Estimation of Shadow-Rate Term Structure Models Near the Zero-Lower Bound

Shabbirhussein Esmail

Supervisor: A/Prof. Peter Ouwehand

A dissertation submitted to the Faculty of Commerce, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

August 29, 2019

MPhil in Mathematical Finance,
University of Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy to the University of Cape Town. It has not before been submitted for any degree or examination.

Signed by candidate

Shabbirhussein Esmail

August 29, 2019

Abstract

Though it is customary to use standard Gaussian term structure models for term structure modelling, this becomes theoretically implausible in cases when nominal interest rates are near zero: Gaussian models can have arbitrarily large negative rates, whereas arbitrage considerations dictate that rates should remain positive (or very slightly negative at most). Black (1995) suggests that interest rates include an optionality which restricts them to non-negative values. This introduces a non-linearity at the zero-lower bound that makes these so-called shadow-rate models a computational challenge. This dissertation analyses the shadow-rate approximations suggested by Krippner (2013) and Priebsch (2013) for the Vasiček and arbitrage-free Nelson-Siegel (AFNS) models. We also investigate and compare the accuracy of the iterated extended Kalman filter (IEKF) with that of the unscented Kalman filter (UKF). We find that Krippner's approach approximates interest rates within reasonable bounds for both the 1-factor Vasiček and AFNS models. Priebsch's first-cumulant method is more accurate than Krippner's method for a 1-factor Vasiček model, while Priebsch's second-cumulant method is deemed impractical because of the computational time it takes. In a multi-factor AFNS model, only Krippner's framework is feasible. Moreover, the IEKF outperforms the UKF in terms of filtering with no significant difference in run-time.

Acknowledgements

"And whoever is grateful, he is only grateful for the benefit of his own self." (Luqman 31:12). "For surely if it was not for the grace of God on you and His mercy, you would have been among the losers" (Al-Baqarah 2:64).

I would like to thank Professor Peter Ouwehand for making my Masters dissertation a great learning experience. From the excellent recommendations to creating a platform for me to finish my dissertation timeously, I was indeed fortunate to have him as my supervisor.

Further, my sincere thanks goes to AIFMRM and the National Research Foundation (NRF) for providing me with the resources to make this journey possible. Also, to my friends who spent many days and nights working in solidarity, thank you for making this a memorable experience!

Finally, I would like to thank my family for supporting, advising and uplifting me during the tough time. My sisters, Rooshna and Arzina, thank you so much for all the support. Most importantly, I would like to send prayers to my parents, without whom, this journey would not have been possible.

My sincere thanks is also offered to all unmentioned people who have contributed towards the completion of this dissertation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The affine Gaussian dynamic term structure model is regarded as the "workhorse" for bond pricing. Closed-form solutions for yields exist in this type of model, which allow for efficient calibration (Christensen and Rudebusch (2014)). However, these models are only appropriate when nominal interest rates are relatively high since the probability of negative rates is very small (Christensen and Rudebusch (2014)). When nominal rates approach zero or go negative (as in Japan, Europe and the U.S – after the financial crisis towards the end of 2008 (Christensen and Rudebusch (2014) and Kim and Priebsch (2013))), Gaussian term structure models assign non-negligible positive probabilities to negative rates. The flaw arises when the option of currency is ignored, i.e. an investor has an option to hold cash ("mattress arbitrage") at zero interest rate (Black (1995)). This means that even though negative interest rates may exist in a real economy due to market frictions, theoretically nominal rates must be non-negative to eliminate arbitrage opportunities.

Tractable short rate models that limit interest rates to non-negative values do exist, e.g. affine model of square-root type and quadratic-Gaussian models. Nonetheless, these models suffer from diminishing volatilities when interest rates approach the zero-lower bound (ZLB) (Gorovoi and Linetsky (2004)). Additionally, the ZLB in these models is a reflecting barrier and not a sticky one – as in reality (Christensen and Rudebusch (2014)).

To account for these empirical problems as well as accommodate for the existence of currency as an option in bond pricing, Black (1995) introduced a notion of a shadow-rate s_t by setting the risk-less rate r_t equal to the shadow-rate except limiting the risk-less rate to zero when the shadow-rate goes negative i.e.

$$r_t = \max\{s_t, 0\}$$

The major factor that has limited the use of the shadow-rate structure is the difficulty in estimating these non-linear models (Christensen and Rudebusch (2014)).

Gorovoi and Linetsky (2004) use the method of eigenfunction expansion to derive analytical solutions for bond prices in a one-factor shadow-rate model. Un-

fortunately, their method does not generalize to multi-factor models. Some researchers (Kim and Singleton (2012) and Ichiue and Ueno (2007)) approximate shadow-rate models using two factors. However, their approach is prone to the curse of dimensionality. Bauer and Rudebusch (2013) use Monte Carlo simulation to evaluate bond prices for model parameters obtained from a Gaussian term structure model without any constraints, but their estimation is limited to a short rate model due to the computational burden of shadow-rate version of the model. Christensen and Rudebusch (2014) estimate three factor shadow-rate Nelson-Siegel models using an option-based yield formula proposed by Krippner (2013). However, the option-based derivation does not comply with the usual arbitrage-free approach. Priebsch (2013) proposes a cumulant-based approach to approximate arbitrage-free yields in the shadow-rate model by expanding a first- and second-order cumulant-generating function. Unfortunately, his approach requires the computation of a double integral with an integrand that involves a bivariate normal distribution function, which makes the computation time-consuming, in comparison to the single integral required to fit the yield curve in the Krippner (2013) framework.

This study intends to further analyse the arbitrage-free Nelson-Siegel models implemented by Christensen and Rudebusch (2016) on Krippner's shadow-rate approximation. We compare the accuracy and efficiency of the iterated extended Kalman filter to the unscented Kalman filter on Krippner's framework. Moreover, we investigate Priebsch's cumulant-based approach and compare its results to Krippner's approximation. Lastly, we intend to expand the aforementioned investigation to one- and three-factor Vasiček models.

Following this introduction, a review of the relevant literature on shadow-rate models is provided in Chapter 2. An overview of the Kalman filter as well as its implementation is presented in Chapter 3. Chapter 4 discusses the computational results obtained, and Chapter 5 concludes the paper.

Chapter 2

Shadow-Rate Term Structure Models

In this section, we start by introducing the dynamics of the state process as proposed by Christensen *et al.* (2011). Black (1995)'s concept of shadow-rate model is detailed. A brief overview on the arbitrage-free Nelson-Siegel (AFNS) model is presented. Finally, the approximations of Priebsch (2013) and Krippner (2013) are derived.

2.1 Underlying State Dynamics

We begin by describing the standard Gaussian affine term structure model, which has closed-form solutions for yields. As with Kim and Priebsch (2013), consider the standard N_{dim}-factor Gaussian term structure model that assumes the short rate is governed by a state process X_t, which follows an affine diffusion. Let $W_t^{\mathbb{P}}$ be N_{dim}-dimensional standard Brownian motion on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Assume that there is a risk-neutral measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$ to obtain $W_t^{\mathbb{Q}}$ under $\mathbb{Q}$. The multivariate Ornstein-Uhlenbeck (or Vasiček) process is of the form

$$dX_t = K(\Theta - X_t)dt + \Sigma dW_t^{\mathbb{Q}}$$

where K and Σ are $N_{dim} \times N_{dim}$-matrices representing the rate of mean reversion and the volatility while Θ is a $N_{dim} \times 1$-vector representing the mean reversion level.

Applying Itô's formula to this stochastic differential equation gives the solution

$$X_{t+h} = e^{-Kh}X_t + (I - e^{-Kh})\Theta + \sqrt{Q}Z \tag{2.1}$$

where h is the time step and Z is a $N_{dim} \times 1$-vector of independent standard normal

random variables with

$$Q = \text{Cov}(X_{t+h}|X_t) = \int_0^h e^{-K(h-u)}\Sigma\Sigma^\top e^{-K^\top(h-u)}\,du$$

leading to

$$\sqrt{Q} = \text{Cholesky}(Q)$$

Both the Vasiček process and the AFNS model have the same form for Σ and Θ. The difference comes about in the K-matrix, as illustrated for a three-factor model:

$$X_t = \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \qquad \Theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}$$

whereas

$$K = \begin{pmatrix} \kappa_1 & -\kappa_1 & 0 \\ 0 & \kappa_2 & -\kappa_2 \\ 0 & 0 & \kappa_3 \end{pmatrix}$$

for the Vasiček model and

$$K = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \kappa & -\kappa \\ 0 & 0 & \kappa \end{pmatrix}$$

for the AFNS model.

2.2 Black Shadow-Rate Model

Whilst Gaussian term structure models are most commonly used for bond pricing, they allow nominal interest rates to go negative. In practice, interest rates do go negative, however, Black (1995) suggested that the existence of currency should limit the *observed* nominal short rates to non-negative values. To account for this option of storing currency, he set the lower bound at zero on the nominal short rate by introducing the notion of a **shadow-rate** s_t. The shadow-rate can be positive or negative, and is used to define the **observed short rate** r_t as

$$r_t = \max\{s_t, 0\}$$

i.e. the risk-free rate equals the shadow-rate with a zero-lower bound (ZLB). Note that the choice of the lower bound results from a zero nominal return when storing currency (by "putting it under the mattress") as indicated by Black (1995). Additionally, the U.S. treasury yield data also supports the choice of a zero-lower bound (Christensen and Rudebusch (2016)).

The observed short rate r_t will be used in the zero-coupon bond (ZCB) price, yield and forward rate formulae, denoted by $\underline{P}, \underline{y}$ and $\underline{f}$ respectively.

2.3 Shadow Bond Prices

The shadow short rate is defined to be an affine function of the state variables as

$$s_t = \mathrm{w}^\top X_t$$

where w is a $N_{dim} \times 1$-vector of weights describing the dependency of the short rate on each of the state factors. We assume $s_t = X_t^1$ for the Vasiček model, and $s_t = X_t^1 + X_t^2$ for the AFNS, so $\mathrm{w} = (1\ 0\ 0)^\top$ and $\mathrm{w} = (1\ 1\ 0)^\top$ respectively. Before further delving into term structure modelling, it is necessary that we take note of the assumptions made in implementing these models. We give a brief overview on term structure modelling from Piazzesi (2010) by employing martingale pricing theory.

