Insurance Mathematics and Economics 114 (2024) 156–175

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Construct Smith-Wilson risk-free interest rate curves with endogenous and positive ultimate forward rates

Chaoyi Zhao [a,*], Zijian Jia [a], Lan Wu [b]

[a] School of Mathematical Sciences, Peking University, Beijing, China
[b] LMEQF, School of Mathematical Sciences, Peking University, Beijing, China

ARTICLE INFO

JEL classification:
L51
E4

Keywords:
Ultimate forward rate (UFR)
Smith-Wilson method
Risk-free interest rate curve
Endogenous and positive
Solvency II
Chinese government bond
EURIBOR swap

ABSTRACT

We propose several methods for obtaining endogenous and positive ultimate forward rates (UFRs) for risk-free interest rate curves based on the Smith-Wilson method. The Smith-Wilson method, which is adopted by Solvency II, can both interpolate the market price data and extrapolate to the UFR. However, the method requires an exogenously-chosen UFR. To obtain an endogenous UFR, de Kort and Vellekoop (2016) proposed an optimization framework based on the Smith-Wilson method. In this paper, we prove the existence of an optimal endogenous UFR to their optimization problem under the condition that the cash flow matrix is square and invertible. In addition, to ensure the positivity of the optimal endogenous UFR during extreme time periods such as the COVID-19 pandemic, we extend their optimization framework by including non-negative constraints. Furthermore, we also propose a new optimization framework that can not only generate endogenous and positive UFRs but also incorporate practitioners' prior knowledge. We prove the feasibility of our frameworks, and conduct empirical studies for both the Chinese government bonds and the EURIBOR swaps to illustrate the capabilities of our methods.

1. Introduction

The risk-free interest rate term structure (RFR-TS) is a fundamental tool for trading, valuation, and risk management for financial institutions. As the accounting principles and solvency regulations evolve in the past two decades, the RFR-TS has become a material factor affecting the valuation of assets and liabilities as well as the profit accounting for life insurance companies. Unfortunately, the RFR-TS cannot be directly observed in the market. While most central banks compile and publish yield curves with maturities less than 30 years for their sovereign currencies regularly based on market data of government bonds, life insurance companies need to assess the value of cash flows with maturities much longer than 30 years. Therefore, they are concerned with the "far-end" forward interest rate, or the ultimate forward rate (UFR) named by the European Insurance and Occupational Pensions Authority (EIOPA) for Solvency II, which is the forward rate in the distant future (European Systemic Risk Board, 2017).

There are generally three categories of methods used to construct the RFR-TS: parametric models, interpolating models, and dynamic models. Nelson and Siegel (1987) and Svensson (1994) proposed famous parsimonious parametric models using a given function with several parameters to fit the observed prices of treasury bonds. Natural cubic splines and Hermite cubic splines are typical interpolating models (Hagan and West, 2006). There are also many dynamic models for short-term rates, such as those proposed by Hull and White (1990) and Heath et al. (1992), which are mainly used for calibrating the prices of liquid interest rate derivatives. All three categories of models are widely used in modelling the RFR-TS. Models in the first two categories, which use functions to fit both short and long rates directly, are more commonly used for asset-liability assessments, solvency regulations, and risk management. Models in the third category generate the RFR-TS by fitting the dynamics of short rates, and they are more applicable for the pricing and trading related to interest rates.

The EIOPA derives and publishes the regulatory RFR-TS for 33 currencies monthly (European Systemic Risk Board, 2017).[1] These published yield curves are used by European insurers to evaluate their insurance and reinsurance liabilities. The curves published by the EIOPA are derived using the Smith-Wilson method (Smith and Wilson, 2000),

* Corresponding author.
E-mail address: zhaochaoyi@pku.edu.cn (C. Zhao).

[1] The RFR-TS and related technical information published by the EIOPA can be found at https://www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en.

https://doi.org/10.1016/j.insmatheco.2023.11.003
Received September 2021; Received in revised form September 2023; Accepted 15 November 2023
Available online 30 November 2023

which is an interpolating RFR-TS model that can both interpolate observable interest rates and extrapolate the term structure beyond the maturities of observable market instruments. In particular, the extrapolation ability of the Smith-Wilson method allows insurers to assess the present value of long-term cash flows. The Smith-Wilson method consists of four major elements: the market values of instruments within the liquid part; the last liquid point (LLP), which is the maturity beyond which market rates are not used[2]; the level of the UFR; and the extrapolation method, which connects the forward rate at the LLP with the UFR.[3]

Among the four elements listed above, the UFR is one of the most significant parameters for the Smith-Wilson method (Jørgensen, 2018). The UFR is defined as the sum of the expected real interest rate and the expected inflation rate in the distant future (EIOPA, 2019, Paragraph 353). For example, in 2017, the UFR was set to be 4.20% for the Eurozone and 4.50% for China (EIOPA, 2019, Paragraph 351) by the EIOPA. Since then, the EIOPA announces the UFR annually and limits its fluctuation within 15 basis points (EIOPA, 2019, Paragraph 352).[4]

In the EIOPA's current RFR-TS framework, a positive UFR is chosen exogenously before fitting the curve. However, this approach has some limitations. Lagerås and Lindholm (2016) pointed out that using a (nearly) fixed UFR is inconsistent with the risk-free interest rate stress scenarios specified in Solvency II. In addition, adjusting the UFR annually may not capture changes in the risk-free interest rate environment timely. Balter et al. (2016) found empirically that long-term rates are indeed uncertain, and it is not reasonable to use a fixed UFR. To address these issues, de Kort and Vellekoop (2016, Section 5) proposed a method for estimating the UFR *endogenously* from market data. They characterized the Smith-Wilson family of interpolating functions as a solution to a functional optimization problem, and represented their endogenous UFR as a solution to a parameter optimization problem. de Kort and Vellekoop's (2016) method not only can generate smoother term structures than the Smith-Wilson method adopted by the EIOPA, but also provides a way to estimate endogenous UFRs using market data.

When finding an endogenous UFR, it is natural to impose a non-negative constraint on the UFR. Hagan and West (2006) discussed various interpolating methods, and argued that a good interpolating method should ensure that the term structure converges to a *positive* level. Although short-term interest rates for several currencies, such as the Euro, the US dollar, and the Japanese Yen, have been negative in recent years, the UFRs for these currencies published by the EIOPA are still significantly greater than zero. In addition, the EIOPA defines the UFR as the forward rate in the distant future, which should be positive because negative interest rate policies are considered to be temporary tools used only in extreme situations (Marques et al., 2021, Chapter 1).

Therefore, we need to extend the method proposed in de Kort and Vellekoop (2016) and develop a mathematical framework to obtain *endogenous* and *positive* UFRs based on the Smith-Wilson family. Our study focuses on exploring the mathematical theory behind generating such UFRs. In this article, we first prove the existence of a solution to the first-order condition of the optimization problem proposed in de Kort and Vellekoop (2016, Section 5). The result not only guarantees the feasibility of their method, but also enhances our understanding of the first-order condition.

Then, we propose two new feasible methods for constructing the RFR-TS with endogenous and positive UFRs. Our first method (Method 1) is based on the framework proposed in de Kort and Vellekoop (2016, Section 5) and guarantees the positivity of the endogenous UFRs. We prove the feasibility of the method under several sufficient conditions, and demonstrate its effectiveness through empirical studies for both the Chinese and the European markets. We also show that, if the market observed data are generated by several conventional interest rate models—the Vasicek model and the Cox-Ingersoll-Ross (CIR) model, the sufficient conditions can be satisfied. In addition, we prove that the method is consistent with the risk-free interest rate stress scenarios specified in Solvency II.

We further propose our second method (Method 2), which can not only find endogenous and positive UFRs, but also include practitioners' prior knowledge regarding the UFR into the model. Method 2 solves a new optimization problem whose objective function contains a regularization term that accounts for the distance between the endogenous UFR and the prior knowledge. We derive the first-order condition for this new objective function and prove the feasibility of Method 2. In particular, different from Method 1, we show that Method 2 is always practicable. Therefore, one can always use Method 2 as long as he/she has some prior knowledge regarding the UFR. Our empirical studies also demonstrate the stability and efficacy of Method 2.

The outline of this paper is as follows. Section 2 introduces the framework of the Smith-Wilson method and specifies the notations used in this paper. Section 3 introduces the method proposed by de Kort and Vellekoop (2016, Section 5), and proves the existence of a solution to their first-order condition. In addition, we extend their method to generate both endogenous and positive UFRs (Method 1). The theoretical properties of the extended method are also studied. Section 4 proposes a new optimization framework allowing practitioners to generate endogenous and positive UFRs with prior knowledge (Method 2). Section 5 presents the empirical results of our methods using Chinese government bond data. Section 6 concludes. Appendix A gives the proofs of our theoretical results. We also discuss potential extensions of our methods in Appendix B, and provide more empirical studies for the European market in Appendix C.

2. The Smith-Wilson method

For completeness, we first present the Smith-Wilson method adopted by the EIOPA.[5]

We assume that there are N risk-free fixed income instruments traded in a single market at time 0, and we fit the risk-free interest rate term structure at time 0. There are T possible payment times in the future for these instruments, say $u_1, u_2, \dots, u_T$, respectively, where $0 < u_1 < u_2 < \cdots < u_T$. For $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, T$, the payment of the i-th instrument at time u_j is $c_{ij} \geq 0$. At time 0, the market price of the i-th instrument is m_i. Let $p(0,t)$ denote the discount curve (discount factor) to be constructed, which represents the price of zero-coupon bonds with one unit payment at the maturity time t.

Smith and Wilson (2000) represented the discount factor $p(0,t)$ by

$$p(0,t) = (1 + g(t))e^{-f_\infty t}, \quad t \geq 0. \tag{1}$$

Here f_∞ is an exogenous parameter representing the UFR, and function $g(\cdot)$ satisfies $g \in \mathcal{G}$, where

[2] The LLP is the longest maturity for which the bond market and the market of relevant financial instruments are deemed to be deep, liquid, and transparent. For example, the LLP is set to be 20 years for the Eurozone and 10 years for China (EIOPA, 2019, Tables 7–8).

[3] The Smith-Wilson method allows the speed of convergence towards the UFR to be controlled and thereby the speed of convergence is chosen in such a way that the forward rates are, up to an immaterial difference, equal to the UFR for maturities at a specified convergence point (CP). Currently, the CP is set to be 60 years for both the Eurozone and China (EIOPA, 2019, Paragraph 120).

[4] More information about the UFR can be found in EIOPA (2016) and EIOPA (2019).

[5] For the consistency of notations in this article, we use the Smith-Wilson framework and notations used in de Kort and Vellekoop (2016). The framework and notations used in the original paper, Smith and Wilson (2000), are slightly different.

$$\mathcal{G} := \{g \in C^2(\mathbb{R}^+) : g(0) = 0, g''(0) = 0, \lim_{t\to+\infty} g'(t) = 0, \lim_{t\to+\infty} g''(t) = 0\}. \tag{2}$$

For $g \in \mathcal{G}$, (1) implies that the instantaneous forward rate can be represented by

$$f(t) = -\frac{\partial \ln p(0,t)}{\partial t} = -\frac{g'(t)}{1+g(t)} + f_\infty, \quad t \geq 0, \tag{3}$$

which will converge to the UFR (f_∞) as the term t increases without bound.

The function g used by Smith and Wilson (2000) satisfies the following optimization problem

$$\arg\min_{g\in\mathcal{G}} \int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right] \mathrm{d}s, \tag{4}$$

$$\text{s.t. } m_i = \sum_{j=1}^{T} c_{ij} p(0,u_j), \quad i = 1,2,\ldots,N, \tag{5}$$

where $\alpha > 0$ is a given tuning parameter, which is used for the tradeoff between the slope (g') and the curvature (g''). The optimal solution takes the form[6]

$$g(t) = \sum_{i=1}^{N} \zeta_i \sum_{j=1}^{T} c_{ij} e^{-f_\infty u_j} W_\alpha(t,u_j), \quad t \geq 0, \tag{6}$$

and the optimal discount curve is

$$p(0,t) = e^{-f_\infty t}\left[1 + \sum_{i=1}^{N} \zeta_i \sum_{j=1}^{T} c_{ij} e^{-f_\infty u_j} W_\alpha(t,u_j)\right], \quad t \geq 0, \tag{7}$$

where ζ_i are constants that can be determined by substituting (7) into (5), and

$$W_\alpha(t,u) := \alpha \min(t,u) - \frac{1}{2}e^{-\alpha|t-u|} + \frac{1}{2}e^{-\alpha(t+u)}, \quad t,u \geq 0. \tag{8}$$

The function $W_\alpha(\cdot,\cdot)$, which is also known as the exponential tension spline function, plays a crucial role in the Smith-Wilson method.

There are two parameters that need to be predetermined before using the Smith-Wilson method: the UFR f_∞ and the tuning parameter α. The UFR is exogenously determined by the EIOPA based on the expected real interest rate and the expected inflation rate of the currency in the distant future. The tuning parameter α is chosen using the following rule (EIOPA, 2019, Paragraphs 121 and 158):

$$\alpha^* = \inf\left\{\ \alpha : \alpha \geq \alpha_{\min},\ |f(\mathrm{CP}) - f_\infty| \leq \tau\ \right\}, \tag{9}$$

where $f(\cdot)$ is the forward rate function (3) induced by the optimal function g (the optimal solution to (4)) given parameter α, $\alpha_{\min}$ is the lower bound of α, CP is the convergence point satisfying $\mathrm{CP} > u_T$, and τ is the maximum threshold of the convergence gap at the convergence point. For example, in EIOPA (2019, Paragraphs 120–121), the EIOPA sets these hyperparameters to be $\alpha_{\min} = 0.05$, $\mathrm{CP} = 60$ years, and $\tau = 0.0001$ for both the Euro and the Chinese yuan renminbi.[7]

In summary, the Smith-Wilson method adopted by the EIOPA has the following features:

- *Interpolation.* The discount curve matches the market prices of instruments with different maturities exactly, see (5);
- *Extrapolation.* The forward rate converges to the exogenous UFR f_∞ as the term t increases without bound, see (2) and (3);
- *Smoothness.* The discount curve is smooth enough, see (4).

[6] See, for example, de Kort and Vellekoop (2016, Theorem 1).

[7] All empirical studies in this paper use these values.

However, in this framework, the UFR is chosen exogenously without using data of the current fixed income instrument market. As a result, the exogenous UFR set by the EIOPA may not reflect the current market environment. In the following sections, we introduce the framework proposed in de Kort and Vellekoop (2016, Section 5), which allows for the determination of endogenous UFRs, and further propose several methods for generating UFRs that are both endogenous and positive.

3. Endogenous and positive UFRs using the de Kort-Vellekoop method

In this section, we propose a method for obtaining UFRs that are both endogenous and positive, building upon the work of de Kort and Vellekoop (2016), who proposed a framework for finding endogenous UFRs without non-negative constraints. We refer to their approach as the de Kort-Vellekoop method. Section 3.1 provides a brief overview of their optimization problem. In Section 3.2, we prove the existence of a solution to their optimization problem under certain regularity conditions. A real data example is also given to illustrate that the optimal endogenous UFR obtained by their method may not be positive. To address this issue, in Section 3.3, we propose a new method for generating UFRs that are both endogenous and positive, and prove the feasibility of our method.

3.1. The de Kort-Vellekoop method

For a given parameter $\alpha > 0$, de Kort and Vellekoop (2016, Section 5) proposed the following optimization problem for solving endogenous UFRs:

$$\arg\min_{f_\infty} \min_{g\in\mathcal{H}(f_\infty)} \int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right] \mathrm{d}s, \tag{10}$$

where

$$\mathcal{H}(f_\infty) = \Bigg\{ g \in \mathcal{G} : \sum_{j=1}^{T} c_{ij} e^{-f_\infty u_j} g(u_j) = m_i - \sum_{j=1}^{T} c_{ij} e^{-f_\infty u_j}, i = 1,2,\ldots,N \Bigg\}. \tag{11}$$

Compared with the original Smith-Wilson optimization problem (4), there are two steps to solve (10): First, for any given f_∞, we solve an optimization problem the same as problem (4) in the original Smith-Wilson method. Second, we find an optimal f_∞ to minimize the objective function. Problem (10) not only allows us to obtain a curve smoother than the curve generated by the original Smith-Wilson method, but also provides a way to obtain an endogenous UFR.

