Consistent Yield Curve Prediction

Josef Teichmann*  Mario V. Wüthrich*

February 6, 2013

Abstract

We present an arbitrage-free, non-parametric yield curve prediction model which takes the full discretized yield curve data as input state variable. The absence of arbitrage is a particular important model feature for prediction models in case of highly correlated data as, e.g., for interest rates. Furthermore, the model structure allows to separate constructing the daily yield curve from estimating the 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 Vasicek short rate model.

2010 Mathematics Subject Classification (MSC2010): 91G30, 91G70

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 has 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 up to $t = 5$ years into the future, however, we are dealing with the whole yield curve, i.e. maturities up to $T = 30$ years are considered.

We introduce some 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

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

Aim and scope.

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

---

*ETH Zurich, Department of Mathematics, 8092 Zurich, Switzerland.
The first author gratefully acknowledges support by the ETH foundation.
(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 the 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

\[
dP(t,T) = \Sigma(t,T) \, d\epsilon_t + \delta(t,T) \, d\tilde{N}_t + \mu(t,T) \, dt,
\]

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 \( (\tilde{N}_t)_t \) and a drift of the ZCB evolution under the constraint \( P(T,T) = 1 \). In the 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, 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, i.e. that \( t \) runs 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 \( t \to T \) and, most importantly, that

\[
\mu(t,T) = \Sigma(t,T) \lambda_t + r_t,
\]

where \( (\lambda_t)_t \) is an appropriately integrable predictable vector-valued process, the 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 \( \Sigma \) 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, [7], [8], [9], [17], [29] and [30].

The second best observable quantities are the components of the instantaneous covariance matrix \( \Sigma \) 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 [2] and
[28], together with a parametric approach for the specification of the local volatility impact of yield levels based on the work [20].

What is hard to be identified is the market price of risk \( \lambda \), i.e. the deviation of \( \mu \) from the risk-free short rate \( r \) in terms of volatility \( \Sigma \), see (1.2). These are good and bad news: it means that for pricing purposes (i.e. under the martingale measure) we are practically able to identify the model from time series data under fairly weak assumptions, since \( \lambda \) is not needed in the martingale measure; but for prediction purposes we have to deal with one quantity that is hard to measure, namely the market price of risk \( \lambda \). 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 \( \Sigma \) is the crucial step towards a measurement of \( \mu \) since, due to no-arbitrage considerations, \( \mu - r \) has to lie in the range of \( \Sigma \), see (1.2).

The presented Heath-Jarrow-Morton (HJM)-type approach [21, 22] 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. [8], p. 139 in [7], or the recent contribution [1], often starts with yield curves directly 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 martingale measure of the market, namely exactly the HJM-type drift correction, which is due to the (non-linear) logarithmic transform between prices and yields, see (1.1). Due to 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 [18], where the general situation is clarified following ideas of [4] and [5].

In contrast, factor models, see [7], [9] and [17], 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 [23, 24] 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. which 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 \( \sigma \) and \( \kappa \) (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 P&L distributions. More precisely, if a prediction model admits arbitrage then implementing long-short arbitrage portfolios yield always positive P&L structures. In practice adding such an arbitrage portfolio can then be used to shift P&L distributions of general portfolios, which is an undesired effect from the point of view of valuation and risk management, see Figure 21 and Section 4.4, below.

**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 [21, 22]. Choose a fixed grid size \( \Delta = 1/n \) for \( n \in \mathbb{N} \). We consider the discrete time points \( t \in \Delta \mathbb{N}_0 = \{0, \Delta, 2\Delta, 3\Delta, \ldots\} \) and the maturity dates \( T \in t + \Delta \mathbb{N} \). For example, the 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 \((\Omega, \mathcal{F}, P, \mathcal{F})\) with real world probability measure \( P \) and (discrete time) filtration \( \mathcal{F} = (\mathcal{F}_t)_{t \in \Delta \mathbb{N}_0} \).

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

\[
Y_t = (Y(t, t + m))'_{m \in \Delta \mathbb{N}}
\]

for all time points \( t \in \Delta \mathbb{N}_0 \). Assume that \((Y_t)_{t \in \Delta \mathbb{N}_0}\) is \( \mathcal{F}\)-adapted, that is, \((Y_s)_{s \leq t}\) is observable at time \( t \) and this information is contained in the \( \sigma\)-field \( \mathcal{F}_t \). Our aim is (as described above) to model and predict \((Y_t)_{t \in \Delta \mathbb{N}_0}\). 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 \((Y_t)_{t \in \Delta \mathbb{N}_0}\) first directly under this equivalent martingale measure \( \mathbb{P}^* \). Notice here that the bank account numeraire is actually a discrete time roll-over portfolio, as will be seen in the next section.
Remark. The assumption that the yield curve is given at any moment \( t \in \Delta N_0 \) for sufficiently many maturities is a very strong one. In practice the yield curve is inter- and extrapolated every day from quite different traded quantities like coupon bearing bonds, swap rates, etc. This inter- and extrapolation allows for a lot of freedom, often parametric families are used, e.g. the Nelson-Siegel \[27\] or the Svensson \[29, 30\] family, but also non-parametric approaches such as splines are applied, see also \[3\], \[8\], \[10, 11\] and \[15, 16, 17\].

Assume the initial yield curve \( Y_0 = (Y(0, m))_{m \in \Delta N} \) at time \( t = 0 \) is given. For \( t, m \in \Delta N \) we make the following model assumptions: assume there exist deterministic functions \( \alpha(\cdot, \cdot, \cdot) \) and \( v(\cdot, \cdot, \cdot) \) such that the yield curve has the following stochastic representation

\[
m Y(t, t + m) = (m + \Delta) Y(t - \Delta, t + m) - \Delta Y(t - \Delta, t) + \alpha(t, m, (Y_s)_{s \leq t - \Delta}) + v(t, m, (Y_s)_{s \leq t - \Delta}) \varepsilon^*_t,
\]

where the innovations \( \varepsilon^*_t \) are \( \mathcal{F}_t \)-measurable, independent of \( \mathcal{F}_{t-\Delta} \), centered and have unit variance under \( \mathbb{P}^* \). In general, the innovations \( \varepsilon^*_t \) 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 \[17\]). The fourth term on the right-hand side of (2.1) described by \( v(t, m, (Y_s)_{s \leq t - \Delta}) \varepsilon^*_t \) adds the stochastic part to the future yield curve development. Finally, the third term \( \alpha(t, m, (Y_s)_{s \leq t - \Delta}) \) 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 \( v_\Delta \), and estimating the market price of risk, i.e. the difference between \( \mathbb{P} \) and \( \mathbb{P}^* \).