We assume that there is an equivalent martingale measure (EMM) $\mathbb{Q}$ and that it is unique. This dictates that the market under consideration is complete and free of arbitrage opportunities, forming the basis for bond pricing. Zero-coupon bond prices are the foundation to term structure modelling. We express zero-coupon bond prices at time t maturing at time T under this EMM as

$$P(t,T) := \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^T s_u^*\, du\right)\middle|\mathcal{F}_t\right]$$

where $\mathcal{F}_t$ is the filtration containing all the information up to time t. The associated zero-coupon bond yields are given by

$$y(t,T) = -\frac{\log P(t,T)}{T-t}$$

Since we are interested in affine term structure models, the ZCB yields are affine in the unobservable state factors X_t

$$y(t,\tau) = -\frac{\log P(t,\tau)}{\tau} = -\frac{1}{\tau}\big(A(\tau) + B(\tau)^{\mathrm{T}} X_t\big) \tag{2.2}$$

with associated ZCB prices

$$P(t,\tau) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^{t+\tau} s_u\, du\right)\right] = \exp\big(A(\tau) + B(\tau)^{\mathrm{T}} X_t\big)$$

where $\tau = T - t$ and $A(\tau), B(\tau)$ are functions that can be obtained by solving the Riccati equations

$$\frac{dA(\tau)}{d\tau} = \frac{1}{2}B(\tau)\Sigma\Sigma^\top B(\tau)^\top + B(\tau)K\Theta \qquad \frac{dB(\tau)}{d\tau} = -B(\tau)K - \mathrm{w}$$

with initial conditions $A(T,T) = 0$ and $B(T,T) = 0$, since $P(T,T) = 1$ (see Christensen *et al.* (2011)).

A number of restrictions need to be applied to obtain strong solutions for the aforementioned differential equations. See Dai and Singleton (2000) for details. In the case of the Vasiček model

$$B(\tau) = -\mathrm{w}^\top K^{-1}(I_m - e^{-K\tau})$$

where I_m is a $N_{dim} \times N_{dim}$-identity matrix. Details on the AFNS are given in Section 2.4.

Alternatively, bond prices (and yields) can be expressed in terms of forward rates

$$P(t,T) = \exp\left(-\int_t^T f(t,s)\,ds\right) \iff f(t,T) = -\frac{\partial}{\partial T}\log P(t,T)$$

where $f(t,T)$ denotes the instantaneous forward rate applicable between time t and T.

Moreover, we require the distribution of X_t and $P(t,T)$ under the $\mathbb{P}$-world to use real data. However, the model cannot be calibrated to real data in a reasonable time, and thus the market price of risk required in adjusting the market data to the model dynamics have been ignored. For this reason, we will generate the measurement yields instead of using market data. For completeness, we will present the transition from the real-world measure to the risk-neutral measure.

Note: Adding price of risk to the model adds 3 parameters for a 3-factor model, which simply equates to higher computation time during optimization.

Assume that the market price of risk Λ_t is affine in the state factors, i.e.

$$\Lambda_t = (\Sigma)^{-1}(\lambda_0 + \lambda_1 X_t)$$

where λ_0 is a $N_{dim} \times 1$-vector of constants and λ_1 is a $N_{dim} \times N_{dim}$-matrix of constants. We then have

$$dW_t^{\mathbb{P}} = dW_t^{\mathbb{Q}} + (\Sigma)^{-1}(\lambda_0 + \lambda_1 X_t)dt$$

As a result of this pricing kernel, we obtain the risk-neutral dynamics:

$$dX_t = K(\Theta - X_t)dt + \Sigma dW_t^{\mathbb{Q}}$$

where $K = K^{\mathbb{P}} - \lambda_1$ and $\Theta = (K^{\mathbb{P}} - \lambda_1)^{-1}(K^{\mathbb{P}}\Theta^{\mathbb{P}} + \lambda_0)$.

2.4 Arbitrage-Free Nelson-Siegel Model

The arbitrage-free Nelson-Siegel (AFNS) model is an affine term structure model that takes its inspiration from the Nelson-Siegel parametrization. The Nelson-Siegel yield curve model takes the form

$$y(t,\tau) = \beta_0 + \beta_1\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right) + \beta_2\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau} - e^{-\kappa\tau}\right)$$

where κ is the rate of mean-reversion and $\beta_0, \beta_1, \beta_2$ are constants. In order to understand the dynamic evolution of yield curves over time, Diebold and Li (2006) interpret the β coefficients as level, slope and curvature respectively, i.e.

$$y(t,\tau) = L_t + S_t\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right) + C_t\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau} - e^{-\kappa\tau}\right)$$

Now recall from Equation 2.2 that zero-coupon yields can be expressed as

$$y(t,\tau) = -\frac{B(\tau)}{\tau}X_t - \frac{A(\tau)}{\tau}$$

Since the zero-coupon yield expression has a similar functional form to that of the Nelson-Siegel model, we can approximate the yield curve for a three-factor affine model by setting $X_t = (X_t^1, X_t^2, X_t^3) = (L_t, S_t, C_t)$

$$y(t,\tau) = X_t^1 + X_t^2\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right) + X_t^3\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau} - e^{-\kappa\tau}\right) - \frac{A(\tau)}{\tau}$$

Christensen *et al.* (2011) present the following proposition to prove that there exists a class of affine models that solve the ODEs for the $B(\tau)$-matrix.

Proposition 2.1. *Assume that the short rate is given by*

$$r_t = X_t^1 + X_t^2$$

where the unobserved state dynamics under the risk-neutral measure are described by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \kappa & -\kappa \\ 0 & 0 & \kappa \end{pmatrix} \left[\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_t^{1,\mathbb{Q}} \\ dW_t^{2,\mathbb{Q}} \\ dW_t^{3,\mathbb{Q}} \end{pmatrix}, \kappa > 0$$

The zero-coupon bond prices are then given by

$$P(t,\tau) = \exp\left(A(\tau) + B^1(\tau)X_t^1 + B^2(\tau)X_t^2 + B^3(\tau)X_t^3\right)$$

where $B^1(\tau), B^2(\tau), B^3(\tau)$ and $A(\tau)$ are solutions to the subsequent system of ODEs:

$$\frac{dB(\tau)}{dt} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \kappa & -\kappa \\ 0 & 0 & \kappa \end{pmatrix} B(\tau)$$

where

$$B(\tau) = \begin{pmatrix} B^1(\tau) \\ B^2(\tau) \\ B^3(\tau) \end{pmatrix}$$

and

$$\frac{dA(\tau)}{dt} = -\frac{1}{2}B(\tau)\Sigma\Sigma^\top B(\tau)^\top - B(t,T)K\Theta$$

with initial conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = A(T,T) = 0$ *given that* $P(T,T) = 1$. *The solution for this set of ODEs is:*

$$\begin{aligned} B^1(\tau) &= -\tau \\ B^2(\tau) &= -\frac{1-e^{-\kappa\tau}}{\kappa} \\ B^3(\tau) &= \tau e^{-\kappa\tau} - \frac{1-e^{-\kappa\tau}}{\kappa} \end{aligned}$$

and[1]

$$A(t,T) = \frac{1}{2}\int_t^T \left(B(s,T)\Sigma\Sigma^\top B(s,T)^\top\right) + (K\Theta)_2 \int_t^T B^2(s,T)ds + (K\Theta)_3 \int_t^T B^3(s,T)ds$$

Thus, zero-coupon yields can be presented as

$$y(t,\tau) = X_t^1 + X_t^2\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau}\right) + X_t^3\left(\frac{1-e^{-\kappa\tau}}{\kappa\tau} - e^{-\kappa\tau}\right) - \frac{A(\tau)}{\tau}$$

The extra term $-\frac{A(\tau)}{\tau}$ is known as the "yield-adjustment" term, and this leads to the difference between the arbitrage-free Nelson-Siegel and the conventional Nelson-Siegel models. For further details, see Christensen *et al.* (2011).

2.5 Cumulant-Based Shadow-Rate Model

Priebsch (2013) proposed a cumulant-based approach for estimating yields in Gaussian shadow-rate models under $\mathbb{Q}$ by taking the logarithm of the zero-coupon bond price[2]

$$\log \underline{P}(t,T) = \log \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u\, du\right)\right]$$

By definition, the logarithm of a random variable's moment-generating function is the cumulant-generating function. As with Rosenbaum (1961), the computation of the first and second moments of truncated Gaussian variables involves the following Lemmas:

[1] For clarity, we express $A(\tau)$ and $B(\tau)$ as $A(t,T)$ and $B(t,T)$ here.

[2] Recall that here $\underline{P}(t,T)$ is the real bond price, and r_t is the real rate (not the shadow versions of these).

Define $X^+(t) := \max\{X(t), 0\}$.

Lemma 2.2. *If $X \sim N(\mu, \sigma^2)$, then*

$$\mathbb{E}[X^+] = \mu\Phi\Big(\frac{\mu}{\sigma}\Big) + \sigma\varphi\Big(\frac{\mu}{\sigma}\Big)$$

where Φ, φ are the standard normal cdf and pdf respectively.

