambda_1, \dots, \lambda_d] \in \mathbb{R}^{d \times d}$, where $\lambda_1, \dots, \lambda_d \in \mathbb{R}^d$ specify the (*raw*) *return directions* of the innovations \mathbf{e}_t^* . To be precise, (i) we choose invertible linear maps/matrices $\zeta(\mathbf{y})$, defined for every $\mathbf{y} \in \mathbb{R}^d$, providing

$$\zeta(\mathbf{y}) : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \lambda \mapsto \zeta(\mathbf{y})\lambda, \tag{3.2}$$

and (ii) we choose vectors $\lambda_1, \dots, \lambda_d \in \mathbb{R}^d$ which define the matrix $\Lambda = [\lambda_1, \dots, \lambda_d] \in \mathbb{R}^{d \times d}$. We think of volatility $\zeta(\mathbf{y})\lambda_i$ acting on the i th coordinate of \mathbf{e}_t^* , justifying the notion of (possible) return direction for λ_i , see (3.3) below. For $\mathbf{y} \in \mathbb{R}^d$, we form the covariance structure

$$\Sigma_\Lambda(\mathbf{y}) = \zeta(\mathbf{y}) \Lambda \Lambda' \zeta'(\mathbf{y}) \in \mathbb{R}^{d \times d}.$$

Using this notation, we proceed towards the following model specification for (3.1):

Model Assumptions 3.1. *We choose the following model for the yield curve at time $t \in \Delta\mathbb{N}$ with time to maturity dates \mathcal{M} :*

$$\Upsilon_t = \Delta \left[-Y(t - \Delta, t) + \frac{1}{2} \text{sp}(\Sigma_\Lambda(\mathbf{Y}_{t-\Delta}^-)) \right] + \sqrt{\Delta} \zeta(\mathbf{Y}_{t-\Delta}^-) \Lambda \mathbf{e}_t^*,$$

with $\mathbf{Y}(t - \Delta, t) = (Y(t - \Delta, t), \dots, Y(t - \Delta, t))' \in \mathbb{R}^d$, $\text{sp}(\Sigma_\Lambda) \in \mathbb{R}^d$ denotes the d -dimensional vector containing the diagonal elements of covariance matrix $\Sigma_\Lambda \in \mathbb{R}^{d \times d}$, and with $\mathbf{Y}_{t-\Delta}^- = (Y(t - \Delta, t + m))'_{m \in \mathcal{M}} \in \mathbb{R}^d$.

For the j th time to maturity $m_j \in \mathcal{M}$, we have done the following choice:

$$\boldsymbol{\sigma}(t, m_j, \mathbf{Y}_{t-\Delta}^-) \boldsymbol{\varepsilon}_t^* = \sum_{i=1}^d \sigma_i(t, m_j, \mathbf{Y}_{t-\Delta}^-) \varepsilon_{t,i}^* = \sum_{i=1}^d [\zeta(\mathbf{Y}_{t-\Delta}^-) \boldsymbol{\lambda}_i]_j \varepsilon_{t,i}^*. \tag{3.3}$$

Our aim is to estimate the volatility scaling factors ζ and the return directions $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_d \in \mathbb{R}^d$. This is done in Subsection 3.2. Note that these choices do not depend on the grid size Δ .

Remark 3.2. The volatility scaling factors $\zeta(\cdot)$ mimic how volatility for different times to maturity scales with the level of yields at this maturity date. Several approaches have been discussed in the literature, see Deguillaume *et al.* (2013). The choice of a square-root dependence of volatilities on yield levels seems to be quite robust over different maturity and interest rate regimes, but for small rates — as we face it for the Swiss Franc (CHF) — linear dependence seems to be a good choice, too, see choice (3.8). In particular, these choices may provide $(\Upsilon_{t,m})_t$ that look stationary over different time to maturity choices $m \in \mathcal{M}$ and different yield levels.

Lemma 3.3. *Under Model Assumptions 3.1, the random vector $\boldsymbol{\Upsilon}_t |_{\mathcal{F}_{t-\Delta}}$ has a d -dimensional conditional Gaussian distribution with the first two conditional moments given by*

$$\begin{aligned} \mathbb{E}_{t-\Delta}^*[\boldsymbol{\Upsilon}_t] &= \Delta \left[-\mathbf{Y}(t - \Delta, t) + \frac{1}{2} \text{sp}(\Sigma_\Lambda(\mathbf{Y}_{t-\Delta}^-)) \right], \\ \text{Cov}_{t-\Delta}^*(\boldsymbol{\Upsilon}_t) &= \Delta \Sigma_\Lambda(\mathbf{Y}_{t-\Delta}^-). \end{aligned}$$

3.2. Calibration procedure

In order to calibrate our model, we need to choose the volatility scaling factors $\zeta(\cdot)$ and we need to specify the return directions $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_d \in \mathbb{R}^d$ which provide matrix Λ . Since we are only interested in the law of the process, we do not need to specify the directions $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_d \in \mathbb{R}^d$ themselves but, due to Lemma 3.3, rather the covariance structure Σ_Λ .

Assume we have the following time series of observations $(\boldsymbol{\Upsilon}_t)_{t=\Delta, \dots, \Delta K}$, $(Y(t - \Delta, t))_{t=\Delta, \dots, \Delta(K+1)}$, and $(\mathbf{Y}_{t-\Delta}^-)_{t=\Delta, \dots, \Delta(K+1)}$. We use these observations to predict/approximate the random vector $\boldsymbol{\Upsilon}_{\Delta(K+1)}$ at time ΔK , see also Lemma 3.3. For $\mathbf{y} \in \mathbb{R}^d$, we define the matrices

$$\begin{aligned} C_{(K)} &= \frac{1}{\sqrt{K}} \left(\left[\zeta(\mathbf{Y}_{\Delta(k-1)}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k} \right]_j \right)_{j=1, \dots, d; k=1, \dots, K} \in \mathbb{R}^{d \times K}, \\ S_{(K)}(\mathbf{y}) &= \zeta(\mathbf{y}) C_{(K)} C'_{(K)} \zeta'(\mathbf{y}) \in \mathbb{R}^{d \times d}. \end{aligned}$$

Choose $t = \Delta(K + 1)$. Note that $C_{(K)}$ is $\mathcal{F}_{t-\Delta}$ -measurable. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we define the d -dimensional random vector

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_t(\mathbf{x}, \mathbf{y}) = -\Delta \mathbf{x} + \frac{1}{2} \text{sp} (S_{(K)}(\mathbf{y})) + \zeta(\mathbf{y}) C_{(K)} \mathbf{W}_t^*, \tag{3.4}$$

with \mathbf{W}_t^* being independent of $\mathcal{F}_{t-\Delta}$, \mathcal{F}_t -measurable, independent of $\boldsymbol{\varepsilon}_t^*$ and having a K -dimensional standard Gaussian distribution with independent components under \mathbb{P}^* .

Lemma 3.4. *The random vector $\boldsymbol{\kappa}_t |_{\mathcal{F}_{t-\Delta}}$ has a d -dimensional Gaussian distribution with the first two conditional moments given by*

$$\begin{aligned} \mathbb{E}_{t-\Delta}^* [\boldsymbol{\kappa}_t] &= -\Delta \mathbf{x} + \frac{1}{2} \text{sp} (S_{(K)}(\mathbf{y})), \\ \text{Cov}_{t-\Delta}^* (\boldsymbol{\kappa}_t) &= S_{(K)}(\mathbf{y}). \end{aligned}$$

Our aim is to show that the matrix $S_{(K)}(\mathbf{y})$ is an appropriate estimator for $\Delta \Sigma_\Delta(\mathbf{y})$ and then Lemmas 3.3 and 3.4 say that $\boldsymbol{\kappa}_t$ is an appropriate approximation in law to $\boldsymbol{\Upsilon}_t$, conditionally given $\mathcal{F}_{t-\Delta}$.

Remark. The random vector $\boldsymbol{\kappa}_t$ can be seen as a filtered historical simulation of $\boldsymbol{\Upsilon}$, where \mathbf{W}_t^* re-simulates the K observations which are appropriately historically scaled through ζ , see Barone-Adesi *et al.* (1998). One may raise the question about the curse of dimensionality in (3.4). However, note that choice (3.4) does not really specify a high-dimensional model, but rather means to generate increments with covariance structure equal to the ones of $(\boldsymbol{\Upsilon}_t)_{t=\Delta, \dots, \Delta K}$ — up to proper re-scaling.

We assume for the moment $\mathbb{P} = \mathbb{P}^*$ and we calculate the expected value of $S_{(K)}(\mathbf{y})$ under \mathbb{P}^* to understand the estimator’s bias and consistency properties. Choose $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$ and define the function

$$f_\Delta(\mathbf{z}, \mathbf{y}) = \zeta(\mathbf{y})^{-1} \left[-\mathbf{z} + \frac{1}{2} \text{sp} (\Sigma_\Delta(\mathbf{y})) \right] \left[-\mathbf{z} + \frac{1}{2} \text{sp} (\Sigma_\Delta(\mathbf{y})) \right]' (\zeta(\mathbf{y})^{-1})'.$$

Note that this function does *not* depend on the grid size Δ . Lemma 3.3 then implies that

$$f_\Delta(\mathbf{Y}(t - \Delta, t), \mathbf{Y}_{t-\Delta}^-) = \Delta^{-2} \zeta(\mathbf{Y}_{t-\Delta}^-)^{-1} \mathbb{E}_{t-\Delta}^* [\boldsymbol{\Upsilon}_t] \mathbb{E}_{t-\Delta}^* [\boldsymbol{\Upsilon}_t]' (\zeta(\mathbf{Y}_{t-\Delta}^-)^{-1})', \tag{3.5}$$

where the left-hand side only depends on Δ through the fact that the yield curve $\mathbf{Y}_{t-\Delta}$ is observed at time $t - \Delta$. The proof of the next theorem is provided in the appendix.

Theorem 3.5. *Under Model Assumptions 3.1, we obtain for all $K \in \mathbb{N}$ and $\mathbf{y} \in \mathbb{R}^d$*

$$\begin{aligned} & \mathbb{E}_0^* [S_{(K)}(\mathbf{y})] \\ &= \Delta \Sigma_\Delta(\mathbf{y}) + \Delta^2 \varsigma(\mathbf{y}) \left(\frac{1}{K} \sum_{k=1}^K \mathbb{E}_0^* [f_\Delta(\mathbf{Y}(\Delta(k-1), \Delta k), \mathbf{Y}_{\Delta k-\Delta}^-)] \right) \varsigma(\mathbf{y})'. \end{aligned}$$

3.2.1. *Interpretation.* Using $S_{(K)}(\mathbf{y})$ as estimator for $\Delta \Sigma_\Delta(\mathbf{y})$ provides, under \mathbb{P}_0^* , a bias given by

$$\Delta^2 \varsigma(\mathbf{y}) \left(\frac{1}{K} \sum_{k=1}^K \mathbb{E}_0^* [f_\Delta(\mathbf{Y}(\Delta(k-1), \Delta k), \mathbf{Y}_{\Delta k-\Delta}^-)] \right) \varsigma(\mathbf{y})'.$$

If we choose $t = \Delta K$ fixed and assume that the term in the bracket is uniformly bounded for $\Delta \rightarrow 0$ then we see that

$$\mathbb{E}_0^* [S_{(K)}(\mathbf{y})] = \Delta \Sigma_\Delta(\mathbf{y}) + \Delta^2 \mathcal{O}(\mathbf{1}), \quad \text{for } \Delta \rightarrow 0. \tag{3.6}$$

That is, for small grid size Δ the second term becomes negligible, and we obtain an asymptotically unbiased estimate.