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

\[
P(t, t + m) = \frac{P(t - \Delta, t + m)}{P(t - \Delta, t)} \exp \left\{ -\alpha(t, m, (Y_s)_{s \leq t - \Delta}) - v(t, m, (Y_s)_{s \leq t - \Delta}) \varepsilon^*_t \right\}.
\]

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

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

The process \( B = (B_t)_{t \in \Delta N_0} \) considers the roll-over of an initial investment 1 into the (discrete time) bank account with grid size \( \Delta \). Note that \( B \) is previsible, i.e. \( B_t \) is \( \mathcal{F}_{t-\Delta} \)-measurable for all \( t \in \Delta N \).
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 \(t, m \in \Delta N\)
\[
\mathbb{E}^* \left[ B_t^{-1} P(t, t + m) \, | \, \mathcal{F}_{t-\Delta} \right] \equiv B_{t-\Delta}^{-1} P(t - \Delta, t + m). \tag{2.2}
\]
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 [14]. For notational convenience we set \(\mathbb{E}^*_t [\cdot] = \mathbb{E}^* [\cdot | \mathcal{F}_t]\) for \(t \in \Delta N_0\). The no-arbitrage condition (2.2) immediately provides the following lemma, see also [21, 22] and [19].

**Lemma 2.1** Under the above assumptions the absence of arbitrage condition (2.2) implies
\[
\alpha_\Delta(t, m, (Y_s)_{s \leq t-\Delta}) = \log \mathbb{E}^*_t \left[ \exp \left\{ -v_\Delta(t, m, (Y_s)_{s \leq t-\Delta}) \varepsilon^*_t \right\} \right].
\]
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 \(v_\Delta\) on the term structure \(Y\) 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 need to discuss the explicit choices \(v_\Delta(t, m, (Y_s)_{s \leq t-\Delta})\) and \(\varepsilon^*_t\) under \(\mathbb{P}^*\) as well as the description of the change of measure \(\mathbb{P}^* \sim \mathbb{P}\). Then, the 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 would like to study a finite set \(\mathcal{M} \subset \Delta N\) of times to maturity. We specify below necessary properties of \(\mathcal{M}\) for the yield curve prediction. In view of (2.1) we define for these times to maturity choices and \(t \in \Delta N\) the random vectors
\[
\Upsilon_t = (\Upsilon_{t,m})'_{m \in \mathcal{M}} = (m Y(t, t + m) - (m + \Delta) Y(t - \Delta, t + m))'_{m \in \mathcal{M}} = \left( -\log \frac{P(t, t + m)}{P(t - \Delta, t + m)} \right)'_{m \in \mathcal{M}}.
\]
We set the dimension \(d = |\mathcal{M}|\) and for \(\varepsilon^*_t|_{\mathcal{F}_{t-\Delta}}\) 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 \(\varepsilon^*_t\) 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 martingale measure is an additional strong assumption which could be weakened.

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

$$v_{\Delta}(t,m,(Y_s)_{s\leq t-\Delta}) = \sqrt{\Delta} \sigma(t,m,Y_{t-\Delta}),$$

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

$$a_{\Delta}(t,m,(Y_s)_{s\leq t-\Delta}) = \log E^*_{t-\Delta} \left[ \exp \left\{ -\sqrt{\Delta} \sigma(t,m,Y_{t-\Delta}) \varepsilon_t^* \right\} \right] = \frac{\Delta}{2} \|\sigma(t,m,Y_{t-\Delta})\|^2.$$

From (2.1) we then obtain for $t \in \Delta N$ and $m \in M$ under $\mathbb{P}^*$(3.1) holds:

$$Y_{t,m} = \Delta \left[ -Y(t-\Delta,t) + \frac{1}{2} \|\sigma(t,m,Y_{t-\Delta})\|^2 \right] + \sqrt{\Delta} \sigma(t,m,Y_{t-\Delta}) \varepsilon_t^*.$$

Note that $(Y_t)_{t \in \Delta N}$ is a $d$-dimensional process, thus, we need a $d$-dimensional Gaussian random vector $\varepsilon_t^*|_{\mathcal{F}_{t-\Delta}}$ for obtaining sufficiently rich stochastic structures. Next, we specify explicitly the $d$-dimensional vector-valued function $\sigma(\cdot,\cdot,\cdot)$. We proceed similar to [28], i.e. we directly model volatilities and return directions. In particular, we choose (i) linear maps $\varsigma$ which describe the volatility scaling factors; and (ii) a matrix $\Lambda = [\lambda_1, \ldots, \lambda_d] \in \mathbb{R}^{d \times d}$, where $\lambda_1, \ldots, \lambda_d \in \mathbb{R}^d$ specify the (raw) return directions of the innovations $\varepsilon_t^*$.

To be precise we choose invertible linear maps/matrices $\varsigma(y)$, defined for every $y \in \mathbb{R}^d$, providing

$$\varsigma(y) : \mathbb{R}^d \to \mathbb{R}^d, \quad \lambda \mapsto \varsigma(y) \lambda,$$

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

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

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 $t \in \Delta N$ with time to maturity dates $\mathcal{M}$:

$$Y_t = \Delta \left[ -Y(t-\Delta,t) + \frac{1}{2} \operatorname{sp} (\Sigma_{\Lambda}(Y_{t-\Delta}^-)) \right] + \sqrt{\Delta} \varsigma(Y_{t-\Delta}^-) \Lambda \varepsilon_t^*,$$

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

7
For the $j$-th maturity $m_j \in M$ we have done the following choice

$$
\sigma(t, m_j, Y_{t-\Delta}^i) \varepsilon^i_t = \sum_{i=1}^{d} \sigma_i(t, m_j, Y_{t-\Delta}^i) \varepsilon^i_{t,i} = \sum_{i=1}^{d} \left[ \varsigma(Y_{t-\Delta}) \lambda_i \right]_j \varepsilon^i_{t,i}. \tag{3.3}
$$

Our aim is to estimate the volatility scaling factors $\varsigma$ and the return directions $\lambda_1, \ldots, \lambda_d \in \mathbb{R}^d$. This is done in Subsection 3.2. Note that these choices do not depend on the grid size $\Delta$.

**Remark.** The volatility scaling factors $\varsigma(\cdot)$ mimic how volatility for different times to maturity scales with the level of yield at this maturity. Several approaches have been discussed in the literature, see [20]. 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 currency CHF – linear dependence seems to be a good choice, too, see choice (3.8). In particular, this choice provides appropriate stationarity in $(\Upsilon_{t,m})_t$ over different times to maturity choices $m \in M$ and different yield levels.

**Lemma 3.2** Under Model Assumptions 3.1, the random vector $\Upsilon_t|\mathcal{F}_{t-\Delta}$ has a $d$-dimensional conditional Gaussian distribution with the first two conditional moments given by

$$
\mathbb{E}_{t-\Delta}[\Upsilon_t] = \Delta \left[ -Y(t-\Delta, t) + \frac{1}{2} \text{sp}(\Sigma_{\Upsilon}(Y_{t-\Delta})) \right],
$$

$$
\text{Cov}_{t-\Delta}(\Upsilon_t) = \Delta \Sigma_{\Upsilon}(Y_{t-\Delta}).
$$

**3.2 Calibration procedure**

In order to calibrate our model we need to choose the volatility scaling factors $\varsigma(\cdot)$ and we need to specify the return directions $\lambda_1, \ldots, \lambda_d \in \mathbb{R}^d$ which provide the matrix $\Lambda$. Since we are only interested in the law of the process we do not need to specify the direction $\lambda_1, \ldots, \lambda_d \in \mathbb{R}^d$ themselves but, due to Lemma 3.2, rather the covariance structure $\Sigma_{\Upsilon}$.