Lemma 2.3. *If* $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right)$*, then*

$$\begin{aligned}\mathbb{E}[X_1^+ X_2^+] = (\mu_1\mu_2 + \sigma_{12})\Phi_2^d(-\zeta_1, -\zeta_2; \chi) \\ + \mu_1\sigma_2\varphi(\zeta_2)\Phi\left(\frac{\zeta_1 - \chi\zeta_2}{\sqrt{1-\chi^2}}\right) + \mu_2\sigma_1\varphi(\zeta_1)\Phi\left(\frac{\zeta_2 - \chi\zeta_1}{\sqrt{1-\chi^2}}\right) \\ + \sigma_1\sigma_2\sqrt{\frac{1-\chi^2}{2\pi}}\varphi\left(\sqrt{\frac{\zeta_1^2 - 2\chi\zeta_1\zeta_2 + \zeta_2^2}{1-\chi^2}}\right)\end{aligned}$$

where Φ, φ are the standard normal cdf and pdf respectively, and Φ_2^d is the decumulative bivariate standard normal distribution function. Moreover,

$$\zeta_i := \frac{\mu_i}{\sigma_i} \qquad \textit{for } \ i = 1, 2 \qquad\qquad \chi := \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

The truncated processes $X_i^+(t)$ are as follows:

$$\mathbb{E}_t^{\mathbb{Q}}[X_i^+(t)] = G_i(t)$$

$$\mathbb{E}_t^{\mathbb{Q}}[X_i^+(s)X_j^+(t)] = F_{ij}(s,t)$$

where $G_i(t)$ and $F_{ij}(s,t)$ are functions defined by

$$G_i(t) := \mu_i(t)\Phi(\zeta_i(t)) - \sigma_i(t)\varphi(\zeta_i(t)) \tag{2.3}$$

and

$$\begin{aligned}F_{ij}(s,t) := (\mu_i(s)\mu_j(t) + \sigma_{ij}(s,t))\Phi_2^d(-\zeta_i(s), -\zeta_j(t); \chi_{ij}(s,t)) \\ + \sigma_i(s)\mu_j(t)\varphi(\zeta_i(s))\Phi(D_{1,ij}(s,t)) \\ + \sigma_j(t)\mu_i(s)\varphi(\zeta_j(t))\Phi(D_{2,ij}(s,t)) \\ + \sigma_i(s)\sigma_j(t)\sqrt{\frac{1-\chi_{ij}(s,t)^2}{2\pi}}\varphi(D_{3,ij}(s,t))\end{aligned} \tag{2.4}$$

where φ denotes the standard univariate normal density function, and Φ its cumulative distribution function. $\Phi_2(x, y; \rho)$ denotes the joint bivariate normal distribution function of two standard normal random variables with correlation ρ, and $\Phi_2^d(x, y; \rho)$ the decumulative bivariate normal distribution, given by

$$\Phi_2^d(x, y; \rho) := 1 - \Phi(x) - \Phi(y) + \Phi_2(x, y; \rho)$$

We have[3]

$$\mu_i(s) = \mathbb{E}^{\mathbb{Q}}[X_i(s)] \qquad \sigma_i^2(s) = \text{Var}^{\mathbb{Q}}[X_i(s)] \qquad \sigma_{ij}(s,t) = \text{Cov}^{\mathbb{Q}}[X_i(s), X_j(t)]$$

We also have

$$\zeta_i(t) := \frac{\mu_i(t)}{\sigma_i(t)} \qquad \text{and} \qquad \chi_{ij}(s,t) := \frac{\sigma_{ij}(s,t)}{\sigma_i(s)\sigma_j(t)}$$

and

$$D_{1,ij}(s,t) := \frac{\zeta_i(s) - \chi_{ij}(s,t)\zeta_j(t)}{\sqrt{1-\chi_{ij}(s,t)^2}}$$

$$D_{2,ij}(s,t) := \frac{\zeta_j(t) - \chi_{ij}(s,t)\zeta_i(s)}{\sqrt{1-\chi_{ij}(s,t)^2}} = D_{1,ij}(s,t)$$

$$D_{3,ij}(s,t) := \sqrt{\frac{\zeta_i(s)^2 - 2\chi_{ij}(s,t)\zeta_i(s)\zeta_j(t) + \zeta_j(t)^2}{1-\chi_{ij}(s,t)^2}}$$

Having established the form of the first two moments, we can now derive Priebsch (2013)'s approximation. The quantity $\log \underline{P}(t,T) = \log \mathbb{E}_t^{\mathbb{Q}}\big[\exp(-\int_t^T r_u\, du)\big]$ is the conditional cumulant-generating function evaluated at -1 of the random variable

$$\underline{R}(t,T) := \int_t^T r_u\, du$$

with the series representation

$$\log \mathbb{E}_t^{\mathbb{Q}}\big[\exp(-\underline{R}(t,T))\big] = \sum_{j=1}^{\infty}(-1)^j \frac{\eta_j}{j!}$$

where η_j is the j^{th} cumulant of $\underline{R}(t,T)$ under $\mathbb{Q}$. The first-order and second-order[4] approximations for the zero-coupon yield are obtained by computing the first two terms in the above expansion as shown by Priebsch (2013):

$$\underline{\tilde{y}}(t,T) = \frac{1}{T-t}\eta_1 = \frac{1}{T-t}\mathbb{E}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big]$$

$$\underline{\tilde{\tilde{y}}}(t,T) = \frac{1}{T-t}\left(\eta_1 - \frac{1}{2}\eta_2\right) = \frac{1}{T-t}\left(\mathbb{E}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big] - \frac{1}{2}\text{Var}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big]\right)$$

[3] Note that the σ here belongs to the Q-matrix and must not be confused with the quantities in the Σ-matrix.

[4] Recall that the first and second cumulants are just the mean and variance.

2.5.1 Priebsch Approximation Under Vasiček Dynamics

Following Priebsch (2013), we compute the first two moments of $\underline{R}(t,T)$ for the Vasiček model. The mean is given by[5]

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big] &= \mathbb{E}_t^{\mathbb{Q}}\bigg[\int_t^T r_u\,du\bigg]\\ &= \int_t^T \mathbb{E}_t^{\mathbb{Q}}\big[r_u\big]\,du\\ &= \int_t^T \mathbb{E}_t^{\mathbb{Q}}\big[X_1^+(u)\big]\,du\\ &= \int_t^T G_1(u)\,du\end{aligned}$$

where $G_i(t)$ is the function defined in Equation 2.3.
To compute the variance, we have

$$\begin{aligned}\mathrm{Var}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big] &= \mathrm{Var}_t^{\mathbb{Q}}\bigg[\int_t^T r_u\,du\bigg]\\ &= \mathbb{E}_t^{\mathbb{Q}}\bigg[\bigg(\int_t^T r_u\,du\bigg)^2\bigg] - \big(\mathbb{E}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big]\big)^2\end{aligned}$$

Using Lemma 2.3, Priebsch (2013) computes

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\bigg[\bigg(\int_t^T r_u\,du\bigg)^2\bigg] &= \int_t^T\int_t^T \mathbb{E}_t^{\mathbb{Q}}\big[X_1^+(u)X_1^+(s)\big]\,du\,ds\\ &= \int_t^T\int_t^T F_{11}(u,s)\,du\,ds\end{aligned}$$

where $F_{ij}(u,s)$ is the function defined in Equation 2.4.
Hence

$$\underline{\tilde{y}}(t,T) \approx \frac{1}{T-t}\bigg[\int_t^T G_1(u)\,du\bigg]$$

$$\underline{\tilde{\tilde{y}}}(t,T) \approx \frac{1}{T-t}\bigg[\int_t^T G_1(u)\,du - \frac{1}{2}\int_t^T\int_t^T F_{11}(u,s)\,du\,ds + \frac{1}{2}\bigg(\int_t^T G_1(u)\,du\bigg)^2\bigg]$$

Priebsch (2013) found that, for the Vasiček model, the second-order approximation was precise in approximating yields for both short and long maturities - by comparing it to the "exact" yields obtained via Monte Carlo simulation. On the other hand, the first-order approximation provided acceptable results for only short maturities.

[5] Recall that $r_t = X_1^+(t)$ under the Vasiček model.

2.5.2 Priebsch Approximation Under AFNS Dynamics

In the case of the AFNS model, we have $s_t = (X_1(t) + X_2(t))$. Then

$$\mathbb{E}^{\mathbb{Q}}\big[X_1(s) + X_2(s)\big] = \mu_1(s) + \mu_2(s) =: \mu_a$$

Similarly,

$$\mathrm{Var}^{\mathbb{Q}}\big[X_1(s) + X_2(s)\big] = \sigma_1^2(s) + 2\sigma_{12}(s,s) + \sigma_2^2(s) =: \sigma_a$$

Thus, by referring to Lemma 2.2, we compute the mean of $\underline{R}(t,T)$ as

$$\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\big[\underline{R}(t,T)\big] &= \mathbb{E}_t^{\mathbb{Q}}\bigg[\int_t^T r_u\, du\bigg] \\
&= \int_t^T \mathbb{E}_t^{\mathbb{Q}}\big[r_u\big]\, du \\
&= \int_t^T \mathbb{E}_t^{\mathbb{Q}}\big[(X_1(u) + X_2(u))^+\big]\, du \\
&= \int_t^T G_a(u)\, du
\end{aligned}$$

where $G_i(t)$ is the function defined in Equation 2.3 with $\mu_i = \mu_a$ and $\sigma_i = \sigma_a$.
To compute the second moment of $\underline{R}(t,T)$, we require to calculate

$$\begin{aligned}
\mathrm{Cov}^{\mathbb{Q}}\Big[\big(X_i(s) + X_j(s)\big), \big(X_i(t) + X_j(t)\big)\Big] &= \mathrm{Cov}^{\mathbb{Q}}[X_i(s), X_i(t)] + \mathrm{Cov}^{\mathbb{Q}}[X_i(s), X_j(t)] \\
&\quad + \mathrm{Cov}^{\mathbb{Q}}[X_j(s), X_i(t)] + \mathrm{Cov}^{\mathbb{Q}}[X_j(s), X_j(t)] \\
&=: \sigma_{ab}(s,t)
\end{aligned}$$

Thus, using Lemma 2.3 we have

$$\begin{aligned}
\mathbb{E}_t^{\mathbb{Q}}\bigg[\bigg(\int_t^T r_u\, du\bigg)^2\bigg] &= \int_t^T \int_t^T \mathbb{E}_t^{\mathbb{Q}}\Big[\big(X_1(u) + X_2(u)\big)^+\big(X_1(s) + X_2(s)\big)^+\Big]\, du\, ds \\
&= \int_t^T \int_t^T F_{ab}(u,s)\, du\, ds
\end{aligned}$$

where $F_{ij}(u,s)$ is the function defined in Equation 2.4 with the subscripts $i,j = a,b$.
Hence

$$\underline{\tilde{y}}(t,T) \approx \frac{1}{T-t}\bigg[\int_t^T G_a(u)\, du\bigg]$$

$$\underline{\tilde{\tilde{y}}}(t,T) \approx \frac{1}{T-t}\bigg[\int_t^T G_a(u)\, du - \frac{1}{2}\int_t^T \int_t^T F_{ab}(u,s)\, du\, ds + \frac{1}{2}\bigg(\int_t^T G_a(u)\, du\bigg)^2\bigg]$$

2.6 Option-Based Shadow-Rate Model

Krippner (2013) suggests an option-based approach which accounts for the option of currency, like that of Black (1995). If the option of currency did not exist, the short rate may go negative leading to a shadow-rate ZCB to trade above par. If we consider the shortest maturity time δ, then the time-t value of a standard observed ZCB $P(t,T)$ restricted by the ZLB is given by

$$\begin{aligned}\underline{P}(t,t+\delta) &= \min\{P(t,t+\delta),1\}\\ &= P(t,t+\delta)-(P(t,t+\delta)-1)^{+}\\ &= P(t,t+\delta)-C(t,t,t+\delta)\end{aligned}$$

where a European call option at time t, maturing at T with a strike of 1, on an underlying shadow ZCB $P(t,S)$ maturing at $S \geq T$, is denoted by $C(t,T,S)$. Krippner (2013) looks at

$$\underline{P}(t,T+\delta) = P(t,T+\delta)-C(t,T,T+\delta)$$

and goes onto examining what happens as $\delta \to 0$. As described by Christensen and Rudebusch (2014), as $\delta \to 0$, the idea for the option-based scheme is better highlighted by setting the price of the unrestricted shadow ZCB, minus the price of a call option on an underlying shadow bond, equal to the price of a standard observed bond restricted by the ZLB, i.e.

$$\underline{P}(t,T) = P(t,T)-C^{A}(t,T,T)$$

where an American call option at time t, maturing at T with a strike of 1, on an underlying shadow ZCB maturing at T, is denoted by $C^{A}(t,T,T)$. This intuitively means that the probability mass linked to the shadow ZCB trading above par, would need to be sold off by the issuer of the shadow bond to equal the price of the standard bond.