Given the observed market data, we define an $N \times T$ cash flow matrix $\mathbf{C}$ as:

$$[\mathbf{C}]_{ij} = c_{ij}, \quad i = 1,2,\ldots,N; \ \ j = 1,2,\ldots,T,$$

and a $T \times T$ matrix $\mathbf{W}_\alpha$ as

$$[\mathbf{W}_\alpha]_{ij} = W_\alpha(u_i,u_j), \quad i,j = 1,2,\ldots,T.$$

Furthermore, let

$$\mathbf{D}^{f_\infty} = \mathrm{diag}\{e^{-f_\infty u_1}, e^{-f_\infty u_2}, \ldots, e^{-f_\infty u_T}\}, \ \mathbf{U} = \mathrm{diag}\{u_1,u_2,\ldots,u_T\}$$

be two $T \times T$ diagonal matrices, and let

$$m = (m_1 \ \ m_2 \ \ \cdots \ \ m_N)^\top, \quad \mathbf{1} = (1 \ \ 1 \ \ \cdots \ \ 1)^\top$$

be an N-dim vector and a T-dim vector, respectively. de Kort and Vellekoop (2016) concluded that the optimal f_∞ for problem (10) should satisfy the following first-order condition:

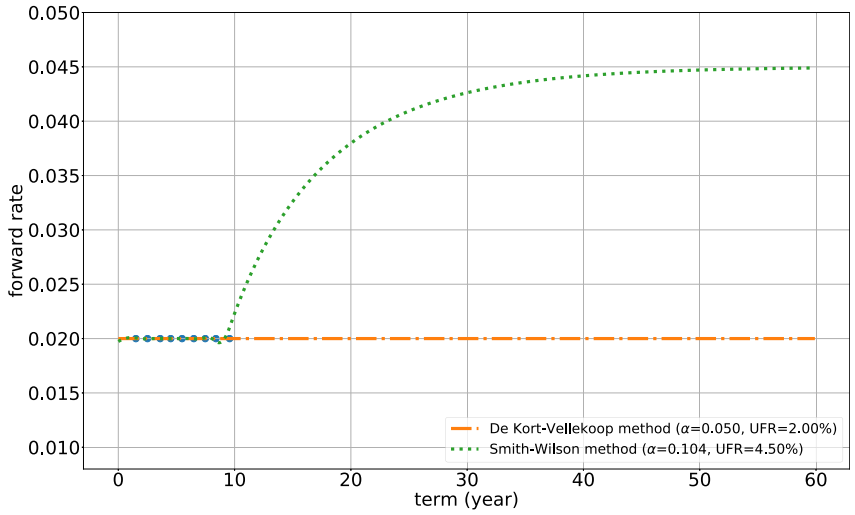


Fig. 1. Forward curves obtained by two methods when $\pi_i = e^{-ru_i}$ for all $i = 1, 2, \ldots, T$. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

$$(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{U}$$
$$\times \left(\mathbf{1} + \mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top(\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})\right) = 0. \tag{12}$$

In particular, when $N = T$ and the cash flow matrix $\mathbf{C}$ is invertible, (12) simplifies to

$$\sum_{i=1}^{T}\sum_{j=1}^{T}(u_i\pi_i e^{f_\infty u_i})[\mathbf{W}_\alpha^{-1}]_{ij}(\pi_j e^{f_\infty u_j} - 1) = 0, \tag{13}$$

where

$$\pi_i = \sum_{j=1}^{T}[\mathbf{C}^{-1}]_{ij}m_j \tag{14}$$

can be considered as the implied price of zero-coupon bonds with maturity u_i calibrated using the observed data. Hereinafter, to highlight the central idea of our paper, we always assume that $N = T$ and the cash flow matrix $\mathbf{C}$ is invertible, and we relax this assumption in Appendix B.[8]

3.2. Existence of a solution to de Kort-Vellekoop's first-order condition

In this section, we establish the existence of a solution to de Kort and Vellekoop's (2016) first-order condition (13) for the endogenous UFR.

Before presenting the theorem (Theorem 1), let us first explore the intuition behind the first-order condition (13). The following Proposition 1 demonstrates that the solutions to (13) form a centrosymmetric quadratic surface.

Proposition 1. *Let*

$$X = (\pi_1 e^{f_\infty u_1} \quad \pi_2 e^{f_\infty u_2} \quad \cdots \quad \pi_T e^{f_\infty u_T})^\top. \tag{15}$$

The first-order condition (13) *is equivalent to*

$$\left(X - \frac{\mathbf{B}^{-1}\mathbf{A}\mathbf{1}}{2}\right)^\top \mathbf{B}\left(X - \frac{\mathbf{B}^{-1}\mathbf{A}\mathbf{1}}{2}\right) = \frac{\mathbf{1}^\top\mathbf{A}^\top\mathbf{B}^{-1}\mathbf{A}\mathbf{1}}{4}, \tag{16}$$

where

$$\mathbf{A} = \mathbf{U}\mathbf{W}_\alpha^{-1} \quad \textit{and} \quad \mathbf{B} = (\mathbf{A} + \mathbf{A}^\top)/2. \tag{17}$$

In particular, (16) *is a centrosymmetric quadratic surface with respect to* X *that centred at* $\dfrac{\mathbf{B}^{-1}\mathbf{A}\mathbf{1}}{2}$.

With the help of (16), an equivalent form of the first-order condition (13), we can immediately obtain the following Corollary 1.

Corollary 1. *Equation* (13) *with respect to* X *has at least four solutions:*

$$X_1 = \mathbf{0}, \quad X_2 = \mathbf{1}, \quad X_3 = \mathbf{B}^{-1}\mathbf{A}\mathbf{1}, \quad X_4 = (\mathbf{I} - \mathbf{B}^{-1}\mathbf{A})\mathbf{1},$$

where $\mathbf{0}$ *is a* T*-dim zero vector,* $\mathbf{I}$ *is a* $T \times T$ *identity matrix, and* $\mathbf{A}$ *and* $\mathbf{B}$ *are as defined in* (17).

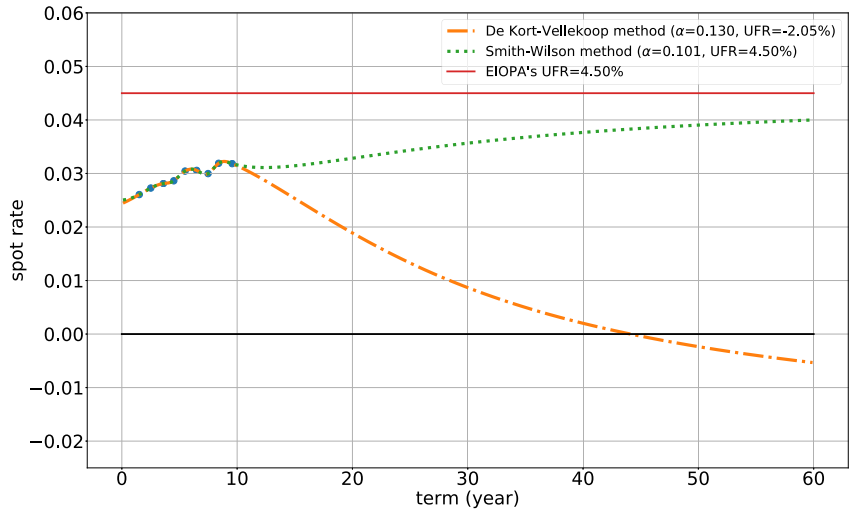
Corollary 1 provides some important insights into the endogenous UFR obtained by the first-order condition (13). The first solution $X_1 = \mathbf{0}$ implies $f_\infty = -\infty$, which is trivial. The second solution $X_2 = \mathbf{1}$ implies that, when the spot curve is flat, i.e., there is a constant r such that $\pi_i = e^{-ru_i}$ for all $i = 1, 2, \ldots, T$, the endogenous UFR f_∞ is exactly the same as the flat rate r. The third and fourth solutions satisfy $X_4 = \mathbf{1} - X_3$. Both the third and the fourth solutions are nontrivial, and the explanations for X_3 and X_4 need to be studied further.

The second solution X_2 motivates us to study the special case where the curve is flat, i.e., there exists a constant r such that $\pi_i = e^{-ru_i}$ for all $i = 1, 2, \ldots, T$. Fig. 1 illustrates the forward curves obtained by both the Smith-Wilson method adopted by the EIOPA and the de Kort-Vellekoop method when $r = 2.00\%$, $T = 9$, and $u_i = 1.51, 2.51, 3.60, 4.50, 5.48, 6.49, 7.48, 8.39, 9.57$, for $i = 1, 2, \ldots, T$, respectively.[9] The green dotted curve is the forward curve constructed by the Smith-Wilson method, which distorts the curve to the exogenous UFR (4.50% in this example), while the orange curve constructed by the de Kort-Vellekoop method stays at the level of 2.00% without distortion.[10] In this case, the curve generated by the Smith-Wilson method with an exogenous UFR is counterintuitive, and the result of the de Kort-Vellekoop method is more reasonable.

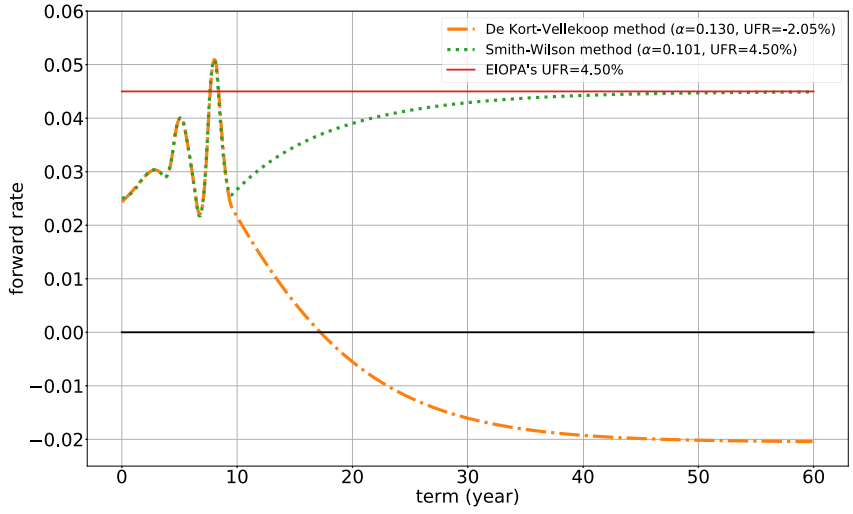
[8] In practice, practitioners can always modify a rectangular $\mathbf{C}$ (i.e., $N \neq T$) into a square matrix by compressing the cash flows, see Section 5.1 for an example. In addition, $\mathbf{C}$ is often required to be a square matrix by the EIOPA when practitioners are using zero-coupon bond data or swap data to construct risk-free interest rate curves, see EIOPA (2019, Table 9).

[9] The values of u_i are set to be the terms of the cash flow matrix generated using Chinese government bonds traded in 2020 Quarter 3, see Table 1 in Section 5.1.

[10] The exogenous UFR is set to be 4.50% for China, see EIOPA (2019, Paragraph 351). In this example, parameter α is set to be 0.104 for the Smith-Wilson method and 0.050 for the de Kort-Vellekoop method, according to (9).



(a) Spot rate curves, 2020 Quarter 3



(b) Forward rate curves, 2020 Quarter 3

Fig. 2. An example of spot and forward curves with an endogenous and negative UFR generated by the de Kort-Vellekoop method.

One should note that Corollary 1 shows the existence of vector X but not the optimal endogenous f_∞. In the following Theorem 1, we propose a regularity condition for the matrix $\mathbf{W}_\alpha$, under which we can establish the existence of the optimal f_∞.

Theorem 1. *Given $\alpha > 0$, the first-order condition (13) with respect to f_∞ has at least one solution if the regularity condition*

$$\sum_{j=1}^{T} [\mathbf{W}_\alpha^{-1}]_{1j} > 0 \tag{18}$$

holds.

Remark 1. The regularity condition (18) states that the sum of the first row of the inverse matrix $\mathbf{W}_\alpha^{-1}$ should be greater than zero. Our simulation and empirical studies both indicate that $\mathbf{W}_\alpha^{-1}$ almost always satisfies the regularity condition (18). In addition, Lemma A.1 in Appendix A implies that the regularity condition holds for large values of α.

To prove Theorem 1, we first give the following two lemmas.

Lemma 1. *The matrix $\mathbf{W}_\alpha^{-1}$ is positive definite, and the diagonal entries of $\mathbf{W}_\alpha^{-1}$ are all positive. More precisely, the i-th diagonal entry $[\mathbf{W}_\alpha^{-1}]_{ii}$ satisfies $[\mathbf{W}_\alpha^{-1}]_{ii} > 1/[\mathbf{W}_\alpha]_{ii} > 1/(\alpha u_i)$.*

Lemma 2. *Given $\alpha > 0$, denote the left-hand side of the first-order condition (13) by a function of f_∞:*

$$h_\alpha(f_\infty) = \sum_{i=1}^{T}\sum_{j=1}^{T} (u_i \pi_i e^{f_\infty u_i})[\mathbf{W}_\alpha^{-1}]_{ij}(\pi_j e^{f_\infty u_j} - 1), \qquad f_\infty \in \mathbb{R}. \tag{19}$$

Then,

(a) $$\lim_{f_\infty \to -\infty} \frac{h_\alpha(f_\infty)}{e^{f_\infty u_1}} = -u_1 \pi_1 \sum_{j=1}^{T} [\mathbf{W}_\alpha^{-1}]_{1j}, \quad \text{and}$$

(b) $$\lim_{f_\infty \to +\infty} \frac{h_\alpha(f_\infty)}{e^{2 f_\infty u_T}} = u_T \pi_T^2 [\mathbf{W}_\alpha^{-1}]_{TT}.$$

Remark 2. Result (a) implies that $h_\alpha(f_\infty)$ converges to 0 exponentially as $f_\infty \to -\infty$. In addition, when $\sum_{j=1}^{T} [\mathbf{W}_\alpha^{-1}]_{1j} > 0$, $h_\alpha(f_\infty)$ will be negative as f_∞ approaches $-\infty$. Result (b) implies that $h_\alpha(f_\infty)$ goes to $+\infty$ exponentially as $f_\infty \to +\infty$ because Lemma 1 implies that $[\mathbf{W}_\alpha^{-1}]_{TT} > 0$.

Thanks to Lemma 1 and Lemma 2, it is now easy to prove Theorem 1.

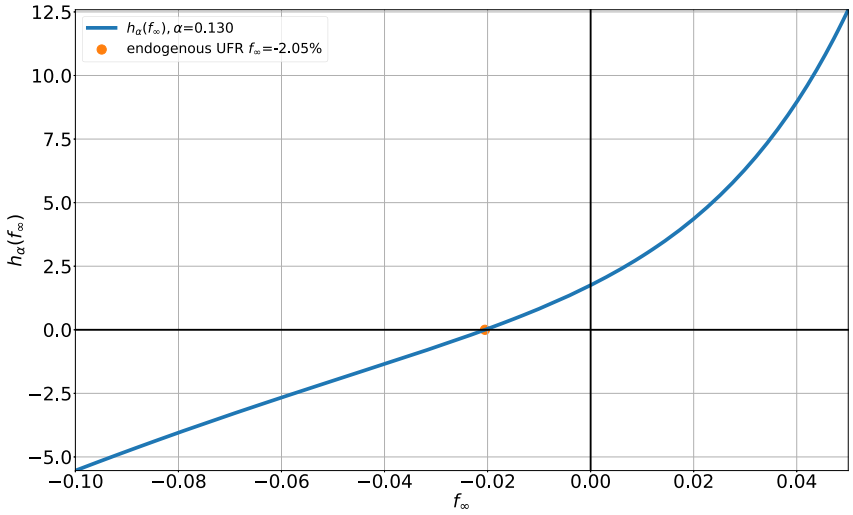


Fig. 3. The graph of the function $h_\alpha(f_\infty)$ with $\alpha = 0.130$, 2020 Quarter 3.