Assume we have the following time series of observations $(\Upsilon_t)_{t=\Delta, \ldots, \Delta K}, (Y(t-\Delta, t))_{t=\Delta, \ldots, \Delta(K+1)}$, and $(Y_{t-\Delta})_{t=\Delta, \ldots, \Delta(K+1)}$. We use these observations to predict/approximate the random vector $\Upsilon_{\Delta(K+1)}$ at time $\Delta K$, see also Lemma 3.2. For $y \in \mathbb{R}^d$ we define the matrices

$$
C_{(K)} = \frac{1}{\sqrt{K}} \left( \varsigma(Y_{\Delta(K-1)})^{-1} Y_{\Delta k} \right)_{j=1, \ldots, d; k=1, \ldots, K} \in \mathbb{R}^{d \times K},
$$

$$
S_{(K)}(y) = \varsigma(y) C_{(K)} C_{(K)}^t \varsigma'(y) \in \mathbb{R}^{d \times d}.
$$

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

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

with $\mathbf{W}^*_t$ is independent of $\mathcal{F}_{t-\Delta}, \mathcal{F}_t$-measurable, independent of $\varepsilon^*_t$ and a $K$-dimensional standard Gaussian random vector with independent components under $\mathbb{P}^*$. 
Lemma 3.3 The random vector $\kappa_t | F_{t-\Delta}$ has a $d$-dimensional Gaussian distribution with the first two conditional moments given by

$$
\mathbb{E}_{t-\Delta}^*[\kappa_t] = -\Delta \mathbf{x} + \frac{1}{2} \text{sp} \left( S_{(K)}(y) \right),
$$

$$
\text{Cov}_{t-\Delta}^*[\kappa_t] = S_{(K)}(y).
$$

Our aim is to show that the matrix $S_{(K)}(y)$ is an appropriate estimator for $\Delta \Sigma(y)$ and then Lemmas 3.2 and 3.3 say that $\kappa_t$ is an appropriate approximation in law to $\Upsilon_t$, conditionally given $F_{t-\Delta}$.

Remark. The random vector $\kappa_t$ can be seen as a filtered historical simulation of $\Upsilon_t$, where $W_t^*$ re-simulates the $K$ observations which are appropriately historically scaled through $\varsigma_t$, see, e.g., [2]. 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 $(\Upsilon_t)_{t=\Delta, \ldots, \Delta K}$ up to proper re-scaling.

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

$$
f_{\Lambda}(z, y) = \varsigma(y)^{-1} \left( -z + \frac{1}{2} \text{sp} \left( \Sigma(y) \right) \right) \left( -z + \frac{1}{2} \text{sp} \left( \Sigma(y) \right) \right)' \varsigma(y)^{-1}'.
$$

Note that this function does not depend on the grid size $\Delta$. Lemma 3.2 then implies that

$$
f_{\Lambda}(\Upsilon(t-\Delta, t), \Upsilon_{t-\Delta}) = \Delta^{-2} \varsigma(\Upsilon_{t-\Delta})^{-1} \mathbb{E}_{t-\Delta}^*[\Upsilon_t] \mathbb{E}_{t-\Delta}^*[\Upsilon_t]' \varsigma(\Upsilon_{t-\Delta})^{-1}',
$$

(3.5)

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

Theorem 3.4 Under Model Assumptions 3.1 we obtain for all $K \in \mathbb{N}$ and $y \in \mathbb{R}^d$

$$
\mathbb{E}_0^* \left[ S_{(K)}(y) \right] = \Delta \Sigma(y) + \Delta^2 \varsigma(y) \left( \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}_0^* \left[ f_{\Lambda}(\Upsilon(\Delta(k - 1), \Delta k), \Upsilon_{\Delta k-\Delta}) \right] \right) \varsigma(y)'.
$$

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

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

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

$$
\mathbb{E}_0^* \left[ S_{(K)}(y) \right] = \Delta \Sigma(y) + \Delta^2 O(1), \quad \text{for } \Delta \to 0.
$$

(3.6)

That is, for small grid size $\Delta$ the second term becomes negligible.
In the general situation $\mathbb{P} \neq \mathbb{P}^*$ the previous calculation can be performed under the assumption that $\varepsilon_t$ under $\mathbb{P}$ shares all the properties of $\varepsilon^*_t$ under $\mathbb{P}^*$, except that we allow for a drift of order $O(\Delta)$, 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 $\varsigma(y)$, i.e. the volatility scaling factors. For $\vartheta \geq 0$ we define the function

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

As already remarked in Subsection 3.1, in the literature one often finds the square-root scaling, see [20], however for small rates a linear scaling can also be appropriate. For the Swiss currency CHF it turns out that the linear scaling is appropriate for a threshold of $\vartheta = 2.5\%$. In addition, we define the function $h(\cdot)$ as above to guarantee that the processes do not explode for large volatilities and small grid sizes.

We set for $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$

$$\varsigma(y) = \text{diag}(h(y_1), \ldots, h(y_d)). \hspace{1cm} (3.8)$$

Basically, volatility is scaled according to the actual observation $y$. This choice implies

$$\varsigma(y) \ KY = \frac{1}{\sqrt{K}} \varsigma(y) \left( \left( \varsigma(Y_{Dk-\Delta})^{-1} \ Y_{Dk} \right)_{j=1,\ldots,d; \ k=1,\ldots,K} \right) = \frac{1}{\sqrt{K}} \text{diag}(h(y_1), \ldots, h(y_d)) \left( \left( \text{diag}(h(Y_{Dk-\Delta}))^{-1} \ Y_{Dk} \right)_{j=1,\ldots,d; \ k=1,\ldots,K} \right).$$

These assumptions now allow to directly analyze the bias term given in (3.6). Therefore, we need to evaluate the function $f_\Lambda$ in Theorem 3.4. However, to this end we would need to know $\Sigma_\Lambda$, i.e. we obtain from Theorem 3.4 an implicit solution (quadratic form) that can be solved for $\Sigma_\Lambda$. We set $y = 1$ and obtain then from Theorem 3.4

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

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

$$\Delta^{-1} \left( \mathbb{E}_0^* \left[ S_{(K)}(1) \right] \right)_{ii} = s_{ii} + \frac{\Delta}{K} \sum_{k=1}^K \left( \mathbb{E}_0^* \left[ \frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i))} \right]^2 \right) + \frac{1}{4} \mathbb{E}_0^* \left[ h(Y(\Delta(k-1), \Delta k + m_i))^2 \right] s_{ii} - \mathbb{E}_0^* \left[ Y(\Delta(k-1), \Delta k) \right] s_{ii}. \hspace{1cm}$$
This is a quadratic equation that can be solved for \( s_{ii} \). Define

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

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

\[
c_i = -\Delta^{-1} \left( \mathbb{E}_0^*[S(K)(1)] \right)_{ii} + \frac{\Delta}{K} \sum_{k=1}^{K} \mathbb{E}_0^* \left[ \left( \frac{Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i))} \right)^2 \right],
\]

then we have \( a_i s_{ii}^2 + b s_{ii} + c_i = 0 \) which provides the solution

\[
s_{ii} = \frac{-b + \sqrt{b^2 - 4a_i c_i}}{2a_i}.
\]