Unfortunately, valuing American options is a challenge, so Krippner (2013) provides an estimate to the correct solution by making use of European options. As a result, Christensen and Rudebusch (2014) present an auxiliary bond price

$$P_a(t,T+\delta) = P(t,T+\delta)-C(t,T,T+\delta)$$

where a European call option at time t, maturing at T with a strike of 1, on an underlying shadow ZCB maturing at $T+\delta$, is denoted by $C(t,T,T+\delta)$. Note that the observed bond price $\underline{P}(t,T+\delta)$ is not identical to the auxiliary bond price $P_a(t,T+\delta)$. Krippner (2013) then goes on to relate the auxiliary bond price to the instantaneous forward rate by setting

$$\underline{f}(t,T) = -\lim_{\delta\to 0}\frac{\partial \ln P_a(t,T+\delta)}{\partial T}$$

where $\underline{f}(t,T)$ can be calculated as follows

$$\begin{aligned}
\underline{f}(t,T) &= -\lim_{\delta\to 0}\left(\frac{1}{P_a(t,T+\delta)}\frac{\partial P_a(t,T+\delta)}{\partial T}\right)\\
&= -\frac{1}{P(t,T)}\frac{\partial P(t,T)}{\partial T} + \frac{1}{P(t,T)}\lim_{\delta\to 0}\frac{\partial C(t,T,T+\delta)}{\partial T}\\
&= f(t,T) + \frac{1}{P(t,T)}\left(\lim_{\delta\to 0}\partial_2 C(t,T,T+\delta) + \lim_{\delta\to 0}\partial_3 C(t,T,T+\delta)\right)\\
&= f(t,T) + \frac{1}{P(t,T)}\lim_{\delta\to 0}\frac{\partial C(t,T,T+\delta)}{\partial \delta}
\end{aligned}$$

using the fact that $\lim_{\delta\to 0} C(t,T,T+\delta) = 0$ for all T, and where ∂_2, ∂_3 denote the partial derivatives with the second and third variables respectively. Hence

$$\underline{f}(t,T) = f(t,T) + z(t,T)$$

where

$$z(t,T) := \frac{1}{P(t,T)}\lim_{\delta\to 0}\frac{\partial C(t,T,T+\delta)}{\partial \delta}$$

In Gaussian models, the shadow bond prices, forward rates and call prices are often available in closed form. Now the real ZCB prices are defined by

$$\underline{P}(t,T) = e^{-\int_t^T \underline{f}(t,u)\,du}$$

Moreover,

$$r_t := \underline{f}(t,t) = f(t,t) + z(t,t) = s_t + z(t,t) = \max\{s_t, 0\}$$

since

$$\begin{aligned}
z(t,t) &= \lim_{\delta\to 0}\frac{\partial C(t,t,t+\delta)}{\partial \delta}\\
&= \lim_{\delta\to 0}\frac{\partial}{\partial \delta}(P(t,t+\delta)-1)^+\\
&= \begin{cases} \lim_{\delta\to 0}\frac{d}{d\delta}P(t,t+\delta) & \text{when } P(t,t+\delta) > 1 \text{ for small } \delta > 0\\ 0 & \text{else}\end{cases}\\
&= \begin{cases} -\lim_{\delta\to 0} P(t,t+\delta) f(t,t+\delta) & \text{when } P(t,t+\delta) > 1 \text{ for small } \delta > 0\\ 0 & \text{else}\end{cases}\\
&= \begin{cases} -s_t & \text{when } s_t < 0\\ 0 & \text{else}\end{cases}\\
&= -\min\{s_t, 0\}
\end{aligned}$$

Following Krippner (2013), we assume generalized Vasiček dynamics with $s_t = \mathrm{w}^\top X_t$ (recall that w is a $N_{dim} \times 1$-vector of weights describing the dependency of the short rate on each of the state factors). Under the T-forward measure $\mathbb{Q}^T$, we have that s_T is Gaussian with mean

$$\mathbb{E}_{\mathbb{Q}^T}[s_T] = f(0,T)$$

and variance

$$\mathrm{Var}(s_T) = \mathrm{w}^\top \mathrm{Cov}(X_T)\mathrm{w} = \mathrm{w}^\top Q(T)\mathrm{w}$$

Now

$$\underline{f}(0,T) = \mathbb{E}_{\mathbb{Q}^T}[r_T] = \mathbb{E}_{\mathbb{Q}^T}[s_T^+]$$

Recall from Lemma 2.2 if $X \sim N(\mu, \sigma^2)$, then

$$\mathbb{E}[X^+] = \mu\Phi\Big(\frac{\mu}{\sigma}\Big) + \sigma\varphi\Big(\frac{\mu}{\sigma}\Big)$$

Thus

$$\underline{f}(0,T) = f(0,T)\Phi\Big(\frac{f(0,T)}{\omega}\Big) + \omega\varphi\Big(\frac{f(0,T)}{\omega}\Big)$$

where

$$\omega = \sqrt{\mathrm{w}^\top Q(T)\mathrm{w}}$$

The generalized Vasiček model has explicit formulas for the shadow forward rates $f(0,T)$ and the covariance matrix Q, and thus the standard deviation ω. Finally, the ZLB restricted zero-coupon yield formulation is given by

$$\underline{\hat{y}}(t,T) = \frac{1}{T-t}\int_t^T \underline{f}(t,u)du = \frac{1}{T-t}\int_t^T \left[f(t,u)\Phi\left(\frac{f(t,u)}{\omega(u)}\right) + \omega(u)\varphi\left(\frac{f(t,u)}{\omega(u)}\right)\right]du$$

Krippner (2013) compares the option-based results to analytical ones for a calibrated Gaussian one-factor model and concludes that the approximation is practicable. Nevertheless, Krippner (2013)'s approximation is not arbitrage-free, especially when the ZLB is binding.

2.7 Bond Option Pricing

An analytical formula for the price of a European call option written on a shadow bond is required in order to complete the derivation of Krippner (2013)'s option-based approach. We know from theory that the value of a European call option with maturity T and strike price K written on a zero-coupon bond maturing at $T+\delta$ is given by

$$C(t,T,T+\delta) = \mathbb{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T s_u\,du}\max\{P(t,T+\delta) - K, 0\}\right]$$

which can be simplified to

$$C(t,T,T+\delta) = P(t,T+\delta)\Phi(d_1) - KP(t,T)\Phi(d_2)$$

where Φ is the standard normal distribution function with

$$d_1 = \frac{\ln \frac{P(t,T+\delta)}{P(t,T)K} + \frac{1}{2}v(t,T,T+\delta)}{\sqrt{v(t,T,T+\delta)}} \qquad \text{and} \qquad d_2 = d_1 - \sqrt{v(t,T,T+\delta)}$$

and v being the the conditional variance of the logarithm of the zero-coupon bond price (see Christensen *et al.* (2011) for details).

Chapter 3

Kalman Filter

This section is based mostly on Simon (2006) and Wan and Van Der Merwe (2000). The Kalman filter (KF) is used in estimating *the state* based on the understanding of the system dynamics, as well as the availability of the noisy measurements observed over time. The system dynamics (F and H) express the propagation of the state mean and covariance using a time-update and a measurement-update equations. In the context of the affine diffusions, the time-update or **state transition equation** is given by

$$\mathrm{x}_k = F(\mathrm{x}_{k-1}, \mathrm{v}_{k-1})$$

where x_k denotes the unobserved state, and v_k is the state process noise which is assumed to be Gaussian $\mathrm{v}_k \sim (0, \mathrm{Q})$. Alternatively, applying the results of Equation 2.1, we can express the transition equation as

$$\mathrm{x}_{k+1} = \mathrm{A}\mathrm{x}_k + \mathrm{a} + \sqrt{\mathrm{Q}}\mathrm{w}_k$$

For our affine term structure models, we have

$$\mathrm{A} = e^{-\mathrm{K}h}, \quad \mathrm{a} = (I - e^{-\mathrm{K}h})\Theta, \quad \sqrt{\mathrm{Q}} = \text{Cholesky}\left(\int_0^h e^{-\mathrm{K}(h-u)} \Sigma\Sigma^\top e^{-\mathrm{K}^\top (h-u)}\, du \right)$$

where h is the time step, Q is a covariance matrix of x_k and w_k is a vector of independent standard normal random variables.

The **measurement equation** is expressed as

$$\mathrm{y}_k = H(\mathrm{x}_k, \mathrm{n}_k)$$

where y_k is the observed signal, and n_k is the measurement noise which is assumed to be Gaussian $\mathrm{n}_k \sim (0, \mathrm{R})$. We express the measurement process as

$$\mathrm{y}_k = \mathrm{H}\mathrm{x}_k + \mathrm{h} + \mathrm{n}_k$$

where for our affine term structure models

$$\mathrm{H} = -\begin{pmatrix} B(\tau_1)/\tau_1 \\ B(\tau_2)/\tau_2 \\ \vdots \\ B(\tau_N)/\tau_N \end{pmatrix} \qquad \mathrm{h} = -\begin{pmatrix} A(\tau_1)/\tau_1 \\ A(\tau_2)/\tau_2 \\ \vdots \\ A(\tau_N)/\tau_N \end{pmatrix}$$

$A(\tau), B(\tau)$ being the solutions to ODEs as shown in Section 2.3, N is the number of maturities, and R is a diagonal covariance matrix of y_k. We assume for tractability that the measurement noises are uncorrelated, and that the measurement noise is the same for all maturities. For the shadow-rate models, the measurement equation is given by the approximation schemes derived in Sections 2.5 and 2.6.