Proof of Theorem 1. The result holds because of Lemma 1, Lemma 2, and the continuity of $h_\alpha(f_\infty)$ with respect to f_∞. □

Theorem 1 provides a sufficient condition for the de Kort-Vellekoop method to guarantee the existence of an endogenous f_∞. However, the method cannot guarantee the *positivity* of the endogenous f_∞ in extreme time periods. For instance, Fig. 2 presents an example using the data of Chinese government bonds traded in 2020 Quarter 3 during the COVID-19 pandemic.[11] The forward rate curves show that the short-term interest rates fluctuated wildly at that time. In this case, the orange lines converge to a negative and endogenous UFR of −2.05%, which is not reasonable because it is commonly believed that the forward rate should be positive when the term is long enough (Hagan and West, 2006). Fig. 3 shows the corresponding graph of the function $h_\alpha(f_\infty)$, and one can observe that the equation $h_\alpha(f_\infty) = 0$ has a negative solution −2.05%. Therefore, it is necessary to improve the method proposed by de Kort and Vellekoop (2016), and give feasible ways to obtain UFRs that are both endogenous and positive.

3.3. The extended de Kort-Vellekoop method

In this section, we extend the de Kort-Vellekoop method to find UFRs that are both endogenous and positive. We first propose our extended de Kort-Vellekoop method that ensures the positivity of the endogenous UFR (Section 3.3.1). Then, we show the feasibility of our method (Section 3.3.2). Finally, we show that our method is consistent with the stress scenarios specified in Solvency II (Section 3.3.3).

3.3.1. The method

The following Method 1 provides a feasible way to obtain endogenous and positive UFRs from market data. The input data includes the times of payments, $u_1, u_2, \ldots, u_T$; a $T \times T$ cash flow matrix, $\mathbf{C}$; and the market prices of the instruments, $m_1, m_2, \ldots, m_T$. There are also several model parameters: the lower bound of α, $\alpha_{\min}$; the convergence point, CP; and the maximum threshold of the convergence gap at the convergence point, τ.

Method 1. ***(Endogenous and positive UFRs)***

Step 1: Derive the feasible region of α:

$$\mathcal{A} := \left\{ \alpha : \alpha > 0, \sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i\right) \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j - 1\right) < 0 \right\}, \tag{20}$$

$$\mathcal{B} := \left\{ \alpha : \alpha \in \mathcal{A},\ \alpha \geq \alpha_{\min},\ |f^\alpha(\mathrm{CP}) - f_\infty^\alpha| \leq \tau \right\}, \tag{21}$$

where $\pi_1, \pi_2, \ldots, \pi_T$ are the prices of zero-coupon bonds as defined in (14), f_∞^α is the solution to the first-order condition (13) given α, and $f^\alpha(\cdot)$ is the forward curve generated by the Smith-Wilson method (3) given α and $f_\infty = f_\infty^\alpha$.

Step 2: Choose an optimal α:

$$\alpha^* = \inf \mathcal{B}.$$

Step 3: Solve the following equation for f_∞:

$$\sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i e^{f_\infty u_i}\right) \left[\mathbf{W}_{\alpha^*}^{-1}\right]_{ij} \left(\pi_j e^{f_\infty u_j} - 1\right) = 0.$$

Remark 3. Step 2 is similar to EIOPA (2019, Paragraph 158), see (9).

In summary, we first construct the feasible region of α, denoted by $\mathcal{B}$. Then, we find the optimal α^* in $\mathcal{B}$. Finally, we solve for the corresponding f_∞ using the first-order condition given $\alpha = \alpha^*$.

Now we explain why Method 1 can guarantee the positivity of the endogenous UFR. In fact, the definition of $\mathcal{A}$ given by (20) is equivalent to $\mathcal{A} = \{\alpha : \alpha > 0, h_\alpha(0) < 0\}$, where $h_\alpha(\cdot)$ is given by (19). Hence, $h_{\alpha^*}(0) \leq 0$. In addition, Remark 2 shows that $h_{\alpha^*}(f_\infty)$ must be positive as $f_\infty \to +\infty$. Therefore, the continuity of $h_{\alpha^*}(f_\infty)$ guarantees the existence of a non-negative solution for Step 3.

Remark 4. The main difference between Method 1 and the de Kort-Vellekoop method is the selection of α^*. In particular, Method 1 uses the sets $\mathcal{A}$ and $\mathcal{B}$ as defined by (20) and (21) to determine α^*, while the de Kort-Vellekoop method and the original Smith-Wilson method use (9). If the value of α^* derived from (9) falls within the set $\mathcal{A}$, then the α^* chosen by Method 1 and the de Kort-Vellekoop method will be the same, and both methods will yield the same endogenous positive UFR. Our empirical study in Section 5 shows that their results are the same in most cases, with Method 1 additionally ensuring the positivity of the UFR during extreme periods, such as when interest rates are wildly fluctuating during the COVID-19 pandemic. Therefore, in general, practitioners only need to apply Method 1 during extreme time periods.

[11] In this example, we set $N = T = 9$. Parameter α is set to be 0.130 for the de Kort-Vellekoop method and 0.101 for the Smith-Wilson method, according to (9). See Section 5 for more details.

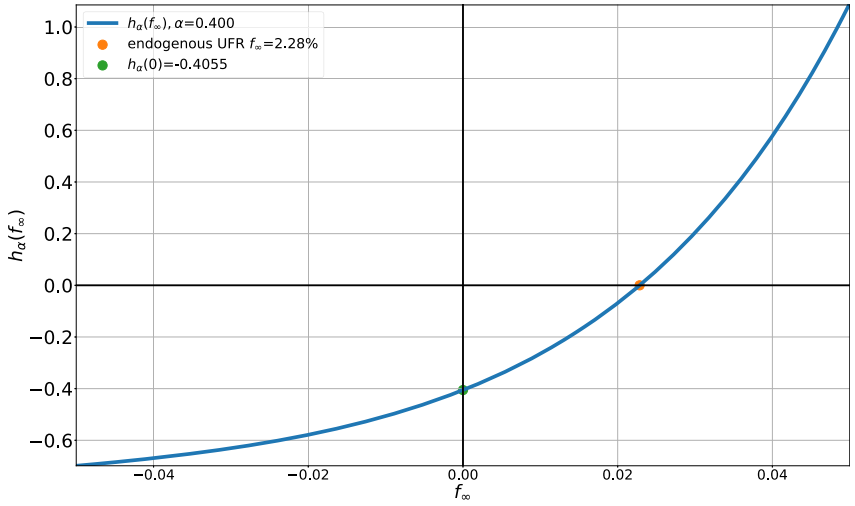


Fig. 4. The graph of the function $h_\alpha(f_\infty)$ with $\alpha = 0.400$, 2020 Quarter 3.

Remark 5. Method 1 is applicable when the cash flow matrix $\mathbf{C}$ is both square and invertible. This is because, as we discussed in Section 3.1, $h_\alpha(f_\infty) = 0$ is the first-order condition only under these conditions. When $\mathbf{C}$ is not square or not invertible, Method 1 cannot be directly employed. We provide a further discussion of this scenario in Appendix B.

Remark 6. In practice, ensuring positivity may not be the only possible constraint imposed on the endogenous UFR. If one wishes to impose other lower bounds on the UFR, she can modify Method 1 by defining $\mathcal{A} = \{\alpha : \alpha > 0, h_\alpha(f_\infty^{\min}) < 0\}$ in Step 1, where $f_\infty^{\min}$ is the specified lower bound. Then, using similar arguments above, a solution no less than $f_\infty^{\min}$ must exist for Step 3. For example, in negative interest rate environments, if one only wishes to introduce a negative lower bound to the UFR, she can set $f_\infty^{\min}$ to be a negative number. Theoretical analysis of these extended constraints is beyond the scope of this paper, and we only consider the positivity constraint in the following discussion.

Fig. 4 provides an example of $h_\alpha(f_\infty)$. The data used in Fig. 4 is the same as in Fig. 2, and the only difference is that we set $\alpha = 0.400$ in Fig. 4. In this example, we have $h_\alpha(0) = -0.4055 < 0$, which implies that $0.400 \in \mathcal{A}$. The endogenous UFR obtained using Method 1 given $\alpha = 0.400$ is 2.28%, which is positive.

Now we show the theoretical properties of Method 1. In Section 3.3.2, we prove that both sets $\mathcal{A}$ and $\mathcal{B}$ as defined in (20) and (21) are not empty under certain conditions. In Section 3.3.3, we show that our method is consistent with the risk-free interest rate scenarios specified in Solvency II.

3.3.2. Feasibility of the method

The following Theorem 2 provides conditions that are sufficient for ensuring that $\mathcal{A}$ and $\mathcal{B}$ are not empty. We set $\pi_0 := 1$ and $u_0 := 0$.

Theorem 2. *The regions $\mathcal{A}$ and $\mathcal{B}$ as defined in (20) and (21) are not empty if the following condition holds:*

$$\sum_{i=1}^{T} \left(u_i\pi_i - u_{i-1}\pi_{i-1}\right) \frac{\pi_i - \pi_{i-1}}{u_i - u_{i-1}} < 0. \tag{22}$$

In particular, (22) holds when both of the following conditions are satisfied for $i = 0, 1, \ldots, T$:

a) π_i is monotonically decreasing with respect to i;
b) $u_i\pi_i$ is monotonically increasing with respect to i.

Remark 7. Condition a) states that the prices of the zero-coupon bonds (i.e., the discount factors), π_i, should decrease as their maturities increase. The rationale behind this condition is that longer-term bonds should have lower present values than shorter-term bonds. Condition b) requires that the product of the payment time and the discount factor, $u_i\pi_i$, should increase as the maturity increases. In fact, $u_i\pi_i$ is known as the *dollar duration*[12] of a zero-coupon bond with maturity u_i, which is a risk measure. The intuition behind this condition is that longer-term bonds are generally considered riskier than shorter-term bonds, and their prices are more sensitive to changes in interest rates.

Our empirical study in Section 5 for the Chinese government bonds shows that both Condition a) and Condition b) are satisfied by our data. Unfortunately, in low (or even negative) interest rate environments, the two conditions may not hold. For example, if the spot rate at maturity time u_1, r_1, is negative, we will have $\pi_1 = e^{-u_1 r_1} > 1 = \pi_0$, which violates Condition a). Despite this, we show in Appendix C that (22) still holds in most cases when we use the EURIBOR swap data, even when Conditions a) and b) do not hold.

The following two examples illustrate that Conditions a) and b) in Theorem 2 can be satisfied if the spot rates are generated by two widely-used dynamic short-term interest rate models—the Vasicek model and the CIR model.

Example 1 *(Vasicek model).* Suppose that, under the risk-neutral measure, the instantaneous spot rate $r(t)$ follows

$$\mathrm{d}r(t) = \kappa(\theta - r(t))\mathrm{d}t + \sigma \mathrm{d}W_t, \quad r(0) = r_0 > 0,$$

where $\kappa, \theta, \sigma > 0$ and W_t is a standard Brownian motion. Then, the discount factor generated by the Vasicek model, $p^{\text{Vasicek}}(0,t)$, can be represented by (Brigo and Mercurio, 2007, Section 3.2)

$$p^{\text{Vasicek}}(0,t) = \exp\left[\frac{1}{\kappa}\left(\theta - \frac{\sigma^2}{2\kappa^2} - r_0\right)\left(1 - e^{-\kappa t}\right) - \frac{\sigma^2}{4\kappa^3}\left(1 - e^{-\kappa t}\right)^2 - \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)t\right]. \tag{23}$$

[12] The dollar duration of a bond is defined to be the opposite value of the partial derivative of its price with respect to its yield, and it measures the risk of the change in bond value when there is a small variation in the yield. In particular, for a zero coupon bond with maturity u_i and yield r_i, its price is $\pi_i = e^{-u_i r_i}$, and its dollar duration is $-\partial\pi_i/\partial r_i = u_i e^{-u_i r_i} = u_i\pi_i$.

Direct calculation shows that the derivatives of $p^{\text{Vasicek}}(0,t)$ and $t \cdot p^{\text{Vasicek}}(0,t)$ with respect to term t are

$$\frac{\mathrm{d}p^{\text{Vasicek}}(0,t)}{\mathrm{d}t} = -\left[r_0 e^{-\kappa t} + \theta(1-e^{-\kappa t}) - \frac{\sigma^2}{2\kappa^2}(1-e^{-\kappa t})^2\right] p^{\text{Vasicek}}(0,t), \quad \text{and}$$

$$\frac{\mathrm{d}[t \cdot p^{\text{Vasicek}}(0,t)]}{\mathrm{d}t} = p^{\text{Vasicek}}(0,t) + \frac{\mathrm{d}p^{\text{Vasicek}}(0,t)}{\mathrm{d}t} \cdot t,$$

respectively. Hence, at $t=0$, we have

$$\frac{\mathrm{d}p^{\text{Vasicek}}(0,t)}{\mathrm{d}t}\Big|_{t=0} = -r_0 < 0, \qquad \frac{\mathrm{d}[t \cdot p^{\text{Vasicek}}(0,t)]}{\mathrm{d}t}\Big|_{t=0} = 1 > 0.$$

Therefore, under the Vasicek dynamic, both conditions hold—$p^{\text{Vasicek}}(0,t)$ is decreasing and $t \cdot p^{\text{Vasicek}}(0,t)$ is increasing—when the term t is small.

If we additionally require that the discount factor converges to zero as the term goes to infinity, i.e., $\lim_{t\to+\infty} p^{\text{Vasicek}}(0,t) = 0$, then (23) implies that $\theta - \frac{\sigma^2}{2\kappa^2} > 0$. In this case, we have

$$\begin{aligned} & r_0 e^{-\kappa t} + \theta(1-e^{-\kappa t}) - \frac{\sigma^2}{2\kappa^2}(1-e^{-\kappa t})^2 \\ & > r_0 e^{-\kappa t} + \theta(1-e^{-\kappa t}) - \frac{\sigma^2}{2\kappa^2}(1-e^{-\kappa t}) \\ & = r_0 e^{-\kappa t} + \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)(1-e^{-\kappa t}) > 0, \end{aligned}$$

which implies that $\frac{\mathrm{d}p^{\text{Vasicek}}(0,t)}{\mathrm{d}t} < 0$ for all t, and therefore, Condition a) always holds.[13] □

Example 2 *(CIR model).* Suppose that, under the risk-neutral measure, the instantaneous spot rate $r(t)$ follows

$$\mathrm{d}r(t) = \kappa(\theta - r(t))\mathrm{d}t + \sigma\sqrt{r(t)}\mathrm{d}W_t, \quad r(0) = r_0 > 0,$$

where $\kappa, \theta, \sigma > 0$, and W_t is a standard Brownian motion. Then the discount factor generated by the CIR model, $p^{\text{CIR}}(0,t)$, can be represented by (Brigo and Mercurio, 2007, Section 3.2)

$$p^{\text{CIR}}(0,t) = \left[\frac{2h\exp((\kappa+h)t/2)}{2h+(\kappa+h)(\exp(th)-1)}\right]^{2\kappa\theta/\sigma^2} \cdot \exp\left[-r_0 \cdot \frac{2(\exp(th)-1)}{2h+(\kappa+h)(\exp(th)-1)}\right],$$

where $h = \sqrt{\kappa^2+2\sigma^2}$. Direct calculation shows that the derivatives of $p^{\text{CIR}}(0,t)$ and $t \cdot p^{\text{CIR}}(0,t)$ with respect to term t are

$$\begin{aligned} & \frac{\mathrm{d}p^{\text{CIR}}(0,t)}{\mathrm{d}t} \\ & = \Bigg(\frac{\kappa(\kappa+h)}{\sigma^2}\theta\left[1 - \frac{2he^{ht}}{2h+(\kappa+h)(e^{ht}-1)}\right] \\ & \quad - r_0 \frac{2he^{ht}[2h+(\kappa+h)(e^{ht}-1)] - [2(e^{ht}-1)][(k+h)he^{ht}]}{[2h+(\kappa+h)(e^{ht}-1)]^2}\Bigg) p^{\text{CIR}}(0,t), \end{aligned}$$

and

$$\frac{\mathrm{d}[t \cdot p^{\text{CIR}}(0,t)]}{\mathrm{d}t} = p^{\text{CIR}}(0,t) + \frac{\mathrm{d}p^{\text{CIR}}(0,t)}{\mathrm{d}t} \cdot t,$$

respectively. Similar to the Vasicek model, for the CIR model, at $t=0$, we also have

$$\frac{\mathrm{d}p^{\text{CIR}}(0,t)}{\mathrm{d}t}\Big|_{t=0} = -r_0 < 0, \qquad \frac{\mathrm{d}[t \cdot p^{\text{CIR}}(0,t)]}{\mathrm{d}t}\Big|_{t=0} = 1 > 0.$$

Therefore, under the CIR model, both conditions hold—$p^{\text{CIR}}(0,t)$ is decreasing and $t \cdot p^{\text{CIR}}(0,t)$ is increasing—when the term t is small. In fact, Condition a) always holds for the CIR model since the CIR model can guarantee the positivity of interest rates, which ensures that the discount factor is decreasing. □

3.3.3. Consistency with the stress scenarios in Solvency II

In this section, we demonstrate that Method 1 is consistent with the stress scenarios specified in Solvency II. Lagerås and Lindholm (2016) pointed out that using a (nearly) fixed UFR is inconsistent with the risk-free interest rate stress scenarios. They illustrated that, for long tenors, the stress scenarios can be regarded as a parallel shift in the spot rate curve. The following Proposition 2 considers the results when there is a parallel shift in the spot rate curve.