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

\[
\beta_{ii} = \Delta^{-1} \left( \mathbb{E}_0^*[S(K)(1)] \right)_{ii} - s_{ii},
\]

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

\[
\Delta^{-1} \left( \mathbb{E}_0^*[S(K)(1)] \right)_{ij} = s_{ij} + \frac{\Delta}{K} \sum_{k=1}^{K} \mathbb{E}_0^* \left[ \frac{Y(\Delta(k-1), \Delta k) Y(\Delta(k-1), \Delta k)}{h(Y(\Delta(k-1), \Delta k + m_i)) h(Y(\Delta(k-1), \Delta k + m_j))} \right] \\
+ \frac{1}{4} \mathbb{E}_0^* \left[ h(Y(\Delta(k-1), \Delta k + m_i)) h(Y(\Delta(k-1), \Delta k + m_j)) \right] s_{ii} s_{jj} \\
- \frac{1}{2} \mathbb{E}_0^* \left[ Y(\Delta(k-1), \Delta k) h(Y(\Delta(k-1), \Delta k + m_i)) \right] s_{ij} \\
- \frac{1}{2} \mathbb{E}_0^* \left[ Y(\Delta(k-1), \Delta k) h(Y(\Delta(k-1), \Delta k + m_i)) \right] s_{jj}.
\]

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 Calibration

We assume that \( \mathbb{P} = \mathbb{P}^* \), i.e. we set the market price of risk identical 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 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 LIBOR\(^1\) (London InterBank Offered Rate) or the SAR\(^1\) (Swiss

\(^1\)data source: Swiss National Bank (SNB), www.snb.ch, for more information see also [26]
Average Rate), see Jordan [25], as (almost) risk-free financial instruments. The LIBOR is the rate at which highly-credited banks borrow and lend money at the inter-bank market. The SAR is the rate at which highly-credited banks borrow and lend money at the inter-bank market. The SAR is a rate determined by the Swiss National Bank (SNB) at which highly-credited institutions borrow and lend money with securitization. We display the yields of these two financial time series for instruments of a time to maturity of 3 months, see Figure 1. We see that the SAR yield typically lies below the LIBOR yield (due to securitization). 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 yields of sufficiently highly rated countries) or from swap rates, the former is described in [26] for the Swiss currency market. In Figure 2 we give the time series of the Swiss government bond yield and the CHF swap yield both for a time to maturity of 5 years. We see that the rates of the Swiss government bonds are below the 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 yield curve as approximation to the risk-free yield curve data for long times to maturity.

We mention that these short terms and long terms data are not completely compatible which may give some difficulties in the calibration. We will also see this in the correlation matrices below.

Thus, for our analysis we choose the SAR for times to maturity \( m \in \{1/52, 1/26, 1/12, 1/4\} \), and for times to maturity \( m \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30\} \) we choose the Swiss government bond inferred yield curve. We choose grid size \( \Delta = 1/52 \) (i.e. a weekly time grid) and then we calculate \( \Upsilon_t \) 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

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

In Figure 3 we give the time series of these estimated \( (\Upsilon_t)_t \) and in Figure 4 we give the component-wise ordered time series obtained from \( (\Upsilon_t)_t \). We observe that the volatility is increasing in the time to maturity due to scaling with time to maturity. Using (3.8) we calculate

\[
\sqrt{K} C(K) = \left( [\varsigma(Y_{\Delta k - \Delta})^{-1} \Upsilon_{\Delta k}]_j \right)_{j=1,\ldots,d; k=1,\ldots,K} \in \mathbb{R}^{d \times K}
\]

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 maturities \( m = 1/52 \) and \( m = 5 \). We observe that the scaling \( \varsigma(Y_{\Delta k - \Delta})^{-1} \) gives more stationarity for short times to maturity, however in financial stress periods it substantially increases the volatility of the observations, see Figure 5. In view of Figure 6 one might discuss or even question the scaling for longer times to maturity.

---

2 data source: Swiss National Bank (SNB), www.snb.ch, for more information see also [26]

3 data source: Bloomberg
because it is less obvious from the graph whether it is needed. Next figures will show that this scaling is also needed for longer times to maturity. We then calculate the observed matrix

\[
\left( \hat{s}_{ij}(K) \right)_{i,j=1,...,d} = \Delta^{-1} S(K)(1)
\]

as a function of the number of observations \( K \) (we set \( \mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^d \)). 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 \( \hat{s}_{ij}(K) \) for \( s_{ij} \) as a function of the number of observations \( K \). The bias correction term is estimated by

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

We expect that for short times to maturity the bias correction term is larger due to more dramatic drifts. The results for selected times to maturity \( m \in \{1/52, 1/4, 1, 5, 20\} \) are presented in Figures 7-11. Let us comment these figures:

- Times to maturity in the set \( \mathcal{M}_1 = \{1/52, 1/26, 1/12\} \) look similar to \( m = 1/52 \) (Figure 7); \( \mathcal{M}_2 = \{1/4\} \) corresponds to Figure 8; times to maturity in the set \( \mathcal{M}_3 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15\} \) look similar to \( m = 1, 5 \) (Figures 9-10); times to maturity \( m \in \mathcal{M}_4 = \{20, 30\} \) look similar to \( m = 20 \) (Figure 11).

- Times to maturity in \( \mathcal{M}_1 \cup \mathcal{M}_3 \) seem to have converged, for \( \mathcal{M}_2 \) the convergence picture is distorted by the last financial crisis, where volatilities relative to yields have substantially increased, see also Figure 5. One might ask whether during financial crisis 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, this 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 maturities in \( \mathcal{M}_3 \cup \mathcal{M}_4 \) we observe that the bias term given in (3.6) is negligible, see Figures 9-11, that is, \( \Delta = 1/52 \) is sufficiently small for times to maturity \( m \geq 1 \). For times to maturities in \( \mathcal{M}_1 \cup \mathcal{M}_2 \) it is however essential that we do a bias correction, see Figure 7-8. This comes from the fact that for small times to maturity the bias term is driven by \( z \) in \( f_\Lambda(z, y) \) which then is of similar order as \( s_{ii} \).

In Table 1 we present the resulting estimated matrix \( \hat{\Sigma}_\Lambda(\mathbf{1}) = (\hat{s}_{ij}(K))_{i,j=1,...,d} \) which is based on all observations in \( \{01/2000, \ldots, 05/2011\} \). We observe that the diagonal \( \hat{s}_{ii}(K) \) 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)

\[
\hat{\Xi} = (\hat{\rho}_{ij})_{i,j=1,...,d} = \left( \frac{\hat{s}_{ij}(K)}{\sqrt{\hat{s}_{ii}(K)\hat{s}_{jj}(K)}} \right)_{i,j=1,...,d}.
\]

Now all entries \( \hat{\rho}_{ij} \) live on the same scale and the result is presented in Figure 12. We observe two different structures, one for times to maturity less than 1 year, i.e. \( m \in \tilde{\mathcal{M}}_1 = \mathcal{M}_1 \cup \mathcal{M}_2 \), and
one for times to maturity \( m \in \tilde{M}_2 = M_3 \cup M_4 \). The former times to maturity \( m \in \tilde{M}_1 \) were modeled using the observations from the SAR, the latter \( m \in \widetilde{M}_2 \) with observations from the Swiss government bond yields. This separation shows that these two data sets are not completely compatible which gives some “additional independence” (diversification) between \( \tilde{M}_1 \) and \( \widetilde{M}_2 \).