Using these system dynamics, the KF updates the state mean and covariance each time a new measurement is observed, yielding a better estimate of the state. The state estimates (for the next time in the sequence) can be obtained using all the measurements up to and including the current time, leading to an *a posteriori* estimate of the state. However, if the current time measurement is not available (i.e. only previous measurements are available), then an *a priori* estimate of the state is formed. KF can update the mean and covariance of the state exactly if the underlying system is linear. However, this problem deals with non-linear systems. Linear estimation techniques can be used by linearizing a non-linear problem. These techniques are described in the following sections.
Note: Since we intend to implement this problem in MATLAB, only discrete-time KF is derived.

3.1 Iterated Extended Kalman Filter

The extended Kalman filter (EKF) linearizes a non-linear system around a nominal state trajectory. The state estimate obtained via KF can be used as the nominal state trajectory. For highly non-linear systems, the error in this linearization technique can be significant. A better way to estimate the state is by iterating through the new estimate obtained via EKF. The iteration can be done for as many times as desired to reduce the linearization error. This is known as the iterated extended Kalman filter (IEKF).

3.1.1 Overview

1. We have a **state transition equation**

$$\mathrm{x}_{k+1} = \mathrm{A}\mathrm{x}_k + \mathrm{a} + \sqrt{\mathrm{Q}}\mathrm{w}_k$$

where x_k denotes the unobserved state, Q is a covariance matrix of x_k, and w_k is a vector of independent standard normal random variables.

2. We have a **measurement equation**

$$y_k = Hx_k + n_k$$

where y_k is the observed signal, and n_k is the measurement noise with a diagonal covariance matrix R.

3. We approximate the terms required in estimating $\hat{x}_k$:

$$\begin{aligned} F_k &= HP_k^- H^\top + R \\ K_k &= P_k^- H^\top F_k^{-1} \\ d_k &= y_k - h(\hat{x}_k^-, 0) \end{aligned}$$

where K_k is known as the Kalman gain and P_k is the state covariance matrix. In the case where we have a non-linear measurement equation, we compute the Jacobian matrix H as:

$$H = \left.\frac{\partial h}{\partial x}\right|_{\hat{x}_k^+}$$

4. Given the observed yields y_k, we can estimate the state(s) $\hat{x}_k$ as:

$$\hat{x}_k = (\text{predicted } \hat{x}_k) + K_k\left[y_k - (\text{predicted } y_k)\right]$$

where the predicted $\hat{x}_k$, the Kalman gain K_k and the predicted y_k are approximated using 3:

5. We then update the error state covariance matrix P_k:

$$P_k^+ = P_k^- - K_k F_k K_k^\top$$

Finally, we iterate through steps 3-5 to reduce the linearization error.

3.1.2 Algorithm

Step I: Initialize

- Set

$$\begin{aligned} \hat{x}_0^+ &:= \mathbb{E}[x_0] && (N_{dim} \times 1) \\ P_0^+ &:= \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^\top] && (N_{dim} \times N_{dim}) \end{aligned}$$

Step II: Loop

For $k = 1$ to number of observations:

Predict:

- Predict x_k and state process noise P_k:

$$\begin{aligned} \hat{\mathrm{x}}_k^- &= F(\hat{\mathrm{x}}_{k-1}^+, 0) && (N_{dim} \times 1) \\ \mathrm{P}_k^- &= \mathrm{A}\mathrm{P}_{k-1}^+\mathrm{A}^\top + \mathrm{Q} && (N_{dim} \times N_{dim}) \end{aligned}$$

  where Q is the variance matrix of the noise v_k driving the state process.

  Note: The transition equation is linear, hence we do not require partial derivatives with respect to its parameters to update the state.

Iterate:

- Initialise iteration:

$$\begin{aligned} \hat{\mathrm{x}}_{k,0}^+ &= \hat{\mathrm{x}}_k^- \\ \mathrm{P}_{k,0}^+ &= \mathrm{P}_k^- \end{aligned}$$

- For $i = 0$ to number of iterations, M:

$$\mathrm{H}_{k,i} = \left.\frac{\partial H}{\partial \mathrm{x}}\right|_{\hat{\mathrm{x}}_{k,i}^+} \qquad (N \times N_{dim})$$

  where N is the number of bond yield maturities.

- Update x_k and state process noise P_k:

$$\begin{aligned} \hat{\mathrm{x}}_{k,i+1}^+ &= \hat{\mathrm{x}}_k^- + \mathrm{K}_{k,i}\mathrm{d}_{k,i} && (N_{dim} \times 1) \\ \mathrm{P}_{k,i+1}^+ &= \mathrm{P}_k^- - \mathrm{K}_{k,i}\mathrm{F}_{k,i}\mathrm{K}_{k,i}^\top && (N_{dim} \times N_{dim}) \end{aligned}$$

  where

$$\begin{aligned} \mathrm{F}_{k,i} &= \mathrm{H}_{k,i}\mathrm{P}_k^-\mathrm{H}_{k,i}^\top + \mathrm{R} && (N \times N) \\ \mathrm{K}_{k,i} &= \mathrm{P}_k^-\mathrm{H}_{k,i}^\top\mathrm{F}_{k,i}^{-1} && (N_{dim} \times N) \\ \mathrm{d}_{k,i} &= \mathrm{y}_k - H(\hat{\mathrm{x}}_k^-, 0) && (N \times 1) \end{aligned}$$

  and R is the covariance matrix of the measurement noise n_k.

- Loop to iterate

Update:

- Update the estimate of the state and its associated covariance matrix:

$$\hat{\mathrm{x}}_k^+ = \hat{\mathrm{x}}_{k,M+1}^+ \qquad (N_{dim} \times 1)$$
$$\mathrm{P}_k^+ = \mathrm{P}_{k,M+1}^+ \qquad (N_{dim} \times N_{dim})$$

- Loop

3.2 Unscented Kalman Filter

While the EKF linearizes a non-linear system, the unscented Kalman filter (UKF) makes use of unscented transforms to propagate means and covariances more accurately, and at a comparable level of complexity (Wan and Van Der Merwe (2000)). Using a deterministic sampling technique, the UKF samples points close to the mean. These sampled points (known as sigma points) are used to update the mean and covariance via non-linear functions. The estimate of the state would, therefore, depend on the unscented transform and the sigma points used.

Furthermore, the calculation of Jacobians is not required in the UKF, which is computationally demanding for complex systems. Some systems' Jacobians require numerical computation (additional computational cost), while other systems may not be continuous or differentiable and, therefore, impossible to compute Jacobians.

Since the system in consideration is a highly non-linear one, we employ the UKF with additive noise to estimate the state process more efficiently. The subsections that follow give a brief overview of the UKF with additive noise, followed by step-by-step description of the algorithm implemented to estimate the states from the observed yields (sourced from Van Der Merwe and Wan (2001)).

3.2.1 Overview

1. We have a **state equation**

$$\mathrm{x}_k = F(\mathrm{x}_{k-1}) + \mathrm{v}_{k-1}$$

where x_k denotes the underlying state, and v_k its associated noise, both being column vectors. We define the state process and state noise dimensions as:

$$N_x := \dim(\mathrm{x}) \qquad N_v := \dim(\mathrm{v})$$

2. We have a **measurement equation**

$$\mathrm{y}_k = H(\mathrm{x}_k) + \mathrm{n}_{k-1}$$

where y_k is the observed signal, and n_k is the measurement noise, both being column vectors. We define dimensions of the measurement process and its associated noise as:

$$N_y := \dim(\mathrm{y}) \qquad N_n := \dim(\mathrm{n})$$

3. We set $L := N_x$, so that the state vector x_k is an L-dimensional column vector.

4. Given a random vector x of dimension L, with mean $\bar{\mathrm{x}}$ and covariance matrix P_x, we construct an $L \times (2L+1)$-matrix of **sigma points** $\chi = [\chi_0, \chi_1, ..., \chi_{2L}]$, with columns defined by

$$\begin{aligned}
\chi_0 &:= \bar{\mathrm{x}} \\
\chi_i &:= \bar{\mathrm{x}} + \sqrt{(L+\lambda)\mathrm{P}_\mathrm{x}} \qquad \text{for} \quad i = 1, ..., L \\
\chi_i &:= \bar{\mathrm{x}} - \sqrt{(L+\lambda)\mathrm{P}_\mathrm{x}} \qquad \text{for} \quad i = L, ..., 2L
\end{aligned}$$

where λ is defined in 5.

5. We introduce three parameters α, β, κ and define

$$\lambda := \alpha^2(L+\kappa) - L$$

which according to Wan and Van Der Merwe (2000) is explained as:

- λ is a scaling parameter.
- α determines the "spread" of sigma points, and is set as small as possible, for example $\alpha = 10^{-3}$.
- κ is a secondary scaling parameter, and usually set to $\kappa = 0$.
- β is used to incorporate prior knowledge of the distribution of x with $\beta = 2$ being optimal for Gaussian distributions.

6. With each sigma point χ_i is associated a set of two weights $W_i^{(m)}, W_i^{(c)}$ –which will be used to predict means (m) and covariance matrices (c) in the Kalman filter recursive algorithm. The definitions are:

$$W_0^{(m)} := \frac{\lambda}{L+\lambda} \qquad W_0^{(c)} := \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta$$

$$W_i^{(m)} := \frac{1}{2(L+\lambda)} \qquad W_i^{(c)} := \frac{1}{2(L+\lambda)} \qquad (i = 1, ..., 2L)$$

7. In 4, we take x to be the state vector, so that $L = N_x$. The matrix χ of sigma points is then $N_x \times (2N_x + 1)$.