In the general situation $\mathbb{P} \neq \mathbb{P}^*$, the previous calculation can be performed under the assumption that $\boldsymbol{\varepsilon}_t$ under \mathbb{P} shares all the properties of $\boldsymbol{\varepsilon}_t^*$ under \mathbb{P}^* , except that we allow for a drift of order $\mathcal{O}(\Delta)$, i.e. it is multivariate Gaussian, identically distributed and independent of previous information. Then, the conclusions of the theorem remain unchanged. It is worth pointing out that only the bias changes under this change of measure.

The only term that still needs to be chosen is the invertible and linear map $\varsigma(\mathbf{y})$, i.e. the volatility scaling factors. For $\vartheta \geq 0$, we define the function

$$h : \mathbb{R}_+ \rightarrow \mathbb{R}_+, \quad y \mapsto h(y) = \vartheta^{-1/2} y 1_{\{y \leq \vartheta\}} + y^{1/2} 1_{\{y > \vartheta\}}. \tag{3.7}$$

At the first sight, this choice might look subjective. It is motivated by the empirical findings that interest rates often have a square-root scaling, but for small yields scalings look rather linear, we refer to Deguillaume *et al.* (2013). Note that function $h(\cdot)$ defined as above also guarantees that the processes do not explode for large volatilities and small grid sizes. For these reasons, we choose for the Swiss currency CHF scaling (3.7) with a threshold of $\vartheta = 2.5\%$. This choice is supported by the empirical findings for our data in Section 4.1. We set for $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$

$$\varsigma(\mathbf{y}) = \text{diag}(h(y_1), \dots, h(y_d)). \tag{3.8}$$

This choice implies

$$\begin{aligned} \varsigma(\mathbf{y}) C_{(K)} &= \frac{1}{\sqrt{K}} \varsigma(\mathbf{y}) \left([\varsigma(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \mathbf{Y}_{\Delta k}]_j \right)_{j=1,\dots,d; k=1,\dots,K} \\ &= \frac{1}{\sqrt{K}} \text{diag}(h(y_1), \dots, h(y_d)) \\ &\quad \times \left([\text{diag}(h(\mathbf{Y}_{\Delta k-\Delta}^-))^{-1} \mathbf{Y}_{\Delta k}]_j \right)_{j=1,\dots,d; k=1,\dots,K}. \end{aligned}$$

These assumptions now allow to directly analyze the bias term given in (3.6). Therefore, we need to evaluate the function f_Δ considered in Theorem 3.5. To this end, we need to know Σ_Δ , i.e. we obtain from Theorem 3.5 an implicit solution (quadratic form) that can be solved for Σ_Δ . We set $\mathbf{y} = \mathbf{1}$ and obtain from Theorem 3.5

$$\Delta^{-1} \mathbb{E}_0^* [S_{(K)}(\mathbf{1})] = \Sigma_\Delta(\mathbf{1}) + \Delta \left(\frac{1}{K} \sum_{k=1}^K \mathbb{E}_0^* [f_\Delta(\mathbf{Y}(\Delta(k-1), \Delta k), \mathbf{Y}_{\Delta k-\Delta}^-)] \right).$$

Note that $\Sigma_\Delta(\mathbf{y}) = \varsigma(\mathbf{y}) \Lambda \Lambda' \varsigma(\mathbf{y})$, thus under (3.8) its elements are given by $h(y_i)h(y_j)s_{ij}$, $i, j = 1, \dots, d$, where we have defined $\Lambda \Lambda' = (s_{ij})_{i,j=1,\dots,d}$. Let us first concentrate on the diagonal elements, i.e. $i = j$, and assume that time to maturity m_i corresponds to index i .

$$\begin{aligned} \Delta^{-1} (\mathbb{E}_0^* [S_{(K)}(\mathbf{1})])_{ii} &= s_{ii} + \frac{\Delta}{K} \sum_{k=1}^K \left(\mathbb{E}_0^* \left[\left(\frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i))} \right)^2 \right] \right. \\ &\quad \left. + \frac{1}{4} \mathbb{E}_0^* [h(Y(\Delta(k-1), \Delta k + m_i))^2] s_{ii}^2 \right. \\ &\quad \left. - \mathbb{E}_0^* [Y(\Delta(k-1), \Delta k)] s_{ii} \right). \end{aligned}$$

This is a quadratic equation that can be solved for s_{ii} . Define

$$a_i = \frac{\Delta}{4K} \sum_{k=1}^K \mathbb{E}_0^* [h(Y(\Delta(k-1), \Delta k + m_i))^2], \tag{3.9}$$

$$b = 1 - \frac{\Delta}{K} \sum_{k=1}^K \mathbb{E}_0^* [Y(\Delta(k-1), \Delta k)], \tag{3.10}$$

$$\begin{aligned} c_i &= -\Delta^{-1} (\mathbb{E}_0^* [S_{(K)}(\mathbf{1})])_{ii} \\ &\quad + \frac{\Delta}{K} \sum_{k=1}^K \mathbb{E}_0^* \left[\left(\frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i))} \right)^2 \right], \end{aligned} \tag{3.11}$$

then we have $a_i s_{ii}^2 + b s_{ii} + c_i = 0$ which provides solution

$$s_{ii} = \frac{-b + \sqrt{b^2 - 4a_i c_i}}{2a_i}. \tag{3.12}$$

Thus, the bias terms of the diagonal elements are given by

$$\beta_{ii} = \Delta^{-1} (\mathbb{E}_0^* [S_{(K)}(\mathbf{1})])_{ii} - s_{ii},$$

which we are going to analyze below for the different times to maturity $m_i \in \mathcal{M}$. For the off-diagonals $i \neq j$ and the corresponding maturities m_i and m_j , we obtain

$$\begin{aligned} &\Delta^{-1} (\mathbb{E}_0^* [S_{(K)}(\mathbf{1})])_{ij} \\ &= s_{ij} + \frac{\Delta}{K} \sum_{k=1}^K \left(\mathbb{E}_0^* \left[\frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i))} \frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_j))} \right] \right. \\ &\quad + \frac{1}{4} \mathbb{E}_0^* [h(Y(\Delta(k-1), \Delta k + m_i))h(Y(\Delta(k-1), \Delta k + m_j))] s_{ii} s_{jj} \\ &\quad - \frac{1}{2} \mathbb{E}_0^* \left[Y(\Delta(k-1), \Delta k) \frac{h(Y(\Delta(k-1), \Delta k + m_i))}{h(Y(\Delta(k-1), \Delta k + m_j))} \right] s_{ii} \\ &\quad \left. - \frac{1}{2} \mathbb{E}_0^* \left[Y(\Delta(k-1), \Delta k) \frac{h(Y(\Delta(k-1), \Delta k + m_j))}{h(Y(\Delta(k-1), \Delta k + m_i))} \right] s_{jj} \right). \tag{3.13} \end{aligned}$$

Equation (3.13) can easily be solved for s_{ij} for given s_{ii} and s_{jj} , see also (3.12). In the next section, we apply this calibration to real data and we determine the corresponding bias terms.

4. CALIBRATION TO REAL DATA

4.1. Model calibration

We assume that $\mathbb{P} = \mathbb{P}^*$, i.e. we set the market price of risk identically equal to 0. As a consequence, we can directly work on the observed data. The choice of the drift term will be discussed later.

The first difficulty is the choice of the data. The reason therefore is that risk-free ZCBs do not exist and, thus, the risk-free yield curve needs to be estimated from available data that has different spreads such as a credit spread, a liquidity spread, a long-term premium, etc. We calibrate the model to the Swiss currency CHF. For short times to maturity (below one year), one typically chooses either the LIBOR¹ (London InterBank Offered Rate) or the SAR¹ (Swiss Average Rate), see Jordan (2009), as (almost) risk-free financial instruments. The LIBOR is the rate at which highly credited banks borrow and lend money at the

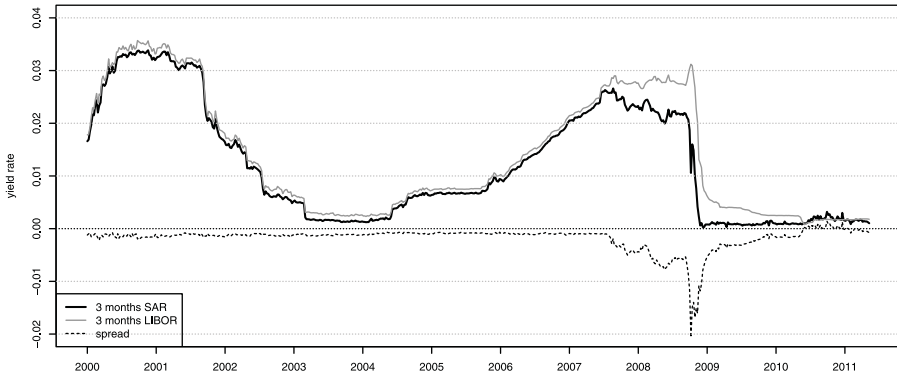


FIGURE 1: Yield curve time series three months SAR and three months CHF LIBOR from 01/2000 until 05/2011. The spread gives the difference between these two time series.

inter-bank market. The SAR is a rate determined by the Swiss National Bank (SNB) at which highly credited institutions borrow and lend collateral money. We display the yields of these two financial time series for instruments of a time to maturity of three months, see Figure 1. We see that the SAR yield typically lies below the LIBOR yield (due to collateralization). Therefore, we consider the SAR to be less risky and we choose it as approximation to a risk-free financial instrument with short times to maturity.

For long times to maturity (above one year), the ZCB yield curve is extracted either from government bond yields² (of sufficiently highly rated countries) or from swap rates³, the former is described in Müller (2002) for the Swiss currency CHF market. In Figure 2, we give the time series of the Swiss government bond extracted yield and the CHF swap yield both for a time to maturity of five years. We see that the rates of the Swiss government bond extracted yields are below the CHF swap rates (due to lower credit risk and maybe an illiquidity premium coming from a high demand) and therefore we choose the Swiss government bond extracted yield curve as approximation to the risk-free yield curve data for long times to maturity.

Concluding, for our numerical analysis below we use the SAR for short times to maturity $m \in \{1/52, 1/26, 1/12, 1/4\}$ and Swiss government bond extracted yield rates for long times to maturity $m \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30\}$. Let us give some remarks concerning the subsequent analysis:

- Short terms and long terms data (SAR and Swiss government bond extracted yields) are not completely compatible. This gives some difficulties in the calibration process. We will see this in the correlation matrix in Figure 10, below.
- As also mentioned in the introduction, the risk-free CHF yield curve is not directly observable at the financial market, but itself estimated using highly credited bonds. Here, we do not deal with the problem of extracting the risk-free yield curve from market information but we use the one extracted by the

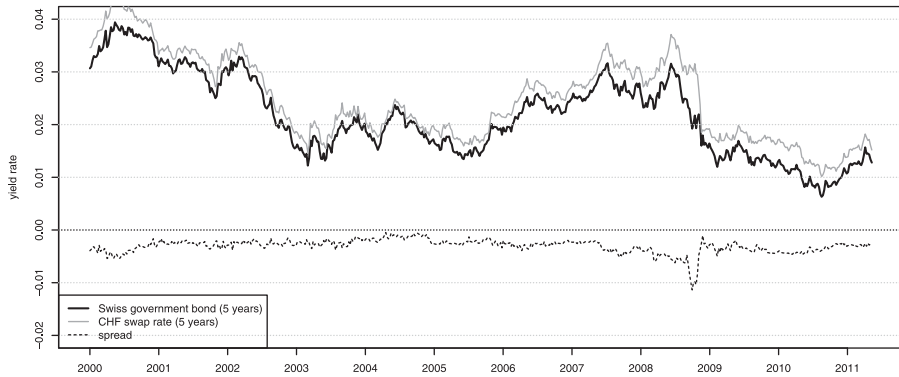


FIGURE 2: Yield curve time series Swiss government bond and CHF swap rate both for time to maturity $m = 5$ years from 01/2000 until 05/2011. The spread gives the difference between these two time series (roughly 30 basis points).