If we calculate the eigenvalues of \( \hat{\Sigma} \) we observe that the first 3 eigenvalues explain about 90% of the total cross-sectional volatility and the first 5 eigenvalues explain about 96% (we have a \( d = 17 \) dimensional space). Thus, a principal component analysis says that we should at least choose a 5-factor model to simultaneously model all SAR and Swiss government bond implied yields in \( M \). These are more factors than typically stated in the literature (see Section 4.1 in [7]). The reason therefore is again that the short end \( \tilde{M}_1 \) and the long end \( \widetilde{M}_2 \) of the estimated yield curve behave more independently due to different sources of the data (see also Figure 12).

If we restrict this principal component analysis to \( \widetilde{M}_2 \) we find the classical result that a 3-factor model explains 95% of the observed cross-sectional volatility.

In the next step we analyze the assumption of the independence of \( \Sigma_\Lambda(1) = \Lambda \Lambda' = (s_{ij})_{i,j=1,\ldots,d} \) from the grid size \( \Delta \). Similar to the analysis above we estimate \( \Sigma_\Lambda(1) \) 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 \( \Delta \) (for illustrative purposes one should compare Figure 9 with \( m = 1 \) and \( \Delta = 1/52 \) and Figure 13 with \( m = 1 \) and \( \Delta = 1/4 \)). Of course, this is exactly the result expected.

In Table 2 we give the differences between the estimated matrices \( \hat{\Sigma}_\Lambda(1) = (\hat{s}_{ij}(K))_{i,j=1,\ldots,d} \) 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 \( m \in \widetilde{M}_2 \cup \tilde{M}_2 \) because in the latter model the times to maturity in \( M_1 \) do not exist. We observe rather small differences within \( \tilde{M}_2 \) which supports the independence assumption from the choice of \( \Delta \) within the Swiss government bond yields. For the SAR in \( M_2 \) 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 8. Thus, we only observe larger differences for covariances that have a bigger difference in times to maturity compared. 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 \( \Sigma_\Lambda(1) \) from \( \Delta \) is not violated by our observations and that the bias terms \( \hat{\beta}_{ij}(K) \) are negligible for maturities \( m_i, m_j \in \tilde{M}_2 \) 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 \( \tilde{M}_2 \).

**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 with nominal payments of size 1 at times to maturity dates \( m \in M_3 \). The present
value of this annuity at time $t$ is given by
\[ \pi_t = \sum_{m \in M_3} P(t, t + m) = \sum_{m \in M_3} \exp \left\{ -m \ Y(t, t + m) \right\} \approx \sum_{m \in M_3} 1 - m \ Y(t, t + m) \overset{\text{def.}}{=} \tilde{\pi}_t. \]

Our (out-of-sample) back-testing setup is such that we try to predict $\tilde{\pi}_t$ based on the 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 $\Delta$ (for $t = \Delta(K + 1)$). Moreover, the Taylor approximation $\tilde{\pi}_t$ to $\pi_t$ 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 – quite large in comparison to possible second order – drift like – correction terms. Such an approximation does not work for short-long portfolios.

For the following approximation (under $\mathbb{P}^*$)
\[ Y_{t|\mathcal{F}_{t-\Delta}} \overset{(d)}{=} \kappa_t(Y(t - \Delta, t), Y_{t-\Delta}^*)|\mathcal{F}_{t-\Delta}, \]

we obtain an approximate forecast to $\tilde{\pi}_t$ given by (denote the cardinality of $M_3$ by $d_3$)
\[ \tilde{\pi}_{t|\mathcal{F}_{t-\Delta}} = d_3 - \sum_{m \in M_3} (m + \Delta) Y(t - \Delta, t + m) + d_3 \Delta Y(t - \Delta, t) - \frac{1}{2} 1_{M_3} \ sp \left( S_{(K)}(Y_{t-\Delta}^*) \right) - 1_{M_3} \ sp \left( Y_{t-\Delta}^* \right) C_{(K)} \ W_t^* |\mathcal{F}_{t-\Delta}, \]

where $1_{M_3} = (1_{\{1 \in M_3\}}, \ldots, 1_{\{d \in M_3\}})^t \in \mathbb{R}^d$. Thus, the conditional distribution of $\tilde{\pi}_t$ under $\mathbb{P}^*$, given $\mathcal{F}_{t-\Delta}$, is a Gaussian distribution with conditional mean and conditional variance given by
\[ \mu_{t-\Delta} = d_3 - \sum_{m \in M_3} (m + \Delta) Y(t - \Delta, t + m) + d_3 \Delta Y(t - \Delta, t) - \frac{1}{2} 1_{M_3} \ sp \left( S_{(K)}(Y_{t-\Delta}^*) \right), \]
\[ \tau_{t-\Delta}^2 = 1_{M_3} S_{(K)}(Y_{t-\Delta}^*) 1_{M_3}. \]

We successively calculate these conditional moments for $t \in \{01/2005, \ldots, 05/2011\}$ based on the $\sigma$-fields $\mathcal{F}_{t-\Delta}$ generated by the data in $\{01/2000, \ldots, t - \Delta\}$, for $\Delta = 1/52, 1/12$ (weekly and monthly grid). From these we can calculate the observable residuals
\[ z_{t-\Delta}^* = \frac{\tilde{\pi}_t - \mu_{t-\Delta}^*}{\tau_{t-\Delta}}, \]

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 14 and for $\Delta = 1/12$ in Figure 15. At the first sight this sequence $(z_{t-\Delta}^*)_t$ seems to fulfill these requirements, thus the out-of-sample back-testing provides the required results. In Figure 16 we also provide the Q-Q-plot of these residuals $(z_{t-\Delta}^*)_t$ 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 suggest that one may relax the Gaussian assumption on $\varepsilon_t^*$ by a more heavy-tailed model (this can also be seen in Figure 14 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 $\Delta$ between the residuals $z^*_t$ we obtain 5% which is a convincingly small value. This supports the assumption having independent residuals. The same holds true if we consider the auto-correlation for time lag $\Delta$ 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 14. 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

$$\xi_t = \prod_{s=1}^{t/\Delta} \exp \left\{ -\frac{1}{2} \| \lambda_{s\Delta} \|^2 + \lambda_{s\Delta} \varepsilon_{s\Delta} \right\}, \quad (4.2)$$

with $\varepsilon_t$ is independent of $\mathcal{F}_{t-\Delta}$, $\mathcal{F}_t$-measurable and a $t/\Delta$-dimensional standard Gaussian random vector with independent components under $\mathbb{P}$. Moreover, it is assumed that $\lambda_t$ 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 $(\mathbb{P}, \mathcal{F})$-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