3.2.2 Algorithm

Step I: Initialize

- Set

$$\begin{aligned} \hat{\mathrm{x}}_0 &:= \mathbb{E}[\mathrm{x}_0] && (L \times 1) \\ \mathrm{P}_0 &:= \mathbb{E}[(\mathrm{x}_0 - \hat{\mathrm{x}}_0)(\mathrm{x}_0 - \hat{\mathrm{x}}_0)^\top] && (L \times L) \end{aligned}$$

- Calculate the weights $W_i^{(m)}, W_i^{(c)}$ for $i = 0, 1, ..., 2L$ as in 6.

Step II: Loop

For $k = 1$ to number of observations:

- Calculate sigma points:

$$\chi_{k-1} := \begin{bmatrix} \hat{\mathrm{x}}_{k-1} & \hat{\mathrm{x}}_{k-1} \pm \sqrt{(L+\lambda)\mathrm{P}_{k-1}} \end{bmatrix} \qquad (L \times (2L+1))$$

  Thus

$$\begin{aligned} \chi_{k-1} &= [\chi_{0,k-1}, \chi_{1,k-1}, ..., \chi_{2L,k-1}] \\ &= \begin{bmatrix} \hat{\mathrm{x}}_{k-1}, \hat{\mathrm{x}}_{k-1} + \mathrm{C}_{1,k-1}, ..., \hat{\mathrm{x}}_{k-1} + \mathrm{C}_{2L,k-1}, ..., \hat{\mathrm{x}}_{k-1} - \mathrm{C}_{1,k-1}, ..., \hat{\mathrm{x}}_{k-1} - \mathrm{C}_{2L,k-1} \end{bmatrix} \end{aligned}$$

  where $\mathrm{C}_{k-1} = \sqrt{(L+\lambda)\mathrm{P}_{k-1}}$ is the Cholesky decomposition of the given matrix, and $\mathrm{C}_{i,k-1}$ denotes the i^{th} column of C_{k-1}.

  Predict:

- Predict "sigma points" of x_k:

$$\begin{aligned} \chi^*_{k|k-1} &:= F(\chi_{k-1}) && (L \times (2L+1)) \\ &= \begin{bmatrix} F(\chi_{0,k-1}, ..., F(\chi_{2L,k-1} \end{bmatrix} \end{aligned}$$

  where $\chi_{i,k-1}$ is the i^{th} column of χ_{k-1}, etc.

  The $\chi_{i,k-1}$ are the sigma points associated with the vectors $\hat{\mathrm{x}}_{k-1}$. Each of these sigma points is transformed according to the state transition equation

$$\mathrm{x}_k = F(\mathrm{x}_{k-1})$$

  to form the "predicted sigma points" $\chi_{i,k|k-1}$ of the vector x_k.

- Predict x_k:

$$\hat{\mathrm{x}}_k^- := \chi_{k|k-1}^* W^{(m)} = \sum_{i=0}^{2L} W_i^{(m)} \chi_{i,k|k-1}^* \qquad (N_x \times 1)$$

where $W^{(m)} := \left(W_0^{(m)}, ..., W_{2L}^{(m)}\right)^\top$. Thus the predicted value $\hat{\mathrm{x}}_k^-$ of x is a weighted sum of the predicted sigma points $\chi_{i,k|k-1}^*$.

- Predict the "covariance matrix" of x_k and add the state process noise $\mathrm{P_v}$:

$$\mathrm{P}_k^- := \sum_{i=0}^{2L} W_i^{(c)} \left(\chi_{i,k|k-1}^* - \hat{\mathrm{x}}_k^-\right)\left(\chi_{i,k|k-1}^* - \hat{\mathrm{x}}_k^-\right)^\top + \mathrm{P_v} \qquad (N_x \times N_x)$$

- Redraw sigma points to take into account the process noise:

$$\chi_{k|k-1} := \left[\hat{\mathrm{x}}_k \quad \hat{\mathrm{x}}_k \pm \sqrt{(L+\lambda)\mathrm{P}_k^-}\right] \qquad (L \times (2L+1))$$

Thus

$$\begin{aligned}\chi_{k|k-1} &= [\chi_{0,k|k-1}, \chi_{1,k|k-1}, ..., \chi_{2L,k|k-1}] \\ &= \left[\hat{\mathrm{x}}_k^-, \hat{\mathrm{x}}_k^- + \mathrm{C}_{1,k}^-, ..., \hat{\mathrm{x}}_k^- + \mathrm{C}_{2L,k}^-, ..., \hat{\mathrm{x}}_k^- \mathrm{C}_{1,k}^-, ..., \hat{\mathrm{x}}_k^- - \mathrm{C}_{2L,k}^-\right]\end{aligned}$$

where $\mathrm{C}_{k-1}^- = \sqrt{(L+\lambda)\mathrm{P}_k^-}$ is the Cholesky decomposition of the given matrix, and $\mathrm{C}_{i,k}^-$ denotes the i^{th} column of C_k^-.

- Predict "sigma points" of y_k:

$$\begin{aligned}\mathcal{Y}_{k|k-1} &:= H(\chi_{k|k-1}) && (N_y \times (2L+1)) \\ &= \left[H(\chi_{0,k|k-1}), ..., H(\chi_{2L,k|k-1})\right]\end{aligned}$$

The $\chi_{i,k|k-1}$ are the sigma points associated with $\hat{\mathrm{x}}_k^-$. These are transformed via the measurement equation

$$\mathrm{y}_k = H(\mathrm{x}_k)$$

to form the "predicted sigma points" of the vector y_k.

- Predict y_k:

$$\hat{\mathrm{y}}_k^- := \mathcal{Y}_{k|k-1} W^{(m)} = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_{i,k|k-1} \qquad (N_y \times 1)$$

where $W^{(m)} := \left(W_0^{(m)}, ..., W_{2L}^{(m)}\right)^\top$. Thus the predicted value $\hat{\mathrm{y}}_k^-$ is a weighted sum of the predicted sigma points $\mathcal{Y}_{i,k|k-1}$.

- Predict the "covariance matrix" of y_k and add the measurement noise $\mathrm{P_n}$:

$$\mathrm{P}_{yy,k} := \sum_{i=0}^{2L} W_i^{(c)} \Big(\mathcal{Y}_{i,k|k-1} - \hat{\mathrm{y}}_k^-\Big)\Big(\mathcal{Y}_{i,k|k-1} - \hat{\mathrm{y}}_k^-\Big)^\top + \mathrm{P_n} \qquad (N_y \times N_y)$$

- Predict the "joint covariances" of $\mathrm{x}_k, \mathrm{y}_k$:

$$\mathrm{P}_{xy,k} := \sum_{i=0}^{2L} W_i^{(c)} \Big(\chi_{i,k|k-1} - \hat{\mathrm{x}}_k^-\Big)\Big(\mathcal{Y}_{i,k|k-1} - \hat{\mathrm{y}}_k^-\Big)^\top \qquad (N_x \times N_y)$$

Update:

- Define the gain matrix

$$\mathrm{K}_k := \mathrm{P}_{xy,k}\mathrm{P}_{yy,k}^{-1}$$

- Correct the estimate x_k using the difference between the observed value of y_k and its predicted value $\hat{\mathrm{y}}_k^-$:

$$\hat{\mathrm{x}}_k := \hat{\mathrm{x}}_k^- + \mathrm{K}_k(\mathrm{y}_k - \hat{\mathrm{y}}_k^-) \qquad (N_x \times 1)$$

- Update the covariance matrix of x_k:

$$\mathrm{P}_k := \mathrm{P}_k^- - \mathrm{K}_k\mathrm{P}_{yy,k}\mathrm{K}_k^\top \qquad (N_x \times N_x)$$

- Loop

3.3 Maximum Likelihood Estimation

Given the observed zero-coupon yield prices, we wish to estimate the value of the state process x_k, as well as estimate the parameters K, Θ, Σ and R that drive the state process dynamics. We, therefore, have a dual estimation problem that can be tackled by combining the Kalman filter with maximum likelihood estimation.

Suppose we have an N_{dim}-dimensional Gaussian model and let $\Omega = \{K, \Theta, \Sigma, \mathrm{R}\}$. Then the log-likelihood has the functional form

$$\begin{aligned}\log l\big(\Omega; \mathrm{y}_{1:T}\big) &= \sum_{k=1}^{T} \log l\big(\Omega; \mathrm{y}_k|\mathrm{y}_{1:(k-1)}\big) \\ &= \sum_{k=1}^{T} \Big[-\frac{N_{dim}}{2}\log(2\pi) - \frac{1}{2}\log|\mathrm{F}_k| - \frac{1}{2}\mathrm{d}_k\mathrm{F}_k^{-1}\mathrm{d}_k\Big]\end{aligned}$$

where

$$\mathrm{d}_k := \mathrm{y}_k - \mathbb{E}\big[\mathrm{y}_k|\mathrm{y}_{1:(k-1)}\big] = \mathrm{y}_k - H(\hat{\mathrm{x}}_k^-, 0) \qquad \mathrm{F}_k = \mathbb{E}\big[\mathrm{d}_k\mathrm{d}_k^\top|\mathrm{y}_{1:(k-1)}\big]$$

This form allows us to integrate the parameter estimation technique easily to the Kalman filter, since F_k and d_k must be computed during filtering anyway[1]. Our aim is to maximize the log-likelihood function numerically to obtain the optimum set of parameters given the noisy yields-data. However, optimizers are minimizers. Thus, instead of maximizing the likelihood we can minimize the likelihood proxy

$$l_{proxy}(\Omega; \mathrm{y}_{1:T}) = \sum_{k=1}^{T} \Big[\log |\mathrm{F}_k| + \mathrm{d}_k \mathrm{F}_k^{-1} \mathrm{d}_k \Big] \tag{3.1}$$

This can be achieved by using the *fmincon* function in MATLAB with the objective function being the likelihood proxy, within reasonable upper and lower bounds. This estimation technique can be implemented to both the AFNS and Vasiček models.

[1] where F_k is the updated covariance matrix of y_k, and d_k is the difference between the observed yield and its predicted value.

Chapter 4

Simulation Based Study of Shadow-Rate Models

Having established the necessary theory on shadow-rate models, we analyse their relative performance when implemented in MATLAB. In particular, we compare the accuracy and computational time of the Krippner and Priebsch approximations to the "exact" solution obtained using Equation 2.2 (as compared by Priebsch (2013)). In addition, we investigate the effectiveness of the IEKF compared to the UKF under the AFNS model on Krippner's framework.