SNB. This is done by fitting the Svensson (1994, 1995) curve to government bonds and local state bonds, for more details we refer to Müller (2002).

- Usually, the fitted Svensson curve does not directly match the government bond yields (in fact, in theory, it should lie slightly below the government bond yields because it should be risk-free). If this is an issue, for instance for pricing cash flows today, Hull and White (1990, 1994) extended model versions should be used that manage to match prices exactly. In prediction models, we are generally less strict about exact match of today’s yield curve because we are mainly interested into the predictive power of future yield curves.
- The Svensson curve is based on six parameters and time series of these parameters can be rather volatile. As described in Cairns (1998), there can be so-called “catastrophic” jumps in the parameter processes (since least-squares optimization may jump non-continuously from one (local) optimum to another, see Figures 1 and 2 in Cairns (1998) for more explanation). This is also observed for the Swiss government bond extracted yields. However, this does not influence our prediction model as long as the resulting yield curves behave sufficiently smooth, because we choose the yield curves themselves as input state variables (and not the Svensson parameters).

We choose grid size $\Delta = 1/52$ (i.e. a weekly time grid) and then we calculate \mathbf{Y}_t for our observations. Note that we cannot directly calculate $\Upsilon_{t,m} = m Y(t, t + m) - (m + \Delta) Y(t - \Delta, t + m)$ for all $m \in \mathcal{M}$ because we have only a limited set of observed times to maturity. Therefore, we make the following interpolation: assume $m + \Delta \in (m, \tilde{m}]$ for $m, \tilde{m} \in \mathcal{M}$, then approximate

$$Y(t - \Delta, t + m) \approx \frac{\tilde{m} - (m + \Delta)}{\tilde{m} - m} Y(t - \Delta, t + m - \Delta) + \frac{\Delta}{\tilde{m} - m} Y(t - \Delta, t + \tilde{m} - \Delta).$$

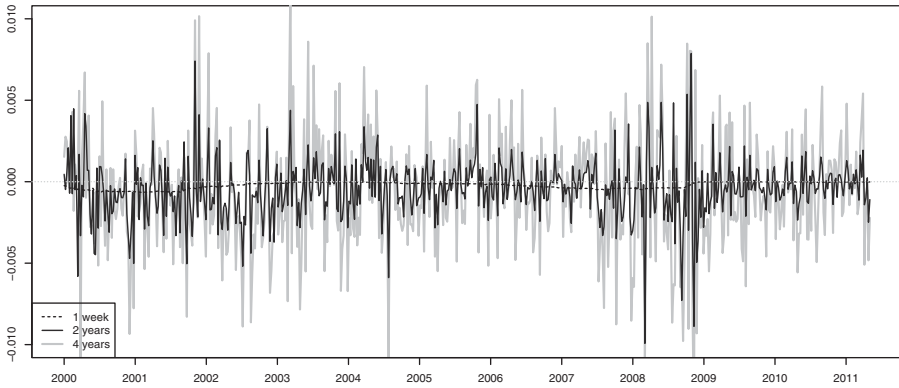


FIGURE 3: Time series Υ_t for $t \in \{01/2000, \dots, 05/2011\}$ and times to maturity selection $m \in \{1/52, 2, 4\}$ on a weekly grid $\Delta = 1/52$.

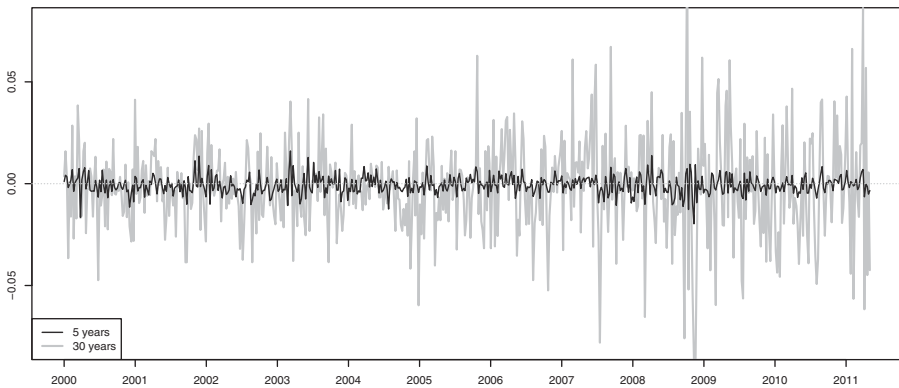


FIGURE 4: Time series Υ_t for $t \in \{01/2000, \dots, 05/2011\}$ and times to maturity selection $m \in \{5, 30\}$ on a weekly grid $\Delta = 1/52$.

In fact, this describes a linear interpolation. For illustrative purposes, we choose here the simplest reasonable interpolation, if this is too crude in real applications splines should be used for interpolation. In Figures 3 and 4, we give the time series of the estimated $(\Upsilon_t)_t$. We see that volatility is increasing in time to maturity due to scaling with time to maturity. Using (3.8), we calculate

$$\sqrt{K} C_{(K)} = \left([\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \Upsilon_{\Delta k}]_j \right)_{j=1, \dots, d; k=1, \dots, K} \in \mathbb{R}^{d \times K}$$

for our observations. In Figures 5 and 6, we plot the time series $\Upsilon_{t,m}$ and $[\sqrt{K} C_{(K)}]_m = \Upsilon_{t,m}/h(Y(t-\Delta, t+m))$ for illustrative purposes only for times to maturity $m = 1/52$ and $m = 5$. We observe that the scaling $\zeta(\mathbf{Y}_{t-\Delta}^-)^{-1}$ gives more stationarity for short times to maturity, however in financial stress periods it substantially increases the volatility of the observations, see Figure 5.

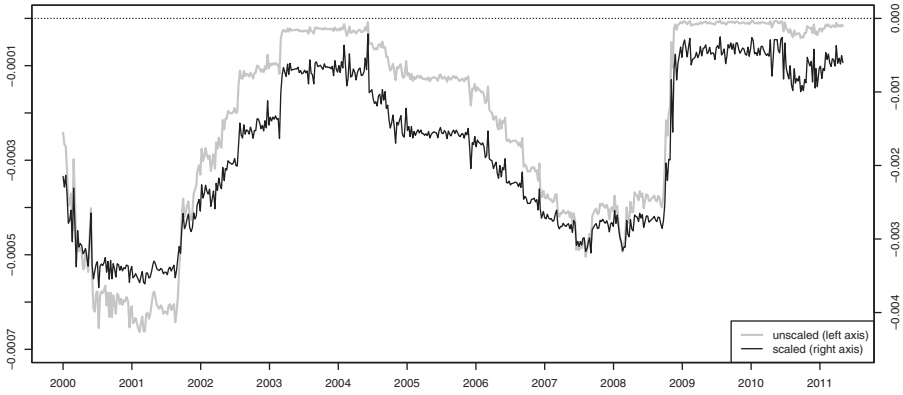


FIGURE 5: Time series $\Upsilon_{t,m}$ (unscaled) and $[\sqrt{K} C_{(K)}]_m = \Upsilon_{t,m}/h(Y(t - \Delta, t + m))$ (scaled) for maturity $m = 1/52$ and $t \in \{01/2000, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

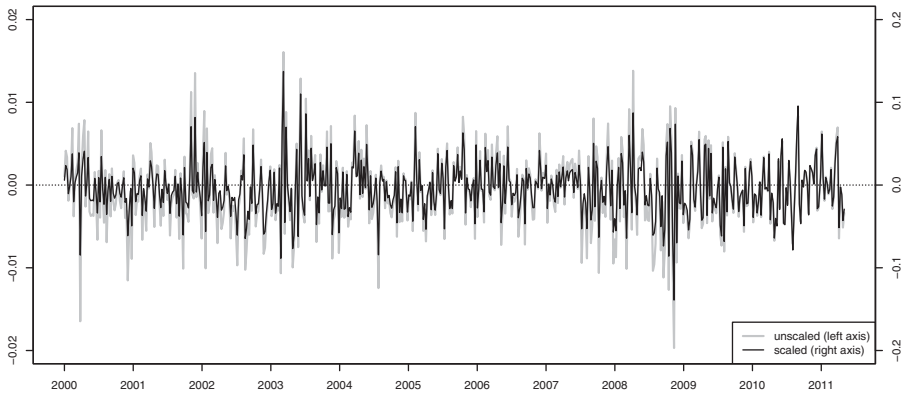


FIGURE 6: Time series $\Upsilon_{t,m}$ (unscaled) and $[\sqrt{K} C_{(K)}]_m = \Upsilon_{t,m}/h(Y(t - \Delta, t + m))$ (scaled) for maturity $m = 5$ and $t \in \{01/2000, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

Moreover, we see that (not surprisingly) the model has difficulties to cope with actions of the SNB because they are beyond our model features. This modeling difficulty holds true in particular for short times to maturity of the Swiss currency CHF. It is at this stage where fine-tuning of function $h(\cdot)$ given in (3.7) can be done. In general, it should be chosen such that volatility is balanced between low and high levels of yield rates. Next figures show that scaling (3.7) is needed for short and long times to maturity. We calculate the observed matrix

$$(\hat{s}_{ij}^{\text{bias}}(K))_{i,j=1,\dots,d} = \Delta^{-1} S_{(K)}(\mathbf{1})$$

as a function of the number of observations K (we set $\mathbf{1} = (1, \dots, 1)' \in \mathbb{R}^d$). Moreover, we calculate the bias correction terms given in (3.9)–(3.11) where we simply replace the expected values on the right-hand sides by the observations. Formulas (3.12), (3.13) then provide the estimates \hat{s}_{ij} as a function of

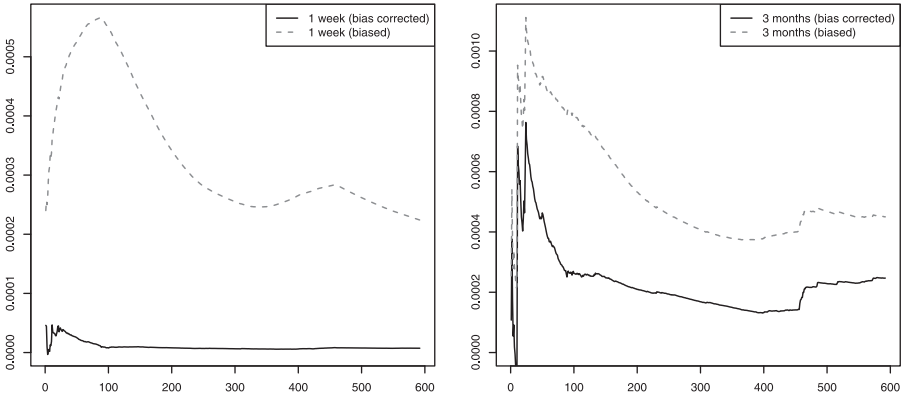


FIGURE 7: Time series $\hat{s}_{ij}^{bias}(K)$ and $\hat{s}_{ij}(K)$, $K = 1, \dots, 600$, on a weekly grid $\Delta = 1/52$ for (lhs) time to maturity $m_i = 1/52$ (1 week) and (rhs) time to maturity $m_i = 1/4$ (three months).