Proposition 2. *Let f_∞ be a solution to the first-order condition (13). Then, for any constant c, $f_\infty + c$ is a solution to the following equation with respect to $\tilde{f}_\infty$:*

$$\sum_{i=1}^{T}\sum_{j=1}^{T}(u_i\tilde{\pi}_i e^{\tilde{f}_\infty u_i})[\mathbf{W}_\alpha^{-1}]_{ij}(\tilde{\pi}_j e^{\tilde{f}_\infty u_j} - 1) = 0,$$

where $\tilde{\pi}_i = \pi_i e^{-cu_i}, i = 1,2,\ldots,T$.

Let us illustrate Proposition 2. For $i = 1,2,\ldots,T$, the discount factor for term u_i with spot rate r_i is $\pi_i = e^{-r_i u_i}$. If the spot rate curve experiences a parallel shift of magnitude c, the corresponding discount factor will change into $\tilde{\pi}_i = e^{-(r_i+c)u_i} = \pi_i e^{-cu_i}$. Hence, Proposition 2 implies that the endogenous and positive UFR generated by Method 1 will shift accordingly with the parallel shift in the spot rates. Therefore, our method is consistent with the stress scenarios specified in Solvency II (Lagerås and Lindholm, 2016).

4. Endogenous and positive UFRs with prior knowledge

Method 1 can be used to solve for an endogenous and positive UFR. However, although an endogenous UFR can reflect the market environment, the optimal solution may not comply with common sense during some periods. Daily variations in fixed income instrument prices can make the optimal UFR sensitive to market fluctuations and noise. Since regulators and practitioners may have certain prior knowledge about the UFR, this knowledge can be incorporated into our model. In this section, we propose Method 2 to generate endogenous and positive UFRs with prior knowledge, and demonstrate its feasibility.

In order to incorporate prior knowledge about the UFR into the framework, we consider a new optimization problem[14]:

$$\arg\min_{f_\infty}\min_{g\in\mathcal{H}(f_\infty)}\left[\int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right]\mathrm{d}s + \frac{\lambda}{2}\alpha^3(f_\infty - f_{\text{prior}})^2\right], \tag{24}$$

where $\mathcal{H}(f_\infty)$ is as defined in (11), $f_{\text{prior}} > 0$ represents the prior knowledge about the UFR (for example, the EIOPA sets the UFR for China to be 4.50% (EIOPA, 2019, Paragraph 351), so we can set $f_{\text{prior}} = 4.50\%$), and $\lambda > 0$ is a tuning parameter. Optimization problem (24) is inspired by the ridge regression, which is a widely-used tool in regression analysis (Hastie et al., 2009). By adding the regularization term $\frac{\lambda}{2}\alpha^3(f_\infty - f_{\text{prior}})^2$, the optimal solution f_∞ to (24) will be closer to f_{prior}

[13] In fact, Condition b) may not hold when t goes to infinity. However, in practice, we always use fixed income instruments with maturities less than the last liquidity point (LLP) (EIOPA, 2019) to construct and extrapolate interest rate curves. For example, the EIOPA sets the LLP to be 20 years for the Eurozone and 10 years for China. Therefore, we only focus on the results when t is small.

[14] The regularization term $\frac{\lambda}{2}\alpha^3(f_\infty - f_{\text{prior}})^2$ includes the factor α^3 to simplify the first-order condition (25).

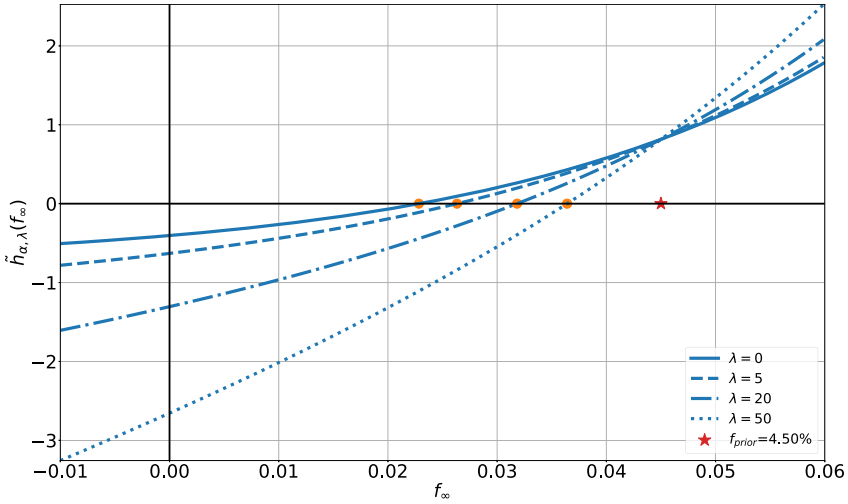


Fig. 5. The graphs of the function $\tilde{h}_{\alpha,\lambda}(f_\infty)$ with $\alpha = 0.400$ and different values of λ.

than the solution to de Kort-Vellekoop's optimization problem (10). The tuning parameter λ represents our confidence in the prior knowledge f_{prior}, and the larger the parameter λ, the closer f_∞ and f_{prior} will be. In particular, when $\lambda = 0$, problem (24) reduces to de Kort-Vellekoop's optimization problem (10).

The following Theorem 3 gives the first-order condition for the optimization problem (24).

Theorem 3. *The optimal solution f_∞ to optimization problem (24) should satisfy the following first-order condition*

$$\sum_{i=1}^{T}\sum_{j=1}^{T}\left(u_i\pi_i e^{f_\infty u_i}\right)\left[\mathbf{W}_\alpha^{-1}\right]_{ij}\left(\pi_j e^{f_\infty u_j}-1\right)+\lambda(f_\infty - f_{\text{prior}})=0. \tag{25}$$

Similar to Theorem 1, the following Theorem 4 states that the solution to the first-order condition (25) also exists.

Theorem 4. *Given $\lambda > 0$, the first-order equation (25) with respect to f_∞ has at least one solution.*

Remark 8. In comparison, Theorem 4 establishes the existence of a solution to the first-order condition (25) without requiring the regularity condition (18) stated in Theorem 1.

Fig. 5 illustrates the relationship between the roots of $\tilde{h}_{\alpha,\lambda}(f_\infty)$ and λ. The four curves in Fig. 5 are the graphs of $\tilde{h}_{\alpha,\lambda}(f_\infty)$ for different values of λ. As we increase the tuning parameter λ, the values of f_∞ (orange circles) will approach the f_{prior} (red star). In practice, we can choose λ according to our confidence in the prior knowledge f_{prior}.

By applying Theorem 3, we propose the following Method 2, which can generate endogenous and positive UFRs and incorporate practitioners' prior knowledge. In addition to the input parameters of Method 1, we also need to predetermine f_{prior} and λ.

Method 2. ***(Endogenous and positive UFRs with prior knowledge)***

Step 1: Derive the feasible region of α:

$$\tilde{\mathcal{A}} := \left\{\alpha : \alpha > 0, \sum_{i=1}^{T}\sum_{j=1}^{T}\left(u_i\pi_i\right)\left[\mathbf{W}_\alpha^{-1}\right]_{ij}\left(\pi_j - 1\right) - \lambda f_{\text{prior}} < 0\right\}, \tag{26}$$

$$\tilde{\mathcal{B}} := \left\{\ \alpha : \alpha \in \tilde{\mathcal{A}},\ \alpha \geq \alpha_{\min},\ |\tilde{f}^\alpha(\text{CP}) - \tilde{f}^\alpha_\infty| \leq \tau\ \right\}, \tag{27}$$

where $\tilde{f}^\alpha_\infty$ is the solution to the new first-order condition (25) given α, and $\tilde{f}^\alpha(\cdot)$ is the forward curve generated by the Smith-Wilson method (3) given α and $f_\infty = \tilde{f}^\alpha_\infty$.

Step 2: Choose an optimal α:

$$\tilde{\alpha}^* = \inf \tilde{\mathcal{B}}.$$

Step 3: Solve the following equation for f_∞:

$$\sum_{i=1}^{T}\sum_{j=1}^{T}\left(u_i\pi_i e^{f_\infty u_i}\right)\left[\mathbf{W}_{\tilde{\alpha}^*}^{-1}\right]_{ij}\left(\pi_j e^{f_\infty u_j}-1\right)+\lambda(f_\infty - f_{\text{prior}})=0.$$

Remark 9. Method 2 reduces to Method 1 when $\lambda = 0$.

Finally, in the following Theorem 5, we show that both $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ as defined in (26) and (27) are not empty, which implies the feasibility of Method 2.

Theorem 5. *Given $\lambda > 0$, the regions $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{B}}$ as defined in (26) and (27) are not empty.*

Theorem 5 shows the feasibility of Method 2. Unlike Theorem 2, which imposes certain conditions for the feasibility of Method 1, Theorem 5 shows that Method 2 is always feasible for any input data.

5. Empirical study

In this section, we illustrate the empirical performance of Method 1 and Method 2 for constructing risk-free interest rate curves of Chinese government bonds. For comparison, we also present the results of the Smith-Wilson method used by the EIOPA and the de Kort-Vellekoop method proposed in de Kort and Vellekoop (2016) (see optimization problem (10)). We demonstrate that our methods are highly effective and practical for constructing smooth curves with endogenous and positive UFRs. We also obtain several useful empirical facts about the risk-free interest rate in the Chinese market. Furthermore, we provide an empirical study in Appendix C to demonstrate the performance of our methods in low (and even negative) interest rate environments using the EURIBOR swap data.

5.1. Data and preprocessing

Our empirical study is based on quarterly data of Chinese government bonds traded on the China Foreign Exchange Trade System

big matrix	0.02943	0.04038	0.04654	0.04723	0.05065	0.05407	0.05886	0.06229	0.0883	0.09001	0.09172	0.09343	0.12252	0.12765	0.13552	0.13723	0.14168	0.14203	0.14339	0.14511	0.17933	0.18275	0.1961
104.1546	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.9725	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
105.8327	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.345	0	0	0	0	0	0	0	0
105.6091	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.7221	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.845
101.4931	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101.533	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102.6118	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.69	0	0	0	0	0
104.4579	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
105.1456	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
105.8219	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106.0693	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
104.704	0	0	0	0	0	0	0	3.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
104.3167	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.3225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.179	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102.1261	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.2279	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101.1411	0	1.495	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103.5214	0	0	0	0	0	0	3.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99.6141	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101.4749	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100.3443	0	0	0	0	0	0	0	0	0	0	0	1.45	0	0	0	0	0	0	0	0	0	0	0
101.0567	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98.257	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100.0201	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102.0168	0	0	0	0	0	2.39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
98.6314	0	0	0	0	0	0	0	0	0	0	1.35	0	0	0	0	0	0	0	0	0	0	0	0
102.7382	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102.5316	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 6. A small fragment of the raw cash flow matrix, 2020 Quarter 3.

Table 1
The u_i, m_i, and cash flow matrix $\mathbf{C}$, 2020 Quarter 3.

i	1	2	3	4	5	6	7	8	9
u_i	1.51	2.51	3.60	4.50	5.48	6.49	7.48	8.39	9.57
m_i	1,783.65	1,291.85	904.13	897.94	687.83	688.48	408.12	298.66	381.23
9×9 cash flow matrix $\mathbf{C}$									
c_{ij}	1	2	3	4	5	6	7	8	9
1	1,855.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	42.26	1,339.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	33.18	33.18	930.97	0.00	0.00	0.00	0.00	0.00	0.00
4	29.50	29.50	29.50	927.62	0.00	0.00	0.00	0.00	0.00
5	21.65	21.65	21.65	21.65	718.73	0.00	0.00	0.00	0.00
6	21.89	21.89	21.89	21.89	21.89	718.84	0.00	0.00	0.00
7	14.90	14.90	14.90	14.90	14.90	14.90	411.07	0.00	0.00
8	10.40	10.40	10.40	10.40	10.40	10.40	10.40	306.85	0.00
9	11.44	11.44	11.44	11.44	11.44	11.44	11.44	11.44	409.88

(CFETS),[15] over the time period from 2016 Quarter 1 to 2021 Quarter 1 (21 quarters in total).[16] Our data comes from the Wind database.[17] Before constructing the curves, we perform three data preprocessing steps: selecting bonds, generating a cash flow matrix, and compressing the cash flow matrix.

First, we select bonds. For each quarter, bonds with maturity less than 1 year or greater than 10 years are excluded,[18] and bonds with no trading volume are also excluded. In addition, we exclude bonds with a relative difference greater than 1% between the average closing price of the last five trading days (excluding the last day) and the closing price of the last trading day.

Second, we generate a raw cash flow matrix for the selected bonds. More precisely, for each bond, we record the future payments and the corresponding times of payments occur. These are recorded according to the basic information (face value, coupon rate, maturity, frequency of coupon payments, etc.) of the bonds. Fig. 6 shows a small fragment of the raw cash flow matrix for 2020 Quarter 3. The green cells in the first row represent payment times, and the blue cells in the first column represent the market prices of each selected bond. From the second to the last column, each row corresponds to the cash flow of a bond.

Third, we compress the raw matrix into a 9×9 matrix $\mathbf{C}$. That is to say, we generate $N=9$ "representative" bonds with $T=9$ different payment times. We first divide the payment times into $T=9$ intervals: $[1,2],(2,3],(3,4],\ldots,(9,10]$ (years). Next, for each interval $(i,i+1]$, we calculate the cash-flow-weighted average payment time u_i, which is the sum of the payment times weighted by their respective cash flows divided by the total cash flow in the interval. For each bond, all cash flows that occur during $(i,i+1]$ are regarded as occurring at u_i. Then, we classify the selected bonds into $N=9$ groups according to their maturities: maturity $\in[1,2],(2,3],(3,4],\ldots,(9,10]$. Finally, we merge the cash flows in each group together to construct $N=9$ representative bonds with payments occurring at $u_1,u_2,\ldots,u_9$, respectively. Now, both the cash flow matrix $\mathbf{C}$ and the prices of the 9 representative bonds $m_1,m_2,\ldots,m_9$ are generated. Table 1 shows an example of u_i, m_i, and the cash flow matrix $\mathbf{C}$ generated using the data of 2020 Quarter 3.