To achieve this numerically, zero-coupon yields were simulated for a period of 5 years in monthly intervals. The bond maturities considered were 6 months, 1 year, 5 year and 10 years. Using reasonably assumed parameters, the state processes were simulated, followed by the computation of bond yields. Gaussian noise with a covariance matrix R was then added to the yield-data to account for the noise inherent in real data.

The noisy yield-data was passed into an optimization routine to search for optimal parameters. This involved the use of the Kalman filter[1] in conjunction with maximum likelihood estimation on the noisy yield-data. Given random but feasible initial parameters, the optimization iterates to achieve a parameter set that gives a minimum likelihood proxy (Equation 3.1). The optimized parameters were then used in recovering the optimized state processes via Kalman filtering.

The computational burden of shadow-rate models lies in the optimization of parameters. Depending on the underlying process, as well as the approximation method being used, optimization times ranged from a few minutes to several days. For this reason, the amount of data attainable was limited. Details will follow later in this chapter. Note that the computer used had a 8GB RAM with a 3.7GHz pro-

[1] Before running the optimization, the non-linear Kalman filters were tested on the noisy yield-data (given the state- and measurement equations). Both the IEKF and UKF recovered the simulated state processes successfully using the initial (known) parameters.

cessor, Intel-core i3-6100.

Moreover, parameter recovery was very much dependent on the initial values used. For each run, initial parameters were randomly selected within a feasible range. It was found that parameters were approximately recovered only when initial values were close to the actual (assumed) values. This observation was only noted in the one-factor Vasiček model. Multi-factor models did not recover any parameters after optimization. However, in all cases, the short rate was recovered.

The three-factor Vasiček model did not yield any results within a span of six days for all three approximation schemes. This is not surprising because the three-factor Vasiček model requires 13 parameters in the optimization, in contrast to 4 parameters in the one-factor case. Under the AFNS model, only Krippner's approach produced results within the stated time.

The quality of the results, and hence the deductions made, can be improved by running the optimization for a larger sample of initial values. The optimized parameters can be cycled into the optimizer for more accurate results. However, these improvements were not possible due to the time constraints and the computing power available.

4.1 One-Factor Black-Vasiček Shadow-Rate Model

Once again, the amount of data attainable was limited by the time consumed in optimizing multi-factor models. Thus, we only have three samples for each of the approximations in the one-factor model. Tables 4.1, 4.2 and 4.3 display results from the optimization. The run-times and parameter recovery varied depending on the initial parameters used. As a test, when actual parameters were used as initial parameters in the optimization of these approximation schemes, all parameters, except the Gaussian noise, were recovered with high accuracy.

Priebsch-1 approximates the interest rate with the least absolute error. It also recovers the rate- and level of mean reversion to a reasonable precision in two of its runs. However, during the third run, the optimization cycled through 3000 parameters in search for the global minimum and failed to find the optimal set of parameters (hence the major leap in the observed time). On the other hand, Krippner underestimates the interest rate in runs 1 and 3 as displayed in Figure 4.1. However, as with Priebsch-1, it recovers the rate- and level of mean reversion with reasonable precision during run 2.

The run-times for Krippner and Priebsch-1 were in a similar range while Priebsch-2 ran for approximately three days, when the initial parameters were completely randomised (as with other approximations). Nevertheless, when the initial

parameters were set close to the actual (known) parameters for Priebsch-2, the run-time halved and the approximation improved significantly (see Table 4.3). This clarifies that Priebsch-2, being the most precise theoretically, is also quantitatively precise, but highly dependent on the initial parameters chosen during optimization. This is most likely due to the computation of the double integral in Priebsch-2 (with an integrand that involves a bivariate normal distribution function), making it computationally time-consuming, as well as difficult for MATLAB to optimize.

Moreover, when the initial parameters were set to the actual parameters, Priebsch-2 not only recovered the rate- and level of mean reversion, but also recovered the standard deviation with a significant improvement in speed (8 times faster than the first two runs).

		Run 1		**Run 2**		**Run 3**	
	True value	Initial	Optimized	Initial	Optimized	Initial	Optimized
Likelihood			1.36410E+03		2.05770E+03		1.36410E+03
κ	0.600	0.52160	0.00001	0.41700	0.60070	0.07630	0.00001
θ	0.020	0.01500	0.09370	0.10800	0.01960	0.11700	0.09570
σ	0.020	-0.06360	-0.10000	0.10000	0.00900	0.01230	0.10000
R	1.000E-08	8.18000E-08	3.72040E-06	3.02000E-08	6.27880E-09	7.23000E-08	3.72156E-06
Run-time			3990		5400		3870

Tab. 4.1: Parameter estimates for a one-factor Vasiček model using Krippner's approximation.

		Run 1		**Run 2**		**Run 3**	
	True value	Initial	Optimized	Initial	Optimized	Initial	Optimized
Likelihood			1.66480E+03		1.66480E+03		1.22070E+03
κ	0.600	0.52160	0.60140	0.41700	0.60140	0.22200	0.30410
θ	0.020	0.01500	0.02000	0.10800	0.02000	0.13100	0.01620
σ	0.020	-0.06360	0.02440	0.10000	0.02440	0.05870	-0.01920
R	1.000E-08	8.18000E-08	4.92580E-16	3.02000E-08	9.85180E-17	9.19000E-08	5.21203E-06
Run-time			3680		4663		59204

Tab. 4.2: Parameter estimates for a one-factor Vasiček model using Priebsch's first cumulant approximation.

		Run 1		Run 2		Run 3	
	True value	Initial	Optimized	Initial	Optimized	Initial	Optimized
Likelihood			1.29382E+03		1.39870E+03		1.66042E+03
κ	0.600	0.52160	0.99640	0.41700	0.43715	0.54000	0.614788
θ	0.020	0.01500	0.02790	0.10800	0.01970	0.01540	0.020639
σ	0.020	-0.06360	-0.09600	0.10000	0.02660	0.02550	0.026867
R	1.000E-08	8.18000E-08	7.11012E-07	3.02000E-08	3.27045E-07	4.88000E-11	8.13084E-11
Run-time			247333		274543		119100

Tab. 4.3: Parameter estimates for a one-factor Vasiček model using Priebsch's second cumulant approximation.

	Run 1	**Run 2**	**Run 3**
Priebsch-1	0.026	0.026	0.778
Priebsch-2	0.591	0.370	0.034
Krippner	0.974	0.032	0.974

Tab. 4.4: Maximum absolute error in the short rate approximation (as a %) for a one-factor Vasiček model.

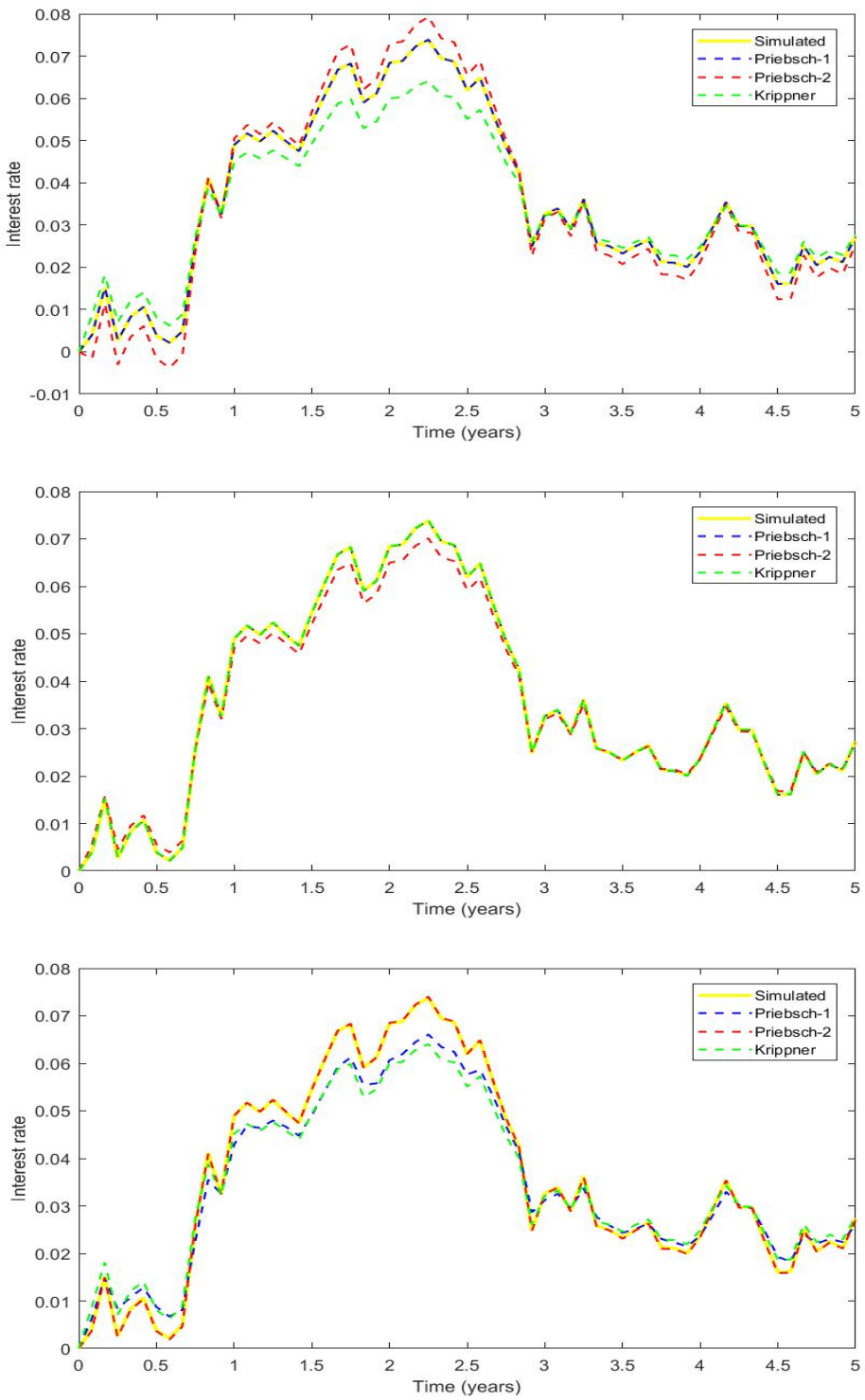

Fig. 4.1: Comparison of the three approximations to the simulated short rate for a one-factor Vasiček model with various initial sets of parameters.