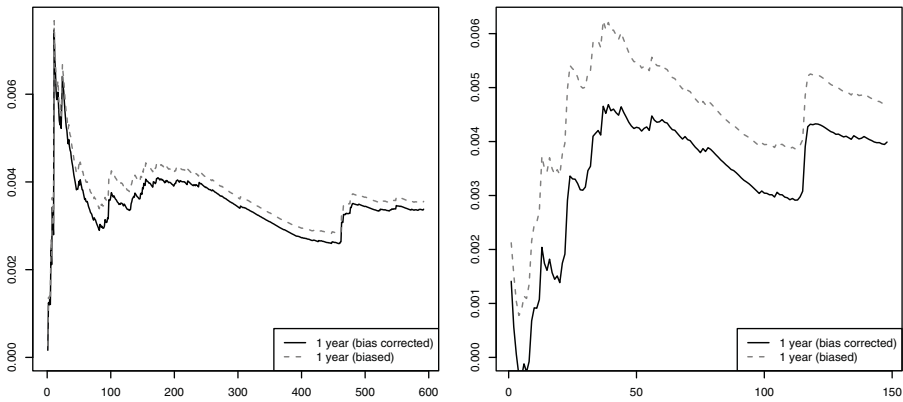


FIGURE 8: Time series $\hat{s}_{ij}^{bias}(K)$ and $\hat{s}_{ij}(K)$, $K = 1, \dots, 600$, for time to maturity $m_i = 1$ year on a (lhs) weekly grid $\Delta = 1/52$ and (rhs) grid $\Delta = 1/13$.

the number of observations K . The bias correction term is estimated by

$$\hat{\beta}_{ij}(K) = \hat{s}_{ij}^{bias}(K) - \hat{s}_{ij}(K).$$

We expect that for short times to maturity the bias correction term is larger due to bigger drifts. The results for selected times to maturity $m \in \{1/52, 1/4, 1, 5, 20\}$ are presented in Figures 7–9. Let us comment on these figures:

- Times to maturity in set $\mathcal{M}_1 = \{1/52, 1/26, 1/12\}$ look similar to $m = 1/52$ (see Figure 7 (lhs)); $\mathcal{M}_2 = \{1/4\}$ corresponds to Figure 7 (rhs); times to maturity in set $\mathcal{M}_3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15\}$ look similar to $m = 1, 5$ (Figures 8 (lhs) and 9 (lhs)); times to maturity $m \in \mathcal{M}_4 = \{20, 30\}$ look similar to $m = 20$ (Figure 9 (rhs)).

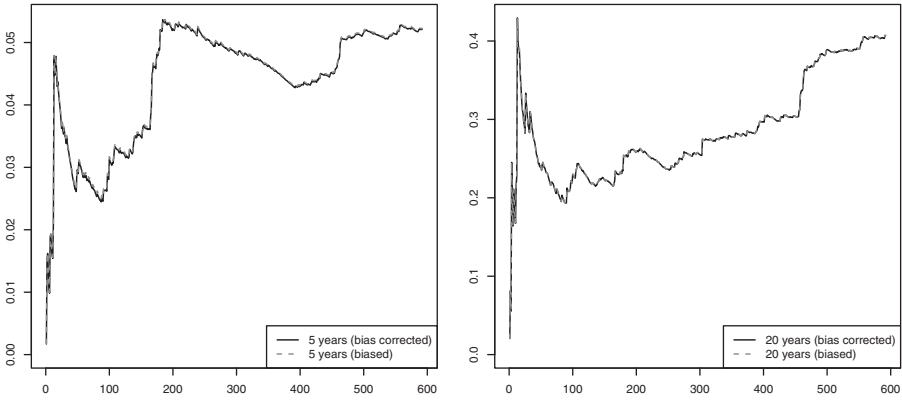


FIGURE 9: Time series $\widehat{s}_{ij}^{bias}(K)$ and $\widehat{s}_{ii}(K)$, $K = 1, \dots, 600$, on a weekly grid $\Delta = 1/52$ for (lhs) time to maturity $m_i = 5$ years and (rhs) time to maturity $m_i = 20$ years; note that the biased and the biased corrected graphs are almost identical.

- Times to maturity in $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$ seem to have converged, but the convergence picture is slightly distorted by the last financial crisis, where volatilities relative to yields have substantially increased, this is also seen in Figure 5. One might therefore ask whether during financial crises we should apply a different scaling (similar to regime switching models). For \mathcal{M}_4 , the convergence picture suggest that we should probably study longer time series (or scaling should be done differently). Concluding, these convergence pictures supports the choice of the function h in (3.7). Only long times to maturity $m \in \mathcal{M}_4$ might suggest a different scaling.
- For times to maturity in $\mathcal{M}_3 \cup \mathcal{M}_4$, we observe that the bias term given in (3.6) is negligible, see Figure 9, that is, $\Delta = 1/52$ is sufficiently small for times to maturity $m \geq 1$. For times to maturity in $\mathcal{M}_1 \cup \mathcal{M}_2$, it is however essential that we do a bias correction, see Figure 7. This comes from the fact that for small times to maturity the bias term is driven by \mathbf{z} in $f_\Lambda(\mathbf{z}, \mathbf{y})$ which then is of similar order as s_{ij} .

In Table 1, we present the resulting estimated matrix $\widehat{\Sigma}_\Lambda(\mathbf{1}) = (\widehat{s}_{ij}(K))_{i,j=1,\dots,d}$ which is based on all observations in $\{01/2000, \dots, 05/2011\}$. We observe that the diagonal $\widehat{s}_{ii}(K)$ is an increasing function in the time to maturity m_i . Therefore, in order to further analyze this matrix, we normalize it as follows (as a correlation matrix)

$$\widehat{\mathbf{E}} = (\widehat{\rho}_{ij})_{i,j=1,\dots,d} = \left(\frac{\widehat{s}_{ij}(K)}{\sqrt{\widehat{s}_{ii}(K)}\sqrt{\widehat{s}_{jj}(K)}} \right)_{i,j=1,\dots,d}.$$

Now all entries $\widehat{\rho}_{ij}$ live on the same scale and the result is presented in Figure 10 (lhs). We observe two different structures, one for times to maturity less than 1 year, i.e. $m \in \widetilde{\mathcal{M}}_1 = \mathcal{M}_1 \cup \mathcal{M}_2$, and one for times to maturity

TABLE 1

ESTIMATED MATRIX $\widehat{\Sigma}_\Lambda(\mathbf{1}) = (\widehat{\sigma}_{ij}(K))_{i,j=1,\dots,d}$ BASED ON ALL OBSERVATIONS IN $\{01/2000, \dots, 05/2011\}$ ON A WEEKLY GRID $\Delta = 1/52$.

	1 Week	2 Weeks	1M	3M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	15Y	20Y	30Y
1 Week	0.0000	0.0001	0.0001	0.0002	0.0004	0.0005	0.0005	0.0006	0.0006	0.0007	0.0007	0.0007	0.0007	0.0007	0.0009	0.0011	0.0017
2 Weeks	0.0001	0.0001	0.0002	0.0003	0.0005	0.0006	0.0006	0.0007	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0011	0.0014	0.0023
1 Month	0.0001	0.0002	0.0004	0.0005	0.0005	0.0007	0.0008	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0010	0.0016	0.0028
3 Months	0.0002	0.0003	0.0005	0.0028	0.0019	0.0024	0.0021	0.0018	0.0017	0.0016	0.0017	0.0017	0.0018	0.0019	0.0025	0.0042	0.0064
1 Year	0.0004	0.0005	0.0005	0.0019	0.0142	0.0164	0.0158	0.0161	0.0168	0.0173	0.0176	0.0176	0.0175	0.0173	0.0173	0.0208	0.0371
2 Years	0.0005	0.0006	0.0007	0.0024	0.0164	0.0309	0.0361	0.0373	0.0376	0.0380	0.0388	0.0397	0.0410	0.0423	0.0500	0.0590	0.0658
3 Years	0.0005	0.0006	0.0008	0.0021	0.0158	0.0361	0.0481	0.0535	0.0562	0.0581	0.0600	0.0619	0.0641	0.0663	0.0779	0.0907	0.1019
4 Years	0.0006	0.0007	0.0009	0.0018	0.0161	0.0373	0.0535	0.0632	0.0693	0.0735	0.0771	0.0801	0.0830	0.0857	0.0990	0.1126	0.1327
5 Years	0.0006	0.0008	0.0009	0.0017	0.0168	0.0376	0.0562	0.0693	0.0787	0.0851	0.0904	0.0947	0.0986	0.1021	0.1173	0.1310	0.1581
6 Years	0.0007	0.0008	0.0009	0.0016	0.0173	0.0380	0.0581	0.0735	0.0851	0.0936	0.1007	0.1065	0.1116	0.1161	0.1342	0.1485	0.1786
7 Years	0.0007	0.0008	0.0009	0.0017	0.0176	0.0388	0.0600	0.0771	0.0904	0.1007	0.1095	0.1169	0.1234	0.1292	0.1520	0.1674	0.1969
8 Years	0.0007	0.0009	0.0009	0.0017	0.0176	0.0397	0.0619	0.0801	0.0947	0.1065	0.1169	0.1259	0.1340	0.1413	0.1700	0.1871	0.2135
9 Years	0.0007	0.0009	0.0009	0.0018	0.0175	0.0410	0.0641	0.0830	0.0986	0.1116	0.1234	0.1340	0.1438	0.1528	0.1883	0.2078	0.2297
10 Years	0.0007	0.0009	0.0009	0.0019	0.0173	0.0423	0.0663	0.0857	0.1021	0.1161	0.1292	0.1413	0.1528	0.1635	0.2064	0.2289	0.2462
15 Years	0.0009	0.0011	0.0010	0.0025	0.0173	0.0500	0.0779	0.0990	0.1173	0.1342	0.1520	0.1700	0.1883	0.2064	0.2869	0.3320	0.3498
20 Years	0.0011	0.0014	0.0016	0.0042	0.0208	0.0590	0.0907	0.1126	0.1310	0.1485	0.1674	0.1871	0.2078	0.2289	0.3320	0.4247	0.5215
30 Years	0.0017	0.0023	0.0028	0.0064	0.0371	0.0658	0.1019	0.1327	0.1581	0.1786	0.1969	0.2135	0.2297	0.2462	0.3498	0.5215	0.9860

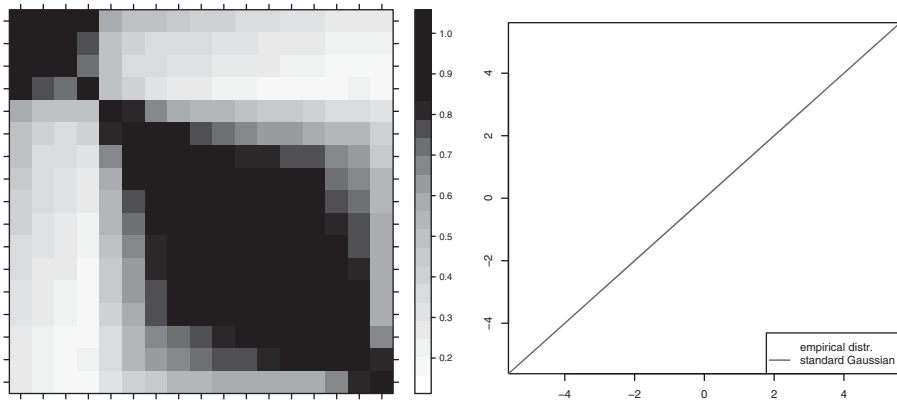


FIGURE 10: (lhs) Estimated correlation matrix $\hat{\Xi} = (\hat{\rho}_{ij})_{i,j=1,\dots,d}$ from all observations in $\{01/2000, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$; (rhs) Q-Q-plot of the residuals $(z_i^*)_i$ against the standard Gaussian distribution for weekly grid $\Delta = 1/52$.