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

$\lambda_t$ 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 $\mathbb{P}$

$$\Upsilon_t = \Delta \left[ -Y(t - \Delta, t) + \frac{1}{2} \text{sp}(\Sigma_{1\Delta}(Y_{t-\Delta}^-)) \right] + \sqrt{\Delta} \varsigma(Y_{t-\Delta}^-) \Lambda \lambda_t + \sqrt{\Delta} \varsigma(Y_{t-\Delta}^-) \Lambda \varepsilon_t,$$

i.e. we have a change of drift given by $\sqrt{\Delta} \varsigma(Y_{t-\Delta}^-) \Lambda \lambda_t$. 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

$$\mu_{t-\Delta} = \mu^*_{t-\Delta} - \sqrt{\Delta} \mathbf{1}_{\mathcal{M}_3} \varsigma(Y_{t-\Delta}^-) \Lambda \lambda_t \quad \text{and} \quad \tau^2_{t-\Delta} = \mathbf{1}'_{\mathcal{M}_3} S_{(K)}(Y_{t-\Delta}^-) \mathbf{1}_{\mathcal{M}_3}.$$

For an appropriate choice of the market price of risk $\lambda_t$ we obtain residuals

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

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 \( \tau_{m,t} \). In Figures 17, 18 and 19 we show the results for \( m = 1, 5, 10 \). The picture is similar to Figure 14, i.e. we observe clustering but not a well-defined drift. This implies that we may set the market price of 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 Vasiček model

We compare our HJM framework to interest rate models based on short rate modeling. We compare our findings to the results of the Vasiček model [31]. The Vasiček model is the simplest short rate model that provides an affine term structure for interest rates (see also [17]), and hence a closed-form solution for ZCB prices. Note that the Vasiček 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-Ingersoll-Ross [12] or the Black-Karasinski [6] models.

The price of the ZCB in the Vasiček 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 the time to maturity \( m \) and the model parameters \( \kappa^*, \theta^* \) and \( g \) (see for instance (3.8) in [7]). The short rate \( r_t \) is then under \( \mathbb{P}^*_{t-\Delta} \) normally distributed with conditional mean and conditional variance given by

\[
\mathbb{E}^*_{t-\Delta}[r_t] = r_{t-\Delta} e^{-\Delta \kappa^*} + \theta^* \left( 1 - e^{-\Delta \kappa^*} \right),
\]

\[
\text{Var}^*_{t-\Delta}(r_t) = \frac{g^2}{2 \kappa^*} \left[ 1 - e^{-2 \kappa^* \Delta} \right].
\]

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

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

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 from a time series of observations \( \kappa^*, \theta^* \) and \( g \), for instance, using maximum likelihood methods (see (3.14)-(3.16) in [7]); (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

\[
v^*_t = \frac{\tilde{\pi}_t - \mathbb{E}^*_{t-\Delta}[\tilde{\pi}_t]}{\text{Var}^*_{t-\Delta}(\tilde{\pi}_t)^{1/2}}.
\]
In Figure 20 we plot the time series $z_t^*$ and $v_t^*$ for $t \in \{01/2005, \ldots, 05/2011\}$. The observation is that $v_t^*$ is far too small! The explanation for this observation lies in the assumption $P^* = \mathbb{P}$, i.e. $\lambda_t = 0$. Since the Vasiček prices are calculated by conditional expectations of the entire future development of the short rate $r_t$ until expiry of the ZCB, the choices of $\tilde{\kappa}^*$, $\tilde{\theta}^*$ and $\tilde{g}$ have a huge influence on the resulting ZCB prices in the Vasiček model. Thus, the calibration of $\hat{A}(\cdot)$ and $\hat{B}(\cdot)$ is completely inappropriate if we set $\lambda_t = 0$. Compare

$$\log P(t, t + m) = -m Y(t, t + m), \quad (4.3)$$

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

The (pricing) functions $A(\cdot)$ and $B(\cdot)$ in (4.4) are calculated completely within the Vasiček model by a forward projection of $r_t$ 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. Hence, the Vasiček model, as any other factor 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).

**4.4 Forward projection of yield curves and arbitrage**

For the 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 that we predict future yield curves within these families $\mathcal{M}$ and then approximate the remaining times to maturity using a parametric family like the Nelson-Siegel [27] or the Svensson [29, 30] family, see [17].

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 $w_t P(t, t + m_1) - P(t, t + m_2)$ for two different times to maturity $m_1$ and $m_2$. We approximate this portfolio by a Taylor expansion up to order 2 and set

$$\tilde{\pi}_t = w_t \left( 1 - m_1 Y(t, t + m_1) + \frac{(m_1 Y(t, t + m_1))^2}{2} \right) - \left( 1 - m_2 Y(t, t + m_2) + \frac{(m_2 Y(t, t + m_2))^2}{2} \right).$$
Under our model assumptions, the returns of both terms \( m_i Y(t, t + m_i) \) in portfolio \( \bar{\pi}_t \) have, conditionally given \( \mathcal{F}_{t-\Delta} \), a Gaussian distribution term with standard deviations given by

\[
\tau_{t-\Delta}^{(i)} = \sqrt{\text{det} \left( \text{Var} \left( Y_{t-\Delta} \right) \right)_{i,i}} \quad \text{for } i = 1, 2.
\]

If we choose \( w_t = \tau_{t-\Delta}^{(2)}/\tau_{t-\Delta}^{(1)} \) then the returns of the Gaussian parts of both terms in portfolio \( \bar{\pi}_t \) 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 correction) we see that the returns of the portfolio \( \bar{\pi}_t \) should provide zero returns conditionally. In Figure 21 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 \( w_t \) are in the range between 1.4 and 1.9. In Figure 21 we plot the aggregated realized gains of the portfolio \( \bar{\pi} \) minus their prognosis including and excluding the HJM correction term. Recall that the predicted gains should be zero conditionally on the current information. We observe that the model without the HJM term clearly 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 [13]. 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 [13] and opens a whole new area of modeling questions to be solved.

### 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)

\[
\exp \{-\Delta Y(t, t+\Delta)\} \mathbb{E}^{t-\Delta}_{t} [P(t, t+m)] = P(t-\Delta, t) \mathbb{E}^{t-\Delta}_{t} [P(t, t+m)] \\
= P(t-\Delta, t+m) \exp \{-\alpha\Delta(t, m_i Y_s)_{s \leq t-\Delta}\} \mathbb{E}^{t-\Delta}_{t} [\exp \{-\nu\Delta(t, m_i Y_s)_{s \leq t-\Delta}\} \mathbb{E}^{t}_{t}]
\]

Solving this requirement proves the claim of Lemma 2.1.

\[ \square \]

**Proof of Theorem 3.4.** In the first step we apply the tower property for conditional expectation which decouples the problem into several steps. We have \( \mathbb{E}^{t}_{t} \left[ S_{(K)}(y) \right] = \mathbb{E}^{t}_{t} \left[ \mathbb{E}^{t}_{t} \left[ S_{(K)}(y) \right] \mathbb{E}^{t}_{t} \left[ S_{(K)}(y) \right] \right] \). Thus, we need to calculate the inner conditional expectation \( \mathbb{E}^{t}_{t} \left[ S_{(K)}(y) \right] \) of the \( d \times d \) matrix \( S_{(K)}(y) \). We define the auxiliary matrix

\[
\tilde{C}_{(K)} = \left( \left( \bar{s}(Y_{t-\Delta})^{-1} Y_{t-\Delta} \right)_{j=1,\ldots,d; k=1,\ldots,K} \right) \in \mathbb{R}^{d \times K}.
\]