4.2 Three-Factor Black-AFNS Shadow-Rate Model

While Krippner's approximation becomes more efficient as we move from the three-factor Vasiček to the AFNS model, Priebsch's approximations tend to worsen for the following two reasons. First, under the AFNS model, Priebsch requires the computation of $\text{Cov}_t^{\mathbb{Q}}\left[\left(X_i(s)+X_j(s)\right)^+,\left(X_i(t)+X_j(t)\right)^+\right]$ for each time point, in contrast to $\text{Cov}_t^{\mathbb{Q}}\left[\left(X_i(s)\right)^+,\left(X_i(t)\right)^+\right]$ under the Vasiček model. Second, under the Vasiček model, Krippner's approach requires the computation of matrix exponentials for the K-matrix, while we only require exponents of the scalar κ under the AFNS model. This is simply because we have one scalar driving the rate of mean-reversion matrix under the AFNS model. For these reasons, only results for Krippner's framework were obtained under the AFNS model.

Although the parameters are never recovered (as shown in Tables 4.5 and 4.6), the likelihood values are similar, particularly when using IEKF in the optimization. This suggests that, either, there are various parameter sets that define the global maximum for the likelihood function or the local maxima are similar to the global maximum. The optimizer, therefore, finds a set of optimal parameters closest to the initial parameters. In some cases, the optimizer fails to find the optimal parameters when the global maximum is not within reach (see run 3 of Priebsch-1 under one-factor Vasiček). In addition, the Kalman filter does not recover the unobserved state processes in multi-factor models but captures the short rate in all cases.

		Run 1		**Run 2**		**Run 3**	
	True value	Initial	Optimized	Initial	Optimized	Initial	Optimized
Likelihood			2.42920E+03		2.43540E+03		2.43450E+03
κ	0.300	0.60730	0.24702	0.41700	0.24679	0.55080	0.24868
θ_1	0.010	0.06800	0.00252	0.10800	4.79073E-07	0.10600	0.00072
θ_2	0.020	0.06900	0.00003	0	5.69355E-10	0.04400	0.00258
θ_3	0.040	0.09900	0.01418	0.04500	0.00008	0.07700	0.02463
σ_{11}	0.020	-0.05410	-0.02671	0.07060	0.03945	-0.07860	-0.04210
σ_{21}	0.015	0.03000	0.04381	0.08150	-0.06582	-0.07930	0.07093
σ_{22}	0.040	-0.03240	-0.04196	0.06270	0.02138	0.07490	0.01687
σ_{31}	0.010	0.01680	0.01964	0.03090	-0.01157	0.05860	0.00812
σ_{32}	0.035	-0.06840	0.08652	0.02060	-0.04744	0.08970	-0.03736
σ_{33}	0.060	-0.06660	0.02819	-0.00780	-0.01753	0.01180	0.01533
R	1.000E-08	2.56000E-08	2.07536E-08	4.19000E-08	1.79066E-08	3.00000E-09	1.63134E-08
Run-time			28597		55945		65653

Tab. 4.5: Parameter estimates for a AFNS model using Krippner's approximation with the IEKF.

		Run 1		Run 2		Run 3	
	True value	Initial	Optimized	Initial	Optimized	Initial	Optimized
Likelihood			1.28995E+03		1.16550E+03		1.20405E+03
κ	0.300	0.28100	0.24848	0.41700	0.47920	0.15420	0.31339
θ_1	0.010	0.06600	1.20788E-08	0.10800	3.73980E-06	0.11100	0.13433
θ_2	0.020	0.07900	0.00632	0	0.02910	0.03900	0.03172
θ_3	0.040	0.06900	8.70360E-11	0.04500	0.05500	0.08000	1.26363E-10
σ_{11}	0.020	-0.07510	-0.02322	0.07060	-0.05550	0.09710	-0.05329
σ_{21}	0.015	-0.00360	-0.02468	0.08150	-0.06920	-0.08370	-0.03291
σ_{22}	0.040	-0.08870	0.03594	0.06270	0.00047	-0.08010	-0.01458
σ_{31}	0.010	-0.02750	0.00563	0.03090	0.01670	0.09330	0.02952
σ_{32}	0.035	-0.09150	0.00503	0.02060	0.02650	-0.09140	-0.04792
σ_{33}	0.060	0.05190	-0.08790	-0.00780	-0.06720	0.07260	0.00009
R	1.000E-08	6.76000E-08	2.37504E-08	4.19000E-08	2.52712E-07	2.84000E-08	1.33916E-07
Run-time			26340		26085		50864

Tab. 4.6: Parameter estimates for a AFNS model using Krippner's approximation with the UKF.

Finally, the IEKF outperforms the UKF in terms of precision, with no significant difference in run-time. Figure 4.2 compares the two filtering methods for the same initial parameters. The IEKF was fixed to three iterations for each time point.

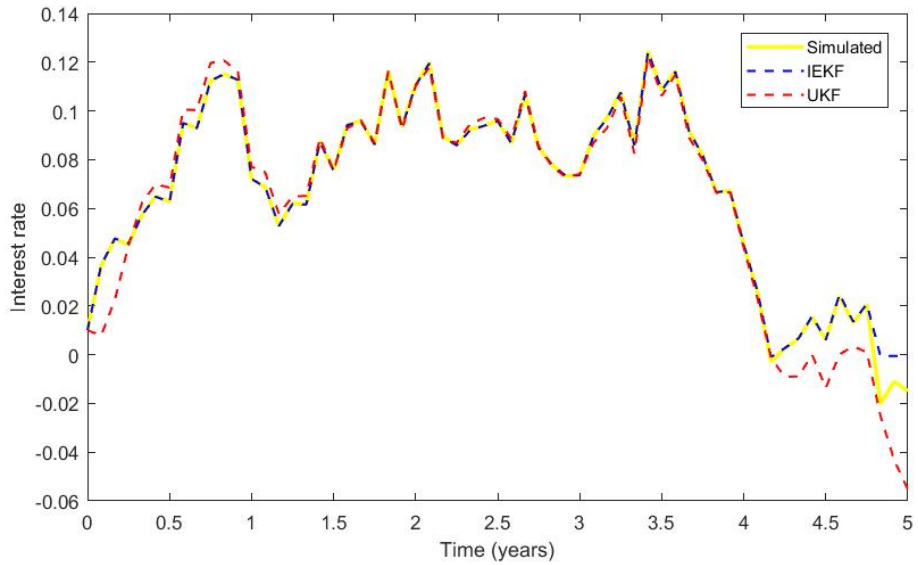

Fig. 4.2: Comparison of the IEKF to the UKF for a AFNS model using Krippner's approximation.

Chapter 5

Conclusions

To estimate Gaussian term structure models near the zero-lower bound, we employ the methodologies suggested by Krippner (2013) and Priebsch (2013), with underlying state dynamics driven by the Vasiček and the arbitrage-free Nelson-Siegel models. For single-factor models, we find that Krippner's approach, which complies with the usual no-arbitrage approach, approximates the short rate within reasonable bounds. Priebsch's first-cumulant method is more accurate than Krippner's framework with identical computational time, and has better parameter recovery. Priebsch's second-cumulant method, that is theoretically the most precise, is not a feasible approach due to the computation of a double integral which makes it time-consuming. When extending to multi-factor models, all approximation schemes suffer during optimization under the Vasiček model and yield no results within reasonable times. However, only Krippner's approximation is found to be practicable under the AFNS dynamics with acceptable accuracy, making it dominant over Priebsch's approximations.

The quality of the results can be improved by running the optimization for a larger sample of initial values, as well as iterating through the optimizer for more accurate parameter values. In addition, high performance computing can be employed to improve optimization times. With added computing power, the risk premiums can be accounted for, and hence, market data can be used instead of simulated yields. Finally, when analysing the non-linear Kalman filters in consideration, we find that the iterated extended Kalman filter recovers the short rate with a better level of accuracy than the unscented Kalman filter, with negligible improvement in run-time. The second-order extended Kalman filter can be used to achieve better precision, by allowing a compromise in run-time.

Bibliography

Bauer, M. D. and Rudebusch, G. D. (2013). Monetary policy expectations at the zero lower bound. Working Paper, Federal Reserve Bank of San Francisco.

Black, F. (1995). Interest rates as options, *the Journal of Finance* **50**(5): 1371–1376.

Christensen, J. H., Diebold, F. X. and Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models, *Journal of Econometrics* **164**(1): 4–20.

Christensen, J. H. and Rudebusch, G. D. (2014). Estimating shadow-rate term structure models with near-zero yields, *Journal of Financial Econometrics* **13**(2): 226–259.

Christensen, J. H. and Rudebusch, G. D. (2016). Modeling yields at the zero lower bound: Are shadow rates the solution?, pp. 75–125.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models, *The Journal of Finance* **55**(5): 1943–1978.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of econometrics* **130**(2): 337–364.

Gorovoi, V. and Linetsky, V. (2004). Black's model of interest rates as options, eigenfunction expansions and Japanese interest rates, *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* **14**(1): 49–78.

Ichiue, H. and Ueno, Y. (2007). Equilibrium interest rate and the yield curve in a low interest rate environment, *Bank of Japan Discussion Paper* (06-E): 18.

Kim, D. H. and Priebsch, M. (2013). Estimation of multi-factor shadow-rate term structure models, *Board of Governors of the Federal Reserve System, Washington DC* **9**.

Kim, D. H. and Singleton, K. J. (2012). Term structure models and the zero bound: an empirical investigation of Japanese yields, *Journal of Econometrics* **170**(1): 32–49.

Krippner, L. (2013). A tractable framework for zero-lower-bound Gaussian term structure models.

Piazzesi, M. (2010). Affine term structure models, *Handbook of financial econometrics* **1**: 691–766.

Priebsch, M. (2013). Computing arbitrage-free yields in multi-factor Gaussian shadow-rate term structure models.

Rosenbaum, S. (1961). Moments of a truncated bivariate normal distribution, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 405–408.

Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*, John Wiley & Sons.

Van Der Merwe, R. and Wan, E. A. (2001). The square-root unscented Kalman filter for state and parameter-estimation, *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, Vol. 6, IEEE, pp. 3461–3464.

Wan, E. A. and Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation, *Adaptive Systems for Signal Processing, Communications, and Control Symposium* pp. 153–158.