$m \in \tilde{\mathcal{M}}_2 = \mathcal{M}_3 \cup \mathcal{M}_4$. The former times to maturity $m \in \tilde{\mathcal{M}}_1$ were modeled using the observations from the SAR, the latter $m \in \tilde{\mathcal{M}}_2$ with observations from the Swiss government bond yields. We observe high correlations within $\tilde{\mathcal{M}}_1$ and $\tilde{\mathcal{M}}_2$, but a separation between these two sets of times to maturity. This separation shows that the two data sets are not completely compatible which gives some “additional independence” (diversification) between $\tilde{\mathcal{M}}_1$ and $\tilde{\mathcal{M}}_2$. If we calculate the eigenvalues of $\hat{\Xi}$ we observe that the first three eigenvalues explain about 90% of the total cross-sectional volatility and the first five eigenvalues explain about 96% (we have a $d = 17$ dimensional space). Thus, a PCA says that we should at least choose a five-factor model to simultaneously model all SAR and Swiss government bond implied yields in \mathcal{M} . These are more factors than typically stated in the literature (see Section 4.1 in Brigo and Mercurio (2006)). The reason therefore is that the short end $\tilde{\mathcal{M}}_1$ and the long end $\tilde{\mathcal{M}}_2$ of the estimated yield curve behave more independently due to different sources of data (see Figure 10 (lhs)) and one might question this composition $\tilde{\mathcal{M}}_1$ and $\tilde{\mathcal{M}}_2$. If we restrict this PCA to $\tilde{\mathcal{M}}_2$, we find the classical result that a three-factor model explains 95% of the observed cross-sectional volatility.

In the next step, we analyze the assumption of the independence of $\Sigma_\Delta(\mathbf{1}) = \Lambda \Lambda' = (s_{ij})_{i,j=1,\dots,d}$ from the grid size Δ . Similar to the analysis above, we estimate $\Sigma_\Delta(\mathbf{1})$ for the grid sizes $\Delta = 1/52, 1/26, 1/13, 1/4$ (weekly, bi-weekly, 4-weekly, quarterly grid size). The first observation is that the bias increases with increasing Δ (for illustrative purposes, one should compare in Figure 8 the left-hand side with $m = 1$ and $\Delta = 1/52$ and right-hand side with $m = 1$ and $\Delta = 1/13$). Of course, this is exactly the result expected.

In Table 2, we give the differences between the estimated matrices $\hat{\Sigma}_\Delta(\mathbf{1}) = (\hat{s}_{ij}(K))_{i,j=1,\dots,d}$ on the weekly grid $\Delta = 1/52$ versus the estimates on a quarterly grid $\Delta = 1/4$ (relative to the estimated values on the quarterly grid). Of course, we can only display these differences for times to maturity $m \in \mathcal{M}_2 \cup \tilde{\mathcal{M}}_2$

TABLE 2

ESTIMATED MATRICES $\widehat{\Sigma}_\Delta(\mathbf{1}) = (\widehat{\beta}_{ij}(K))_{i,j=1,\dots,d}$ BASED ON ALL OBSERVATIONS IN $\{01/2000, \dots, 05/2011\}$. THE TABLE SHOWS THE DIFFERENCES BETWEEN THE ESTIMATES ON A WEEKLY GRID $\Delta = 1/52$ VERSUS THE ESTIMATES ON A QUARTERLY GRID $\Delta = 1/4$ (RELATIVE TO THE ESTIMATED VALUES ON THE QUARTERLY GRID).

	3M	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	15Y	20Y	30Y
3 Months	31%	70%	72%	78%	82%	84%	85%	85%	85%	85%	84%	83%	76%	76%
1 Year	70%	2%	7%	16%	19%	21%	23%	24%	27%	29%	30%	38%	36%	24%
2 Years	72%	7%	-6%	-1%	6%	11%	13%	14%	15%	16%	15%	15%	13%	23%
3 Years	78%	16%	-1%	0%	2%	4%	6%	7%	7%	8%	7%	8%	7%	15%
4 Years	82%	19%	6%	2%	1%	1%	1%	2%	2%	3%	3%	6%	7%	11%
5 Years	84%	21%	11%	4%	1%	-1%	-1%	-1%	-1%	0%	1%	5%	8%	11%
6 Years	85%	23%	13%	6%	1%	-1%	-2%	-2%	-2%	-1%	0%	5%	9%	12%
7 Years	85%	24%	14%	7%	2%	-1%	-2%	-3%	-2%	-1%	-1%	4%	9%	14%
8 Years	85%	27%	15%	7%	2%	-1%	-2%	-2%	-2%	-1%	-1%	3%	9%	16%
9 Years	85%	29%	16%	8%	3%	0%	-1%	-1%	-1%	-1%	-1%	2%	8%	18%
10 Years	84%	30%	15%	7%	3%	1%	0%	-1%	-1%	-1%	-1%	1%	7%	20%
15 Years	83%	38%	15%	8%	6%	5%	5%	4%	3%	2%	1%	0%	5%	20%
20 Years	76%	36%	13%	7%	7%	8%	9%	9%	9%	8%	7%	5%	3%	7%
30 Years	76%	24%	23%	15%	11%	11%	12%	14%	16%	18%	20%	20%	7%	-24%

because in the latter model the times to maturity in \mathcal{M}_1 do not exist. We observe rather small differences within $\widetilde{\mathcal{M}}_2$ which supports the independence assumption from the choice of Δ within the Swiss government bond extracted yields. For the SAR in \mathcal{M}_2 , this picture does not entirely hold true which has also to do with the fact that the model does not completely fit to the data, see Figure 7 (rhs). The pictures for $\Delta = 1/26, 1/13$ are quite similar which justifies our independence choice.

Conclusions 4.1

We conclude that the independence assumption of $\Sigma_\Delta(\mathbf{1})$ from Δ is not violated by our observations and that the bias terms $\widehat{\beta}_{i,j}(K)$ are negligible for maturities $m_i, m_j \in \widetilde{\mathcal{M}}_2$ and time grids $\Delta = 1/52, 1/16, 1/13$, therefore we can directly work with model (3.4) to predict future yields for times to maturity in $\widetilde{\mathcal{M}}_2$.

4.2. Out-of-sample back-testing and market price of risk

In this subsection, we back-test our model against the observations. We therefore choose a fixed-term annuity portfolio with nominal payments of size 1 at times to maturity dates $m \in \mathcal{M}_3$. The present value of this annuity portfolio at

time t is given by

$$\begin{aligned} \pi_t &= \sum_{m \in \mathcal{M}_3} P(t, t + m) = \sum_{m \in \mathcal{M}_3} \exp \{-m Y(t, t + m)\} \\ &\approx \sum_{m \in \mathcal{M}_3} 1 - m Y(t, t + m) \stackrel{\text{def.}}{=} \tilde{\pi}_t. \end{aligned}$$

Our (out-of-sample) back-testing setup is such that we try to predict $\tilde{\pi}_t$ based on observations $\mathcal{F}_{t-\Delta}$ and then (one period later) we compare this forecast with the realization of $\tilde{\pi}_t$. In view of Conclusions 4.1, we directly work with $C_{(K)}$ for small time grids Δ (for $t = \Delta(K + 1)$). Moreover, the Taylor approximation $\tilde{\pi}_t$ to π_t is used in order to avoid (time-consuming) simulations. Here, a first order Taylor expansion is sufficient since the portfolio’s variance will be — due to high positive correlation, see Figure 10 (lhs) — quite large in comparison to possible second order — drift like — correction terms. Remark that such an approximation does not work for short–long portfolios.

For the following approximation (under \mathbb{P}^*)

$$\Upsilon_t |_{\mathcal{F}_{t-\Delta}} \stackrel{(d)}{\approx} \kappa_t(\mathbf{Y}(t - \Delta, t), \mathbf{Y}_{t-\Delta}^-) |_{\mathcal{F}_{t-\Delta}},$$

we obtain an approximate forecast to $\tilde{\pi}_t$ given by (denote the cardinality of \mathcal{M}_3 by d_3)

$$\begin{aligned} \tilde{\tilde{\pi}}_t |_{\mathcal{F}_{t-\Delta}} &= d_3 - \sum_{m \in \mathcal{M}_3} (m + \Delta) Y(t - \Delta, t + m) + d_3 \Delta Y(t - \Delta, t) \\ &\quad - \frac{1}{2} \mathbf{1}'_{\mathcal{M}_3} \text{sp}(\mathcal{S}_{(K)}(\mathbf{Y}_{t-\Delta}^-)) - \mathbf{1}'_{\mathcal{M}_3} \mathcal{S}(\mathbf{Y}_{t-\Delta}^-) C_{(k)} \mathbf{W}_t^* |_{\mathcal{F}_{t-\Delta}}, \quad (4.1) \end{aligned}$$

where $\mathbf{1}_{\mathcal{M}_3} = (1_{\{1 \in \mathcal{M}_3\}}, \dots, 1_{\{d \in \mathcal{M}_3\}})' \in \mathbb{R}^d$. Thus, the conditional distribution of $\tilde{\tilde{\pi}}_t$ under \mathbb{P}^* , given $\mathcal{F}_{t-\Delta}$, is a Gaussian distribution with conditional mean and conditional variance given by

$$\begin{aligned} \mu_{t-\Delta}^* &= d_3 - \sum_{m \in \mathcal{M}_3} (m + \Delta) Y(t - \Delta, t + m) + d_3 \Delta Y(t - \Delta, t) \\ &\quad - \frac{1}{2} \mathbf{1}'_{\mathcal{M}_3} \text{sp}(\mathcal{S}_{(K)}(\mathbf{Y}_{t-\Delta}^-)), \\ \tau_{t-\Delta}^2 &= \mathbf{1}'_{\mathcal{M}_3} \mathcal{S}_{(K)}(\mathbf{Y}_{t-\Delta}^-) \mathbf{1}_{\mathcal{M}_3}. \end{aligned}$$

We successively calculate these conditional moments for $t \in \{01/2005, \dots, 05/2011\}$ based on the σ -fields $\mathcal{F}_{t-\Delta}$ generated by the data in $\{01/2000, \dots, t - \Delta\}$, for $\Delta = 1/52, 1/12$ (weekly and monthly grid). From these, we can

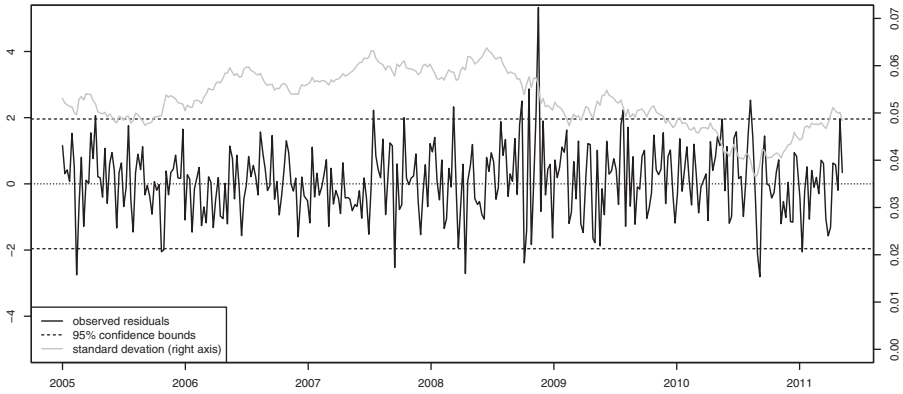


FIGURE 11: Time series of residuals z_t^* for $t \in \{01/2005, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{t-\Delta}$.