5.2. Empirical results

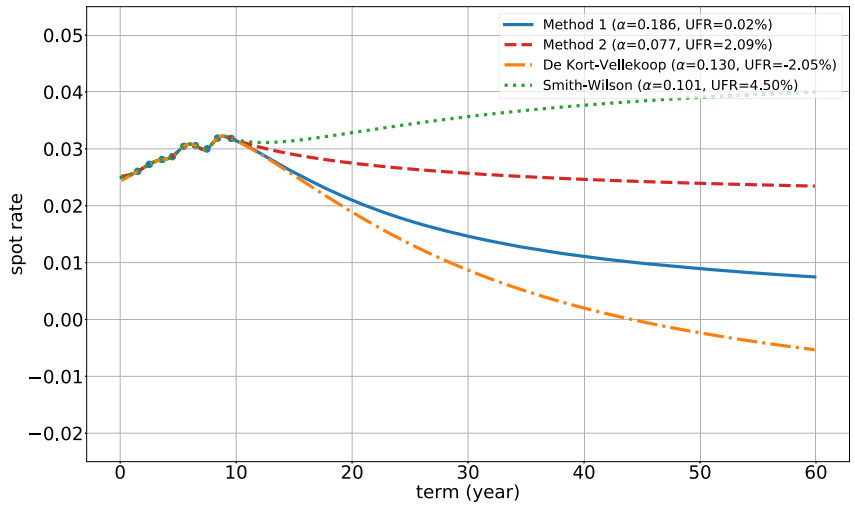
In this section, we present empirical results and demonstrate the capabilities of our methods for modelling the Chinese risk-free interest rate. Section 5.2.1 compares the performance of curves generated by different methods for several specific quarters. This comparison highlights the advantages of our methods over existing techniques. Section 5.2.2 compares the UFRs obtained using different methods from a

[15] See http://www.chinamoney.com.cn/english/mdtdp/.

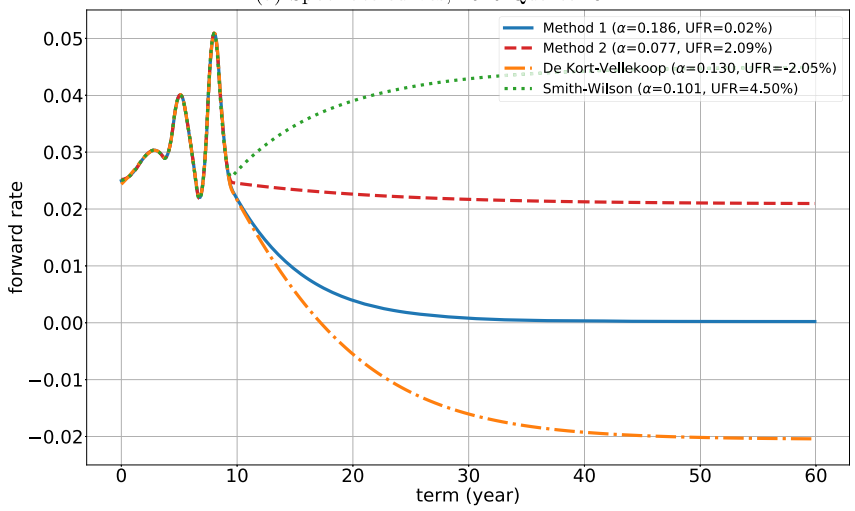
[16] We select this period because the China Risk-Oriented Solvency System was officially implemented in 2016, see, for example, Fung et al. (2018).

[17] See https://www.wind.com.cn/en/Default.html.

[18] This is because the EIOPA recommends excluding financial instruments with maturities less than 1 year when constructing risk-free rate term structures (EIOPA, 2019, Paragraph 15). In addition, the last liquid point is set to be 10 years for China (EIOPA, 2019, Table 8).



(a) Spot rate curves, 2020 Quarter 3



(b) Forward rate curves, 2020 Quarter 3

Fig. 7. Risk-free interest rate curves generated by different methods, 2020 Quarter 3.

time series perspective, and reveals several useful facts about the behaviour of the Chinese risk-free interest rate over time.

5.2.1. Comparison of different methods

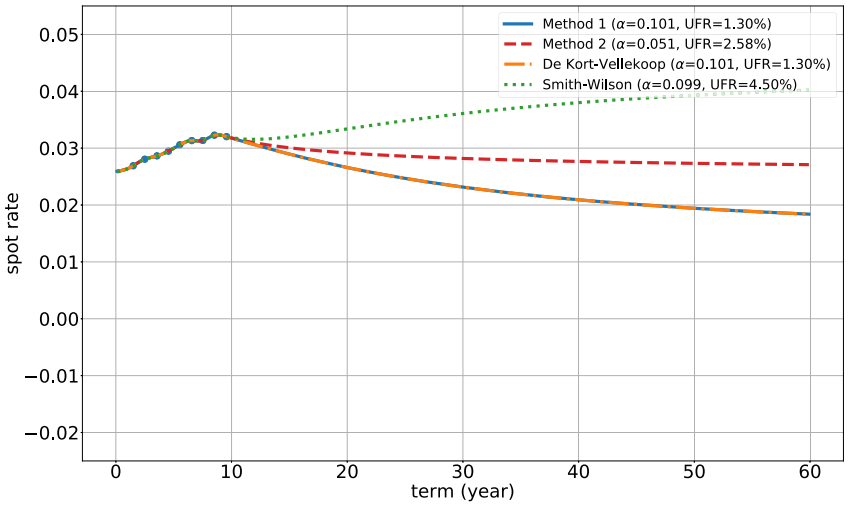
We compare different methods by studying two typical quarters: 2020 Quarter 3 and 2020 Quarter 4. These quarters were chosen because they represent periods of significant volatility in the Chinese risk-free interest rate, making them ideal for evaluating the performance of our methods and comparing them to existing techniques.

Fig. 7 and Fig. 8 present the spot and forward curves for 2020 Quarter 3 and 2020 Quarter 4, respectively. Each figure shows the results of four methods: Method 1 (the blue solid lines), Method 2 (the red dashed lines, with $f_{\text{prior}} = 4.50\%$ and $\lambda = 1,000$), the de Kort-Vellekoop method (the orange dash-dotted lines), and the original Smith-Wilson method adopted by the EIOPA (the green dotted lines, with an exogenous UFR of 4.50%). The convergence point for all methods is set to be 60 years (EIOPA, 2019, Paragraph 120). From the figures, the comparison of the four methods in 2020 Quarter 3 shows greater differences than in 2020 Quarter 4.

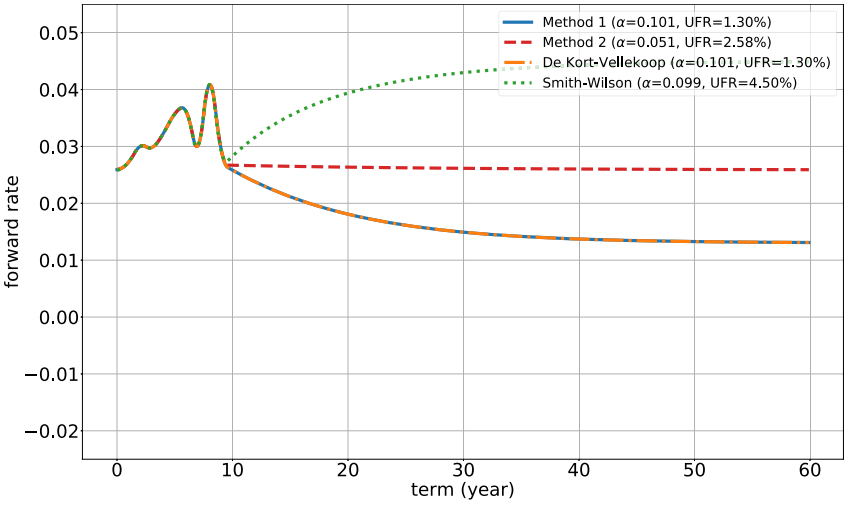
We first analyse the behaviour of the original Smith-Wilson method (the green dotted lines). The UFR is set to be 4.50% exogenously, and α is chosen using (9) (EIOPA, 2019, Paragraphs 121 and 158). One can observe that, for the original Smith-Wilson method, both the spot curves and the forward curves in Fig. 7 and Fig. 8 converge to the exogenous UFR (4.50%) as the term increases without bound, regardless of how the market data of bonds behaves. In addition, it seems that the (green dotted) forward curve is not smooth enough in the neighbourhood of the last liquid point (10 years).

Next, we analyse the behaviour of the de Kort-Vellekoop method (the orange dash-dotted lines). As de Kort and Vellekoop (2016) did not consider the choice of α, we apply the same procedure as the original Smith-Wilson method to determine α using (9). In Fig. 7, both the spot curves and the forward curves become negative as the term increases without bound (the endogenous UFR is -2.05%, and $\alpha = 0.130$), while in Fig. 8, they do not (the endogenous UFR is 1.30%, and $\alpha = 0.101$). That is to say, although the de Kort-Vellekoop method can obtain endogenous UFRs, the endogenous UFRs may not always be positive. In fact, for 2020 Quarter 3, $\alpha = 0.130 \notin \mathcal{A}$; while for 2020 Quarter 4, $\alpha = 0.101$ lies in the set $\mathcal{A}$ of permissible values of α as defined in (20).

Then, we turn to study the results of Method 1 (the blue solid lines). Both the spot and forward curves generated by Method 1 in Fig. 7 and Fig. 8 converge to endogenous and positive UFRs (0.02% for 2020 Quarter 3 and 1.30% for 2020 Quarter 4). This confirms that Method 1 successfully realizes the goal of generating a UFR that is both endogenous and positive. It is worth noting that, in Fig. 8, the results of Method 1 (the blue solid lines) are the same as the results of the de Kort-Vellekoop method (the orange dash-dotted lines). This is because Method 1 is built upon the de Kort-Vellekoop method, and for 2020



(a) Spot rate curves, 2020 Quarter 4



(b) Forward rate curves, 2020 Quarter 4

Fig. 8. Risk-free interest rate curves generated by different methods, 2020 Quarter 4.

Quarter 4, the value of α used by both methods, 0.101, falls within the feasible region $\mathcal{A}$. As a result, the de Kort-Vellekoop method has already found an endogenous and positive UFR, and the results of both methods should therefore be the same.

Finally, we analyse the results of Method 2 (the red dashed lines). The UFR generated by this method can be seen as a mixture of the completely endogenous UFR and the prior knowledge ($f_{\text{prior}} = 4.50\%$) about the UFR. Therefore, the curves generated by Method 2 always stay in the middle of other curves. As explained in Section 4, a completely endogenous UFR may not always be desirable, and Method 2 is proposed to generate endogenous and positive UFRs while incorporating prior knowledge. In addition, the spot curves of Method 2 in both figures appear to be quite smooth.

Therefore, the results of both 2020 Quarter 3 and Quarter 4 show the strengths of our two proposed methods. Method 1 addresses the challenge of obtaining endogenous and positive UFRs, and Method 2 can further incorporate practitioners' prior knowledge.

5.2.2. Time series of the UFRs

Now we compare different methods from the perspective of the time series of the UFRs. We study the data from 2016 Quarter 1 to 2021 Quarter 1, and the time series of the UFRs obtained by four methods are shown in Fig. 9. For the convenience of comparison, the yield rates of 10-year Chinese government bonds are also shown in the figure.

According to Fig. 9, we have the following viewpoints.

First, Method 1 can guarantee the positivity of the endogenous UFR. We can observe that the endogenous UFRs obtained by the de Kort-Vellekoop method are negative in 2016 Quarter 4, 2017 Quarter 1, and 2020 Quarters 1–3, whereas the endogenous UFRs generated by Method 1 (the blue solid line) are always non-negative. This implies that, during normal time periods, the de Kort-Vellekoop method can yield positive UFRs, and Method 1 yields the same results as the de Kort-Vellekoop method (see Remark 4); during idiosyncratic time periods such as 2020, Method 1 can further ensure the positivity of the UFR. In particular, the UFR generated by Method 1 is close to the positive part of the UFR generated by the de Kort-Vellekoop method.[19]

Second, the UFRs generated by Method 2 (the grey dashed lines) exhibit more stability over time than other methods (with the exception of the original Smith-Wilson method, which uses a fixed exogenous UFR, shown as the green dotted line). These UFRs are situated between completely endogenous UFRs and completely exogenous UFRs, and can be viewed as a mixture of the endogenous UFRs and prior knowledge.

[19] The UFR generated using Method 1 does not always precisely equal the positive part of the result obtained from the de Kort-Vellekoop method. For example, Fig. 7 illustrates that while the de Kort-Vellekoop method generates a negative UFR in 2020 Quarter 3, Method 1 produces a UFR of 0.02%, not 0.00%.

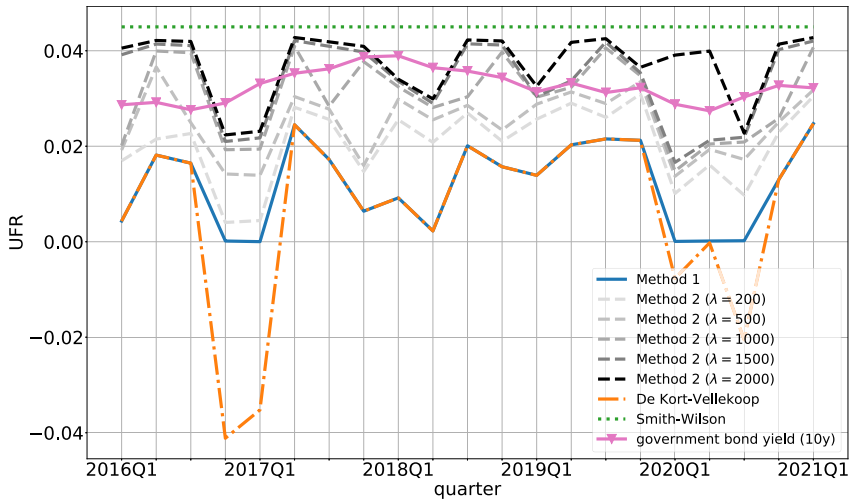


Fig. 9. The time series of UFRs obtained by different methods.

Third, the values of UFRs generated by Method 2 approach $f_{\text{prior}} = 4.50\%$ as λ increases. This is consistent with the fact that λ represents practitioners' confidence in the prior knowledge f_{prior}. The figure also shows the sensitivity of the UFR generated by Method 2 with respect to λ. For example, as λ varies from 200 to 2,000, the UFR change (in absolute value) spans from 0.56% (2019 Quarter 4) to 2.90% (2020 Quarter 1) in our dataset, with an average change of 1.67%. This implies that, on average, for our dataset, a unit change in λ results in a UFR change of $1.67\%/(2,000-200) = 0.000926\%$.

Finally, the results provide insights into the dynamics of the UFRs of Chinese risk-free interest rates. The figure depicts low UFRs in 2016 and 2020, which is consistent with the environment of the Chinese risk-free interest rate market during these years (as indicated by the 10-year yield rates of the Chinese government bonds, shown as the pink line). Therefore, our proposed methods are able to incorporate some information about the current risk-free interest rates into the choice of UFRs. In contrast, the original Smith-Wilson method adopted by the EIOPA relies on a (nearly) constant exogenous UFR, which cannot reflect the information of the market timely.

To summarize, these results demonstrate that our methods can both construct a smooth risk-free interest rate curve with endogenous and positive UFRs and incorporate insights about the market.

6. Conclusions

In this paper, we propose several methods for generating UFRs that are both endogenous and positive, and study their theoretical properties. Our methods are built upon the framework proposed by de Kort and Vellekoop (2016), which can generate endogenous UFRs. Under the assumption that the cash flow matrix is square and invertible, we show the existence of a solution to their first-order condition, and also demonstrate the potential issue of negative UFRs obtained by their framework during extreme time periods such as the COVID-19 pandemic. To address this issue, we introduce Method 1, which generates endogenous and positive UFRs, and give conditions for its feasibility. Moreover, we propose Method 2, a new framework for generating endogenous and positive UFRs with prior knowledge, and demonstrate that it is always feasible. Our methods are also consistent with the stress scenarios specified in Solvency II.

To further illustrate the capabilities of our proposed methods, we perform an empirical analysis using Chinese government bonds traded on the China Foreign Exchange Trade System, covering the time period from 2016 Quarter 1 to 2021 Quarter 1. The results show that our methods can effectively construct smooth risk-free interest rate curves with endogenous and positive UFRs. The endogenous UFRs generated by our methods are more stable than those obtained by other methods, and can also provide useful insights into the dynamics of the Chinese risk-free interest rate.