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

\[
\tilde{C}_{(K)} = \left[ \bar{C}_{(K-1)}; \bar{s}(Y_{t-\Delta})^{-1} Y_{t-\Delta} \right],
\]
with $\tilde{C}_{(K-1)} \in \mathbb{R}^{d \times (K-1)}$ is $\mathcal{F}_{\Delta(K-1)}$-measurable. This implies the following decomposition
\[
S_{(K)}(y) = \frac{1}{K} \langle y \rangle \tilde{C}_{(K)} \tilde{C}_{(K)}' \langle y \rangle' = \frac{1}{K} \langle y \rangle \left[ \tilde{C}_{(K-1)}, \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \left[ \tilde{C}_{(K-1)}, \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right]' \langle y \rangle' \right] = \frac{1}{K} \langle y \rangle \left( \tilde{C}_{(K-1)} \tilde{C}_{(K-1)}' + \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle' \right) \langle y \rangle' \right) = \frac{K-1}{K} S_{(K-1)}(y) + \frac{1}{K} \langle y \rangle \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \mathbf{T}_{\Delta k} \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right)' \langle y \rangle'.
\]
This implies for the conditional expectation of $S_{(K)}(y)$
\[
E_{\Delta(K-1)} [S_{(K)}(y)] = \frac{K-1}{K} S_{(K-1)}(y) + \frac{1}{K} \langle y \rangle \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \mathbf{T}_{\Delta k} \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right)' \langle y \rangle'.
\]

We calculate the conditional expectation in the last term, we start with the conditional covariance. From Lemma 3.2 we obtain
\[
\frac{1}{K} \langle y \rangle \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \operatorname{Cov}_{\Delta(K-1)} \left( \mathbf{Y}_{\Delta k} \right) \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right)' \langle y \rangle' = \frac{\Delta}{K} \langle y \rangle \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \Sigma_{\Delta} \left( \mathbf{Y}_{\Delta k-\Delta} \right) \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right)' \langle y \rangle' = \frac{\Delta}{K} \Sigma_{\Delta}(y).
\]
This implies
\[
E_{0} [S_{(K)}(y)] = \frac{K-1}{K} E_{0} [S_{(K-1)}(y)] + \frac{\Delta}{K} \Sigma_{\Delta}(y)
+ \frac{1}{K} \langle y \rangle E_{0} \left[ \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \mathbf{T}_{\Delta k} \left( \langle \mathbf{Y}_{\Delta k-\Delta}^{-1} \mathbf{Y}_{\Delta k} \rangle \right)' \langle y \rangle' \right] = \frac{K-1}{K} E_{0} [S_{(K-1)}(y)] + \frac{\Delta}{K} \Sigma_{\Delta}(y) + \frac{\Delta^{2}}{K} \langle y \rangle E_{0} [f_{\Delta}(\mathbf{Y}(\Delta(K-1), \Delta K), \mathbf{Y}_{\Delta k-\Delta})] \langle y \rangle'.
\]

Iterating this provides the result.

\[
\square
\]

References


Figure 1: Yield curve time series 3 months SAR and 3 months CHF LIBOR from 01/2000 until 05/2011. The spread gives the difference between these two time series.

Figure 2: Yield curve time series Swiss government bond and CHF swap rate both for time to maturity $m = 5$ years. The spread gives the difference between these two time series (roughly 30 basis points).
Figure 3: Time series $\mathbf{\Upsilon}_t$ for $t \in \{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

Figure 4: Component-wise ordered time series obtained from $\mathbf{\Upsilon}_t$ for $t \in \{01/2000, \ldots, 05/2011\}$, i.e. $\mathbf{\Upsilon}(t,m) \leq \mathbf{\Upsilon}(t+1,m)$ for all $t$ and $m \in \mathcal{M}$ on a weekly grid $\Delta = 1/52$. 

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

Figure 6: Time series $\Upsilon_{t,m}$ and $[\sqrt{K} C_{(K)}]_m = \Upsilon_{t,m}/h(Y(t - \Delta, t + m))$ for maturity $m = 5$ and $t \in \{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. 
Figure 7: Time series $\hat{s}^{\text{bias}}_{ii}(K)$ and $\hat{s}_{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_i = 1$ week and observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

Figure 8: Time series $\hat{s}^{\text{bias}}_{ii}(K)$ and $\hat{s}_{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_i = 3$ months and observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. 
Figure 9: Time series $\hat{s}_{ii}^{\text{bias}}(K)$ and $\hat{s}_{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_i = 1$ year and observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

Figure 10: Time series $\hat{s}_{ii}^{\text{bias}}(K)$ and $\hat{s}_{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_i = 5$ years and observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. 
Figure 11: Time series $\hat{s}_{ii}(K)$ and $\tilde{s}_{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_i = 20$ years and observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$.

Figure 12: Estimated matrix $\hat{\Xi} = (\hat{\rho}_{ij})_{i,j=1,\ldots,d}$ from all observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. 
Figure 13: Time series $\hat{s}_{i}^{\text{bias}}(K)$ and $\hat{s}_{i}^{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_{i} = 1$ year and observations in {01/2000, \ldots, 05/2011} on a monthly grid $\Delta = 1/12$. 

---

**Figure 13**: Time series $\hat{s}_{i}^{\text{bias}}(K)$ and $\hat{s}_{i}^{ii}(K)$, $K = 1, \ldots, 600$, for maturity $m_{i} = 1$ year and observations in {01/2000, \ldots, 05/2011} on a monthly grid $\Delta = 1/12$. 

---

29
Figure 14: Time series of residuals $z_t^*$ for $t \in \{01/2005, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{t-\Delta}$.

Figure 15: Time series of residuals $z_t^*$ for $t \in \{01/2005, \ldots, 05/2011\}$ on a monthly grid $\Delta = 1/12$. The axis on the right-hand side displays the time series of $\tau_{t-\Delta}$. 
Figure 16: Q-Q-plot of the residuals \( (z_t^n) \) against the standard Gaussian distribution for \( \Delta = 1/52 \).

Figure 17: Time series of residuals \( z_{m,t}^\tau \) for time to maturity \( m = 1 \) and \( t \in \{01/2005, \ldots, 05/2011\} \) on a weekly grid \( \Delta = 1/52 \). The axis on the right-hand side displays the time series of \( \tau_{m,t-\Delta} \).
Figure 18: Time series of residuals $z_{m,t}^*$ for time to maturity $m = 5$ and $t \in \{01/2005, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{m,t-\Delta}$.

Figure 19: Time series of residuals $z_{m,t}^*$ for time to maturity $m = 10$ and $t \in \{01/2005, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. The axis on the right-hand side displays the time series of $\tau_{m,t-\Delta}$. 

32
Figure 20: Time series of residuals $z^*_t$ and $v^*_t$ for $t \in \{01/2005, \ldots, 05/2011\}$ on a monthly grid $\Delta = 1/12$ under the assumption $\mathbb{P}^* = \mathbb{P}$.