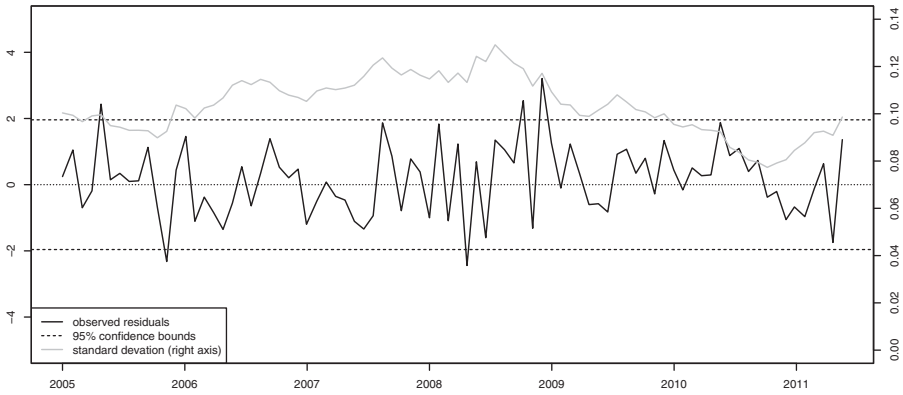


FIGURE 12: Time series of residuals z_t^* for $t \in \{01/2005, \dots, 05/2011\}$ on a monthly grid $\Delta = 1/12$. The axis on the right-hand side displays the time series of $\tau_{t-\Delta}$.

calculate the observable residuals

$$z_t^* = \frac{\tilde{\pi}_t - \mu_{t-\Delta}^*}{\tau_{t-\Delta}},$$

which provides the out-of-sample back-test. The sequence of these observable residuals should approximately look like an i.i.d. standard Gaussian distributed sequence. The result for $\Delta = 1/52$ is given in Figure 11 and for $\Delta = 1/12$ in Figure 12. At the first sight, this sequence $(z_t^*)_t$ seems to fulfill these requirements, thus the out-of-sample back-testing provides the required results. In Figure 10 (rhs), we also provide the Q-Q-plot of these residuals $(z_t^*)_t$ against the standard Gaussian distribution for $\Delta = 1/52$. Also in this plot, we observe a good fit, except for the tails of the distribution. This suggests that one may relax the Gaussian assumption on ϵ_t^* by a more heavy-tailed model (this can also be seen

in Figure 11 where we have a few outliers). This has already been mentioned in Section 2 but for this exposition we keep the Gaussian assumption.

If we calculate the auto-correlation for time lag Δ between the residuals z_t^* we obtain 5% which is a convincingly small value. This supports the assumption of having independent residuals. The same holds true if we consider the auto-correlation for time lag Δ between the absolute values $|z_t^*|$ of the residuals resulting in 11%. The only observation which may contradict the i.i.d. assumption is that we observe slight clustering in Figure 11. This non-stationarity might have to do with the fact that we calculate the residuals under the equivalent martingale measure \mathbb{P}^* , however we make the observations under the real-world probability measure \mathbb{P} . If these measures coincide the statements are the same.

The classical approach is that one assumes that the two probability measures are equivalent, i.e. $\mathbb{P}^* \sim \mathbb{P}$, with density process

$$\xi_t = \prod_{s=1}^{t/\Delta} \exp \left\{ -\frac{1}{2} \|\lambda_{\Delta s}\|^2 + \lambda_{\Delta s} \boldsymbol{\varepsilon}_{\Delta s} \right\}, \tag{4.2}$$

with $\boldsymbol{\varepsilon}_t$ is independent of $\mathcal{F}_{t-\Delta}$, \mathcal{F}_t -measurable and a t/Δ -dimensional standard Gaussian random vector with independent components under \mathbb{P} . Moreover, it is assumed that λ_t is d -dimensional and previsible, i.e. $\mathcal{F}_{t-\Delta}$ -measurable. Note that this density process $(\xi_t)_t$ is a strictly positive and normalized (\mathbb{P}, \mathbb{F}) -martingale. For any \mathbb{P}^* -integrable and \mathcal{F}_t -measurable random variable X_t , we have, \mathbb{P} -a.s.,

$$\mathbb{E}_{t-\Delta}^* [X_t] = \frac{1}{\xi_{t-\Delta}} \mathbb{E}_{t-\Delta} [\xi_t X_t].$$

This implies that

$$\boldsymbol{\varepsilon}_t - \lambda_t \stackrel{(d)}{=} \boldsymbol{\varepsilon}_t^* \quad \text{under } \mathbb{P}_{t-\Delta}^*.$$

λ_t is the market price of risk at time t which explains the drift term in (1.2) and which reflects the difference between $\mathbb{P}_{t-\Delta}^*$ and $\mathbb{P}_{t-\Delta}$. Under Model Assumptions 3.1, we then obtain under the real-world probability measure \mathbb{P}

$$\begin{aligned} \mathbf{Y}_t &= \Delta \left[-\mathbf{Y}(t - \Delta, t) + \frac{1}{2} \text{sp}(\Sigma_\Lambda(\mathbf{Y}_{t-\Delta}^-)) \right] + \sqrt{\Delta} \zeta(\mathbf{Y}_{t-\Delta}^-) \wedge \lambda_t \\ &\quad + \sqrt{\Delta} \zeta(\mathbf{Y}_{t-\Delta}^-) \wedge \boldsymbol{\varepsilon}_t, \end{aligned}$$

i.e. we have a change of drift given by $\sqrt{\Delta} \zeta(\mathbf{Y}_{t-\Delta}^-) \wedge \lambda_t$. Thus, under the (conditional) real-world probability measure $\mathbb{P}_{t-\Delta}$ the approximate forecast $\tilde{\pi}_t$ has a Gaussian distribution with conditional mean and conditional covariance given by

$$\mu_{t-\Delta} = \mu_{t-\Delta}^* - \sqrt{\Delta} \mathbf{1}'_{\mathcal{M}_3} \zeta(\mathbf{Y}_{t-\Delta}^-) \wedge \lambda_t \quad \text{and} \quad \tau_{t-\Delta}^2 = \mathbf{1}'_{\mathcal{M}_3} \mathbf{S}_{(K)}(\mathbf{Y}_{t-\Delta}^-) \mathbf{1}_{\mathcal{M}_3}.$$

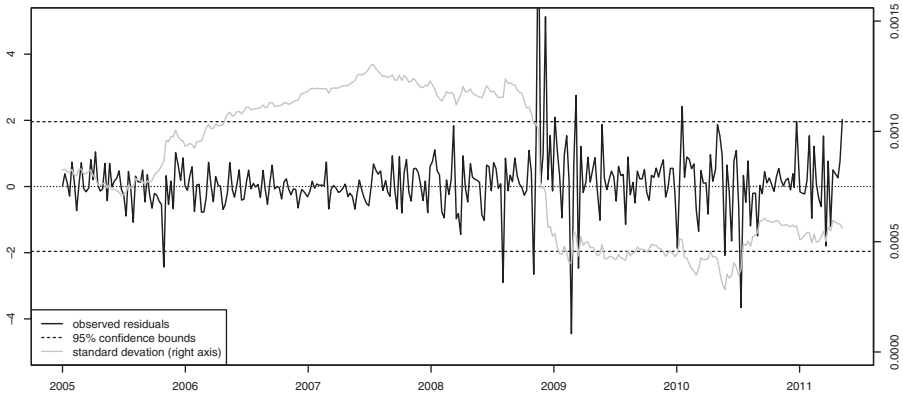


FIGURE 13: Time series of residuals $z_{m,t}^*$ for time to maturity $m = 1$ and $t \in \{01/2005, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{m,t-\Delta}$.

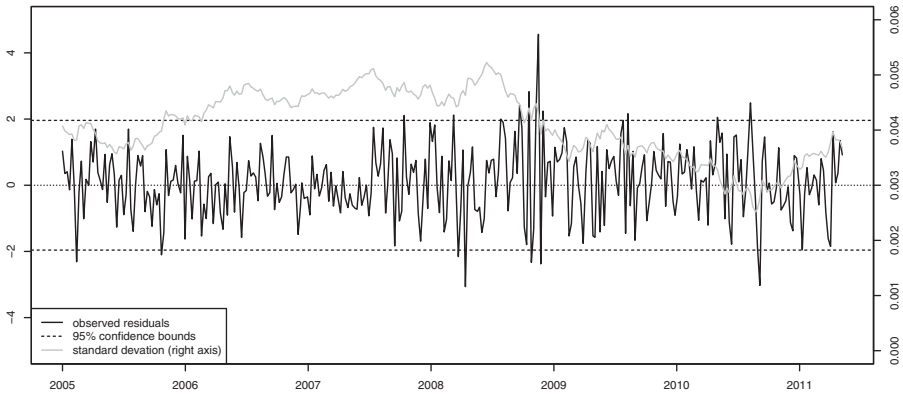


FIGURE 14: Time series of residuals $z_{m,t}^*$ for time to maturity $m = 5$ and $t \in \{01/2005, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{m,t-\Delta}$.

For an appropriate choice of the market price of risk λ_t , we obtain residuals

$$z_t = \frac{\tilde{\pi}_t - \mu_{t-\Delta}}{\tau_{t-\Delta}},$$

which should then form an i.i.d. standard Gaussian distributed sequence under the real-world probability measure \mathbb{P} .

In order to detect the market price of risk term, we look at residuals for individual times to maturity $m \in \mathcal{M}$, i.e. we replace the indicators $\mathbf{1}_{\mathcal{M}_3}$ in (4.1) by indicators $\mathbf{1}_{\{m\}}$.

We denote the resulting residuals by $z_{m,t}^*$ and the corresponding volatilities by $\tau_{m,t-\Delta}$. In Figures 13–15, we show the results for selected times to maturity $m = 1, 5, 10$. The picture is similar to Figure 11, i.e. we observe clustering but not a well-defined drift. This implies that we may set the market price of

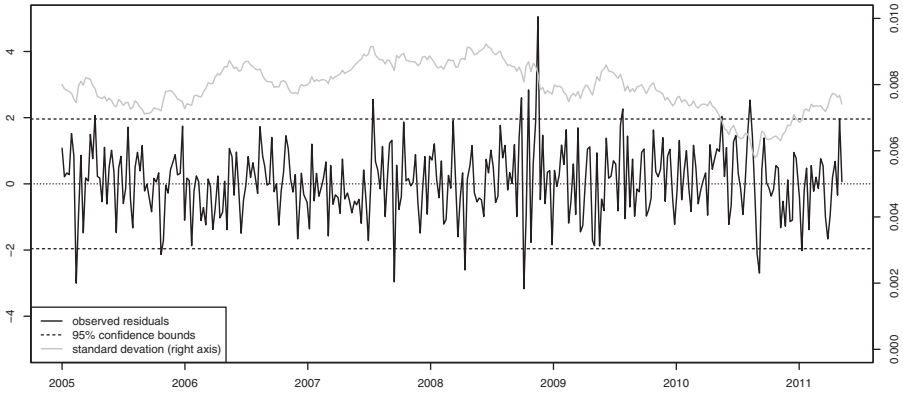


FIGURE 15: Time series of residuals $z_{m,t}^*$ for time to maturity $m = 10$ and $t \in \{01/2005, \dots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{m,t-\Delta}$.

risk $\lambda_t = 0$ for the prediction of future yield curves (we come back to this in Section 4.3).