There are several topics related to the Smith-Wilson method that could be further explored. First, the matrix $\mathbf{W}_\alpha$ (and its inverse) plays a crucial role in the methods proposed in this paper. While we prove the existence of a solution to de Kort and Vellekoop's (2016) first-order condition under the regularity condition (18), it would be valuable to investigate when this condition is satisfied. Second, under the Smith-Wilson framework, the optimal interest rate curves are generated by the exponential tension spline function (8). By further studying the properties of this function, we can gain a deeper understanding of the relationship between the endogenous UFR obtained by (13) and the observed market data. Third, in Appendix B, we describe the challenge faced by our framework when the cash flow matrix is not square and invertible, which requires further investigation. Fourth, the results generated by Method 2 depend on the choice of the tuning parameter λ, and how to calibrate this parameter in practice can be further discussed.

Overall, our research presents new methods for generating endogenous and positive UFRs. It provides a new framework for constructing risk-free interest rate curves, which can motivate further research on the UFR, the Smith-Wilson method, and curve construction methods in general.

Declaration of competing interest

There is no competing interest.

Data availability

Data will be made available on request.

Appendix A. Lemmas and proofs

This appendix gives lemmas and proofs of the theoretical results in the main article.

A.1. Lemmas

We first provide the proofs of Lemmas 1 and 2 in the main article. We then propose two additional lemmas, Lemmas A.1 and A.2, and give their proofs.

Proof of Lemma 1. de Kort and Vellekoop (2016, Section 3) showed that $\mathbf{W}_\alpha$ is the covariance matrix of a non-degenerate Gaussian vector.

Therefore, both $\mathbf{W}_\alpha$ and its inverse $\mathbf{W}_\alpha^{-1}$ are positive definite. Hence, $[\mathbf{W}_\alpha^{-1}]_{ii}$ is positive for all $i = 1, 2, \ldots, T$. Since $[\mathbf{W}_\alpha]_{ii} = W_\alpha(u_i, u_i) = \alpha u_i - \frac{1}{2} + \frac{1}{2}e^{-2\alpha u_i} < \alpha u_i$, we only need to prove $[\mathbf{W}_\alpha^{-1}]_{ii} \cdot [\mathbf{W}_\alpha]_{ii} > 1$.

Let the eigenvalue decomposition of $\mathbf{W}_\alpha$ be $\mathbf{W}_\alpha = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$, where $\mathbf{Q}$ is orthogonal and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_T\}$ with $\lambda_i > 0, i = 1, 2, \ldots, T$. Therefore, $\mathbf{W}_\alpha^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^\top$. The i-th diagonal entries of $\mathbf{W}_\alpha$ and $\mathbf{W}_\alpha^{-1}$ are $[\mathbf{W}_\alpha]_{ii} = \sum_{j=1}^T \lambda_j \mathbf{Q}_{ji}^2$ and $[\mathbf{W}_\alpha^{-1}]_{ii} = \sum_{j=1}^T \frac{1}{\lambda_j}\mathbf{Q}_{ji}^2$, respectively. The Cauchy-Schwarz inequality implies that

$$[\mathbf{W}_\alpha^{-1}]_{ii} \cdot [\mathbf{W}_\alpha]_{ii} = \left[\sum_{j=1}^T \left(\sqrt{\lambda_j \mathbf{Q}_{ji}^2}\right)^2\right]\left[\sum_{j=1}^T \left(\sqrt{\frac{1}{\lambda_j}\mathbf{Q}_{ji}^2}\right)^2\right]$$

$$\geq \left[\sum_{j=1}^T \left(\sqrt{\lambda_j \mathbf{Q}_{ji}^2}\right)\left(\sqrt{\frac{1}{\lambda_j}\mathbf{Q}_{ji}^2}\right)\right]^2 = \left[\sum_{j=1}^T \mathbf{Q}_{ji}^2\right]^2 = 1.$$

Note that the equality holds if and only if $\sqrt{\lambda_j \mathbf{Q}_{ji}^2} = c\sqrt{\frac{1}{\lambda_j}\mathbf{Q}_{ji}^2}$ for some constant c, which implies that

$$[\mathbf{W}_\alpha]_{ij} = \sum_{k=1}^T \lambda_k \mathbf{Q}_{ki}\mathbf{Q}_{kj} = \sum_{k=1}^T c\mathbf{Q}_{ki}\mathbf{Q}_{kj} = c\sum_{k=1}^T \mathbf{Q}_{ki}\mathbf{Q}_{kj} = 0$$

for all $i \neq j$. But this is contradicting with the definition of $\mathbf{W}_\alpha$. Therefore, $[\mathbf{W}_\alpha^{-1}]_{ii} > 1/[\mathbf{W}_\alpha]_{ii} > 1/(\alpha u_i)$. □

Proof of Lemma 2. Proof of (a): Note that

$$\frac{h_\alpha(f_\infty)}{e^{f_\infty u_1}} = \sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\pi_j[\mathbf{W}_\alpha^{-1}]_{ij}\right) e^{f_\infty(u_i+u_j-u_1)}$$

$$- \sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i[\mathbf{W}_\alpha^{-1}]_{ij}\right) e^{f_\infty(u_i-u_1)}.$$

Since $u_i + u_j - u_1 > 0$ for $i, j = 1, 2, \ldots, T$ and $u_i - u_1 > 0$ for $i = 2, 3, \ldots, T$, all terms in the equation above will converge to 0 as $f_\infty \to -\infty$. Thus, the result holds.

Proof of (b): Note that

$$\frac{h_\alpha(f_\infty)}{e^{2f_\infty u_T}} = \sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\pi_j[\mathbf{W}_\alpha^{-1}]_{ij}\right) e^{f_\infty(u_i+u_j-2u_T)}$$

$$- \sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i[\mathbf{W}_\alpha^{-1}]_{ij}\right) e^{f_\infty(u_i-2u_T)}.$$

Since $u_i + u_j - 2u_T < 0$ if i and j are not both T, and $u_i - 2u_T < 0$ for $i = 1, 2, \ldots, T$, all terms in the equation above will converge to 0 as $f_\infty \to +\infty$. Thus, the result holds. □

Lemma A.1. *The matrix $\mathbf{W}_\alpha$ satisfies*

$$\lim_{\alpha\to+\infty} \frac{\mathbf{W}_\alpha}{\alpha} = \begin{pmatrix} u_1 & u_1 & \cdots & u_1 & u_1 \\ u_1 & u_2 & \cdots & u_2 & u_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_1 & u_2 & \cdots & u_{T-1} & u_{T-1} \\ u_1 & u_2 & \cdots & u_{T-1} & u_T \end{pmatrix} =: \tilde{\mathbf{W}}_\infty, \tag{A.1}$$

and the inverse of $\tilde{\mathbf{W}}_\infty$ is

$$\tilde{\mathbf{W}}_\infty^{-1} = \begin{pmatrix} \frac{1}{\Delta u_1} + \frac{1}{\Delta u_2} & -\frac{1}{\Delta u_2} & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{\Delta u_2} & \frac{1}{\Delta u_2} + \frac{1}{\Delta u_3} & -\frac{1}{\Delta u_3} & \cdots & 0 & 0 & 0 \\ 0 & -\frac{1}{\Delta u_3} & \frac{1}{\Delta u_3} + \frac{1}{\Delta u_4} & \cdots & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\Delta u_4} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{\Delta u_{T-1}} & \frac{1}{\Delta u_{T-1}} + \frac{1}{\Delta u_T} & -\frac{1}{\Delta u_T} \\ 0 & 0 & 0 & \cdots & 0 & -\frac{1}{\Delta u_T} & \frac{1}{\Delta u_T} \end{pmatrix}, \tag{A.2}$$

where $\Delta u_i = u_i - u_{i-1}$ for $i = 1, 2, \ldots, T$.

Proof of Lemma A.1. By definition, we have $W_\alpha(t, u) = \alpha\min(t, u) - \frac{1}{2}e^{-\alpha|t-u|} + \frac{1}{2}e^{-\alpha(t+u)}$, and therefore

$$\lim_{\alpha\to+\infty} \frac{W_\alpha(t, u)}{\alpha} = \min(t, u).$$

Hence, (A.1) holds. Then, one can directly check that the inverse of $\tilde{\mathbf{W}}_\infty$ is given by (A.2). □

Lemma A.2. *As $\alpha \to +\infty$, the solution to the first-order condition (13), f_∞^α, converges to the solution to the following equation with respect to f_∞:*

$$\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i e^{f_\infty u_i}\right)\left[\tilde{\mathbf{W}}_\infty^{-1}\right]_{ij}\left(\pi_j e^{f_\infty u_j} - 1\right) = 0. \tag{A.3}$$

Proof of Lemma A.2. This lemma holds because multiplying both sides of the first-order condition (13) by α does not change the solution to the equation. Equation (A.3) exists at least one solution because one can directly check that the matrix $\tilde{\mathbf{W}}_\infty^{-1}$ as defined in (A.2) always satisfies the regularity condition (18) proposed in Theorem 1. □

A.2. Proofs

Proof of Proposition 1. The first-order condition (13) can be rewritten as

$$(\mathbf{U}X)^\top \mathbf{W}_\alpha^{-1}(X - \mathbf{1}) = 0. \tag{A.4}$$

Then, direct calculation shows that (A.4) is equivalent to (16). □

Proof of Corollary 1. This is a direct corollary of Proposition 1. □

Proof of Theorem 2. We first prove that $\mathcal{A}$ as defined in (20) is not empty. In fact, if we can show $\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\right)\left[\mathbf{W}_\alpha^{-1}\right]_{ij}\left(\pi_j - 1\right) < 0$ when $\alpha \to +\infty$, the result will be true due to the continuity of $\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\right)\left[\mathbf{W}_\alpha^{-1}\right]_{ij}\left(\pi_j - 1\right)$ with respect to α. Therefore, it suffices to prove

$$\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\right)\left[\tilde{\mathbf{W}}_\infty^{-1}\right]_{ij}\left(\pi_j - 1\right) < 0.$$

According to (A.2) in Lemma A.1, we have

$$\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\right)\left[\tilde{\mathbf{W}}_\infty^{-1}\right]_{ij}\left(\pi_j - 1\right) = \sum_{i=1}^{T-1}(u_i\pi_i)\left(\frac{\pi_i - \pi_{i-1}}{u_i - u_{i-1}} - \frac{\pi_{i+1} - \pi_i}{u_{i+1} - u_i}\right)$$

$$+ (u_T\pi_T)\frac{\pi_T - \pi_{T-1}}{u_T - u_{T-1}}$$

$$= \sum_{i=1}^T \left(u_i\pi_i - u_{i-1}\pi_{i-1}\right)\frac{\pi_i - \pi_{i-1}}{u_i - u_{i-1}}.$$

Hence, $\sum_{i=1}^T\sum_{j=1}^T \left(u_i\pi_i\right)\left[\tilde{\mathbf{W}}_\infty^{-1}\right]_{ij}\left(\pi_j - 1\right) < 0$ when (22) holds, which proves that $\mathcal{A}$ is not empty.

Now we prove that, for sufficiently large values of α, we also have $\alpha \in \mathcal{B}$. The gap between the forward curve and the UFR at the convergence point, $|f^\alpha(\mathrm{CP}) - f^\alpha_\infty|$, takes the form of (EIOPA, 2019, Paragraph 158)

$$|f^\alpha(\mathrm{CP}) - f^\alpha_\infty| = \frac{\alpha}{|1-\kappa_\alpha e^{\alpha \mathrm{CP}}|},$$

where

$$\kappa_\alpha = \frac{1 + \sum_{i=1}^{T}\sum_{j=1}^{T} \alpha u_i \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f^\alpha_\infty} - 1\right)}{\sum_{i=1}^{T}\sum_{j=1}^{T} \sinh(\alpha u_i) \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f^\alpha_\infty} - 1\right)}.$$

Hence, we only need to prove

$$\lim_{\alpha\to+\infty} |f^\alpha(\mathrm{CP}) - f^\alpha_\infty| = \lim_{\alpha\to+\infty} \frac{\alpha}{|1-\kappa_\alpha e^{\alpha \mathrm{CP}}|} = 0.$$

Since $\frac{\alpha}{|1-\kappa_\alpha e^{\alpha \mathrm{CP}}|} = \frac{1}{|1/\alpha - \kappa_\alpha e^{\alpha \mathrm{CP}}/\alpha|}$, we consider

$$\kappa_\alpha e^{\alpha \mathrm{CP}}/\alpha = \frac{1 + \sum_{i=1}^{T}\sum_{j=1}^{T} \alpha u_i \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f^\alpha_\infty} - 1\right)}{\sum_{i=1}^{T}\sum_{j=1}^{T} \frac{\sinh(\alpha u_i)}{\exp(\alpha \mathrm{CP})} \alpha \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f^\alpha_\infty} - 1\right)}. \tag{A.5}$$

For the denominator of (A.5), Lemma A.1 implies that $\lim_{\alpha\to+\infty} \alpha \left[\mathbf{W}_\alpha^{-1}\right]_{ij}$ exists, and Lemma A.2 establishes the convergence of f^α_∞. In addition, $\lim_{\alpha\to+\infty} \frac{\sinh(\alpha u_i)}{\exp(\alpha \mathrm{CP})} = 0$ since $\mathrm{CP} > u_i$ for all $i = 1, 2, \ldots, T$. Therefore, the denominator of (A.5) converges to 0 as $\alpha \to +\infty$.

Now we consider the numerator of (A.5). By using the result of Lemma A.1, we have

$$\begin{aligned}
&\lim_{\alpha\to+\infty} \left[1 + \sum_{i=1}^{T}\sum_{j=1}^{T} \alpha u_i \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f^\alpha_\infty} - 1\right)\right] \\
&= 1 + \sum_{i=1}^{T}\sum_{j=1}^{T} u_i \left[\tilde{\mathbf{W}}_\infty^{-1}\right]_{ij} \left(\pi_j e^{f^\infty_\infty} - 1\right) \\
&= 1 + \pi_T e^{f^\infty_\infty} - 1 \\
&= \pi_T e^{f^\infty_\infty} > 0,
\end{aligned}$$

where f^∞_∞ is the root of (A.3). Hence, we have

$$\lim_{\alpha\to+\infty} \kappa_\alpha e^{\alpha \mathrm{CP}}/\alpha = +\infty,$$

which implies that

$$\begin{aligned}
\lim_{\alpha\to+\infty} |f^\alpha(\mathrm{CP}) - f^\alpha_\infty| &= \lim_{\alpha\to+\infty} \frac{\alpha}{|1-\kappa_\alpha e^{\alpha \mathrm{CP}}|} \\
&= \lim_{\alpha\to+\infty} \frac{1}{|1/\alpha - \kappa_\alpha e^{\alpha \mathrm{CP}}/\alpha|} = 0.
\end{aligned}$$

Therefore, for sufficiently large values of α, we have $\alpha \in \mathcal{B}$. This completes the proof. □

Proof of Proposition 2. One can check that the first-order condition (13) is equivalent to

$$\sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i e^{-cu_i} e^{(f_\infty + c)u_i}\right) [\mathbf{W}_\alpha^{-1}]_{ij} \left(\pi_j e^{-cu_j} e^{(f_\infty + c)u_j} - 1\right) = 0,$$

and hence the result holds. □

Proof of Theorem 3. This result can be regarded as a corollary of the proof given in the appendix of de Kort and Vellekoop (2016, Page 118). Denote by

$$\mathcal{L}_\alpha(f_\infty) = \min_{g\in\mathcal{H}(f_\infty)} \int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right] \mathrm{d}s.$$

de Kort and Vellekoop (2016, Page 118) proved that, when the cash flow matrix $\mathbf{C}$ is invertible, we have

$$\frac{\partial \mathcal{L}_\alpha(f_\infty)}{\partial f_\infty} = \alpha^3 \sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i e^{f_\infty u_i}\right) \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f_\infty u_j} - 1\right).$$