Figure 21: Back testing the difference of aggregated realized gains of portfolio $\bar{\pi}_t$ for $w_t = \frac{\tau_t^{(2)}}{\tau_t^{(1)}}$ and their model prognosis with and without the no-arbitrage HJM correction term.
<table>
<thead>
<tr>
<th>Time Period</th>
<th>1 week</th>
<th>2 weeks</th>
<th>1M</th>
<th>3M</th>
<th>1Y</th>
<th>2Y</th>
<th>3Y</th>
<th>4Y</th>
<th>5Y</th>
<th>6Y</th>
<th>7Y</th>
<th>8Y</th>
<th>9Y</th>
<th>10Y</th>
<th>15Y</th>
<th>20Y</th>
<th>30Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 week</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0011</td>
</tr>
<tr>
<td>2 weeks</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0011</td>
</tr>
<tr>
<td>1 month</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0010</td>
</tr>
<tr>
<td>3 months</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0008</td>
<td>0.0019</td>
<td>0.0024</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0018</td>
<td>0.0019</td>
<td>0.0025</td>
<td>0.0042</td>
</tr>
<tr>
<td>1 year</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0019</td>
<td>0.0142</td>
<td>0.0164</td>
<td>0.0164</td>
<td>0.0158</td>
<td>0.0148</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
<td>0.0153</td>
</tr>
<tr>
<td>2 years</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0024</td>
<td>0.0164</td>
<td>0.0309</td>
<td>0.0361</td>
<td>0.0373</td>
<td>0.0376</td>
<td>0.0380</td>
<td>0.0388</td>
<td>0.0397</td>
<td>0.0410</td>
<td>0.0423</td>
<td>0.0500</td>
<td>0.0590</td>
<td>0.0658</td>
</tr>
<tr>
<td>3 years</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0021</td>
<td>0.0158</td>
<td>0.0361</td>
<td>0.0481</td>
<td>0.0535</td>
<td>0.0562</td>
<td>0.0581</td>
<td>0.0600</td>
<td>0.0619</td>
<td>0.0641</td>
<td>0.0663</td>
<td>0.0719</td>
<td>0.0907</td>
<td>0.1019</td>
</tr>
<tr>
<td>4 years</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0018</td>
<td>0.0161</td>
<td>0.0373</td>
<td>0.0535</td>
<td>0.0632</td>
<td>0.0693</td>
<td>0.0735</td>
<td>0.0771</td>
<td>0.0801</td>
<td>0.0830</td>
<td>0.0857</td>
<td>0.0990</td>
<td>0.1126</td>
<td>0.1327</td>
</tr>
<tr>
<td>5 years</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0017</td>
<td>0.0168</td>
<td>0.0376</td>
<td>0.0562</td>
<td>0.0693</td>
<td>0.0787</td>
<td>0.0851</td>
<td>0.0904</td>
<td>0.0947</td>
<td>0.0986</td>
<td>0.1021</td>
<td>0.1173</td>
<td>0.1310</td>
<td>0.1581</td>
</tr>
<tr>
<td>6 years</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0016</td>
<td>0.0173</td>
<td>0.0380</td>
<td>0.0581</td>
<td>0.0735</td>
<td>0.0851</td>
<td>0.0936</td>
<td>0.1007</td>
<td>0.1065</td>
<td>0.1116</td>
<td>0.1161</td>
<td>0.1342</td>
<td>0.1485</td>
<td>0.1786</td>
</tr>
<tr>
<td>7 years</td>
<td>0.0007</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0017</td>
<td>0.0176</td>
<td>0.0388</td>
<td>0.0600</td>
<td>0.0771</td>
<td>0.0904</td>
<td>0.1007</td>
<td>0.1095</td>
<td>0.1169</td>
<td>0.1234</td>
<td>0.1292</td>
<td>0.1520</td>
<td>0.1674</td>
<td>0.1969</td>
</tr>
<tr>
<td>8 years</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0017</td>
<td>0.0176</td>
<td>0.0397</td>
<td>0.0619</td>
<td>0.0801</td>
<td>0.0947</td>
<td>0.1065</td>
<td>0.1169</td>
<td>0.1259</td>
<td>0.1340</td>
<td>0.1413</td>
<td>0.1700</td>
<td>0.1871</td>
<td>0.2135</td>
</tr>
<tr>
<td>9 years</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0018</td>
<td>0.0175</td>
<td>0.0410</td>
<td>0.0641</td>
<td>0.0830</td>
<td>0.0986</td>
<td>0.1116</td>
<td>0.1234</td>
<td>0.1340</td>
<td>0.1438</td>
<td>0.1528</td>
<td>0.1883</td>
<td>0.2078</td>
<td>0.2297</td>
</tr>
<tr>
<td>10 years</td>
<td>0.0007</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0019</td>
<td>0.0173</td>
<td>0.0423</td>
<td>0.0663</td>
<td>0.0857</td>
<td>0.1021</td>
<td>0.1161</td>
<td>0.1292</td>
<td>0.1413</td>
<td>0.1528</td>
<td>0.1635</td>
<td>0.2064</td>
<td>0.2289</td>
<td>0.2462</td>
</tr>
<tr>
<td>15 years</td>
<td>0.0009</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0025</td>
<td>0.0173</td>
<td>0.0500</td>
<td>0.0779</td>
<td>0.0990</td>
<td>0.1173</td>
<td>0.1342</td>
<td>0.1520</td>
<td>0.1700</td>
<td>0.1883</td>
<td>0.2064</td>
<td>0.2869</td>
<td>0.3320</td>
<td>0.3498</td>
</tr>
<tr>
<td>20 years</td>
<td>0.0011</td>
<td>0.0014</td>
<td>0.0016</td>
<td>0.0042</td>
<td>0.0208</td>
<td>0.0590</td>
<td>0.0907</td>
<td>0.1126</td>
<td>0.1310</td>
<td>0.1485</td>
<td>0.1674</td>
<td>0.1871</td>
<td>0.2078</td>
<td>0.2289</td>
<td>0.3320</td>
<td>0.4247</td>
<td>0.5215</td>
</tr>
<tr>
<td>30 years</td>
<td>0.0017</td>
<td>0.0023</td>
<td>0.0028</td>
<td>0.0064</td>
<td>0.0371</td>
<td>0.0658</td>
<td>0.1019</td>
<td>0.1327</td>
<td>0.1581</td>
<td>0.1786</td>
<td>0.1969</td>
<td>0.2135</td>
<td>0.2297</td>
<td>0.2462</td>
<td>0.3498</td>
<td>0.5215</td>
<td>0.9860</td>
</tr>
</tbody>
</table>

Table 1: Estimated matrix $\hat{\Sigma}_A(1) = (\hat{s}_{ij}(K))_{i,j=1,\ldots,d}$ based on all observations in $\{01/2000, \ldots, 05/2011\}$ on a weekly grid $\Delta = 1/52$. 
Table 2: Estimated matrices $\hat{\Sigma}_\Lambda(1) = (\hat{s}_{ij}(K))_{i,j=1,...,d}$ based on all observations in \{01/2000, \ldots, 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).