4.3. Comparison to the Vasicek model

We compare our HJM framework to interest-rate models based on short-rate modeling, in particular, to the results of the Vasicek (1977) model. The Vasicek model is the simplest short-rate model that provides an affine term structure for interest rates, see also Filipović (2009), and hence a closed-form solution for ZCB prices. Note that the Vasicek model is known to have a weak performance, however, we would like to emphasize that the findings of this subsection are common to all short-rate models, such as the Cox *et al.* (1985), or the Black and Karasinski (1991) models.

The price of the ZCB in the Vasicek model takes the following form:

$$P(t, t + m) = \exp \{ A(m) - r_t B(m) \},$$

where the short-rate process $(r_t)_t$ evolves as an Ornstein–Uhlenbeck process under \mathbb{P}^* , and $A(m)$ and $B(m)$ are constants only depending on time to maturity m and the model parameters κ^* , θ^* and g (see for instance (3.8) in Brigo and Mercurio (2006)). The short-rate r_t is then under $\mathbb{P}_{t-\Delta}^*$ normally distributed with conditional mean and conditional variance given by

$$\begin{aligned} \mathbb{E}_{t-\Delta}^*[r_t] &= r_{t-\Delta} e^{-\Delta\kappa^*} + \theta^* (1 - e^{-\Delta\kappa^*}), \\ \text{Var}_{t-\Delta}^*(r_t) &= \frac{g^2}{2\kappa^*} [1 - e^{-2\kappa^*\Delta}]. \end{aligned}$$

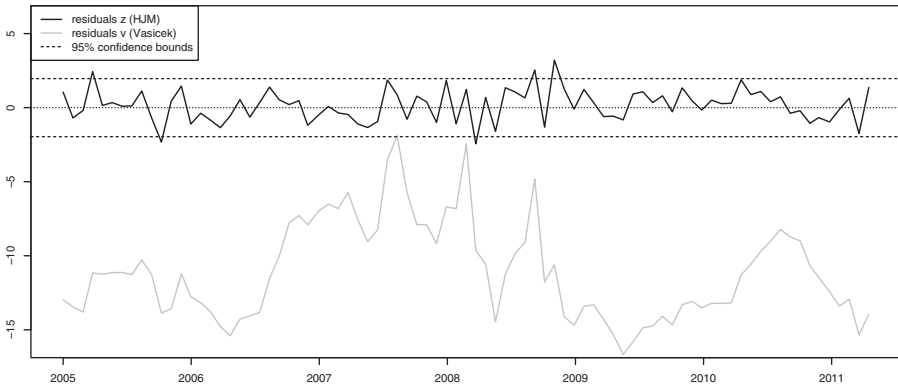


FIGURE 16: Time series of residuals z_t^* and v_t^* for $t \in \{01/2005, \dots, 05/2011\}$ on a monthly grid $\Delta = 1/12$ under the assumption $\mathbb{P}^* = \mathbb{P}$.

Thus, the approximation $\tilde{\pi}_t$ has under $\mathbb{P}_{t-\Delta}^*$ a normal distribution with conditional mean

$$\mathbb{E}_{t-\Delta}^*[\tilde{\pi}_t] = \sum_{m \in \mathcal{M}_3} (1 + A(m) - \mathbb{E}_{t-\Delta}^*[r_t] B(m)),$$

and conditional variance

$$\text{Var}_{t-\Delta}^*(\tilde{\pi}_t) = \text{Var}_{t-\Delta}^*(r_t) \left(\sum_{m \in \mathcal{M}_3} B(m) \right)^2.$$

As in the previous section, we first assume $\mathbb{P}^* = \mathbb{P}$, i.e. we set the market price of risk $\lambda_t = 0$: (i) this allows to estimate the model parameters κ^* , θ^* and g from a time series of observations, for instance, using maximum likelihood methods (see (3.14)–(3.16) in Brigo and Mercurio (2006)); (ii) makes the model comparable to the calibration of our model. We will comment on this “comparability” below.

Thus, we estimate these parameters and obtain parameter estimates $\hat{\kappa}^*$, $\hat{\theta}^*$ and \hat{g} from which we get the estimated functions $\hat{A}(\cdot)$ and $\hat{B}(\cdot)$. This then allows to estimate the conditional mean and variance of $\tilde{\pi}_t$, given $\mathcal{F}_{t-\Delta}$. From these, we calculate the observable residuals

$$v_t^* = \frac{\tilde{\pi}_t - \mathbb{E}_{t-\Delta}^*[\tilde{\pi}_t]}{\widehat{\text{Var}}_{t-\Delta}^*(\tilde{\pi}_t)^{1/2}}.$$

In Figure 16, we plot the time series z_t^* and v_t^* for $t \in \{01/2005, \dots, 05/2011\}$. The observation is that v_t^* is far too small! The explanation for this observation lies in the assumption $\mathbb{P}^* = \mathbb{P}$, i.e. $\lambda_t = 0$. Since the Vasicek prices are calculated

by conditional expectations of the *entire* future development of the short-rate process $(r_t)_t$ until expiry of the ZCB, the choices of $\widehat{\kappa}^*$, $\widehat{\theta}^*$ and \widehat{g} have a huge influence on the resulting ZCB prices in the Vasicek model. Thus, the calibration of $\widehat{A}(\cdot)$ and $\widehat{B}(\cdot)$ is completely inappropriate if we set $\lambda_t = 0$. Compare

$$\log P(t, t + m) = -m Y(t, t + m), \tag{4.3}$$

$$\log P(t, t + m) = A(m) - r_t B(m). \tag{4.4}$$

The (pricing) functions $A(\cdot)$ and $B(\cdot)$ in (4.4) are calculated completely within the Vasicek model by a forward projection of $(r_t)_t$ until maturity date $t + m$. If this forward projection is done under the wrong measure \mathbb{P} , then these pricing components completely miss the market risk dynamics and hence are not appropriate. Therefore, the Vasicek model, as any other short-rate model, is not robust against unavoidable inappropriate choices of market prices of risk.

Conclusions 4.2

- We conclude that the HJM models (similar to Model Assumptions 3.1) are much more robust against inappropriate choices of the market price of risk compared to short-rate models, because in the former we only need to choose the market price of risk for the one-step ahead for the prediction of the ZCB prices at the end of the period (i.e. from $t - \Delta$ to t) whereas for short-rate models we need to choose the market price of risk appropriately for the entire life time of the ZCB (i.e. from $t - \Delta$ to $t + m$).
- Our HJM model (Model Assumptions 3.1) always captures the actual yield curve, whereas this is not necessarily the case for short-rate models, see (4.3) versus (4.4). Hull and White (1990, 1994) extended version may be used instead, see Harms *et al.* (2015) and Wüthrich (2015) for consistent re-calibration of such short-rate models.

4.4. Forward projection of yield curves and arbitrage

For calibration of the model and for yield curve prediction, we have chosen a restricted set \mathcal{M} of times to maturity. In most applied cases, one has to stay within such a restricted set because there do not exist observations for all times to maturity. We propose to predict future yield curves within these families \mathcal{M} and then approximate the remaining times to maturity using a parametric family like the Nelson and Siegel (1987) or the Svensson (1994, 1995).

Finally, we demonstrate the absence of arbitrage condition given in Lemma 2.1. At the end of Section 1, we have emphasized the importance of the no-arbitrage property of the prediction model. Let us choose an asset portfolio $w_{t-\Delta} P(t, t + m_1) - P(t, t + m_2)$ for two different times to maturity m_1 and m_2 .



FIGURE 17: Back-testing the difference of aggregated realized gains of portfolios $\tilde{\pi}_t$ for $w_{t-\Delta} = \tau_{t-\Delta}^{(2)}/\tau_{t-\Delta}^{(1)}$ and the their model prognoses with and without the no-arbitrage HJM correction term.

We approximate this portfolio by a second order Taylor expansion and set

$$\tilde{\pi}_t = w_{t-\Delta} \left(1 - m_1 Y(t, t + m_1) + \frac{(m_1 Y(t, t + m_1))^2}{2} \right) - \left(1 - m_2 Y(t, t + m_2) + \frac{(m_2 Y(t, t + m_2))^2}{2} \right).$$

Under our model assumptions, the returns of both terms $m_i Y(t, t + m_i)$ in portfolio $\tilde{\pi}_t$ have, conditionally given $\mathcal{F}_{t-\Delta}$, a Gaussian distribution term with standard deviations given by

$$\tau_{t-\Delta}^{(i)} = \sqrt{\mathbf{1}'_{\{m_i\}} S_{(K)}(\mathbf{Y}_{t-\Delta}^-) \mathbf{1}_{\{m_i\}}} \quad \text{for } i = 1, 2.$$

If we choose $w_{t-\Delta} = \tau_{t-\Delta}^{(2)}/\tau_{t-\Delta}^{(1)}$ then the returns of the Gaussian parts of both terms in portfolio $\tilde{\pi}_t$ have the same variance and, thus, under the Gaussian assumption have the same marginal distributions. Since the conditional expectation of the second order term in the Taylor expansion cancels the no-arbitrage drift term (up to a small short-rate related correction), we see that the returns of the portfolio $\tilde{\pi}_t$ should approximately provide zero returns, conditionally. In Figure 17, we give an example for times to maturity $m_1 = 10$ and $m_2 = 20$. The correlation between the prices of these ZCBs is high, about 85%, i.e. their prices tend to move simultaneously. The resulting weights $w_{t-\Delta}$ are in the range between 1.4 and 1.9. In Figure 17, we plot the aggregated realized gains of the portfolio $\tilde{\pi}$ minus their prognosis including and excluding the HJM correction term. Recall that the predicted should be zero conditionally on the current information. We observe that the model without the HJM term drifts away from zero, which opens the possibility of arbitrage. Therefore, we insist on a prediction model that is free of arbitrage.

We close with the remark that the model presented in this paper is also a first step towards to extrapolation of the yield curve beyond the maximal observed maturity date. This extrapolation is a main open problem in Solvency II on which only little mathematical research has been done in the literature, see Dahl (2007). To achieve this task, our model needs an extra feature that describes how new ZCBs are launched in the future, this is the reinvestment risk described in Dahl (2007) and Stefanovits and Wüthrich (2014), and opens a whole new area of modeling questions to be solved.

ACKNOWLEDGEMENT

The authors gratefully acknowledge support by the ETH foundation and by the Swiss Finance Institute.

NOTES

1. data source: Swiss National Bank (SNB), www.snb.ch.
2. data source: Swiss National Bank (SNB), www.snb.ch, for more information see also Müller (2002).
3. data source: Bloomberg.