As for our optimization problem (24), let

$$\begin{aligned}
\tilde{\mathcal{L}}_\alpha(f_\infty) &= \min_{g\in\mathcal{H}(f_\infty)} \left[\int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right] \mathrm{d}s + \frac{\lambda}{2}\alpha^3 (f_\infty - f_{\mathrm{prior}})^2\right] \\
&= \min_{g\in\mathcal{H}(f_\infty)} \left[\int_0^\infty \left[g''(s)^2 + \alpha^2 g'(s)^2\right] \mathrm{d}s\right] + \frac{\lambda}{2}\alpha^3 (f_\infty - f_{\mathrm{prior}})^2,
\end{aligned}$$

then we immediately have

$$\tilde{\mathcal{L}}_\alpha(f_\infty) = \mathcal{L}_\alpha(f_\infty) + \frac{\lambda}{2}\alpha^3 (f_\infty - f_{\mathrm{prior}})^2.$$

Therefore, the first-order condition for problem (24) with respect to f_∞ is

$$\begin{aligned}
&\frac{\partial[\tilde{\mathcal{L}}_\alpha(f_\infty)]}{\partial f_\infty} = \frac{\partial[\mathcal{L}_\alpha(f_\infty) + \lambda\alpha^3 (f_\infty - f_{\mathrm{prior}})^2/2]}{\partial f_\infty} \\
&= \alpha^3 \sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i e^{f_\infty u_i}\right) \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j e^{f_\infty u_j} - 1\right) + \lambda\alpha^3 (f_\infty - f_{\mathrm{prior}}) = 0,
\end{aligned}$$

which proves the result. □

Proof of Theorem 4. The left-hand side of the first-order equation (25) can be rewritten as

$$\tilde{h}_{\alpha,\lambda}(f_\infty) = h_\alpha(f_\infty) + \lambda(f_\infty - f_{\mathrm{prior}}),$$

where $h_\alpha(f_\infty)$ is given by (19). Remark 2 of Lemma 2 states that $\lim_{f_\infty\to-\infty} h_\alpha(f_\infty) = 0$ and $\lim_{f_\infty\to+\infty} h_\alpha(f_\infty) = +\infty$. Therefore, we have

$$\lim_{f_\infty\to-\infty} \tilde{h}_{\alpha,\lambda}(f_\infty) = -\infty, \quad \lim_{f_\infty\to+\infty} \tilde{h}_{\alpha,\lambda}(f_\infty) = +\infty.$$

Then the result holds because of the continuity of $\tilde{h}_{\alpha,\lambda}(f_\infty)$ with respect to f_∞. □

Proof of Theorem 5. Lemma A.1 implies that, for any $i, j = 1, 2, \ldots, T$, we have

$$\lim_{\alpha\to+\infty} \left[\mathbf{W}_\alpha^{-1}\right]_{ij} = 0. \tag{A.6}$$

Therefore, $\tilde{\mathcal{A}}$ is not empty because

$$\lim_{\alpha\to+\infty} \sum_{i=1}^{T}\sum_{j=1}^{T} \left(u_i \pi_i\right) \left[\mathbf{W}_\alpha^{-1}\right]_{ij} \left(\pi_j - 1\right) - \lambda f_{\mathrm{prior}} = -\lambda f_{\mathrm{prior}} < 0.$$

Now we prove $\alpha \in \tilde{\mathcal{B}}$ for sufficiently large values of α. By using (A.6), one can easily find that the solution to the new first-order condition (25) converges to f_{prior} as $\alpha \to +\infty$. We can then use a similar approach to the proof of Theorem 2 to conclude that $\alpha \in \tilde{\mathcal{B}}$ for sufficiently large values of α. This completes the proof. □

Appendix B. Rectangular and non-invertible cash flow matrices

When $\mathbf{C}$ is not square and invertible, the theories presented in our main article do not apply. The fundamental issue lies in the fact that our methods and theories rely on the simplified first-order condition (13). When $\mathbf{C}$ is not square or not invertible, we must resort to the more

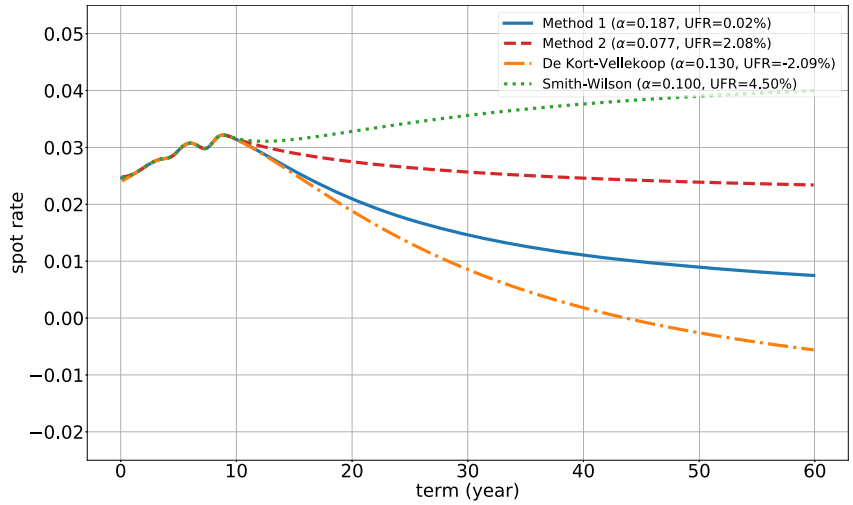
Table B.1
The u_i, m_i, and the rectangular cash flow matrix $\mathbf{C}$, 2020 Quarter 3.

j	1	2	3	4	5	6	7	8	9	10
u_j	0.50	1.51	2.51	3.60	4.50	5.48	6.49	7.48	8.39	9.57

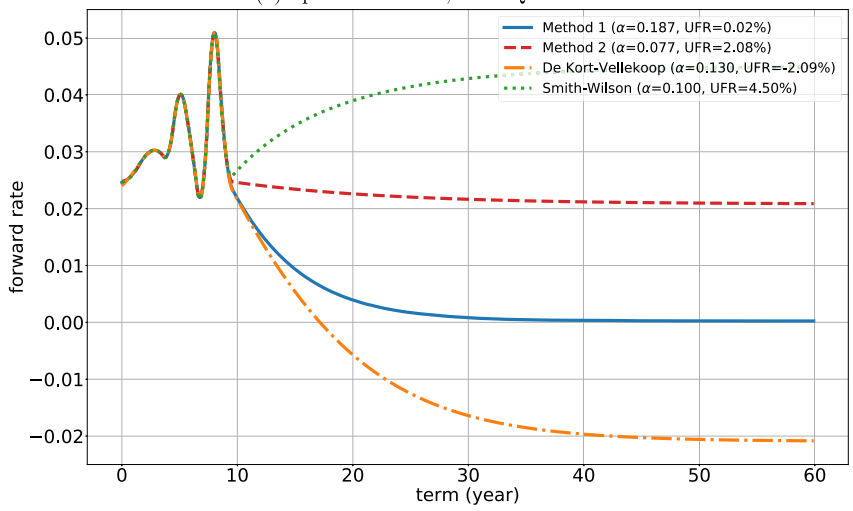
i	1	2	3	4	5	6	7	8	9
m_i	1,840.73	1,334.11	937.31	927.44	709.48	710.37	423.02	309.06	392.67

9×10 cash flow matrix $\mathbf{C}$

c_{ij}	1	2	3	4	5	6	7	8	9	10
1	57.08	1,855.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	42.26	42.26	1,339.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	33.18	33.18	33.18	930.97	0.00	0.00	0.00	0.00	0.00	0.00
4	29.50	29.50	29.50	29.50	927.62	0.00	0.00	0.00	0.00	0.00
5	21.65	21.65	21.65	21.65	21.65	718.73	0.00	0.00	0.00	0.00
6	21.89	21.89	21.89	21.89	21.89	21.89	718.84	0.00	0.00	0.00
7	14.90	14.90	14.90	14.90	14.90	14.90	14.90	411.07	0.00	0.00
8	10.40	10.40	10.40	10.40	10.40	10.40	10.40	10.40	306.85	0.00
9	11.44	11.44	11.44	11.44	11.44	11.44	11.44	11.44	11.44	409.88



(a) Spot rate curves, 2020 Quarter 3

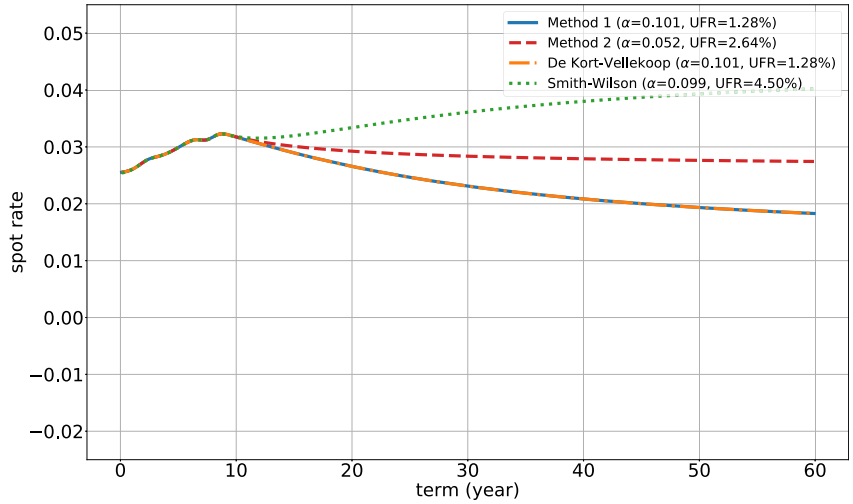


(b) Forward rate curves, 2020 Quarter 3

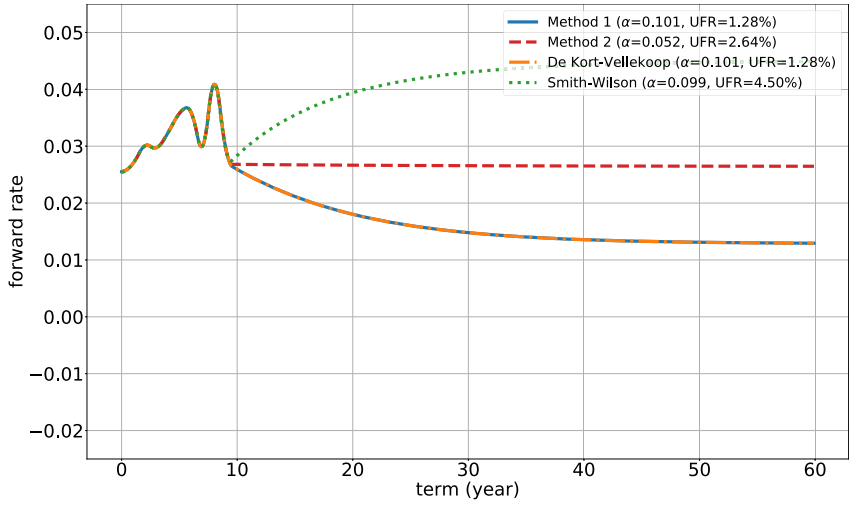
Fig. B.1. Risk-free interest rate curves generated by different methods under a rectangular cash flow matrix, 2020 Quarter 3.

general version of the first-order condition (12). Therefore, to enable our theoretical results to hold for $\mathbf{C}$ that is not square or not invertible, a more detailed analysis of (12) is required. This is beyond the scope of this paper, and a detailed analysis is left for future work.

Despite the absence of a theoretical guarantee, we endeavor to extend both Method 1 and Method 2 using (12) as follows. The ideas behind these generalized methods align with those of Method 1 and Method 2 as presented in our main article.



(a) Spot rate curves, 2020 Quarter 4



(b) Forward rate curves, 2020 Quarter 4

Fig. B.2. Risk-free interest rate curves generated by different methods under a rectangular cash flow matrix, 2020 Quarter 4.

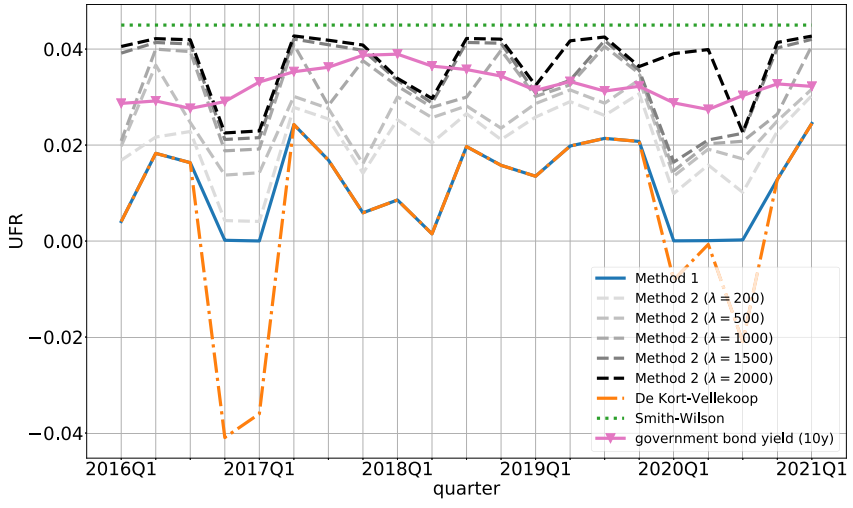
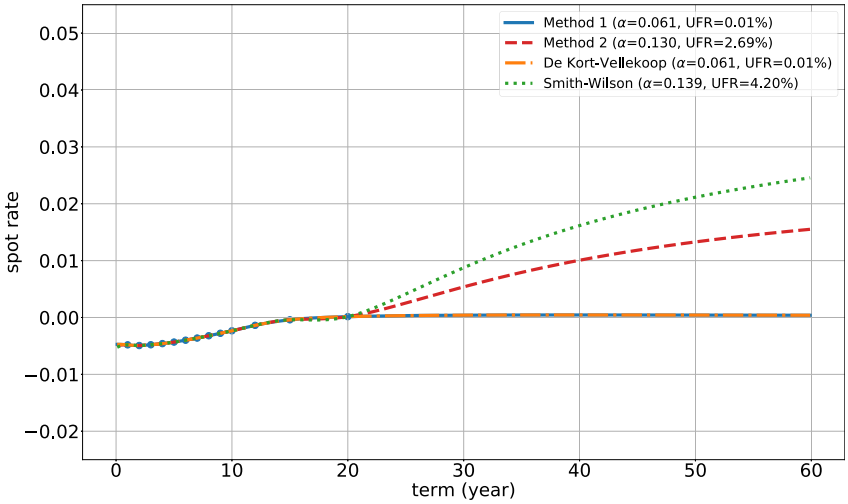
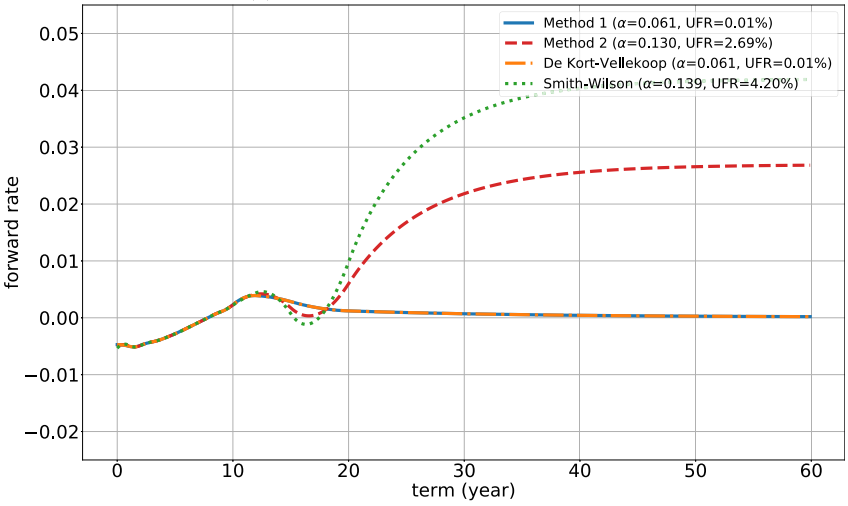


Fig. B.3. The time series of UFRs obtained by different methods with rectangular cash flow matrices.



(a) Spot rate curves, September 2020



(b) Forward rate curves, September 2020

Fig. C.4. Risk-free interest rate curves generated by different methods based the EURIBOR swap data, September 2020.

Method B.1. ***(Endogenous and positive UFRs: Generalized)***

Step 1: Derive the feasible region of α:

$$\mathcal{A}_g := \Big\{ \alpha : \alpha > 0, (m - \mathbf{C1})^\top (\mathbf{C}\mathbf{W}_\alpha \mathbf{C}^\top)^{-1} \mathbf{C}\mathbf{U} \times \big(1 + \mathbf{W}_\alpha \mathbf{C}^\top (\mathbf{C}\mathbf{W}_\alpha \mathbf{C}^\top)^{-1} (m - \mathbf{C1})\big) < 0 \Big\},$$

$$\mathcal{B}_g := \{ \alpha : \alpha \in \mathcal{A}_g,\ \alpha \geq \alpha_{\min},\ |f^\alpha(\mathrm{CP}) - f^\alpha_\infty| \leq \tau \},$$

where f^α_∞ is the solution to the first-order condition (12) given α, and $f^\alpha(\cdot)$ is the forward curve generated by the Smith-Wilson method (3) given α and $f_\infty = f^\alpha_\infty$.

Step 2: Choose an optimal α:

$$\alpha^*_g = \inf \mathcal{B}_g.$$

Step 3: Solve the following equation for f_∞:

$$(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_{\alpha^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{U} \times \Big(1 + \mathbf{W}_{\alpha^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_{\alpha^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})\Big) = 0.$$

Method B.2. ***(Endogenous and positive UFRs with prior knowledge: Generalized)***

Step 1: Derive the feasible region of α:

$$\tilde{\mathcal{A}}_g := \Big\{ \alpha : \alpha > 0, (m - \mathbf{C1})^\top (\mathbf{C}\mathbf{W}_\alpha \mathbf{C}^\top)^{-1} \mathbf{C}\mathbf{U} \times \big(1 + \mathbf{W}_\alpha \mathbf{C}^\top (\mathbf{C}\mathbf{W}_\alpha \mathbf{C}^\top)^{-1} (m - \mathbf{C1})\big) - \lambda f_{\text{prior}} < 0 \Big\},$$

$$\tilde{\mathcal{B}}_g := \{ \alpha : \alpha \in \tilde{\mathcal{A}}_g,\ \alpha \geq \alpha_{\min},\ |\tilde{f}^\alpha(\mathrm{CP}) - \tilde{f}^\alpha_\infty| \leq \tau \},$$

where $\tilde{f}^\alpha_\infty$ is the solution to the following new first-order condition with respect to f_∞ given α:

$$(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{U} \times \big(1 + \mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_\alpha\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})\big) + \lambda(f_\infty - f_{\text{prior}}) = 0,$$

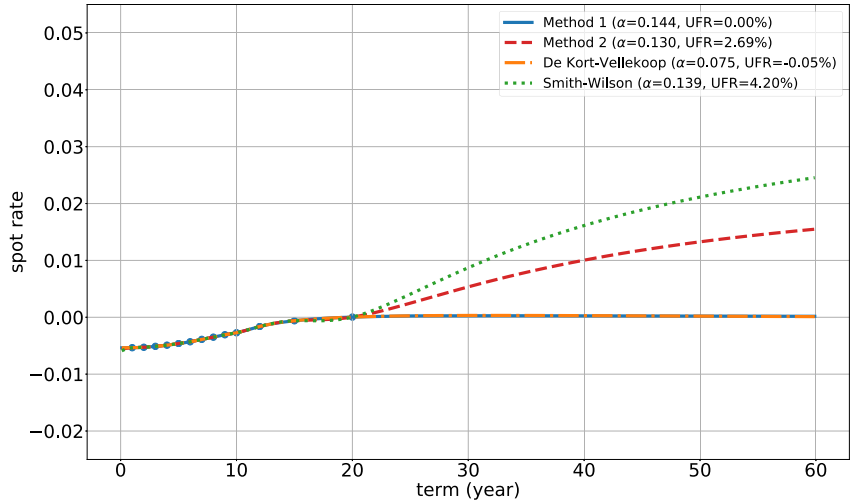
and $\tilde{f}^\alpha(\cdot)$ is the forward curve generated by the Smith-Wilson method (3) given α and $f_\infty = \tilde{f}^\alpha_\infty$.

Step 2: Choose an optimal α:

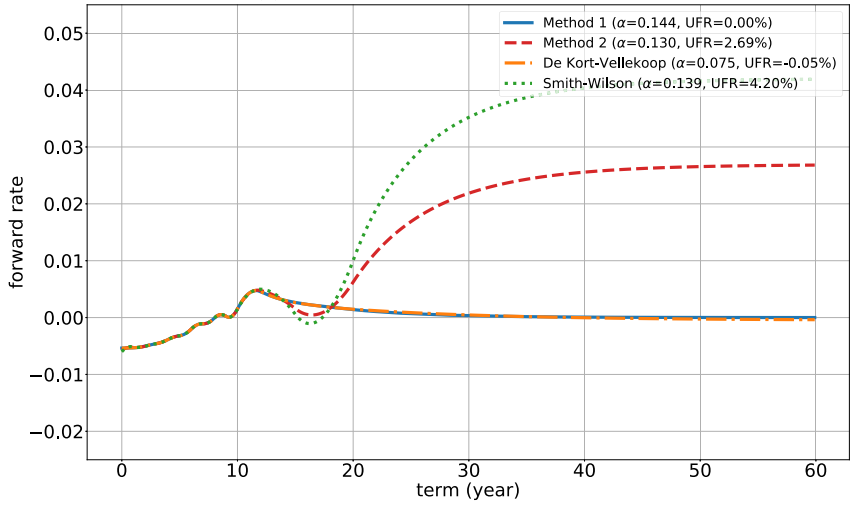
$$\tilde{\alpha}^*_g = \inf \tilde{\mathcal{B}}_g.$$

Step 3: Solve the following equation for f_∞:

$$(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_{\tilde{\alpha}^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{U} \times \Big(1 + \mathbf{W}_{\tilde{\alpha}^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top (\mathbf{C}\mathbf{D}^{f_\infty}\mathbf{W}_{\tilde{\alpha}^*_g}\mathbf{D}^{f_\infty}\mathbf{C}^\top)^{-1}(m - \mathbf{C}\mathbf{D}^{f_\infty}\mathbf{1})\Big)$$



(a) Spot rate curves, December 2020



(b) Forward rate curves, December 2020

Fig. C.5. Risk-free interest rate curves generated by different methods based on the EURIBOR swap data, December 2020.

$$+\lambda(f_{\infty}-f_{\text{prior}})=0.$$

The sets defined in Methods B.1 and B.2, $\mathcal{A}_g$, $\mathcal{B}_g$, $\tilde{\mathcal{A}}_g$, and $\tilde{\mathcal{B}}_g$ are more complex, and we cannot guarantee their non-emptiness as we did for Methods 1 and 2 in our main article.

Let us use an example to illustrate that Methods B.1 and B.2 may still be applied in practice. We use the same data as in Section 5. The main difference is that, here we take the payments of the $N=9$ representative bonds that occur within the first year into consideration. Therefore, we have $T=10$ payment intervals: $(0,1],(1,2],(2,3],(3,4],\ldots,(9,10]$. We denote the cash-flow-weighted average payment times within each interval by $u_1,u_2,\ldots,u_{10}$, respectively. In this setup, the cash flow matrix of the 9 representative bonds, denoted by $\mathbf{C}$, is a 9×10 rectangular matrix. Table B.1 provides an example of u_i, m_i, and the rectangular cash flow matrix $\mathbf{C}$ generated using the data of 2020 Quarter 3.

Similar to Section 5.2.1, we apply different methods to construct the risk-free interest rates based on the rectangular cash flow matrices. Fig. B.1 and Fig. B.2 show the spot rate curves and forward rate curves for 2020 Quarters 3 and 4, respectively. Four methods are applied: Method B.1 (the blue solid lines), Method B.2 (the red dashed lines), the de Kort-Vellekoop method (the orange dash-dotted lines), and the original Smith-Wilson method adopted by the EIOPA (the green dotted lines). The parameter setups are the same as in Section 5.2.1. We can find that the results are quite similar to those obtained with square cash flow matrices (see Figs. 7,8). The UFRs obtained by Method B.1 and Method B.2 for 2020 Quarter 3 (2020 Quarter 4) are 0.02% and 2.08% (1.28% and 2.64%), respectively, which are very close to those obtained with square cash flow matrices: 0.02% and 2.09% (1.30% and 2.58%), respectively.

Fig. B.3 shows the time series of UFRs obtained using the four methods from 2016 Quarter 1 to 2021 Quarter 1 when the cash flow matrices are rectangular. The yield rates of 10-year Chinese government bonds are also shown in the figure. The figure is similar to Fig. 9, which displays the time series of UFRs under square cash flow matrices. Therefore, the generalized methods, Methods B.1 and B.2, can provide results similar to those obtained by Methods 1 and 2 for our data.

Appendix C. Empirical study for EURIBOR swap rates

In this appendix, we present the results of an empirical study of our methods using the EURIBOR swap rates. Unlike the Chinese government bond interest rates, it is well known that short-term interest rates in the Eurozone have gone negative in recent years. This motivates us to study the performance of our methods in low or negative interest rate environments.

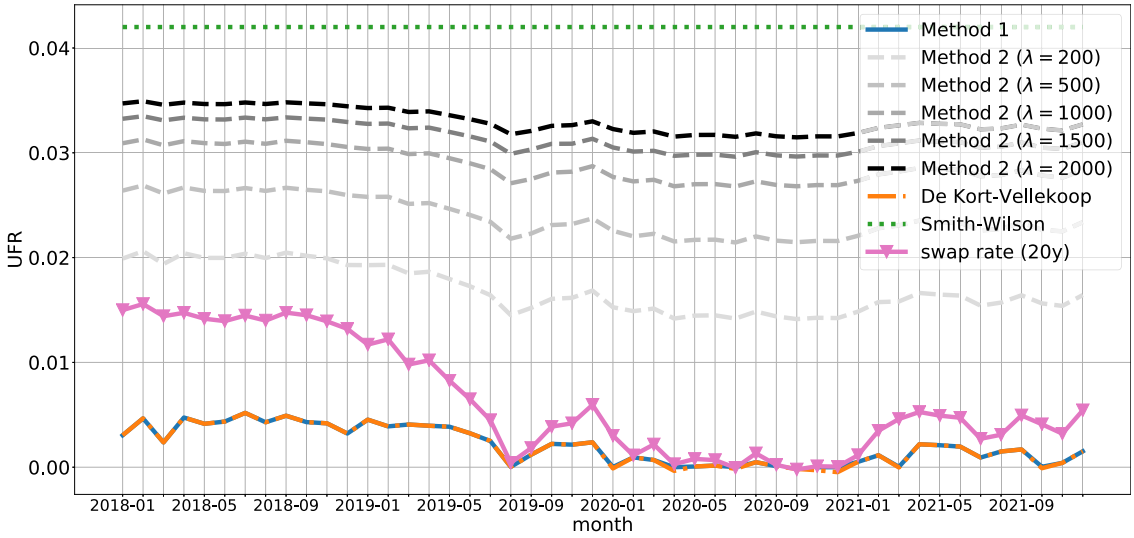


Fig. C.6. The time series of UFRs obtained by different methods based on the EURIBOR swap data.

We obtain the EURIBOR swap rates data from Refinitiv,[20] which is the same as the data vendor used by the EIOPA (EIOPA, 2019, Paragraph 60). The data includes monthly closing prices for EURIBOR swaps with various maturities from 2018 to 2021. According to the EIOPA (EIOPA, 2019, Paragraph 345), we use swaps with maturities of 1–10, 12, 15, and 20 years ($T = 13$), resulting in 13×13 cash flow matrices $\mathbf{C}$ for the swaps. The cash flow matrices $\mathbf{C}$ are generated based on the methodology proposed in EIOPA (2019, Table 9).

As two examples, Fig. C.4 and Fig. C.5 show the results obtained by Method 1 (the blue solid lines), Method 2 (the red dashed lines, with $f_{\text{prior}} = 4.20\%$), the de Kort-Vellekoop method (the orange dash-dotted lines), and the original Smith-Wilson method adopted by the EIOPA (the green dotted lines, with an exogenous UFR of 4.20%), based on the EURIBOR swap data of September and December 2020, respectively. Other parameter setups are the same as in Section 5.2.1. It is worth noting that almost all short-term swap rates are negative in both months, and the endogenous UFRs generated by the de Kort-Vellekoop method are 0.01% and −0.05% for September and December 2020, respectively.

We now examine the performance of our methods. Method 1 is successful in finding positive UFRs for both September 2020 (0.01%) and December 2020 (0.00%). Although Method 1 is practical in both months, the sufficient condition (22) for Theorem 2 may not hold. In September 2020, the left-hand side of (22) equals −0.0016, while in December 2020, it equals 0.0008. However, (22) holds for most of the months in our dataset.

Fig. C.4 and Fig. C.5 also imply that the curves generated by Method 2 generally fall between the curves obtained by Method 1 and those produced by the original Smith-Wilson method. In addition, the endogenous UFRs obtained by Method 2 are both 2.69% for September and December 2020.

Fig. C.6 shows the time series of UFRs obtained using different methods based on the EURIBOR swap data. Despite the low and even negative short-term swap rates during this period, Method 1 and Method 2 consistently generate positive UFRs. In addition, the curve for the time series of endogenous UFRs generated by Method 2 has a similar trend to that for the time series of 20-year swap rates (pink line). This suggests that Method 2 incorporates information from both the prior knowledge ($f_{\text{prior}} = 4.20\%$) and the current market data.

[20] See https://www.refinitiv.com/en.

References

Balter, A., Pelsser, A., Schotman, P.C., 2016. What does a term structure model imply about very long-term interest rates?. Available at SSRN 2386699.

Brigo, D., Mercurio, F., 2007. Interest Rate Models—Theory and Practice: with Smile, Inflation and Credit. Springer Science & Business Media.

de Kort, J., Vellekoop, M., 2016. Term structure extrapolation and asymptotic forward rates. Insurance: Mathematics and Economics 67, 107–119.

EIOPA, 2016. Consultation paper on the methodology to derive the UFR and its implementation. EIOPA-CP-16/03.

EIOPA, 2019. Technical documentation of the methodology to derive EIOPA's risk-free interest rate term structures. EIOPA-BoS-19/408.

European Systemic Risk Board, 2017. Regulatory risk-free yield curve properties and macroprudential consequences. Technical report. The ATC Expert Group on Insurance.

Fung, D.W., Jou, D., Shao, A.J., Yeh, J.J., 2018. The China Risk-Oriented Solvency System: a comparative assessment with other risk-based supervisory frameworks. The Geneva Papers on Risk and Insurance – Issues and Practice 43, 16–36.

Hagan, P.S., West, G., 2006. Interpolation methods for curve construction. Applied Mathematical Finance 13 (2), 89–129.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2 ed. Springer.

Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. Econometrica 60 (1), 77–105.

Hull, J., White, A., 1990. Pricing interest-rate-derivative securities. The Review of Financial Studies 3 (4), 573–592.

Jørgensen, P.L., 2018. An analysis of the Solvency II regulatory framework's Smith-Wilson model for the term structure of risk-free interest rates. Journal of Banking & Finance 97, 219–237.

Lagerås, A., Lindholm, M., 2016. Issues with the Smith-Wilson method. Insurance: Mathematics and Economics 71, 93–102.

Marques, L.B., Casiraghi, M., Gelos, G., Kamber, G., Meeks, R., 2021. Negative Interest Rates: Taking Stock of the Experience so Far. International Monetary Fund.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. Journal of Business 60 (4), 473–489.

Smith, A., Wilson, T., 2000. Fitting yield curves with long term constraints. Bacon & Woodrow Research Notes.

Svensson, L.E., 1994. Estimating and interpreting forward interest rates. Sweden 1992–1994. Technical report. National Bureau of Economic Research.