REFERENCES

- AUDRINO, F. and FILIPOVA, K. (2009) Yield curve predictability, regimes, and macroeconomic information: A data driven approach. University of St. Gallen. Discussion Paper no. 2009–10.
- BARONE-ADESI, G., BOURGOIN, F. and GIANOPOULOS, K. (1998) Don't look back. *Risk* 08/1998, 100–103.
- BJÖRK, T. (1998) Interest rate theory. In *Financial Mathematics, Bressanone 1996* (ed. W. Runggaldier), pp. 53–122. Lecture Notes in Mathematics 1656, Springer.
- BJÖRK, T. and LANDÉN, C. (2000) On the construction of finite dimensional realizations for non-linear forward rate models. Working paper, Stockholm School of Economics.
- BJÖRK, T. and SVENSSON, L. (2001) On the existence of finite dimensional realizations for nonlinear forward rate models. *Mathematical Finance*, 11(2), 205–243.
- BLACK, F. and KARASINSKI, P. (1991) Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47(4), 52–59.
- BRIGO, D. and MERCURIO, F. (2006) *Interest Rate Models - Theory and Practice*. Berlin, Heidelberg: Springer.
- CAIRNS, A.J.G. (1998) Descriptive bond-yield and forward-rate models for the British government securities' market. *British Act. J.*, 4(2), 265–321.
- CAIRNS, A.J.G. (2004) *Interest Rate Models: An Introduction*. Princeton, New Jersey: Princeton University Press.
- CHRISTENSEN, J.H.E., DIEBOLD, F.X. and RUDEBUSCH, G.D. (2007) The affine arbitrage-free class of Nelson-Siegel term structure models. Working Paper Series 2007-20, Federal Reserve Bank of San Francisco.
- CHRISTENSEN, J.H.E., DIEBOLD, F.X. and RUDEBUSCH, G.D. (2009) An arbitrage-free generalized Nelson-Siegel term structure model. *Econometrics Journal, Royal Economic Society*, 12(3), C33–C64.
- COX, J.C., INGERSOLL, J.E. and ROSS, S.A. (1985) A theory of the term structure of interest rates. *Econometrica*, 53(2), 385–407.

- DAHL, M. (2007) A discrete-time model for reinvestment risk in the bond market. *ASTIN Bulletin*, **37**(2), 235–264.
- DEGUILLAUME, N., REBONATO, R. and POGUDIN, A. (2013) The nature of the dependence of the magnitude of rate moves on the level of rates: A universal relationship. *Quantitative Finance*, **13**(3), 351–367.
- DELBAEN, F. and SCHACHERMAYER, W. (1994) A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, **300**, 463–520.
- FILIPOVIĆ, D. (1999) A note on the Nelson-Siegel family. *Mathematical Finance*, **9**(4), 349–359.
- FILIPOVIĆ, D. (2000) Exponential-polynomial families and the term structure of interest rates. *Bernoulli*, **6**(6), 1–27.
- FILIPOVIĆ, D. (2009) *Term-Structure Models*. Berlin, Heidelberg: Springer.
- FILIPOVIĆ, D. and TEICHMANN, J. (2004) On the geometry of the term structure of interest rates. *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, **460**, 129–167.
- FILIPOVIĆ, D. and ZABCZYK, J. (2002) Markovian term structure models in discrete time. *Annals of Applied Probability*, **12**(2), 710–729.
- HARMS, P., STEFANOVITS, D., TEICHMANN, J. and WÜTHRICH, M.V. (2015) Consistent recalibration of yield curve models. Preprint on arXiv 1502.02926.
- HEATH, D., JARROW, R. and MORTON, A. (1990) Bond pricing and the term structure of interest rates: A discrete time approximation. *Journal of Financial Quantitative Analysis*, **25**(4), 419–440.
- HEATH, D., JARROW, R. and MORTON, A. (1992) Bond pricing and the term structure of interest rates: A new methodology. *Econometrica*, **60**(1), 77–105.
- HULL, J. and WHITE, A. (1990) Pricing interest-rate-derivative securities. *Review of Financial Studies*, **3**(4), 573–592.
- HULL, J. and WHITE, A. (1994) Branching out. *Risk*, **7**, 34–37.
- JORDAN, T.J. (2009) SARON - an innovation for the financial markets. Launch event for Swiss Reference Rates, Zurich, August 25, 2009.
- MÜLLER, R. (2002) Zur Berechnung der Obligationenrenditen im Statistischen Monatsheft der SNB. *Swiss National Bank Quartalsheft*, **2**, 64–73.
- NELSON, C.R. and SIEGEL, A.F. (1987) Parsimonious modeling of yield curves. *J. Business*, **60**(4), 473–489.
- ORTEGA, J.P., PULLIRSCH, R., TEICHMANN, J. and WERGIELUK, J. (2009) A dynamic approach for scenario generation in risk management. Preprint on arXiv 0904.0624.
- STEFANOVITS, D. and WÜTHRICH, M.V. (2014) Hedging of long term zero-coupon bonds in a market model with reinvestment risk. *European Actuarial Journal*, **4**(1), 49–75.
- SVENSSON, L.E.O. (1994) Estimating and interpreting forward interest rates: Sweden 1992–1994. NBER Working Paper Series Nr. 4871.
- SVENSSON, L.E.O. (1995) Estimating forward interest rates with the extended Nelson & Siegel method. *Sveriges Riksbank Quarterly Review*, **3**, 13–26.
- VASIČEK, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**(2), 177–188.
- WÜTHRICH, M.V. (2016) Consistent re-calibration in yield curve modeling: An example. In: Causal Inference in Econometrics, 57–82. V.-N. Huynh, V. Kreinovich, S. Sriboonchitta (eds.), Studies in computational Intelligence 622, Springer.

JOSEF TEICHMANN (Corresponding author)

ETH Zurich, Department of Mathematics, 8092 Zurich, Switzerland

E-Mail: jteichma@math.ethz.ch

MARIO V. WÜTHRICH

ETH Zurich, RiskLab, Department of Mathematics

8092 Zurich, Switzerland

E-Mail: mario.wuethrich@math.ethz.ch

APPENDIX A. PROOFS

Proof of Lemma 2.1. We rewrite (2.2) as follows (where we use assumption (2.1) of the yield curve development and the appropriate measurability properties)

$$\begin{aligned} \exp\{-\Delta Y(t - \Delta, t)\} \mathbb{E}_{t-\Delta}^* [P(t, t + m)] &= P(t - \Delta, t) \mathbb{E}_{t-\Delta}^* [P(t, t + m)] \\ &= P(t - \Delta, t + m) \exp\{-\alpha_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta})\} \mathbb{E}_{t-\Delta}^* [\exp\{-\mathbf{v}_\Delta(t, m, (\mathbf{Y}_s)_{s \leq t-\Delta}) \boldsymbol{\varepsilon}_t^*\}] \\ &\stackrel{!}{=} P(t - \Delta, t + m). \end{aligned}$$

Solving this requirement proves the claim of Lemma 2.1.

Proof of Theorem 3.5. In the first step, we apply the tower property for conditional expectation which decouples the problem into several steps. We have $\mathbb{E}_0^*[S_{(K)}(\mathbf{y})] = \mathbb{E}_0^*[\mathbb{E}_{\Delta(K-1)}^*[S_{(K)}(\mathbf{y})]]$. Thus, we need to calculate the inner conditional expectation $\mathbb{E}_{\Delta(K-1)}^*[\cdot]$ of the $d \times d$ matrix $S_{(K)}(\mathbf{y})$. We define the auxiliary matrix

$$\tilde{C}_{(K)} = \left([\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k}]_j \right)_{j=1, \dots, d; k=1, \dots, K} \in \mathbb{R}^{d \times K}.$$

This implies that we can rewrite $C_{(K)} = K^{-1/2} \tilde{C}_{(K)}$. Moreover, we rewrite the matrix $\tilde{C}_{(K)}$ as follows:

$$\tilde{C}_{(K)} = [\tilde{C}_{(K-1)}, \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k}],$$

with $\tilde{C}_{(K-1)} \in \mathbb{R}^{d \times (K-1)}$ is $\mathcal{F}_{\Delta(K-1)}$ -measurable. This implies the following decomposition:

$$\begin{aligned} S_{(K)}(\mathbf{y}) &= \frac{1}{K} \zeta(\mathbf{y}) \tilde{C}_{(K)} \tilde{C}'_{(K)} \zeta(\mathbf{y})' \\ &= \frac{1}{K} \zeta(\mathbf{y}) [\tilde{C}_{(K-1)}, \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k}] [\tilde{C}_{(K-1)}, \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k}]' \zeta(\mathbf{y})' \\ &= \frac{1}{K} \zeta(\mathbf{y}) \left(\tilde{C}_{(K-1)} \tilde{C}'_{(K-1)} + (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k}) (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k})' \right) \zeta(\mathbf{y})' \\ &= \frac{K-1}{K} S_{(K-1)}(\mathbf{y}) + \frac{1}{K} \zeta(\mathbf{y}) \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \boldsymbol{\Upsilon}_{\Delta k} \boldsymbol{\Upsilon}'_{\Delta k} (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1})' \zeta(\mathbf{y})'. \end{aligned}$$

This implies for the conditional expectation of $S_{(K)}(\mathbf{y})$

$$\begin{aligned} \mathbb{E}_{\Delta(K-1)}^* [S_{(K)}(\mathbf{y})] &= \frac{K-1}{K} S_{(K-1)}(\mathbf{y}) \\ &\quad + \frac{1}{K} \zeta(\mathbf{y}) \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \mathbb{E}_{\Delta(K-1)}^* [\boldsymbol{\Upsilon}_{\Delta k} \boldsymbol{\Upsilon}'_{\Delta k}] (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1})' \zeta(\mathbf{y})'. \end{aligned}$$

We calculate the conditional expectation in the last term, we start with the conditional covariance. From Lemma 3.3, we obtain

$$\begin{aligned} &\frac{1}{K} \zeta(\mathbf{y}) \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \text{Cov}_{\Delta(K-1)}^* (\boldsymbol{\Upsilon}_{\Delta k}) (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1})' \zeta(\mathbf{y})' \\ &= \frac{\Delta}{K} \zeta(\mathbf{y}) \zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \Sigma_\Lambda(\mathbf{Y}_{\Delta k, -}) (\zeta(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1})' \zeta(\mathbf{y})' = \frac{\Delta}{K} \Sigma_\Lambda(\mathbf{y}). \end{aligned}$$

This implies

$$\begin{aligned}
 \mathbb{E}_0^* [S_{(K)}(\mathbf{y})] &= \frac{K-1}{K} \mathbb{E}_0^* [S_{(K-1)}(\mathbf{y})] + \frac{\Delta}{K} \Sigma_{\Lambda}(\mathbf{y}) \\
 &\quad + \frac{1}{K} \varsigma(\mathbf{y}) \mathbb{E}_0^* \left[\varsigma(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1} \mathbb{E}_{\Delta(K-1)}^* [\mathbf{Y}_{\Delta k}] \mathbb{E}_{\Delta(K-1)}^* [\mathbf{Y}_{\Delta k}]' (\varsigma(\mathbf{Y}_{\Delta k-\Delta}^-)^{-1})' \right] \varsigma(\mathbf{y})' \\
 &= \frac{K-1}{K} \mathbb{E}_0^* [S_{(K-1)}(\mathbf{y})] + \frac{\Delta}{K} \Sigma_{\Lambda}(\mathbf{y}) \\
 &\quad + \frac{\Delta^2}{K} \varsigma(\mathbf{y}) \mathbb{E}_0^* [f_{\Lambda}(\mathbf{Y}(\Delta(K-1), \Delta K), \mathbf{Y}_{\Delta k-\Delta}^-)] \varsigma(\mathbf{y})'.
 \end{aligned}$$

Iterating this provides the result.